14th International Conference on

FORMAL POWER SERIES & ALGEBRAIC COMBINATORICS

SÉRIES FORMELLES ET COMBINATOIRE ALGÉBRIQUE

CONFERENCE PROCEEDINGS

The University of Melbourne
Australia
8th – 12th July
2002

Editors:
Richard Brak, Omar Foda, Catherine Greenhill, Tony Guttmann, Aleks Owczarek
May 2002

ISBN: 0 7340 2215 8

Formal Power Series and Algebraic Combinatorics

FPSAC 2002

14th Conference

Melbourne, 8–12 July 2002

Séries Formelles et Combinatoire Algébrique

SFCA 2002

14éme Colloque

Melbourne, 8–12 Juillet 2002

Proceedings • Actes

The University of Melbourne, Australia

Eds.: R. Brak, O. Foda, C. Greenhill, T. Guttmann, A. Owczarek

FPSAC 2002 – SFCA 2002

Program Committee

Susumu Ariki (Tokyo)
Sara Billey (MIT)
Maylis Delest (Bordeaux)
Art Duval (Texas-El Paso)
Omar Foda (Co-chair; Melbourne)
Sergey Fomin (Michigan)
Vesselin Gasharov (Cornell)
Tony Guttmann (Co-chair; Melbourne)
Angele Hamel (Wilfrid Laurier)
Ron King (Southampton)
Gilbert Labelle (UQAM)
Jean-Christophe Novelli (Lille)
Renzo Pinzani (Florence)
Andrew Rechnitzer (Toronto)
Frank Sottile (Massachusetts)
Itaru Terada (Tokyo)
Jean-Yves Thibon (Marne-la-Valle)
Dominic Welsh (Oxford)
Trevor Welsh (Melbourne)
Nick Wormald (Melbourne)

Organising Committee

Nantel Bergeron (Toronto)
Richard Brak (Melbourne)
Catherine Greenhill (Melbourne)
Tony Guttmann (Chair; Melbourne)
Aleks Owczarek (Melbourne)

Sponsors

The Department of Mathematics and Statistics
The University of Melbourne, Australia

Microsoft Corporation
Redmond, Washington, USA

National Science Foundation
Arlington, Virginia, USA

State Government Victoria AUSTRALIA

Department of Innovation, Industry and Regional Development,
The Victorian Government,
Australia

Invited Speakers

Hélène Barcelo	Arizona State University
Jan De Gier	The University of Melbourne
Philippe Di Francesco	Commissariat à l'Energie Atomique
Peter Forrester	The University of Melbourne
Christian Krattenthaler	University of Vienna
Brendan McKay	Australian National University
Thomas Prellberg	Technische Universität Clausthal
Alan Sokal	New York University
Ole Warnaar	The University of Melbourne
David Wilson	Microsoft Research

Preface

Welcome to the 14th International Conference on Formal Power Series and Algebraic Combinatorics FPSAC 2002, Séries Formelles et Combinatoire Algébrique, SFCA 2002, held in Melbourne, Australia, in the week 8 – 12 July 2002. We are particularly pleased because this is the first time that FPSAC/SFCA has been held in the southern hemisphere, and also the first time that it has been held in the Asia–Pacific region. We hope that all overseas participants (especially first-time visitors) enjoy their stay in Australia.

The FPSAC 2002 program consists of invited and contributed talks and posters involving algebraic and enumerative combinatorics, applied to problems drawn from pure mathematics, applied mathematics, theoretical computer science and physics. It is hoped that the cross-disciplinary nature of the contributions will benefit researchers in all areas.

This proceedings contains the extended abstracts of all papers presented at the conference. There are ten invited lectures, presented by Hélène Barcelo, Jan de Gier, Philippe Di Francesco, Peter Forrester, Christian Krattenthaler, Brendan McKay, Thomas Prellberg, Alan Sokal, Ole Warnaar and David Wilson. This year the extended abstracts will be distributed to participants in the form of a PDF file on a CD-ROM. Please inform members of the permanent committee of your response to this innovation.

Thanks are due to all members of the Program Committee who worked speedily and judiciously, and also to the additional referees called upon to review the papers. Following the FPSAC custom, all abstracts are given in English and French. Many thanks to all those who helped with the translations.

We are also grateful to the following institutions for their financial support: The Department of Mathematics of The University of Melbourne, Microsoft Corporation, the National Science Foundation (USA) and the Department of Innovation Industry and Regional Development, (The Victorian Government, Australia).

Richard Brak, Omar Foda, Catherine Greenhill, Tony Guttmann, Aleks Owczarek

Melbourne, May 2002

Table of Contents – Table des Matières

A LOWER BOUND FOR THE ORDER OF TELESCOPERS FOR A HYPERGEOMETRIC TERM

S. A. ABRAMOV* AND H. Q. LE†

Abstract. We present an algorithm to compute a lower bound for the order of the minimal telescoper for a given hypergeometric term. We also describe a Maple implementation of the algorithm and show the efficiency improvement it provides to Zeilberger's algorithm in the construction of the telescopers.

Résumé. Cet article présente un algorithme de calcul d'une borne inférieure pour le téléscopeur minimal associé à un terme hypergéométrique. Une implantation en Maple est également décrite, qui permet d'observer le gain de notre méthode par rapport à l'algorithme de Zeilberger lors de la construction des téléscopeurs.

1. Preliminaries

Let K be an algebraically closed field of characteristic 0, the variables n, k be integer-valued, and E_n, E_k be the corresponding shift operators, acting on functions of n and k, by $E_n f(n,k) = f(n+1,k)$, $E_k f(n,k) = f(n,k+1)$. A K-valued function $t(k)$ is a *hypergeometric term* of k over K if the consecutive term ratio $R = E_k t/t$ is a rational function of k over K. This rational function is the *certificate* of $t(k)$. A K-valued function $T(n,k)$ is a hypergeometric term of two variables n and k if the two consecutive term ratios $R_1 = E_n T/T$, and $R_2 = E_k T/T$ are rational functions of n and k over K. They are called the n-certificate and the k-certificate of T, respectively. Given a hypergeometric term $T(n,k)$ as input, Zeilberger's algorithm [13, 15, 16] (which we name hereafter as $\mathcal{Z}$) constructs for $T(n,k)$ a *Z-pair* (L,G), provided that such a pair exists. The computed Z-pair consists of L, a linear recurrence operator with coefficients which are polynomials of n over K

$$L = a_\rho(n)E_n^\rho + \cdots + a_1(n)E_n^1 + a_0(n)E_n^0, \tag{1}$$

and a hypergeometric term $G(n,k)$ such that

$$LT(n,k) = (E_k - 1)G(n,k). \tag{2}$$

The k-free operator L is called a *telescoper*. It is noteworthy that the problem of establishing a necessary and sufficient condition for the applicability of $\mathcal{Z}$ to $T(n,k)$ is solved and presented in [1] (the well-known *fundamental theorem* [15, 16] only provides a sufficient condition). It is proven in [16] that if there exists a Z-pair for $T(n,k)$, then $\mathcal{Z}$ terminates with one of the Z-pairs and the telescoper L in the returned Z-pair is of minimal order. The computed telescoper L is unique up to a left-hand factor $P(n) \in K[n]$, and we name it *the minimal telescoper.*

$\mathcal{Z}$ has a wide range of applications which include finding closed forms of definite sums of hypergeometric terms, verification of combinatorial identities, and asymptotic estimation [13, 16, 12].

*Partially supported by the French-Russian Lyapunov Institute under grant 98-03.

†Partially supported by Natural Sciences and Engineering Research Council of Canada Grant No. CRD215442-98.

The algorithm uses an *item-by-item examination* on the order ρ of the operator L in (1). It starts with the value of 0 for ρ and increases ρ until it is successful in finding a $\mathcal{Z}$-pair (L, G) for T. As a consequence, we waste resources trying to compute without success a telescoper of $\operatorname{ord} L < \rho$ where ρ is the order of the minimal telescoper.

In this paper, we present an algorithm that computes a lower bound for the order of the telescopers. The general approach of the algorithm can be described as follow. If the given hypergeometric term $T(n, k)$ is not k-summable, i.e., there does not exist a hypergeometric term T_1 such that $T = (E_k - 1)T_1$, then we represent T as $(E_k - 1)T_1 + T_2$ where the hypergeometric term T_2 has some specific features each of which ensures that T_2 is not k-summable. It is then easy to show that a telescoper for T exists iff a telescoper for T_2 exists, and the sets of telescopers for T and T_2 are equal. We consider recurrence operators, called crushing operators, with the distinguishing property that if $M \in K[n, E_n]$ is a crushing operator for T_2, then MT_2 does not have at least one of the specific features that T_2 does (this does not guarantee that MT_2 is k-summable, though). It follows that the order of the minimal telescoper for T_2 is always greater than or equal to that of the minimal crushing operator for T_2. We then describe an algorithm to compute a lower bound μ for the order of the crushing operators for T_2. This value is automatically also a lower bound for the order of the telescopers for T.

When the algorithm is used in conjunction with the algorithm to determine the applicability of $\mathcal{Z}$ to $T(n, k)$ [1], it allows one to use $\mathcal{Z}$ to compute a $\mathcal{Z}$-pair only if the existence of such a pair is guaranteed; and in this case, one can use μ as the starting value for the order of L, instead of 0. Additionally, the computation of a lower bound is much less expensive than the construction of a telescoper using $\mathcal{Z}$, especially when the order of the minimal telescoper is high.

Note that for the case where $T(n, k)$ is also a rational function, there exists a direct algorithm to compute the minimal telescoper for T efficiently without using item-by-item examination [8].

The paper is organized in the following manner. In Sections 2 we discuss some known results which are needed in subsequent sections. They include a description of the additive decomposition problem of hypergeometric terms [2, 3], and a criterion for the applicability of $\mathcal{Z}$ [1]. The main result of Section 3 is a theorem to compute a lower bound for the order of the minimal crushing operator M. An algorithm which realizes this theorem is presented in Section 4. We conclude the paper with a description of an implementation of the algorithm in Section 5. Various examples are used to show the advantages of this implementation over other implementations of the original $\mathcal{Z}$.

Throughout the paper, K is an algebraically closed field of characteristic 0, and $\mathbb{N}$ denotes the set of nonnegative integers. Following [13], we write $T_1(n, k) \sim T_2(n, k)$ if two hypergeometric terms $T_1(n, k)$ and $T_2(n, k)$ are *similar*, i.e., their ratio is a rational function of n and k.

2. The Additive Decomposition Problem and the Existence of a Telescoper

We begin this section with the notion of *Rational Normal Forms* (RNF) of rational functions [4]. This concept plays an important role in the follow-up algorithms.

Definition 1. *Let Λ be a field of characteristic 0. Let $R \in \Lambda(x)$ be a nonzero rational function. If there exist nonzero polynomials $f_1, f_2, v_1, v_2 \in \Lambda[x]$ such that*

(i) $R = F \cdot \frac{EV}{V}$ *where* $F = \frac{f_1}{f_2}$, $V = \frac{v_1}{v_2}$, *and* $\gcd(v_1, v_2) = 1$,

(ii) $\gcd(f_1, E^h f_2) = 1$ *for all* $h \in \mathbb{Z}$,

then $F \cdot \frac{EV}{V}$ is an RNF of R.

Note that every rational function has an RNF [3, Thm. 1] which in general is not unique, and the rational function F in (i) with property (ii) is called the *kernel* of the RNF.

2.1. **The Additive Decomposition Problem.** For a hypergeometric term $T(k)$ of k over $K(n)$, the algorithm to solve the additive decomposition problem [2, 3] constructs two hypergeometric terms $T_1(k)$, $T_2(k)$ similar to $T(k)$ such that

$$T(k) = (E_k - 1)\, T_1(k) + T_2(k), \tag{3}$$

and either $T_2 = 0$ or the k-certificate of T_2 has an RNF

$$\frac{f_1}{f_2}\,\frac{E_k(v_1/v_2)}{(v_1/v_2)} \tag{4}$$

with v_2 of minimal possible degree. Note that any RNF of the k-certificate of T_2 has $v_2 \in K(n)[k]$ of the same (minimal possible) degree.

Lemma 1. [2, 3] *Let $T(k)$ be a hypergeometric term over $K(n)$. If (3) is an additive decomposition of $T(k)$, then for any RNF of the form (4) of the k-certificate of $T_2(k)$, and for each irreducible p from $K(n)[k]$ such that $p \mid v_2$, the following three properties hold:*

$$\mathbf{Pa}: E_k^h p \mid v_2 \Rightarrow h = 0, \quad \mathbf{Pb}: E_k^h p \mid f_1 \Rightarrow h < 0, \quad \mathbf{Pc}: E_k^h p \mid f_2 \Rightarrow h > 0. \tag{5}$$

If the hypergeometric term $T_2(k)$ in (3) vanishes, then $T(k)$ is said to be *k-summable.* Otherwise, each irreducible factor p of v_2 has properties **Pa, Pb, Pc**, and T is k-non-summable.

Proposition 1. [2, 3] *Let an RNF of the k-certificate of a given hypergeometric term $T(n,k)$ be of the form (4). If there exists at least one irreducible factor p of v_2 such that all three properties* **Pa, Pb, Pc** *hold, then $T(n,k)$ is k-non-summable.*

Proposition 2. *Let the similar hypergeometric terms $T(n,k)$, $T_1(n,k)$, and $T_2(n,k)$ be as defined in (3). (The algorithm to solve the additive decomposition problem is applied to $T(n,k)$ w.r.t. k over $K(n)$.) Then*

(i) *A* Z*-pair for $T(n,k)$ exists iff a* Z*-pair for $T_2(n,k)$ exists;*
(ii) *The minimal telescopers for T and T_2 are the same.*

Proof:
(i): Let (L,G) be a *Z*-pair for T_2. It follows from (3) that $LT = (E_k - 1)\,(LT_1 + G)$. Since $T_1 \sim T_2, T_2 \sim G$, and $\sim$ is an equivalence relation, $LT_1 + G$ is a hypergeometric term [13, Prop. 5.6.2]. Consequently, $(L, LT_1 + G)$ is a *Z*-pair for T. On the other hand, let (L,G) be a *Z*-pair for T. By following the same argument, one can easily show that $(L, G - LT_1)$ is a *Z*-pair for T_2.
(ii): Let L be the minimal telescoper for T_2. It follows from (i) that L is a telescoper for T. Suppose there exists a telescoper $\tilde{L}$ for T and $\operatorname{ord}\tilde{L} < \operatorname{ord} L$. Then it follows from (i) then $\tilde{L}$ is a telescoper for T_2 and $\operatorname{ord}\tilde{L} < \operatorname{ord} L$. Contradiction. ∎

Definition 2. *A polynomial $p(n,k) \in K[n,k]$ is* integer-linear *if it has the form*

$$\alpha n + \beta k + \gamma \ \text{ where } \ \alpha,\, \beta \in \mathbb{Z} \text{ and } \gamma \in K. \tag{6}$$

Theorem 1. [5, Thm. 8] *For a hypergeometric term $T(n,k)$, let $F, V \in K(n,k)$ be such that*

$$F\,\frac{E_k\,V}{V}$$

is an RNF over $K(n)$ of the k-certificate of T. Then there exists $D \in K(n,k)$ so that the n-certificate of T can be written as

$$D\,\frac{E_n\,V}{V},\ D = \frac{d_1}{d_2},\ \gcd(d_1, d_2) = 1, \tag{7}$$

and F, D both factor into constants and integer-linear polynomials.

2.2. **The Existence of a Telescoper.** Recall that the fundamental theorem [15, 16] provides only a sufficient condition for the termination of $\mathcal{Z}$. It states that if $T(n,k)$ is a *proper* hypergeometric term (see the definition from [15, 16]), then a telescoper for $T(n,k)$ exists. However, it is well-known that the set $\mathcal{S}$ of hypergeometric terms on which $\mathcal{Z}$ terminates is a proper subset of the set of all hypergeometric terms, but a super-set of the set of proper hypergeometric terms. The following theorem [1] gives a complete description of $\mathcal{S}$. It provides a necessary and sufficient condition for the termination of $\mathcal{Z}$.

Theorem 2. *(Criterion for the existence of a telescoper). Let $T(n,k)$ be a hypergeometric term of n and k, and (3) be an additive decomposition of $T(n,k)$. Let (4) be an RNF w.r.t. k over $K(n)$ of the k-certificate of $T_2(n,k)$ with $v_2 \in K[n,k]$. Then a telescoper for $T(n,k)$ exists iff each factor of $v_2(n,k)$ irreducible in $K[n,k]$ is an integer-linear polynomial, i.e., iff $T_2(n,k)$ is proper.*

3. A Lower Bound for the Order of Telescopers for a Minimal k-Non-summable Term

Definition 3. *A* minimal k-non-summable *hypergeometric term $T(n,k)$ is a hypergeometric term where the k-certificate of T has an RNF of the form (4), and for each irreducible p such that $p \mid v_2$, all three properties* **Pa**, **Pb**, **Pc** *hold.*

For the remainder of this section, we assume $T(n,k)$ to be a minimal k-non-summable hypergeometric term. Let us now introduce the notion of *crushing operators*.

Definition 4. *Let $M \in K[n, E_n]$ be such that $MT \neq 0$, and there exists an RNF*

$$F' \frac{E_k V'}{V'}, \; V' = \frac{v_1'}{v_2'} \tag{8}$$

of the k-certificate of the hypergeometric term MT such that each of the irreducible factors of v_2' does not have at least one of the three properties **Pa, Pb, Pc**. *Then M is a* crushing operator *for T. The minimal crushing operator is a crushing operator of minimal possible order.*

Proposition 3. *If L is a telescoper for T, then L is a crushing operator for T.*

Proof: The claim follows from Proposition 1. ∎

Corollary 1. *If there does not exist any crushing operator for T of order less than μ, $\mu \geq 1$, then there does not exist any telescoper for T of order less than μ.*

Hence, the problem of computing a lower bound for the order of the telescopers for T is reduced to the problem of computing a lower bound for the order of the minimal crushing operator for T.

Theorem 3. *Let the k-certificate of T has an RNF $F(E_k V)/V$ of the form (4). Let the n-certificate A of T $(E_n T)/T = D(E_n V)/V$ be as defined in Theorem 1. Suppose that the polynomial $v_2 \in K[n,k]$ factors into a constant and integer-linear polynomials. Let $M \in K[n, E_n]$ be a crushing operator for $T(n,k)$, $\operatorname{ord} M = \rho$. Let p be an integer-linear factor of v_2, $\deg_k p = 1$. Then*

(i) *There exists an integer h such that*

$$E_k^h\, p \mid E_n v_2 \cdot E_n^2 v_2 \cdots E_n^{\rho} v_2 \cdot d_2 \cdot E_n d_2 \cdots E_n^{\rho-1} d_2; \tag{9}$$

(ii) *Let ρ_p be the minimal value of ρ in (i) such that (9) is satisfied. Then the order of the minimal crushing operator for T is not less than $\mu = \max_{p \mid v_2} \rho_p$.*

Proof:
(i): Let

$$M = a_\rho(n)E_n^\rho + \cdots + a_1(n)E_n + a_0(n),\ a_i(n) \in K[n].$$

Then

$$MT = \left(\sum_{m=0}^{\rho} a_m(n)A \cdot E_n A \cdots E_n^{m-1}A\right)T.$$

Therefore, the k-certificate of MT is

$$F\frac{E_k R}{R}, \tag{10}$$

where

$$\begin{aligned} R &= V\sum_{m=0}^{\rho} a_m(n)A \cdot E_n A \cdots E_n^{m-1}A \\ &= V\sum_{m=0}^{\rho} a_m(n)\frac{E_n^m V}{V}D \cdot E_n D \cdots E_n^{m-1}D \\ &= \sum_{m=0}^{\rho} a_m(n)\frac{E_n^m v_1 \cdot d_1 \cdot E_n d_1 \cdots E_n^{m-1}d_1}{E_n^m v_2 \cdot d_2 \cdot E_n d_2 \cdots E_n^{m-1}d_2}. \end{aligned}$$

Rewrite R as

$$R = \frac{r_1}{r_2},\ r_1,\ r_2 \in K[n,k],$$

$$r_2 = v_2 \cdot E_n v_2 \cdots E_n^\rho v_2 \cdot d_2 \cdot E_n d_2 \cdots E_n^{\rho-1}d_2,\ r_1 = s_1 + v_2\, s_2,$$

where s_2 is a polynomial from $K[n,k]$, and $s_1 = a_0(n) \cdot E_n v_2 \cdots E_n^\rho v_2 \cdot d_2 \cdot E_n d_2 \cdots E_n^{\rho-1}d_2$.

If p is *not* a factor of the denominator r_2 of R, then since v_2 is a factor of r_2, p must divide the numerator r_1 of R, i.e.,

$$p \,|\, (s_1 + v_2\, s_2).$$

Since p is a factor of v_2, this implies $p \,|\, s_1$. Additionally, p does not divide $a_0(n)$ since $\deg_k p = 1$. Therefore,

$$p \,|\, E_n v_2 \cdot E_n^2 v_2 \cdots E_n^\rho v_2 \cdot d_2 \cdot E_n d_2 \cdots E_n^{\rho-1}d_2. \tag{11}$$

If p is a factor of the denominator r_2, then since M is a crushing operator for T, at least one of the three properties **Pa**, **Pb**, **Pc** does not hold for p. Notice that the k-certificates of T in (4) and MT in (10) have the same kernel F. It follows from Lemma 1 that for the integer-linear factor p of v_2, properties **Pb** and **Pc** *always* hold. Consequently, property **Pa** does *not* hold, i.e., there exists an $h \in \mathbb{Z} \setminus \{0\}$ such that $E_k^h p$ divides r_2. Additionally, since T is a minimal k-non-summable hypergeometric term, it follows from property **Pa** that there does not exist an $h \in \mathbb{Z} \setminus \{0\}$ such that $E_k^h p | v_2$. This gives

$$E_k^h\, p \,|\, E_n v_2 \cdot E_n^2 v_2 \cdots E_n^\rho v_2 \cdot d_2 \cdot E_n d_2 \cdots E_n^{\rho-1}d_2. \tag{12}$$

It follows from (11) and (12) that (i) is satisfied.
(ii): The claim follows from the fact that for each factor p of v_2, there does not exist any crushing operator for T of order less than ρ_p. ∎

4. A General Algorithm

For a given hypergeometric term $T(n,k)$ of n and k, an algorithm to compute a lower bound μ for the order of the telescopers for T consists of two steps. A check to determine the existence of a telescoper for T is performed in the first step. This is attained by first applying to $T(n,k)$ the algorithm to solve the additive decomposition problem w.r.t. k to construct two hypergeometric terms $T_1(n,k)$, $T_2(n,k)$ such that

$$T(n,k) = (E_k - 1)\,T_1(n,k) + T_2(n,k), \tag{13}$$

and the k-certificate of T_2 has an RNF of the form (4). If v_2 does not factor into integer-linear polynomials, then it follows from Theorem 2 that $\mathcal{Z}$ is not applicable to T, and there is no need to compute a lower bound μ. Otherwise, rewrite v_2 as a product of integer-linear polynomials each of which is of the form (6). An algorithm, based on gcd and resultant computation, to check if $v_2 \in K[n,k]$ factors into integer-linear polynomials, and if this is the case, rewrite v_2 in the desired factored form is described in [6, 7]. Without loss of generality, we can assume that $\gcd(\alpha,\beta)=1$, and $\beta \geq 0$.

In the second step, since T_2 is a minimal k-non-summable hypergeometric term, if follows from Proposition 3 that the existence of the crushing operators for T_2 is guaranteed. Additionally, all the hypotheses required for the computation of a lower bound μ for the order of the telescopers for T_2 exist. Hence, apply Theorem 3 to $T_2(n,k)$ to compute a lower bound μ. It follows from Proposition 2 that one can use μ as a lower bound for the order of the telescopers for T.

For each integer-linear factor p of v_2, $\deg_k p = 1$, the second step requires the computation of the minimal value of ρ in the pair (ρ, h), $h \in \mathbb{Z}$, $\rho \in \mathbb{N}\setminus\{0\}$ such that

(i) $E_k^h p \mid E_n v_2 \cdot E_n^2 v_2 \cdots E_n^{\rho} v_2$, or
(ii) $E_k^h p \mid d_2 \cdot E_n d_2 \cdots E_n^{\rho-1} d_2$.

Consider the following simple algorithm $C_{(i)}$:

`algorithm` $C_{(i)}$;
`input:` $p = \alpha n + \beta k + \gamma$, α, $\beta \in \mathbb{Z}$, $\gcd(\alpha,\beta)=1$, $\beta > 0$, $\gamma \in K$,
$v_2 = \prod_{i=1}^{m}(\alpha_i n + \beta_i k + \gamma_i)$, α_i, $\beta_i \in \mathbb{Z}$, $\gcd(\alpha_i,\beta_i)=1$, $\beta_i \geq 0$, $\gamma_i \in K$;
`output:` the minimal value of $\rho \in \mathbb{N}\setminus\{0\}$ such that (i) is satisfied;

$\rho_{min} := \infty$;
for $i = 1, 2, \ldots, m$ **do**
  if $\alpha = \alpha_i$ **and** $\beta = \beta_i$ **and** $\gamma - \gamma_i \in \mathbb{Z}$ **then**
    find the minimal $\rho \in \mathbb{N}\setminus\{0\}$ and $h \in \mathbb{Z}$ such that
      $\alpha\rho - \beta h = \gamma - \gamma_i$;
    $\rho_{min} := \min\{\rho_{min}, \rho\}$;
  fi;
od;
return ρ_{min}.

For a given integer-linear factor p of v_2, $\deg_k p = 1$, the algorithm $C_{(i)}$ simply iterates through each integer-linear polynomial q of v_2. If $p - q = \sigma \in \mathbb{Z}$, then the algorithm solves the diophantine equation $\alpha\rho - \beta h = \sigma$, and chooses the minimal positive value of ρ. (Note that since $\gcd(\alpha,\beta)=1$, the solution is guaranteed to exist.)

An algorithm $C_{(ii)}$ which finds the minimal value of ρ such that (ii) is satisfied can be described in a very similar manner. Note that it follows from Theorem 1 that the polynomial $d_2 \in K[n,k]$ in (7) factors into integer-linear polynomials.

By iterating through each factor p of v_2, we obtain the requested lower bound μ. This leads to the following algorithm which computes a lower bound for the order of the telescopers for a given hypergeometric term $T(n,k)$.

```
algorithm LowerBound;
input: a hypergeometric term T(n,k) ∈ K[n,k];
output: a lower bound μ for the order of the telescopers for T;

  apply the algorithm to solve the additive decomposition problem
     w.r.t. k to obtain T1(n,k), T2(n,k) in (13);
  if T2 = 0 then return 0; fi;
  at this point, T2 has an RNF of the form (4);
  if the polynomial v2(n,k) in (4) is written as
     v2 = ∏_{i=1}^{s} p_i where p_i = (α_i n + β_i k + γ_i),
     α_i, β_i ∈ ℤ, gcd(α_i, β_i) = 1, β_i ≥ 0, γ_i ∈ K then
     if s = 0 then return 1; fi;
     μ := −∞;
     Rewrite d2 as
        d2 = ∏_{j=1}^{t} q_j where q_j = (α_j n + β_j k + γ_j),
        α_j, β_j ∈ ℤ, gcd(α_j, β_j) = 1, β_j ≥ 0, γ_j ∈ K;
     for i = 1, 2, ..., s do
        if deg_k p_i = 1 then
           μ_min := C_(i)(p_i, v2);
           μ_min := min{μ_min, C_(ii)(p_i, d2)};
           μ := max {μ, μ_min};
        fi;
     od;
     return μ;
  else
      return "Zeilberger's algorithm is not applicable";
  fi;
```

Note that instead of rewriting d_2 as a product of integer-linear polynomials, and using it in the call $C_{(ii)}(p_i, d_2)$ in `LowerBound`, it is possible to use a simpler polynomial which is a divisor of d_2. For a given $f \in K[n,k]$ and $c \in \mathbb{Q}$, there exists an algorithm [7] (called *wc*) to extract the maximal factor $w \in K[n,k]$ from f where w can be written in the form

$$\prod_i (k + c\,n + \gamma_i),\ \gamma_i \in K.$$

Hence, for each factor $p = (\alpha n + \beta k + \gamma)$ of v_2, we call *wc* with d_2 and α/β as input. This also helps reduce the number of integer-linear factors of d_2 to be compared with p.

Example 1 Consider the hypergeometric term

$$T = \frac{1}{(5\,n + 2\,k + 1)\,(-3\,n + 5\,k + 5)}.$$

(T is also a rational function of n and k.) Applying the algorithm to solve the additive decomposition problem yields two hypergeometric terms $T_1(n,k) = 0$ and $T_2(n,k) = T(n,k)$ in (13). Since T is a rational function, the polynomial v_2 in (4), and subsequently d_2 in (7) can be readily rewritten as

$$v_2 = (5\,n + 2\,k + 1)(-3\,n + 5\,k + 5),\ \ d_2 = 1.$$

Since v_2 can be written as a product of integer-linear polynomials, it follows from algorithm *LowerBound* that $\mathcal{Z}$ is applicable to T, and the two possible values for the integer-linear factor p are

$$p_1 = 5n + 2k + 1, \quad p_2 = -3n + 5k + 5.$$

When $p = p_1 = 5\,n + 2\,k + 1$, the diophantine equation to be solved is $5\,\rho - 2\,h = 0$, which yields $(\rho_1, h_1) = (2, 5)$ as the solution. When $p = p_2 = -3\,n + 5\,k + 5$, the diophantine equation to be solved is $-3\,\rho - 5\,h = 0$, which yields $(\rho_2, h_2) = (5, -3)$ as the solution. Therefore, a lower bound μ for the order of the telescopers for T is $\mu = \max\{2, 5\} = 5$. Note that invoking $\mathcal{Z}$ on T results in the minimal telescoper L of order 6 where

$$L = (31\,n + 181)\,{E_n}^6 + (31\,n + 150)\,{E_n}^5 - (31\,n + 26)\,E_n - (31\,n - 5).$$

Example 2 Consider the class of hypergeometric terms of the form

$$T = \frac{1}{(a_1 n + b_1 k + c_1)\,(a_2 n + b_2 k + c_2)!}, \tag{14}$$

where $a_1,\, b_1,\, a_2,\, b_2 \in \mathbb{Z}$, $\gcd(a_1, b_1) = 1$, $b_1 \neq 0$, $a_1 \neq a_2$ or $b_1 \neq b_2$. Without loss of generality, we can assume that $b_1 > 0$. Applying the algorithm to solve the additive decomposition problem yields two hypergeometric terms $T_1(n, k) = 0$ and $T_2(n, k) = T(n, k)$ in (13), and the polynomial v_2 in (4) is

$$a_1 n + b_1 k + c_1,$$

which is also the only possible value of p. Subsequently, the value of d_2 in (7) is

$$\begin{cases} d_2 = (a_2 n + b_2 k + c_2 + 1) \cdots (a_2 n + b_2 k + a_2 + c_2) & \text{if } a_2 > 0, \\ d_2 = 1 & \text{if } a_2 = 0, \\ d_2 = (a_2 n + b_2 k + c_2 + a_2 + 1) \cdots (a_2 n + b_2 k + c_2) & \text{if } a_2 < 0. \end{cases}$$

Since $a_1 \neq a_2$ or $b_1 \neq b_2$, there does not exist any integer h such that $E_k^h p \,|\, d_2 \cdot E_n d_2 \cdots E_n^{\rho-1} d_2$. When $p = a_1 n + b_1 k + c_1$, the diophantine equation to be solved is $a_1 \rho - b_1 h = 0$, which yields $(\rho_1, h_1) = (b_1, a_1)$ as the solution. Therefore, a lower bound μ for the order of the telescopers for T is $\mu = b_1$.

In summary, for the class of hypergeometric terms of the form (14), the polynomial factor $(a_1 n + b_1 k + c_1)$ is the *dominant* factor. It determines the lower bound (which is b_1) for the order of the minimal telescoper for T. As an example, the computed lower bound for the minimal telescoper for

$$T = \frac{1}{(n - 9\,k - 2)\,(2\,n + k + 3)!}$$

is 9, while the order of the minimal telescoper for T is 10. By first computing this lower bound, we can safely avoid the computation of a telescoper of order less than 9 (in addition to the assurance that the telescopers for T do exist). On the other hand, if $b_1 = 1$, then the computed lower bound μ equals 1, i.e., the lowest possible value for μ. As an example, the computed lower bound for the minimal telescoper for

$$T = \frac{1}{(n + k + 1)(n + 5\,k + 2)!}$$

is 1, while the order of the minimal telescoper for T is 6.

Notice that when the factorial term $(a_2 n + b_2 k + c_2)!$ in (14) equals 1, we have b_1 as a lower bound for the order of the minimal telescoper for T. This lower bound also equals the order of the minimal telescoper for T (see [8]).

5. Implementation

The algorithm to compute a lower bound for the order of the telescopers and related functionalities are implemented in Maple 7 [11]. These functions are merged into the module `HypergeometricSum` [9]. They include:

(1) `AdditiveDecomposition` solves the additive decomposition problem;
(2) `IsZApplicable` determines the applicability of Zeilberger's algorithm;
(3) `LowerBound` computes a lower bound for the order of the telescopers.

The function `LowerBound` has the calling sequence

$$\text{LowerBound}(T, n, k, E_n, Zpair);$$

where T is a hypergeometric term of n and k, and E_n denotes the shift operator w.r.t. n. (E_n and $Zpair$ are optional arguments.) If the non-existence of a Z-pair (L, G) for T is guaranteed, then `LowerBound` returns the conclusive error message "Zeilberger's algorithm is not applicable." Otherwise, the output is a non-negative integer μ denoting the value of the computed lower bound for the order of L. In this case, if the optional arguments E_n and $Zpair$ (each of which can be any unassigned name) are given, then the function `Zeilberger` [9] is invoked starting with μ as a lower bound for the order of L, and $Zpair$ will be assigned to the computed Z-pair (L, G).

Note that there exist different Maple implementations of $\mathcal{Z}$ such as `zeil` in the EKHAD package [13], and `sumrecursion` in the `sumtools` package. A Mathematica implementation is presented in [12]. Since the terminating condition that allows a hypergeometric term to have a Z-pair is unknown at the time these functions were implemented, an upper bound for the order of the recurrence operator L in the Z-pair (L, G) needs to be specified in advance (for instance, the default values are 6 for the parameter `MAXORDER` in `zeil`, and 5 for the global parameter `` `sum/zborder` `` in `sumrecursion`). As a consequence, when given a hypergeometric term $T(n, k)$ as input, (1) these programs might fail even if a Z-pair exists, i.e., the maximum order of L is not set "high enough", or (2) they simply "waste" CPU time trying to find a Z-pair when no such Z-pair exists. The function `LowerBound`, on the other hand, first determines the applicability of $\mathcal{Z}$ to $T(n, k)$. If the existence of a Z-pair is guaranteed, then it computes a lower bound μ for the order of L, and if requested, calls $\mathcal{Z}$ using μ as the starting value for the order of L, instead of 0. Since the existence of a Z-pair is guaranteed, there is no need to set an upper bound for the order of L.

Example 3 Consider the hypergeometric term

$$T(n, k) = \frac{1}{(2k-1)(n-8k+1)} \binom{2n-2k}{n-k} \binom{2k}{k}.$$

We first apply `LowerBound` to T. The optional arguments are provided so that the minimal Z-pair can be computed. The time and space required are recorded [1].

```
> T := binomial(2*n-2*k,n-k)*binomial(2*k,k)/
>        ((2*k-1)*(n-8*k+1)):
> t1 := time(): b1 := kernelopts(bytesused):
> LowerBound(T,n,k,En,'Zpair');
```

$$8$$

```
> printf(`time taken: %a seconds, memory used: %a bytes\n`,
          time()-t1, kernelopts(bytesused)-b1);
```

time taken: 30.740 seconds, memory used: 124111132 bytes

[1] All the reported timings were obtained on a 400Mhz SUN SPARC SOLARIS with 1Gb RAM.

In this example the computed lower bound equals the order of the minimal telescoper L for T, and the function `Zeilberger` is called using this lower bound as the starting value for the order of L. We now apply `Zeilberger` directly to T.

```
> Zeilberger(T,n,k,En):
```

Error, (in Zeilberger) No recurrence of order 6 was found

The function `Zeilberger` tries to compute the minimal $\mathcal{Z}$-pair (L, G) for T starting with the value of 0 for the order ρ of L. It reaches the default value for the upper bound for ρ, and returns the above inconclusive error message. If one sets the upper bound to a "high enough" value, then `Zeilberger` will succeed in computing the minimal $\mathcal{Z}$-pair.

```
> t1 := time(): b1 := kernelopts(bytesused):
> _MAXORDER := 8:
> Zeilberger(T,n,k,En):
> printf(`time taken: %a seconds, memory used: %a bytes\n`,
          time()-t1, kernelopts(bytesused)-b1);
```

time taken: 45.260 seconds, memory used: 174678848 bytes

Example 4 Consider the hypergeometric term

$$T(n,k) = \frac{1}{nk+1}\binom{2\,n}{2\,k}.$$

It takes `LowerBound` 0.62 seconds and 3,045 kilobytes to return the error message "Error, (in LowerBound) Zeilberger's algorithm is not applicable". The function recognizes that the polynomial $v_2(n,k)$ in (4) is $(nk+1)$ which does not factor into a product of integer-linear polynomials, and returns the conclusive answer quickly. On the other hand, it takes `Zeilberger` 33.95 seconds and 166,396 kilobytes to return the error message "Error, (in Zeilberger) No recurrence of order 6 was found". The function does not know if a $\mathcal{Z}$-pair (L, G) for T exists. It tries to compute one and returns an inconclusive answer. Since there does not exist a $\mathcal{Z}$-pair for T, the higher the value of the upper bound for the order of L is set, the more time and memory are wasted.

In this example T is not a proper term, and $\mathcal{Z}$ is not applicable to T.

Example 5 Consider the hypergeometric term

$$T(n,k) = \frac{(n+k+2)!}{(n^2+k+2)(k+3)!} - \frac{(n+k+1)!}{(n^2+k+1)(2+k)!} + \frac{(n+k)!}{(n+7\,k-2)\,k!}.$$

```
> T := 1/(n^2+2+k)*(n+2+k)!/(3+k)!-1/(n^2+k+1)*(n+1+k)!/(2+k)!+
>       1/(n+7*k-2)*(n+k)!/k!:
```

We first compute an RNF of the k-certificate of T :

```
> IsHypergeometricTerm(T,k,'Rk'):
> (z,f1,f2,v1,v2) := RationalCanonicalForm[1](Rk,k):
> v2;
```

$$\left(\frac{1}{7}\,n + k - \frac{2}{7}\right)\left(n^2+k+2\right)\left(n^2+k+1\right)$$

Note that the polynomial v_2 has irreducible factors that are not integer-linear. We now apply `LowerBound` to T. The optional arguments are provided so that the computation of the minimal $\mathcal{Z}$-pair is carried out. `infolevel` is used to show the main steps of the function.

```
> t1 := time(): b1 := kernelopts(bytesused):
> LowerBound(T,n,k,En,'Zpair');
```

LowerBound: "check for the applicability of Zeilberger's algorithm"
LowerBound: "Zeilberger's algorithm is applicable"

```
LowerBound: “apply Theorem 3 to compute a lower bound”
LowerBound: “v2 = (n+7*k-2)”
LowerBound: “the candidate set for p is {n+7*k-2}”
LowerBound: “p = n+7*k-2”
LowerBound: “find the minimal positive integer r and integer h such that
        E_k^h p divides E_n v_2 . E_n^2 v_2 ... E_n^r v_2”
LowerBound: “find the minimal positive integer r and integer h such that
        E_k^h p divides d_2. E_n d_2 ...E_n^{r-1} d_2”
                                7
> printf(`time taken: %a seconds, memory used: %a bytes\n`,
        time()-t1, kernelopts(bytesused)-b1);
  time taken: 18.070 seconds, memory used: 69908612 bytes
```

It is shown above that the additive decomposition of T yields T_2 in (13) where the polynomial v_2 in (4) is integer-linear $(n + 7\,k - 2)$. Finally, we apply `Zeilberger` directly to T. Note the difference in the time and space required to complete each function.

```
> t1 := time(): b1 := kernelopts(bytesused):
> _MAXORDER := 7:
> Zeilberger(T,n,k,En):
> printf(`time taken: %a seconds, memory used: %a bytes\n`,
        time()-t1, kernelopts(bytesused)-b1);
  time taken: 3294.680 seconds, memory used: 5936292572 bytes
```

In this example T is not a proper term. However, $\mathcal{Z}$ is applicable to T.

Example 6 For a given hypergeometric term $T(n, k)$, instead of applying $\mathcal{Z}$ to T, we suggest that $\mathcal{Z}$ be applied to the minimal k-non-summable hypergeometric term $T_2(n, k)$ in the decomposition (13). Following Proposition 2, the required Z-pair for $T(n, k)$ can then be easily obtained from the computed Z-pair for $T_2(n, k)$. This in general helps reduce the size of the problem to be solved. As an example, for $b \in \mathbb{N} \setminus \{0\}$, $j \in \{1, 3\}$, let

$$T_1(n, k) = \frac{1}{(nk - 1)(n - bk - 2)^j (2\,n + k + 3)!},$$

$$T_2(n, k) = \frac{1}{(n - bk - 2)(2\,n + k + 3)!}.$$

Consider the hypergeometric term

$$T(n, k) = (E_k - 1)\,T_1(n, k) + T_2(n, k).$$

Since $T_1 \sim T_2$, T is a hypergeometric term. We apply the functions `Zeilberger` (`Z`) and `LowerBound` (`LB`) to T. `LowerBound` is called with the optional arguments so that the minimal Z-pair for T can be computed (it follows from Example 2 that the computed lower bound is $|b|$.) Table 1 shows the time and space requirement. As one can easily notice, as b and/or j increase, the relative performance of `Zeilberger` (compared to `LowerBound`) quickly worsens.

Acknowledgements

The authors wish to express their thanks to K.O. Geddes of the University of Waterloo for his support.

TABLE 1. Time and space requirements for `LowerBound` and `Zeilberger`.

		Timing (seconds)		Memory (kilobytes)	
j	b	LB	Z	LB	Z
1	1	6.49	5.35	27,838	24,702
	2	8.34	34.64	33,066	142,889
	3	11.13	124.53	44,233	535,736
	4	14.46	570.02	56,410	1,882,730
	5	25.79	2999.22	97,506	6,536,309
3	1	14.64	16.40	62,566	73,830
	2	17.24	228.59	68,304	770,529
	3	20.15	1,286.51	78,701	3,074,051
	4	24.08	8,771.08	91,844	10,766,646
	5	38.60	77,663.68	139,823	33,423,168

REFERENCES

[1] S.A. Abramov. Applicability of Zeilberger's Algorithm to Hypergeometric Terms. Submitted to ISSAC'2002.

[2] S.A. Abramov, M. Petkovšek. Minimal Decomposition of Indefinite Hypergeometric Sums. *Proc. ISSAC'2001*, 2001, 7–14.

[3] S.A. Abramov, M. Petkovšek. Rational Normal Forms and Minimal Decompositions of Hypergeometric Terms. Submitted to *J. Symb. Comput.*

[4] S.A. Abramov and M. Petkovšek. Canonical representations of hypergeometric terms. In *Proc. FPSAC'2001*, 2001, 1–10.

[5] S.A. Abramov, M. Petkovšek. Proof of a Conjecture of Wilf and Zeilberger. Preprint series, vol. 39, 2001, 748, University of Ljubljana, ISSN 1318–4865.

[6] S.A. Abramov, H.Q. Le. Applicability of Zeilberger's algorithm to rational functions. *Proc. FPSAC'2000*, Springer–Verlag LNCS, 2000, 91–102.

[7] S.A. Abramov, H.Q. Le. A criterion for the applicability of Zeilberger's algorithm to rational functions. To appear in *Discrete Math.*

[8] H.Q. Le. A Direct Algorithm to Construct Zeilberger's Recurrences for Rational Functions. *Proc. FPSAC'2001*, 2001, 303–312.

[9] H.Q. Le, S.A. Abramov, K.O. Geddes. HypergeometricSum: A Maple Package for Finding Closed Forms of Indefinite and Definite Sums of Hypergeometric Type. *Technical Report CS-2001-24*, Department of Computer Science, University of Waterloo, Ontario, Canada.

[10] Y.K. Man, F.J. Wright. Fast polynomial dispersion computation and its application to indefinite summation. J. von zur Gathen, M. Giesbrecht, Eds., *Proc. ISSAC '94*, ACM Press 1994, 175–180.

[11] M.B. Monagan, K.O. Geddes, K.M. Heal, G. Labahn, S.M. Vorkoetter, J. McCarron, P. DeMarco. *Maple 7 Programming Guide,* Toronto: Waterloo Maple Inc., 2001.

[12] P. Paule, M. Schorn. A Mathematica version of Zeilberger's algorithm for proving binomial coefficient identities. *J. Symb. Comput.* **20**, 1995, 673–698.

[13] M. Petkovšek, H. S. Wilf and D. Zeilberger. $A = B$. A K Peters, 1996.

[14] P. Verbaeten. The automatic construction of pure recurrence relations. *Proc. EUROSAM '74, ACM-SIGSAM Bulletin* **8**, 1974, 96–98.

[15] H. S. Wilf and D. Zeilberger. An algorithmic proof theory for hypergeometric (ordinary and "q") multisum/integral identities. *Invent. Math.* **108** (1992) 575–633.

[16] D. Zeilberger. The method of creative telescoping. *J. Symb. Comput.* **11**, 1991, 195–204.

Russian Academy of Science, Dorodnicyn Computing Centre, Vavilova st. 40, 119991, Moscow, GSP-1, Russia
E-mail address: abramov@ccas.ru

Symbolic Computation Group, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada
E-mail address: hqle@scg.math.uwaterloo.ca

DESCENT REPRESENTATIONS AND MULTIVARIATE STATISTICS (EXTENDED ABSTRACT)

RON M. ADIN, FRANCESCO BRENTI, AND YUVAL ROICHMAN

Abstract. Combinatorial identities on Weyl groups of types A and B are derived from special bases of the corresponding coinvariant algebras. Using the Garsia-Stanton descent basis of the coinvariant algebra of type A we give a new construction of the Solomon descent representations. An extension of the descent basis to type B, using new multivariate statistics on the group, yields a refinement of the descent representations. These constructions are then applied to refine well-known decomposition rules of the coinvariant algebra and to generalize various identities.

Résumé. Nous démontrons certaines identités combinatoires sur les groupes de Weyl de type A et B à partir de bases spéciales des algèbres coinvariantes associées. En utilisant la base de Garsia-Stanton de l'algèbre coinvariante de type A, nous donnons une nouvelle construction des représentations de descentes de Solomon. Une extension de la base des descentes de type B, utilisant de nouvelles statistiques multivariées sur le groupe, mène à un raffinement des représentations des descentes. Nous appliquons ensuite ces constructions pour raffiner des règles de décomposition bien connues de l'algèbre coinvariante, et pour généraliser diverses identités.

1. **Introduction**

This paper studies the interplay between representations of classical Weyl groups of types A and B and combinatorial identities on these groups. New combinatorial statistics on these groups are introduced, which lead to a new construction of representations. The Hilbert series which emerge give rise to multivariate identities generalizing known ones.

The set of elements in a Coxeter group having a fixed descent set carries a natural representation of the group, called a descent representation. Descent representations of Weyl groups were first introduced by Solomon [23] as alternating sums of permutation representations. This concept was extended to arbitrary Coxeter groups, using a different construction, by Kazhdan and Lusztig [20] [19, §7.15].

For Weyl groups of type A, these representations also appear in the top homology of certain (Cohen-Macaulay) rank-selected posets [25]. Another description (for type A) is by means of zig-zag diagrams [17]. In [2] we give a new construction of descent representations for Weyl groups of type A, using the coinvariant algebra as a representation space. This viewpoint gives rise to a new extension for type B, which refines the one by Solomon.

The construction of a basis for the coinvariant algebra is important for many applications, and has been approached from different viewpoints. A geometric approach identifies the coinvariant algebra with the cohomology ring $H^*(G/B)$ of the flag variety. This leads to the Schubert basis, and applies to any Weyl group. This identification also appears in Springer's construction of irreducible representations. Barcelo [7] found bases for the resulting quotients. An algebraic approach, applying Young symmetrizers, was used by

Research of all authors was supported in part by the Israel Science Foundation, founded by the Israel Academy of Sciences and Humanities, and by internal research grants from Bar-Ilan University.

Ariki, Terasoma and Yamada [6] to produce a basis compatible with the decomposition into irreducible representations.

A combinatorial approach, which produces a basis of monomials, was presented by Garsia and Stanton in [15] [13]. They actually presented a basis for a finite dimensional quotient of the Stanley-Reisner ring arising from a finite Weyl group. For type A, unlike other types, this quotient is isomorphic to the coinvariant algebra. The Garsia-Stanton descent basis for type A may be constructed from the coinvariant algebra via a straightening algorithm [5]. Using a reformulation of this algorithm we give a natural construction of Solomon's descent representations as factors of the coinvariant algebra of type A.

An analogue of the descent basis for type B is now given. This analogue (again consisting of monomials) involves extended descent sets and new combinatorial statistics. An extension of the construction of descent representations, using the new basis for type B, gives rise to a family of descent representations, refining Solomon's. A decomposition of these descent representations into irreducibles, refining theorems of Kraskiewicz-Weyman (for type A) [22, Theorem 8.8] and Stembridge (for type B) [30], is carried out using a multivariate version of Stanley's formula for the principal specialization of Schur functions.

This algebraic setting is then applied to obtain new multivariate combinatorial identities. Suitable Hilbert series are computed and compared to each other and to generating functions of multivariate statistics. The resulting identities present a far reaching generalization of bivariate identities from [16], [14], and [1].

2. Preliminaries

2.1. **Notations.** Let $\boldsymbol{P} := \{1, 2, 3, \ldots\}$, $\boldsymbol{N} := \boldsymbol{P} \cup \{0\}$, $\boldsymbol{Z}$ be the ring of integers, and $\mathbf{Q}$ be the field of rational numbers; for $a \in \boldsymbol{N}$ let $[a] := \{1, 2, \ldots, a\}$ (where $[0] := \emptyset$). Given $n, m \in \boldsymbol{Z}$, $n \leq m$, let $[n, m] := \{n, n+1, \ldots, m\}$. The cardinality of a set A will be denoted by $|A|$.

Given a variable q and a commutative ring R, denote by $R[q]$ (respectively, $R[[q]]$) the ring of polynomials (respectively, formal power series) in q with coefficients in R. For $i \in \boldsymbol{N}$ let, as customary, $[i]_q := 1 + q + q^2 + \ldots + q^{i-1}$ (so $[0]_q = 0$).

2.2. **Sequences and Permutations.** Let Σ be a linearly ordered alphabet. Given a sequence $\sigma = (\sigma_1, \ldots, \sigma_n) \in \Sigma^n$ we say that a pair $(i, j) \in [n] \times [n]$ is an *inversion* of σ if $i < j$ and $\sigma_i > \sigma_j$. We say that $i \in [n-1]$ is a *descent* of σ if $\sigma_i > \sigma_{i+1}$;

$$Des(\sigma) := \{1 \leq i \leq n-1 \mid \sigma_i > \sigma_{i+1}\}$$

is the *descent set* of σ. Denote by $inv(\sigma)$ (respectively, $des(\sigma)$) the number of inversions (respectively, descents) of σ. We also let

$$maj(\sigma) := \sum_{i \in Des(\sigma)} i$$

and call it the *major index* of σ.

Given a set T let $S(T)$ be the set of all bijections $\pi : T \to T$, and $S_n := S([n])$. For $\pi \in S_n$ write $\pi = \pi_1 \ldots \pi_n$ to mean that $\pi(i) = \pi_i$, for $i = 1, \ldots, n$. Any $\pi \in S_n$ may also be written in *disjoint cycle form* (see, e.g., [26, p.17]), usually omitting the 1-cycles of π. For example, $\pi = 365492187$ may also be written as $\pi = (9, 7, 1, 3, 5)(2, 6)$. Given $\pi, \tau \in S_n$ let $\pi\tau := \pi \circ \tau$ (composition of functions) so that, for example, $(1, 2)(2, 3) = (1, 2, 3)$.

Denote by B_n the group of all bijections σ of the set $[-n, n] \setminus \{0\}$ onto itself such that

$$\sigma(-a) = -\sigma(a)$$

for all $a \in [-n,n] \setminus \{0\}$, with composition as the group operation. This group is usually known as the group of "signed permutations" on $[n]$, or as the *hyperoctahedral group* of rank n. We identify S_n as a subgroup of B_n, and B_n as a subgroup of S_{2n}, in the natural ways.

For $\sigma \in B_n$ write $\sigma = [a_1, \ldots, a_n]$ to mean that $\sigma(i) = a_i$ for $i = 1, \ldots, n$, and (using the natural linear order on $[-n,n] \setminus \{0\}$) let

$$inv(\sigma) := inv(a_1, \ldots, a_n), \qquad des(\sigma) := des(a_1, \ldots, a_n),$$
$$maj(\sigma) := maj(a_1, \ldots, a_n), \qquad Neg(\sigma) := \{i \in [n] : \ a_i < 0\},$$
$$neg(\sigma) := |Neg(\sigma)|, \qquad fmaj(\sigma) := 2 \cdot maj(\sigma) + neg(\sigma).$$

The statistic *fmaj* was introduced in [3] [4] and further studied in [1].

2.3. **Partitions and Tableaux.** Let n be a nonnegative integer. A *partition* of n is an infinite sequence of nonnegative integers with finitely many nonzero terms $\lambda = (\lambda_1, \lambda_2, \ldots)$, where $\lambda_1 \geq \lambda_2 \geq \ldots$ and $\sum\limits_{i=1}^{\infty} \lambda_i = n$. The sum $\sum \lambda_i = n$ is called the *size* of λ, denoted $|\lambda|$; write also $\lambda \vdash n$. The number of parts of λ, $\ell(\lambda)$, is the maximal j for which $\lambda_j > 0$.

The *dominance* partial order on partitions is defined as follows : For any two partitions μ and λ of the same integer, μ *dominates* λ (denoted $\mu \trianglerighteq \lambda$) if and only if $\sum_{j=1}^{i} \mu_j \geq \sum_{j=1}^{i} \lambda_j$ for all i (and, by assumption, $\sum_{j=1}^{\infty} \mu_j = \sum_{j=1}^{\infty} \lambda_j$).

The subset $\{(i,j) \mid i,j \in \mathbf{P}, \ j \leq \lambda_i\}$ of $\mathbf{P}^2$ is called the *Young diagram* of *shape* λ. (i,j) is the *cell* in row i and column j. A *Young tableau* of shape λ is obtained by inserting the integers $1, 2, \ldots, n$ (where $n = |\lambda|$) as *entries* in the cells of the Young diagram of shape λ, allowing no repetitions. A *standard Young tableau* of shape λ is a Young tableau whose entries increase along rows and columns.

A *descent* in a standard Young tableau T is an entry i such that $i+1$ is strictly south (and hence weakly west) of i. Denote the set of all descents in T by $Des(T)$. The *descent number* and the *major index* (for tableaux) are defined as follows :

$$des(T) := \sum_{i \in Des(T)} 1 \ ; \qquad maj(T) := \sum_{i \in Des(T)} i.$$

A *bipartition* of n is a pair (λ^1, λ^2) of partitions of total size $|\lambda^1| + |\lambda^2| = n$. A skew diagram of shape (λ^1, λ^2) is a disjoint union of a diagram of shape λ^1 and a diagram of shape λ^2, where the second diagram lies southwest of the first. A standard Young tableau $T = (T^1, T^2)$ of shape (λ^1, λ^2) is obtained by inserting the integers $1, 2, \ldots, n$ as entries in the cells, such that the entries increase along rows and columns. The descent set $Des(T)$, the descent number $des(T)$, and the major index $maj(T)$ of T are defined as above. The *negative set*, $Neg(T)$, of such a tableau T is the set of entries in the cells of λ^2. Define $neg(T) := |Neg(T)|$ and

$$fmaj(T) := 2 \cdot maj(T) + neg(T).$$

Example 1. Let T be

$$\begin{array}{ccccc} & & 2 & 5 & 6 \\ & & 3 & & \\ 1 & 7 & & & \\ 4 & 9 & & & \\ 8 & & & & \end{array}$$

T is a standard Young tableau of shape $((3,1),(2,2,1))$, $Des(T)=\{2,3,6,7\}$, $des(T)=4$, $maj(T)=18$, $Neg(T)=\{1,4,7,8,9\}$, $neg(T)=5$ and $fmaj(T)=41$.

Denote by $SYT(\lambda)$ the set of all standard Young tableaux of shape λ, and by $SYT(\lambda^1,\lambda^2)$ the set of all standard Young tableaux of shape (λ^1,λ^2).

2.4. **The Coinvariant Algebra.** The groups S_n and B_n have natural actions on the ring of polynomials P_n (cf. [19, §3.1]). S_n acts by permuting the variables, and B_n acts by permuting the variables and multiplying by ± 1. The ring of S_n-invariant polynomials is Λ_n, the ring of symmetric functions in $x_1,\ldots,x_n$. Similarly, the ring of B_n-invariant polynomials is Λ_n^B, the ring of symmetric functions in $x_1^2,\ldots,x_n^2$. Let I_n, I_n^B be the ideals of P_n generated by the elements of Λ_n, Λ_n^B (respectively) without constant term. The quotient P_n/I_n (P_n/I_n^B) is called the *coinvariant algebra* of S_n (B_n). Each group acts naturally on its coinvariant algebra. The resulting representation is isomorphic to the regular representation. See, e.g., [19, §3.6] and [18, §II.3].

Let R_k ($0\le k\le \binom{n}{2}$) be the k-th homogeneous component of the coinvariant algebra of S_n: $P_n/I_n=\oplus_k R_k$. Each R_k is an S_n-module. The following theorem is attributed by Reutenauer to Kraskiewicz and Weyman [22, p. 215].

Kraskiewicz-Weyman Theorem. [22, Theorem 8.8] *For any $0\le k\le\binom{n}{2}$ and $\mu\vdash n$, the multiplicity in R_k of the irreducible S_n-representation corresponding to μ is*

$$m_{k,\mu}=|\ \{\ T\in SYT(\mu)\ \mid\ maj(T)=k\ \}\ |.$$

The following B-analogue (in different terminology) was proved in [30]. Here R_k^B is the k-th homogeneous component of the coinvariant algebra of B_n.

Stembridge's Theorem. *For any $0\le k\le n^2$ and bipartition (μ^1,μ^2) of n, the multiplicity in R_k^B of the irreducible B_n-representation corresponding to (μ^1,μ^2) is*

$$m_{k,\mu^1,\mu^2}=|\ \{\ T\in SYT(\mu^1,\mu^2)\mid fmaj(T)=k\ \}\ |.$$

3. Main Results

For any partition λ with (at most) n parts, let $P_\lambda^{\trianglelefteq}$ be the subspace of the polynomial ring $P_n=\mathbf{Q}[x_1,\ldots,x_n]$ spanned by all monomials whose exponent partition is dominated by λ, and let R_λ be a distinguished quotient of the image of $P_\lambda^{\trianglelefteq}$ under the projection of P_n onto the coinvariant algebra. We will show that the homogeneous components of the coinvariant algebra decompose as direct sums of certain R_λ-s. This will be done using an explicit construction of a basis for R_λ. The construction of this basis involves combinatorial statistics.

3.1. **New Statistics.** Let Σ be a linearly ordered alphabet. For any finite sequence $\sigma=\sigma_1,\sigma_2,\ldots,\sigma_n$ of letters in Σ define

$$Des(\sigma):=\{i\mid \sigma_i>\sigma_{i+1}\},$$

the descent set of σ, and

$$d_i(\sigma):=|\{j\in Des(\sigma):\ j\ge i\}|,$$

the number of descents in σ from position i on.

If Σ consists of integers, let

$$Neg(\sigma):=\{i\mid \sigma_i<0\};$$

$$n_i(\sigma):=|\{j\in Neg(\sigma):\ j\ge i\}|;\qquad \varepsilon_i(\sigma):=\begin{cases}1, & \text{if } \sigma_i<0,\\ 0, & \text{otherwise;}\end{cases}$$

and

$$f_i(\sigma) := 2d_i(\sigma) + \varepsilon_i(\sigma).$$

The statistics $f_i(\sigma)$ refine the flag-major index $fmaj(\sigma)$, which was introduced and studied in [3] [4] [1].

3.2. **The Garsia-Stanton Descent Basis and its Extension.** To any $\pi \in S_n$ Garsia and Stanton [15] associated the monomial

$$a_\pi := \prod_{i \in Des(\pi)} (x_{\pi(1)} \cdots x_{\pi(i)}).$$

It should be noted that in our notation $a_\pi = \prod_{i=1}^{n} x_{\pi(i)}^{d_i(\pi)}$. Using Stanley-Reisner rings Garsia and Stanton showed that the set $\{a_\pi + I_n \mid \pi \in S_n\}$ forms a basis for the coinvariant algebra of type A [15]. This basis will be called the *descent basis*. The Garsia-Stanton approach is not applicable to the coinvariant algebras of other Weyl groups. In [2] we extend the descent basis to the Weyl groups of type B.

To any $\sigma \in B_n$ we associate the monomial

$$b_\sigma := \prod_{i=1}^{n} x_{|\sigma(i)|}^{f_i(\sigma)}.$$

Theorem 3.1. *The set*

$$\{b_\sigma + I_n^B \mid \sigma \in B_n\}$$

forms a basis for the coinvariant algebra of type B.

3.3. **Descent Representations.** For a monomial m in the polynomial ring $P_n = \mathbf{Q}[x_1, \ldots, x_n]$, let the exponent partition $\lambda(m)$ be the partition obtained by rearranging the exponents in a weakly decreasing order. For any partition λ with at most n parts, let $P_\lambda^{\trianglelefteq}$ be the subspace of P_n spanned by all monomials whose exponent partition is dominated by λ :

$$P_\lambda^{\trianglelefteq} := span_{\mathbf{Q}}\{m \mid \lambda(m) \trianglelefteq \lambda\}.$$

Similarly, define $P_\lambda^{\triangleleft}$ by strict dominance :

$$P_\lambda^{\triangleleft} := span_{\mathbf{Q}}\{m \mid \lambda(m) \triangleleft \lambda\}.$$

Consider now the canonical projection of P_n onto the coinvariant algebra

$$\psi : P_n \longrightarrow P_n/I_n.$$

Define R_λ to be a quotient of images under this map :

$$R_\lambda := \psi(P_\lambda^{\trianglelefteq})/\psi(P_\lambda^{\triangleleft}).$$

Then R_λ is an S_n-module.

Lemma 3.2. *$R_\lambda \neq 0$ if and only if $\lambda = \lambda(a_\pi)$ for some $\pi \in S_n$.*

For any subset $S \subseteq \{1, \ldots, n\}$ define a partition

$$\lambda_S := (\lambda_1, \ldots, \lambda_n)$$

by

$$\lambda_i := |\, S \cap \{i, \ldots, n\}\,|.$$

Using a straightening algorithm for the descent basis it is shown that $R_\lambda \neq 0$ if and only if $\lambda = \lambda_S$ for some $S \subseteq [n-1]$, and that a basis for R_{λ_S} may be indexed by the permutations with descent set equal to S. Let R_k be the k-th homogeneous component of the coinvariant algebra P_n/I_n.

Theorem 3.3. *For every* $0 \le k \le \binom{n}{2}$,

$$R_k \cong \bigoplus_S R_{\lambda_S}$$

as S_n*-modules, where the sum is over all subsets* $S \subseteq [n-1]$ *such that* $\sum_{i \in S} i = k$.

Let

$$R_\lambda^B := \psi^B(P_\lambda^{\unlhd}) \;/\; \psi^B(P_\lambda^{\lhd}),$$

where $\psi^B : P_n \longrightarrow P_n/I_n^B$ is the canonical map from P_n onto the coinvariant algebra of type B.

Lemma 3.4. $R_\lambda^B \neq 0$ *if and only if* $\lambda = \lambda(b_\sigma)$ *for some* $\sigma \in B_n$.

For subsets $S_1 \subseteq [n-1]$ and $S_2 \subseteq [n]$, let λ_{S_1,S_2} be the vector

$$\lambda_{S_1,S_2} := 2\lambda_{S_1} + \mathbf{1}_{S_2},$$

where λ_{S_1} is as above and $\mathbf{1}_{S_2} \in \{0,1\}^n$ is the characteristic vector of S_2. This is not always a partition. Indeed, for a partition λ, $R_\lambda^B \neq 0$ if and only if $\lambda = \lambda_{S_1,S_2}$ for some $S_1 \subseteq [n-1], S_2 \subseteq [n]$. In this case, a basis for $R_{\lambda_{S_1,S_2}}$ may be indexed by the signed permutations $\sigma \in B_n$ with $Des(\sigma) = S_1$ and $Neg(\sigma) = S_2$.

Let R_k^B for the k-th homogeneous component of the coinvariant algebra of type B. The following theorem is a B-analogue of Theorem 3.3.

Theorem 3.5. *For every* $0 \le k \le n^2$,

$$R_k^B \cong \bigoplus_{S_1,S_2} R_{\lambda_{S_1,S_2}}$$

as B_n*-modules, where the sum is over all subsets* $S_1 \subseteq [n-1]$ *and* $S_2 \subseteq [n]$ *such that* λ_{S_1,S_2} *is a partition and*

$$2 \cdot \sum_{i \in S_1} i + |S_2| = k.$$

3.4. **Decomposition into Irreducibles.**

Theorem 3.6. *For any subset* $S \subseteq [n-1]$ *and partition* $\mu \vdash n$*, the multiplicity in* R_{λ_S} *of the irreducible* S_n*-representation corresponding to* μ *is*

$$m_{S,\mu} := |\{\, T \in SYT(\mu) \mid Des(T) = S \,\}|,$$

the number of standard Young tableaux of shape μ *and descent set* S.

This theorem implies the well known Kraskiewicz-Weyman theorem for decomposing the homogeneous components of the coinvariant algebra into irreducibles. See Subsection 2.4.

For type B we have

Theorem 3.7. *For any pair of subsets* $S_1 \subseteq [n-1]$, $S_2 \subseteq [n]$*, and a bipartition* (μ^1, μ^2) *of* n*, the multiplicity of the irreducible* B_n*-representation corresponding to* (μ^1, μ^2) *in* $R_{\lambda_{S_1,S_2}}$ *is*

$$m_{S_1,S_2,\mu^1,\mu^2} := |\{\, T \in SYT(\mu^1, \mu^2) \mid Des(T) = S_1 \text{ and } Neg(T) = S_2 \,\}|,$$

the number of pairs of standard Young tableaux of shapes μ^1 *and* μ^2 *with descent set* S_1 *and sets of entries* $[n] \setminus S_2$ *and* S_2*, respectively.*

The proofs apply multivariate extensions of Stanley's formula for the principal specialization of a Schur function [27, Prop. 7.19.11] to obtain the graded character.

3.5. **Combinatorial Identities.** For any partition $\lambda = (\lambda_1, \ldots, \lambda_n)$ with at most n parts define

$$m_j(\lambda) := |\ \{1 \le i \le n \mid \lambda_i = j\}\ | \qquad (\forall j \ge 0).$$

By considering Hilbert series of the polynomial ring with respect to rearranged multi-degree and applying the Straightening Lemma for the coinvariant algebra of type A we obtain

Theorem 3.8. *For any positive integer n*

$$\sum_{\ell(\lambda)\le n} \binom{n}{m_0(\lambda), m_1(\lambda), \ldots} \prod_{i=1}^{n} q_i^{\lambda_i} = \frac{\sum_{\pi\in S_n} \prod_{i=1}^{n} q_i^{d_i(\pi)}}{\prod_{i=1}^{n}(1 - q_1\cdots q_i)}$$

in $\boldsymbol{Z}[[q_1, \ldots, q_n]]$, where the sum on the left-hand side is taken over all partitions with at most n parts.

Note that the multinomial coefficient in the theorem is the number of monomials with exponent partition λ. Taking $q_1 = qt$ and $q_2 = \ldots = q_n = q$ yields the following well known result (attributed by Garsia [14] to Gessel [16]; see also [10]).

Corollary 3.9. *Let $n \in \boldsymbol{P}$. Then*

$$\frac{\sum_{\pi\in S_n} t^{des(\pi)} q^{maj(\pi)}}{\prod_{i=0}^{n}(1 - tq^i)} = \sum_{r\ge 0} [r+1]_q^n t^r.$$

in $\boldsymbol{Z}[q][[t]]$.

The Hilbert series of P_n rearranged by multi-degree may be computed in a different way, by considering the signed descent basis for the coinvariant algebra of type B and applying the Straightening Lemma for this type.

Theorem 3.10. *With notations as in Theorem 3.8*

$$\sum_{\ell(\lambda)\le n} \binom{n}{m_0(\lambda), m_1(\lambda), \ldots} \prod_{i=1}^{n} q_i^{\lambda_i} = \frac{\sum_{\sigma\in B_n} \prod_{i=1}^{n} q_i^{f_i(\sigma)}}{\prod_{i=1}^{n}(1 - q_1^2\cdots q_i^2)}$$

in $\boldsymbol{Z}[[q_1, \ldots, q_n]]$, where the sum on the left-hand side runs through all partitions with at most n parts.

The main combinatorial result for type B is a far reaching generalization of [1, Corollary 4.5].

Theorem 3.11. *For any positive integer n*

$$\sum_{\sigma\in B_n} \prod_{i=1}^{n} q_i^{d_i(\sigma)+n_i(\sigma^{-1})} = \sum_{\sigma\in B_n} \prod_{i=1}^{n} q_i^{2d_i(\sigma)+\varepsilon_i(\sigma)}.$$

For further identities see [2, Section 6].

Acknowledgments. The authors are indebted to Dominique Foata for posing motivating problems and to Ira Gessel for the idea of using the coinvariant algebra in the study of multivariate statistics.

References

[1] R. M. Adin, F. Brenti and Y. Roichman, *Descent numbers and major indices for the hyperoctahedral group*, Adv. Appl. Math. 27 (2001), 210–224.
[2] R. M. Adin, F. Brenti and Y. Roichman, *Descent representations and multivariate statistics*, preprint.
[3] R. M. Adin and Y. Roichman, *A flag major index for signed permutations*, Proc. 11-th Conference on Formal Power Series and Algebraic Combinatorics, Universitat Politècnica de Catalunya, Barcelona, 1999, 10–17.
[4] R. M. Adin and Y. Roichman, *The flag major index and group actions on polynomial rings*, Europ. J. Combin. 22 (2001), 431–446.
[5] E. E. Allen, *The descent monomials and a basis for the diagonally symmetric polynomials*, J. Alg. Combin. 3 (1994), 5–16.
[6] S. Ariki, T. Terasoma and H. F. Yamada, *Higher Specht polynomials*, Hiroshima Math. J. 27 (1997), 177–188.
[7] H. Barcelo, *Young straightening in a quotient S_n-module*, J. Alg. Combin. 2 (1993), 5–23.
[8] A. Björner and F. Brenti, *Combinatorics of Coxeter Groups*, Graduate Texts in Mathematics, Springer-Verlag, 2001, to appear.
[9] F. Brenti, *q-Eulerian polynomials arising from Coxeter groups*, Europ. J. Combin. 15 (1994), 417-441.
[10] L. Carlitz, *A combinatorial prperty of q-Eulerian numbers*, Amer. Math. Monthly 82 (1975), 51–54.
[11] D. Foata, personal communication, July 2000.
[12] D. Foata and G. N. Han, *Calcul basique des permutations signees. I. Longueur et nombre d'inversions*, Adv. Appl. Math. 18 (1997), 489-509.
[13] A. M. Garsia, *Combinatorial methods in the theory of Cohen-Macaulay rings*, Adv. Math. 38 (1980), 229–266.
[14] A. M. Garsia, *On the "maj" and "inv" q-analogues of Eulerian polynomials*, Linear and Multilinear Algebra 8 (1979/80), 21-34.
[15] A. M. Garsia and D. Stanton, *Group actions of Stanley-Reisner rings and invariants of permutation groups*, Adv. Math. 51 (1984), 107–201.
[16] I. M. Gessel, *Generating functions and enumeration of sequences*, Ph.D. Thesis, M.I.T., 1977.
[17] I. M. Gessel, *Multipartite P-partitions and inner products of Schur functions*, Contemp. Math. 34 (1984), 289–302.
[18] H. L. Hiller, *Geometry of Coxeter Groups*, Res. Notes in Math. 54, Pitman, Boston, 1982.
[19] J. E. Humphreys, *Reflection Groups and Coxeter Groups*, Cambridge Studies in Advanced Mathematics, no.29, Cambridge Univ. Press, Cambridge, 1990.
[20] D. Kazhdan and G. Lusztig, *Representations of Coxeter groups and Hecke algebras*, Invent. Math. 53 (1979), 165-184.
[21] V. Reiner, *Signed permutation statistics*, Europ. J. Combin. 14 (1993), 553-567.
[22] C. Reutenauer, *Free Lie Algebras*, London Math. Soc. Monographs, New Series 7, Oxford Univ. Press, 1993.
[23] L. Solomon, *The orders of the finite Chevalley groups*, J. Algebra 3 (1966), 376-393.
[24] R. P. Stanley, *Binomial posets, Möbius inversion, and permutation enumeration*, J. Combin. Theory Ser. A 20 (1976), 336-356.
[25] R. P. Stanley, *Some aspects of group acting on finite posets*, J. Combin. Theory Ser. A 32 (1982), 132–161.
[26] R. P. Stanley, *Enumerative Combinatorics*, Vol. 1, Wadsworth and Brooks/Cole, Monterey, CA, 1986.
[27] R. P. Stanley, *Enumerative Combinatorics*, Vol. 2, Cambridge Studies in Advanced Mathematics 62, Cambridge Univ. Press, Cambridge, 1999.
[28] R. Steinberg, *On a theorem of Pittie*, Topology 14 (1975), 173–177.
[29] E. Steingrimsson, *Permutation statistics of indexed permutations*, Europ. J. Combin. 15 (1994), 187-205.
[30] J. Stembridge, *On the eigenvalues of representations of reflection groups and wreath products*, Pacific J. Math. 140 (1989), 359–396.

Department of Mathematics and Computer Science, Bar-Ilan University, Ramat Gan 52900, Israel.
E-mail address: radin@math.biu.ac.il

Dipartimento di Matematica, Universitá di Roma "Tor Vergata", Via della Ricerca Scientifica, 00133 Roma, Italy.
E-mail address: brenti@mat.uniroma2.it

Department of Mathematics and Computer Science, Bar-Ilan University, Ramat Gan 52900, Israel.
E-mail address: yuvalr@math.biu.ac.il

LITTLEWOOD-RICHARDSON COEFFICIENTS AND HOOK INTERPOLATIONS (EXTENDED ABSTRACT)

RON M. ADIN, AVITAL FRUMKIN, AND YUVAL ROICHMAN

Abstract. The hook components of $V^{\otimes n}$ interpolate between the symmetric power $\mathrm{Sym}^n(V)$ and the exterior power $\wedge^n(V)$. When V is the vector space of $k \times m$ matrices, a decomposition of the hook components into irreducibles involving convolutions of Littlewood-Richardson coefficients is presented. Classical theorems of Ehresmann, Thrall, Helgason, James, Shimura and others are proved as boundary cases.

Résumé. Les composants d'équerres de $V^{\otimes den}$ interpolent entre la puissance symétrique $\mathrm{Sym}^n(v)$ et la puissance extérieure $\wedge^n(v)$. Quand V est l'espace vectoriel des matrices $k \times m$, une décomposition des composantes d'équerres en composantes irréductibles comprenant des convolutions de coefficients de Littlewood-Richardson est présentée. Des théorèmes classiques d'Ehresmann, de Thrall, de Helgason, de James, de Shimura et de d'autres sont prouvés comme des cas limites.

1. Introduction

The vector space $M_{k,m}$ of $k \times m$ matrices over $\mathbb{C}$ carries a (left) $GL_k(\mathbb{C})$-action and a (right) $GL_m(\mathbb{C})$-action. A classical Theorem of Ehresmann [3] describes the decomposition of an exterior power of $M_{k,m}$ into irreducible bimodules. The symmetric analogue was given later (cf. [7]). See Section 4 below.

In this paper we present a natural interpolation between these theorems, in terms of hook components of the n-th tensor power of $M_{k,m}$. This interpolation involves convolutions of the Littlewood-Richardson coefficients. Duality and asymptotics of the decomposition of hook components follow.

Similar concepts are applied to the diagonal two-sided $GL_k(\mathbb{C})$-action on the vector space of $k \times k$ matrices. Classical theorems of Thrall [19] and James [8] (for the symmetric powers of symmetric matrices), and of Helgason [5], Shimura [15] and Howe [6] (for the symmetric powers of anti-symmetric matrices) are extended, and a bivariate interpolation is presented. This interpolation involves natural extensions of the Littlewood-Richardson coefficients.

Proofs are obtained using the representation theory of the symmetric and hyperoctahedral groups, together with plethysm of symmetric functions and Schur-Weyl duality. The techniques are different in spirit from those used in the classical works cited above, except for [8].

The interpolations presented have surprising combinatorial implications, which will be studied elsewhere.

2. Definitions and Notations

Let n be a positive integer. A *partition* of n is a vector of positive integers $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ and $\lambda_1 + \ldots + \lambda_k = n$. We denote this by

Research was supported in part by the Israel Science Foundation, founded by the Israel Academy of Sciences and Humanities, and by internal research grants from Bar-Ilan University.

$\lambda \vdash n$. The *size* of a partition $\lambda \vdash n$, denoted $|\lambda|$, is n, and its *length*, $\ell(\lambda)$, is the number of parts. The empty partition $\emptyset$ has size and length zero: $|\emptyset| = \ell(\emptyset) = 0$. The set of all partitions of n with at most k parts is denoted by $\mathrm{Par}_k(n)$.

For a partition $\lambda = (\lambda_1, \ldots, \lambda_k)$ define the *conjugate partition* $\lambda' = (\lambda'_1, \ldots, \lambda'_t)$ by letting λ'_i be the number of parts of λ that have size at least i.

A partition $\lambda = (\lambda_1, \ldots, \lambda_k)$ may be viewed as the subset

$$\{(i, j) \mid 1 \le i \le k, 1 \le j \le \lambda_i\} \subseteq \mathbb{Z}^2,$$

the corresponding *Young diagram*. Using this interpretation, we may speak of inclusion $\mu \subseteq \lambda$, intersection $\lambda \cap \mu$ and the set difference $\lambda \setminus \mu$ of any two partitions. The set difference is called a *skew shape*; when $\mu \subseteq \lambda$ it is usually denoted λ/μ.

A *semistandard Young tableau* of shape λ/μ is obtained by inserting positive integers as entries in the cells of the Young diagram of shape λ, so that the entries weakly increase along rows and strictly increase down columns. The *content vector* of a semistandard Young tableau T $cont(T) = (m_1, m_2, \ldots)$ is defined by $m_i := |\{\text{cells in } T \text{ with entry } i\}|$ for all $i \ge 0$.

We shall also use the Frobenius notation for partitions, defined as follows: Let λ be a partition of n and set $d := \max\{i \mid \lambda_i - i \ge 0\}$ (i.e., the length of the main diagonal in the Young diagram of λ). Then the Frobenius notation for λ is $(\lambda_1 - 1, \ldots, \lambda_d - d \mid \lambda'_1 - 1, \ldots, \lambda'_d - d)$.

For any partition $\lambda = (\lambda_1, \ldots, \lambda_k)$ of n define the following doubling operation

$$2 \cdot \lambda := (2\lambda_1, \ldots, 2\lambda_k) \vdash 2n.$$

If all the parts of λ are distinct, define also

$$2 * \lambda := (\lambda_1, \ldots \lambda_k \mid \lambda_1 - 1, \ldots \lambda_k - 1) \vdash 2n$$

in the Frobenius notation.

3. The Littlewood-Richardson Coefficients

Let $\bar{a} = (a_1, a_2, \ldots, a_n)$ be a sequence of positive integers. $\bar{a}$ is called a *reverse ballot sequence* if for every $1 \le i < n$ and $1 \le j \le n$ the number of occurences of i in the prefix $(a_1, \ldots, a_j)$ is not less than the number of occurences of $i+1$ in $(a_1, \ldots, a_j)$.

A semistandard Young tableau of shape λ/μ is *proper* if, when reading its entries from right to left, starting in the topmost row and going down, we obtain a reverse ballot sequence.

The Littlewood-Richardson coefficient $c^{\lambda}_{\mu\nu}$ is the number of proper semistandard Young tableaux of shape λ/μ and content vector ν.

The irreducible S_n-modules (Specht modules) will be denoted by S^λ, and the irreducible $GL_k(\mathbb{C})$-modules (Weyl modules) by V^λ_k. The Littlewood-Richardson coefficients describe the decomposition of tensor products of Weyl modules. Let $\mu \vdash t$ and $\nu \vdash n - t$. Then

$$V^\mu_k \otimes V^\nu_k \cong \bigoplus_{\lambda \vdash n} c^{\lambda}_{\mu\nu} V^\lambda_k,$$

for $k \ge \max\{\ell(\lambda), \ell(\mu), \ell(\nu)\}$ (and the coefficients $c^{\lambda}_{\mu\nu}$ are then independent of k).

By Schur-Weyl duality they are also the coefficients of the outer product of Specht modules. Namely,

$$(S^\mu \otimes S^\nu) \uparrow^{S_n}_{S_t \times S_{n-t}} \cong \bigoplus_{\lambda \vdash n} c^{\lambda}_{\mu\nu} S^\lambda.$$

Let λ and μ be two partitions of the same integer n, and let $0 \leq i \leq n$. Define

$$c^{\lambda\mu}(i) := \sum_{\alpha \vdash n-i,\ \beta \vdash i} c^{\lambda}_{\alpha\beta} c^{\mu}_{\alpha\beta'} \ .$$

Thus $c^{\lambda\mu}(i)$ is the number of pairs of proper semistandard Young tableaux of shapes λ/α, μ/α respectively (where α is some partition of $n-i$) with conjugate content vectors.
Example.

$$c^{\lambda\mu}(0) = \delta_{\lambda\mu} \quad , \qquad c^{\lambda\mu}(n) = \delta_{\lambda\mu'} \ . \tag{3.1}$$

We shall use also the following notation for *extended Littlewood-Richardson coefficients*:

$$c^{\lambda}_{\alpha\beta\gamma\delta} := \sum_{\mu,\nu} c^{\lambda}_{\alpha\mu} c^{\mu}_{\beta\nu} c^{\nu}_{\gamma\delta};$$

so that

$$V^{\alpha}_k \otimes V^{\beta}_k \otimes V^{\gamma}_k \otimes V^{\delta}_k = \bigoplus_{\lambda} c^{\lambda}_{\alpha\beta\gamma\delta} V^{\lambda}_k.$$

4. Symmetric and Exterior Powers of Matrix Spaces

In this section we cite well-known classical theorems, concerning the decomposition into irreducibles of symmetric and exterior powers of matrix spaces, which are to be generalized in this paper.

Let $M_{k,m}$ be the vector space of $k \times m$ matrices over $\mathbb{C}$. Then $M_{k,m}$ carries a (left) $GL_k(\mathbb{C})$-action and a (right) $GL_m(\mathbb{C})$-action. A classical Theorem of Ehresmann [3] (see also [11]) describes the decomposition of an exterior power of $M_{k,m}$ into irreducible $GL_k(\mathbb{C}) \times GL_m(\mathbb{C})$-modules.

Theorem 4.1. *The n-th exterior power of $M_{k,m}$ is isomorphic, as a $GL_k(\mathbb{C}) \times GL_m(\mathbb{C})$-module, to*

$$\wedge^n(M_{k,m}) \cong \bigoplus_{\lambda \vdash n \ \text{and}\ \lambda \subseteq (m^k)} V^{\lambda}_k \otimes V^{\lambda'}_m,$$

where λ' is the partition conjugate to λ.

The following three results on symmetric powers were proved several times independently; these results may be found in [7] and [4].

The symmetric analogue of Theorem 4.1 was studied, for example, in [7, (11.1.1)] and [4, Theorem 5.2.7].

Theorem 4.2. *The n-th symmetric power of $M_{k,m}$ is isomorphic, as a $GL_k(\mathbb{C}) \times GL_m(\mathbb{C})$-module, to*

$$Sym^n(M_{k,m}) \cong \bigoplus_{\lambda \vdash n \ \text{and}\ \ell(\lambda) \leq \min(k,m)} V^{\lambda}_k \otimes V^{\lambda}_m.$$

Let $M^+_{k,k}$ be the vector space of symmetric $k \times k$ matrices over $\mathbb{C}$. This space carries a natural two sided $GL_k(\mathbb{C})$-action. The following theorem describes the decomposition of its symmetric powers into irreducible $GL_k(\mathbb{C})$-modules.

Theorem 4.3. *The n-th symmetric power of $M^+_{k,k}$ is isomorphic, as a $GL_k(\mathbb{C})$-module, to*

$$Sym^n(M^+_{k,k}) \cong \bigoplus_{\lambda \in Par_k(n)} V^{2\cdot\lambda}_k.$$

This theorem was proved by A.T. James [8], but had already appeared in an early work of Thrall [19]. See also [6], [15], [7, (11.2.2)] and [4, Theorem 5.2.9] for further proofs and references.

Let $M^-_{k,k}$ be the vector space of skew symmetric $k \times k$ matrices over $\mathbb{C}$. Then

Theorem 4.4. *The n-th symmetric power of $M^-_{k,k}$ is isomorphic, as a $GL_k(\mathbb{C})$-module, to*

$$Sym^n(M^-_{k,k}) \cong \bigoplus_{(2\cdot\lambda)' \in Par_k(2n)} V_k^{(2\cdot\lambda)'}.$$

This theorem was proved in [5], [6], [15]. See also [7, (11.3.2)] and [4, Theorem 5.2.11].

5. Main Results

Let $M_{k,m}$ be the vector space of $k \times m$ matrices over $\mathbb{C}$. The tensor power $M_{k,m}^{\otimes n}$ carries a natural S_n-action by permuting the factors. This action decomposes the tensor power into irreducible S_n-modules. Let $M_{k,m}^{\otimes n}(i)$ be the isotypic component of $M_{k,m}^{\otimes n}$ corresponding to the irreducible S_n-representation indexed by the hook $(n-i, 1^i)$, where $0 \le i \le n-1$. This component still carries a $GL_k(\mathbb{C}) \times GL_m(\mathbb{C})$-action. Its decomposition into irreducibles is given by a convolution of the Littlewood-Richardson coefficients.

Theorem 5.1. *Let λ and μ be partitions of n, of lengths at most k and m, respectively. For every $0 \le i \le n$ the multiplicity of the irreducible $GL_k(\mathbb{C}) \times GL_m(\mathbb{C})$-module $V_k^\lambda \otimes V_m^\mu$ in $M_{k,m}^{\otimes n}(i-1) \oplus M_{k,m}^{\otimes n}(i)$ is the restricted convolution $c^{\lambda\mu}(i)$, as defined in Section 3 above. By convention, $M_{k,m}^{\otimes n}(-1) = M_{k,m}^{\otimes n}(n) = 0$.*

Theorem 5.1 interpolates between two well-known classical theorems, Theorems 4.1 and 4.2. Indeed, $M_{k,m}^{\otimes n}(0) \cong \mathrm{Sym}^n(M_{k,m})$ and $M_{k,m}^{\otimes n}(-1) = 0$. Substituting $i = 0$ and applying (3.1) shows that the relevant multiplicity is $\delta_{\lambda\mu}$, thus proving Theorem 4.2. Similarly, the substitution $i = n$ gives Theorem 4.1.

The following corollary generalizes the duality between Theorem 4.1 and Theorem 4.2.

Corollary 5.2. *Let $\mu \subseteq (m^m)$ and λ be partitions of n. For every $0 \le i \le n-1$ the multiplicity of $V_k^\lambda \otimes V_m^\mu$ in $M_{k,m}^{\otimes n}(i)$ is equal to the multiplicity of $V_k^\lambda \otimes V_m^{\mu'}$ in $M_{k,m}^{\otimes n}(n-1-i)$.*

Let λ and μ be partitions of n. Define the *distance*

$$d(\lambda, \mu) := \frac{1}{2} \sum_i |\lambda_i - \mu_i|.$$

Theorem 5.1 together with results of Regev [14, Theorem 12] and Dvir [2, Theorem 1.6] imply

Theorem 5.3. *If $V_k^\lambda \otimes V_m^\mu$ appears as a factor in $M_{k,m}^{\otimes n}(t)$ (for some $0 \le t \le n-1$) then*

$$d(\lambda, \mu) < km.$$

This shows that, for $V_k^\lambda \otimes V_m^\mu$ to appear in a hook component, λ and μ must be very "close" to each other (for k and m fixed, n tending to infinity).

Consider now the vector space $M_{k,k}$ of $k \times k$ square matrices over $\mathbb{C}$. Let $M_{k,k}^{\otimes n}(i,j)$ be the component of $M_{k,k}^{\otimes n}(i)$ consisting of tensors with j skew symmetric and $n-j$ symmetric factors. $M_{k,k}^{\otimes n}(i,j)$ carries a $GL_k(\mathbb{C})$ two-sided diagonal action. The following theorem describes its decomposition as a $GL_k(\mathbb{C})$-module.

Theorem 5.4. *Let λ be a partition of $2n$ of length at most k. For every $0 \le i \le n$ and $0 \le j \le n$ the multiplicity of V_k^λ in $M_{k,k}^{\otimes n}(i,j) \oplus M_{k,k}^{\otimes n}(i-1,j)$ is*

$$\sum_{|\alpha|+|\beta|+|\gamma|+|\delta|=n,\ |\beta|+|\delta|=j\ ,\ |\gamma|+|\delta|=i} c^{\lambda}_{2\cdot\alpha,(2\cdot\beta)',2*\gamma,(2*\delta)'},$$

where the sum runs over all partitions $\alpha, \beta, \gamma, \delta$ with total size n such that γ and δ have distinct parts, β and δ have total size j, and γ and δ have total size i. The operations $*$ and $\cdot$ are as defined in Section 2, and the extended Littlewood-Richardson coefficients $c^\lambda_{\alpha\beta\gamma\delta}$ are as defined in Section 3.

The proof of Theorem 5.4 involves results on plethysm of elementary and homogeneous symmetric functions [13, Ch. I §8 Ex 5-6].

Theorem 5.4, for $i = 0$, interpolates between classical results, regarding symmetric powers of the spaces of symmetric and skew symmetric matrices (Theorems 4.3 and 4.4). Another boundary case, $i = n$, gives an interpolation between exterior powers of the same matrix spaces.

Corollary 5.5. *Let $\lambda \subseteq (k^k)$ be a partition of $2n$. For every $0 \le i \le n-1$ and $0 \le j \le n$, the multiplicity of V_k^λ in $M_{k,k}^{\otimes n}(i,j)$ is equal to the multiplicity of $V_k^{\lambda'}$ in $M_{k,k}^{\otimes n}(i,n-j)$.*

For proofs and more details see [1].

Acknowledgments. The authors thank R. Howe and N. Wallach for their useful comments.

References

[1] R. M. Adin, A. Frumkin and Y. Roichman, *Hook Interpolations.* J. Algebra, to appear.
[2] Y. Dvir, *On the Kronecker product of S_n characters.* J. Algebra 154 (1993), 125–140.
[3] C. Ehresmann, *Sur la topologie de certains espaces homogènes.* Ann. of Math. 35 (1934), 396–443.
[4] R. Goodman and N. R. Wallach, *Representations and Invariants of the Classical Groups.* Encyclopedia of Math. and its Appl. Vol. 68, Cambridge University Press, 1998.
[5] S. Helgason, *A duality for symmetric spaces with applications.* Adv. Math. 5 (1970), 1–54.
[6] R. Howe, *Remarks on classical invariant theory.* Trans. Amer. Math. Soc. 313 (1989), 539–570.
[7] R. Howe and T. Umeda, *The Capelli identity, the double commutant theorem, and multiplicity-free actions.* Math. Ann. 290 (1991), 565–619.
[8] A. T. James, *Zonal polynomials of the real positive definite matrices.* Annals of Math. 74 (1961), 456–469.
[9] G. D. James and A. Kerber, *The Representation Theory of the Symmetric Group.* Encyclopedia of Math. and its Appl. Vol. 16, Addison-Wesley, 1981.
[10] A. Kerber, *Algebraic Combinatorics via Finite Group Actions.* Bibliographisches Institut, Mannheim, 1991.
[11] B. Kostant, *Lie algebra cohomology and the generalized Borel Weil theorem.* Ann. of Math. 74 (1961), 329–387.
[12] J. P. Serre, *Linear Representations of Finite Groups.* Springer-Verlag, 1977.
[13] I. G. Macdonald, *Symmetric Functions and Hall Polynomials.* second edition, Oxford Math. Monographs, Oxford Univ. Press, Oxford, 1995.
[14] A. Regev, *The Kronecker product of S_n characters and an $A \otimes B$ theorem for Capelli identities.* J. Algebra 66 (1980), 505–510.
[15] G. Shimura, *On differential operators attached to certain representations of classical groups.* Invent. Math. 77 (1984), 463–488.
[16] R. P. Stanley, *Enumerative Combinatorics, Volume II.* Cambridge Univ. Press, Cambridge, 1999.
[17] J. R. Stembridge, *On Schur Q-functions and the primitive idempotents of commutative Hecke algebra.* J. Alg. Combin. 1 (1992), 71–95.
[18] J. Stembridge, *On the eigenvalues of representations of reflection groups and wreath products.* Pacific J. Math. 140 (1989), 359–396.

[19] R. Thrall, *On symmetrized Kronecker powers and the structure of the free Lie ring.* Amer. J. Math. 64 (1942), 371–388.

Department of Mathematics and Computer Science, Bar-Ilan University, Ramat-Gan 52900, Israel

E-mail address: radin@math.biu.ac.il

Sackler School of Mathematical Sciences, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel

E-mail address: frumkin@math.tau.ac.il

Department of Mathematics and Computer Science, Bar-Ilan University, Ramat-Gan 52900, Israel

E-mail address: yuvalr@math.biu.ac.il

STRUCTURE OF THE MALVENUTO-REUTENAUER HOPF ALGEBRA OF PERMUTATIONS (EXTENDED ABSTRACT)

MARCELO AGUIAR

Abstract. We analyze the structure of the Malvenuto-Reutenauer Hopf algebra of permutations in detail. We give explicit formulas for its antipode, prove that it is a cofree coalgebra, determine its primitive elements and its coradical filtration and show that it decomposes as a crossed product over the Hopf algebra of quasi-symmetric functions. We also describe the structure constants of the multiplication as a certain number of facets of the permutahedron. Our results reveal a close relationship between the structure of this Hopf algebra and the weak order on the symmetric groups.

Résumé. On analyse la structure de l'algèbre de Hopf de Malvenuto et Reutenauer en détail. On donne des formules explicites pour son antipode, on prouve que c'est une coalgèbre colibre, on détermine ses éléments primitifs et sa filtration coradicale et on montre qu'elle se décompose comme un produit croisé sur l'algèbre de Hopf de fonctions quasi-symétriques. On décrit aussi les constantes de structure de la multiplication comme un certain nombre de facettes du permutoèdre. Nos résultats mettent en évidence une forte relation entre la structure de cette algèbre de Hopf et l'ordre faible dans les groupes symétriques.

Introduction

Malvenuto [15] introduced the Hopf algebra $\mathfrak{S}Sym$ of permutations, which has a linear basis $\{\mathcal{F}_u \mid u \in \mathfrak{S}_n, n \geq 0\}$ indexed by permutations in all symmetric groups $\mathfrak{S}_n$. The Hopf algebra $\mathfrak{S}Sym$ is non-commutative, non-cocommutative, self-dual, and graded. Among its sub- and quotient- Hopf algebras are many algebras central to algebraic combinatorics. These include the algebra of symmetric functions [14, 22], Gessel's algebra $\mathcal{Q}Sym$ of quasi-symmetric functions [11], the algebra of non-commutative symmetric functions [10], the Loday-Ronco algebra of planar binary trees [12], Stembridge's algebra of peaks [23], the Billera-Liu algebra of Eulerian enumeration [4], and others. The structure of these combinatorial Hopf algebras with respect to certain distinguished bases has been an important theme in algebraic combinatorics, with applications to the combinatorial problems these algebras were created to study. We give a detailed understanding of the structure of $\mathfrak{S}Sym$, both in algebraic and in combinatorial terms.

Our main tool is a new basis $\{\mathcal{M}_u \mid u \in \mathfrak{S}_n,\ n \geq 0\}$ for $\mathfrak{S}Sym$ related to its original basis by Möbius inversion on the weak order of the symmetric groups. These bases $\{\mathcal{M}_u\}$ and $\{\mathcal{F}_u\}$ are analogous to the monomial and fundamental basis of $\mathcal{Q}Sym$, which are related via Möbius inversion on their index sets, the Boolean posets $\mathcal{Q}_n$.

We give enumerative-combinatorial descriptions of the product, coproduct, and antipode of $\mathfrak{S}Sym$ with respect to the basis $\{\mathcal{M}_u\}$. For example, the coproduct is obtained by

1991 *Mathematics Subject Classification.* Primary 05E05, 06A11, 16W30.

Key words and phrases. Hopf algebras, symmetric group, weak order, quasi-symmetric functions.

Extended abstract for FPSAC'02 in Melbourne. Abridged version of [2].

splitting a permutation at certain special positions that we call global descents. Descents and global descents are left adjoint and right adjoint to a natural map $\mathcal{Q}_n \to \mathfrak{S}_n$.

The structure constants for the product with respect to the basis $\{\mathcal{M}_u\}$ are non-negative integers with the following geometric-combinatorial description. The 1-skeleton of the permutahedron Π_{n-1} is the Hasse diagram of the weak order on $\mathfrak{S}_n$. The facets of the permutahedron are canonically isomorphic to products of lower dimensional permutahedra. Say that a facet isomorphic to $\Pi_{p-1} \times \Pi_{q-1}$ has type (p, q). Given $u \in \mathfrak{S}_p$ and $v \in \mathfrak{S}_q$, such a facet has a distinguished vertex corresponding to (u, v) under the canonical isomorphism. Then, for $w \in \mathfrak{S}_{p+q}$, the coefficient of $\mathcal{M}_w$ in the product $\mathcal{M}_u \cdot \mathcal{M}_v$ is the number of facets of the permutahedron Π_{p+q-1} of type (p, q) with the property that the distinguished vertex is below w and closer to w than to any other vertex in the facet.

We also give explicit formulas for the antipode with respect to both bases. The structure constants with respect to the basis $\{\mathcal{M}_u\}$ have constant sign, as in the case of $\mathcal{Q}Sym$. The situation is more complicated for the basis $\{\mathcal{F}_u\}$, which may explain why no such explicit formulas were previously known.

Elucidating the elementary structure of $\mathfrak{S}Sym$ with respect to the basis reveals further algebraic structures. For example, $\mathfrak{S}Sym$ is a cofree graded coalgebra. A consequence is that the coradical filtration of $\mathfrak{S}Sym$ (which encapsulates the complexity of iterated coproducts) is the algebraic counterpart of a filtration of the symmetric groups by certain lower order ideals. In particular, the space of primitive elements is spanned by the set $\{\mathcal{M}_u \mid u \text{ has no global descents}\}$. Cofreenes was shown by Poirier and Reutenauer [20], in dual form, through the introduction of a different basis. The study of primitive elements was pursued from this point of view by Duchamp, Hivert, and Thibon [7].

There is a well-known morphism of Hopf algebras $\mathfrak{S}Sym \to \mathcal{Q}Sym$ that maps one fundamental basis onto the other, by associating to a permutation u its descent set $\mathrm{Des}(u)$. We describe this map in terms of the bases $\{\mathcal{M}_u\}$ and $\{M_\alpha\}$.

Lastly, $\mathfrak{S}Sym$ decomposes as a crossed product over $\mathcal{Q}Sym$. This construction from the theory of Hopf algebras is a generalization of the notion of group extensions. We provide a combinatorial description for the Hopf kernel of the map $\mathfrak{S}Sym \to \mathcal{Q}Sym$.

These results are expanded on and proven in the manuscript [2] of the same name. For a background on quasi-symmetric functions, see [22, §7.19], for Hopf Algebras, we recommend the book of Montgomery [18]. We also recommend the papers [20] of Poirier and Reutenauer and [7] of Duchamp, Thibon, and Hivert, who studied this same Hopf algebra of permutations from a different perspective, the latter under the name 'free quasi-symmetric functions'.

We thank Swapneel Mahajan, Nantel Bergeron, and the referees for helpful comments.

1. Essential definitions

1.1. Quasi-symmetric functions. Gessel [11] introduced the algebra $\mathcal{Q}Sym$ of quasi-symmetric functions as the natural target for Stanley's P-partition generating function. Subsequent work has shown its centrality, even universality, for generating functions in algebraic combinatorics [9, 3, 1].

A sequence $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ of positive integers is a *composition of* n if $\sum_i \alpha_i = n$. Compositions of n correspond to subsets of $[n-1]$ as follows

$$\alpha \;=\; (\alpha_1, \alpha_2, \dots, \alpha_k) \quad\longleftrightarrow\quad I(\alpha) \;=\; \{\alpha_1, \alpha_1 + \alpha_2, \dots, \alpha_1 + \cdots + \alpha_{k-1}\}\ .$$

These subsets of $[n-1]$ (and thus compositions of n) form the Boolean poset $\mathcal{Q}_n$, and the induced order relation on compositions is called *refinement*.

For $\mathsf{S} \subseteq [n-1]$, the *fundamental quasi-symmetric* function $F_{\mathsf{S},n}$ is

$$F_{\mathsf{S},n} \;:=\; \sum_{\substack{j_1 \le \cdots \le j_n \\ i \in \mathsf{S} \Rightarrow j_i < j_{i+1}}} x_{j_1} x_{j_2} \dots x_{j_n} \;.$$

These form a basis for $\mathcal{Q}Sym$. Another basis is provided by the monomial quasi-symmetric functions M_α, which are indexed by compositions $\alpha = (\alpha_1, \dots, \alpha_k)$

$$M_\alpha \;:=\; \sum_{i_1 < i_2 < \cdots < i_k} x_{i_1}^{\alpha_1} x_{i_2}^{\alpha_2} \cdots x_{i_k}^{\alpha_k} \;.$$

These two bases are related via Möbius inversion on the Boolean poset $\mathcal{Q}_n$.

$$F_\alpha \;=\; \sum_{\alpha \le \beta} M_\beta \qquad \text{and} \qquad M_\alpha \;=\; \sum_{\alpha \le \beta} (-1)^{c(\beta)-c(\alpha)} F_\beta \,,$$

where $c((\alpha_1, \dots, \alpha_k)) = k$ and $(-1)^{c(\beta)-c(\alpha)}$ is the Möbius function of $\mathcal{Q}_n$.

The product of these M_α is given by quasi-shuffles of their indices [9, Lemma 3.3]. A *quasi-shuffle* of compositions α and β is a shuffle of the components of α and β, where in addition we may replace any number of pairs of consecutive components (α_i, β_j) in the shuffle by $\alpha_i + \beta_j$. Then we have

$$M_\alpha \cdot M_\beta \;=\; \sum_\gamma M_\gamma \,, \tag{1.1}$$

where the sum is over all quasi-shuffles γ of the compositions α and β. For instance,

$$M_{(2)} \cdot M_{(1,1)} \;=\; M_{(1,1,2)} + M_{(1,2,1)} + M_{(2,1,1)} + M_{(1,3)} + M_{(3,1)} \,. \tag{1.2}$$

The unit element $1 = M_{(\,)}$ is indexed by the empty composition.

Let X and Y be ordered alphabets with $X < Y$ their disjoint union ordered as indicated. Substitution $f(X) \mapsto f(X < Y)$ induces a coproduct $\Delta\colon \mathcal{Q}Sym \to \mathcal{Q}Sym \otimes \mathcal{Q}Sym$ whose action on a monomial function is as follows.

$$\Delta\big(M_{(\alpha_1,\dots,\alpha_k)}\big) \;=\; \sum_{p=0}^{k} M_{(\alpha_1,\dots,\alpha_p)} \otimes M_{(\alpha_{p+1},\dots,\alpha_k)} \,. \tag{1.3}$$

For instance, $\Delta(M_{(2,1)}) = 1 \otimes M_{(2,1)} + M_{(2)} \otimes M_{(1)} + M_{(2,1)} \otimes 1$.

These definitions give $\mathcal{Q}Sym$ the structure of a graded, connected Hopf algebra. The degree n component is spanned by those M_α where α is a composition of n. It is connected, as its degree 0 component is 1-dimensional, and it is a Hopf algebra. An explicit formula for the antipode was given by Malvenuto [15, corollaire 4.20] and Ehrenborg [9, Proposition 3.4]

$$S(M_\alpha) \;=\; (-1)^{c(\alpha)} \sum_{\beta \le \alpha} M_{\widetilde{\beta}} \,. \tag{1.4}$$

Here, if $\beta = (\beta_1, \beta_2, \dots, \beta_t)$ then $\widetilde{\beta}$ is β written in reverse order $(\beta_t, \dots, \beta_2, \beta_1)$.

1.2. **The Hopf algebra of permutations.** Let $\mathfrak{S}Sym$ be the graded $\mathbb{Q}$-vector space with *fundamental basis* $\{\mathcal{F}_u \mid u \in \mathfrak{S}_n, n \ge 0\}$, graded by n. $\mathfrak{S}Sym$ has a graded Hopf algebra structure first considered in Malvenuto's thesis [15, §5.2] and in her work with Reutenauer [16]. Write 1 for the basis element of degree 0.

The product of two basis elements is obtained by shuffling the corresponding permutations, as in the following example.

$$\begin{aligned}\mathcal{F}_{12}\cdot\mathcal{F}_{312} \;=\; & \mathcal{F}_{12534}+\mathcal{F}_{15234}+\mathcal{F}_{15324}+\mathcal{F}_{15342}+\mathcal{F}_{51234}\\ & +\mathcal{F}_{51324}+\mathcal{F}_{51342}+\mathcal{F}_{53124}+\mathcal{F}_{53142}+\mathcal{F}_{53412}\,.\end{aligned}$$

More precisely, for $p,q>0$, set

$$\mathfrak{S}^{(p,q)} \;:=\; \{\zeta\in\mathfrak{S}_{p+q} \mid \zeta \text{ has at most one descent, at position } p\}\,.$$

This is the collection of minimal (in length) representatives of left cosets of the Young or parabolic subgroup $\mathfrak{S}_p\times\mathfrak{S}_q$ in $\mathfrak{S}_{p+q}$, called *Grassmannian permutations*. With these definitions, we describe the product. For $u\in\mathfrak{S}_p$ and $v\in\mathfrak{S}_q$, set

$$\mathcal{F}_u\cdot\mathcal{F}_v \;=\; \sum_{\zeta\in\mathfrak{S}^{(p,q)}}\mathcal{F}_{(u\times v)\cdot\zeta^{-1}}\,. \tag{1.5}$$

This endows $\mathfrak{S}Sym$ with the structure of a graded algebra with unit 1.

The algebra $\mathfrak{S}Sym$ is also a graded coalgebra with coproduct given by all ways of splitting a permutation. For a sequence $(a_1,\ldots,a_p)$ of distinct integers, let its *standard permutation*[†] $\mathrm{st}(a_1,\ldots,a_p)\in\mathfrak{S}_p$ be the permutation u defined by

$$u_i<u_j \iff a_i<a_j\,. \tag{1.6}$$

For instance, $\mathrm{st}(625)=312$. The coproduct $\Delta\colon\mathfrak{S}Sym\to\mathfrak{S}Sym\otimes\mathfrak{S}Sym$ is defined by

$$\Delta(\mathcal{F}_u) \;=\; \sum_{p=0}^{n}\mathcal{F}_{\mathrm{st}(u_1,\ldots,u_p)}\otimes\mathcal{F}_{\mathrm{st}(u_{p+1},\ldots,u_n)}\,, \tag{1.7}$$

when $u\in\mathfrak{S}_n$. For instance, $\Delta(\mathcal{F}_{42531})$ is

$$1\otimes\mathcal{F}_{42531}+\mathcal{F}_1\otimes\mathcal{F}_{2431}+\mathcal{F}_{21}\otimes\mathcal{F}_{321}+\mathcal{F}_{213}\otimes\mathcal{F}_{21}+\mathcal{F}_{3142}\otimes\mathcal{F}_1+\mathcal{F}_{42531}\otimes 1\,.$$

$\mathfrak{S}Sym$ is a graded connected Hopf algebra [15, théorème 5.4].

This Hopf algebra $\mathfrak{S}Sym$ has been an object of recent interest [21, 16, 20, 19, 8, 7, 12, 13]. We remark that sometimes it is the dual Hopf algebra that is considered. To compare results, one may use that $\mathfrak{S}Sym$ is self-dual under the map $\mathcal{F}_u\mapsto\mathcal{F}^*_{u^{-1}}$, where $\mathcal{F}^*_{u^{-1}}$ is the element of the dual basis that is dual to $\mathcal{F}_{u^{-1}}$.

To define the *monomial* basis $\{\mathcal{M}_u\}$ for $\mathfrak{S}Sym$ (in analogy to the basis $\{M_\alpha\}$ of $\mathcal{Q}Sym$), we use the weak order on the symmetric groups $\mathfrak{S}_n$. Let $\ell(n)$ count the inversions $\{i<j \mid u_i>u_j\}$ of a permutation u. The *weak order* on $\mathfrak{S}_n$ is defined by

$$u\le v \iff \exists w\in\mathfrak{S}_n \text{ such that } v \;=\; wu \text{ and } \ell(v)=\ell(w)+\ell(u)\,.$$

The cover relation $u\lessdot v$ occurs precisely when v is obtained from u by transposing a pair of consecutive *values* of u; a pair (u_i,u_j) such that $i<j$ and $u_j=u_i+1$. The maximum element of $\mathfrak{S}_n$ is $\omega_n=(n,\ldots,2,1)$. Figure 1 shows the weak order on $\mathfrak{S}_4$.

For each $n\ge 0$ and $u\in\mathfrak{S}_n$, define

$$\mathcal{M}_u \;:=\; \sum_{u\le v}\mu_{\mathfrak{S}_n}(u,v)\cdot\mathcal{F}_v\,, \tag{1.8}$$

where $\mu_{\mathfrak{S}_n}(\cdot,\cdot)$ is the Möbius function of the weak order in $\mathfrak{S}_n$. By Möbius inversion,

$$\mathcal{F}_u \;:=\; \sum_{u\le v}\mathcal{M}_v\,, \tag{1.9}$$

[†]Some authors call this flattening.

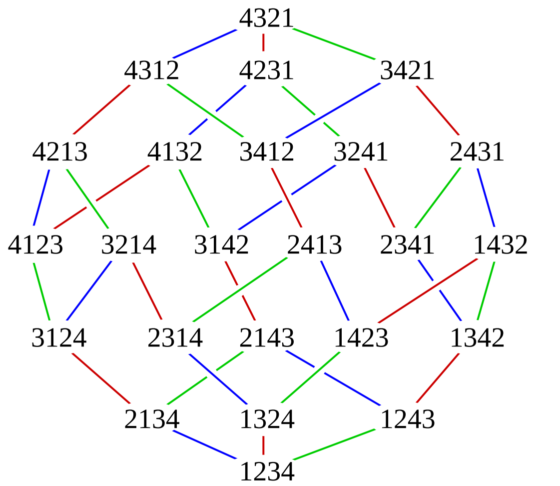


FIGURE 1. The weak order on $\mathfrak{S}_4$

so these elements $\mathcal{M}_u$ indeed form a basis of $\mathfrak{S}Sym$. For instance,

$$\mathcal{M}_{4123} \;=\; \mathcal{F}_{4123} - \mathcal{F}_{4132} - \mathcal{F}_{4213} + \mathcal{F}_{4321}\,.$$

1.3. **The descent map $\mathcal{D}\colon \mathfrak{S}Sym \to \mathcal{Q}Sym$.** The descent set $\mathrm{Des}(u)$ of a permutation $u \in \mathfrak{S}_n$ is the subset of $[n-1]$ recording the descents of u

$$\mathrm{Des}(u) \;:=\; \{p \in [n-1] \mid u_p > u_{p+1}\}\,. \tag{1.10}$$

Thus $\mathrm{Des}(46\,5128\,37) = \{2,3,6\}$. Malvenuto [15, théorèmes 5.12, 5.13, and 5.18] shows that there is a morphism of Hopf algebras

$$\begin{array}{rcccl} \mathcal{D} & : & \mathfrak{S}Sym & \longrightarrow & \mathcal{Q}Sym \\ & & \mathcal{F}_u & \longmapsto & F_{\mathrm{Des}(u)} \end{array} \tag{1.11}$$

1.4. **Galois connections $\mathfrak{S}_n \leftrightarrows \mathcal{Q}_n$.** Underlying our results are combinatorial facts concerning the lattices $\mathfrak{S}_n$ and $\mathcal{Q}_n$. We describe two important conceptual facts. For a subset $\mathsf{S} \subseteq [n-1]$, let $\mathfrak{S}_{\mathsf{S}} \subseteq \mathfrak{S}_n$ be the parabolic subgroup

$$\mathfrak{S}_{\alpha_1} \times \mathfrak{S}_{\alpha_2} \times \cdots \times \mathfrak{S}_{\alpha_k}\,,$$

where $\alpha = (\alpha_1, \ldots, \alpha_k)$ is the composition of n such that $I(\alpha) = \mathsf{S}$. For a subset $\mathsf{S} \subseteq [n-1]$, let $Z(\mathsf{S}) \in \mathfrak{S}_n$ be the maximal permutation with descent set S.

A *Galois connection* between posets P and Q is a pair (f, g) of order preserving maps $f\colon P \to Q$ and $g\colon Q \to P$ such that for any $x \in P$ and $y \in Q$,

$$f(x) \;\le\; y \quad\Longleftrightarrow\quad x \;\le\; g(y)\,. \tag{1.12}$$

Equivalently, the map f is left adjoint to the map g.

Proposition 1.1. *The pair of maps* $(\mathrm{Des}, Z) : \mathfrak{S}_n \rightleftarrows \mathcal{Q}_n$ *is a Galois connection.*

This Galois connection is why the monomial basis of $\mathfrak{S}Sym$ is analogous to that of $\mathcal{Q}Sym$, and is why we consider the weak order on $\mathfrak{S}_n$. The connection between the monomial bases of these two algebras will be elucidated in Theorem 3.4.

A permutation $u \in \mathfrak{S}_n$ has a *global descent* at a position $p \in [n-1]$ if

$$i \le p < j \implies u_i > u_j\,.$$

$$
\begin{tikzcd}[column sep=small, row sep=small]
 & 321 \arrow[dl, dash] \arrow[dr, dash] & & & & \{1,2\} \arrow[dl, dash] \arrow[dr, dash] & \\
312 \arrow[d, dash] & & 231 \arrow[d, dash] & & \{1\} \arrow[d, equal] & & \{2\} \arrow[d, equal] \\
213 & & 132 & \overset{\mathrm{Des}}{\longmapsto} & \{1\} & & \{2\} \\
 & 123 \arrow[ul, dash] \arrow[ur, dash] & & & & \emptyset \arrow[ul, dash] \arrow[ur, dash] &
\end{tikzcd}
\qquad
\begin{tikzcd}[column sep=small, row sep=small]
 & \{1,2\} \arrow[dl, dash] \arrow[dr, dash] & & & & 321 \arrow[dl, dash] \arrow[dr, dash] & \\
\{1\} & & \{2\} & \overset{Z}{\longmapsto} & 312 & & 231 \\
 & \emptyset \arrow[ul, dash] \arrow[ur, dash] & & & & 123 \arrow[ul, dash] \arrow[ur, dash] &
\end{tikzcd}
$$

FIGURE 2. The Galois connection $\mathfrak{S}_3 \rightleftarrows \mathcal{Q}_3$

Equivalently, if $\{u_1, \ldots, u_p\} = \{n, n-1, \ldots, n-p+1\}$. Let $\mathrm{GDes}(u) \subseteq [n-1]$ be the set of global descents of u. Note that $\mathrm{GDes}(u) \subseteq \mathrm{Des}(u)$, but these are not equal in general.

The notion of global descents is a very natural companion of that of (ordinary) descents, in that the map $\mathrm{GDes}\colon \mathfrak{S}_n \to \mathcal{Q}_n$ is *right* adjoint to $Z\colon \mathcal{Q}_n \to \mathfrak{S}_n$.

Proposition 1.2. *The pair of maps* $(Z, \mathrm{GDes}) : \mathcal{Q}_n \rightleftarrows \mathfrak{S}_n$ *is a Galois connection.*

2. Elementary Algebraic Structure of $\mathfrak{S}Sym$

2.1. **The coproduct of $\mathfrak{S}Sym$.** The coproduct of $\mathfrak{S}Sym$ (1.7) takes a simple form on the monomial basis. For a permutation $u \in \mathfrak{S}_n$, define $\overline{\mathrm{GDes}}(u)$ to be $\mathrm{GDes}(u) \cup \{0, n\}$.

Theorem 2.1. *Let* $u \in \mathfrak{S}_n$. *Then*

$$\Delta(\mathcal{M}_u) \;=\; \sum_{p \in \overline{\mathrm{GDes}}(u)} \mathcal{M}_{\mathrm{st}(u_1,\ldots,u_p)} \otimes \mathcal{M}_{\mathrm{st}(u_{p+1},\ldots,u_n)} \,. \tag{2.1}$$

2.2. **The product of $\mathfrak{S}Sym$.** The product of $\mathfrak{S}Sym$ in terms of its monomial basis has non-negative structure constants, which we describe. For instance,

$$\begin{aligned} \mathcal{M}_{12} \cdot \mathcal{M}_{21} \;=\; & \mathcal{M}_{4312} + \mathcal{M}_{4231} + \mathcal{M}_{3421} + \mathcal{M}_{4123} + \mathcal{M}_{2341} \\ & + \mathcal{M}_{1243} + \mathcal{M}_{1423} + \mathcal{M}_{1342} + 3\mathcal{M}_{1432} + 2\mathcal{M}_{2431} + 2\mathcal{M}_{4132} \,. \end{aligned} \tag{2.2}$$

First, for a Grassmannian permutation $\zeta \in \mathfrak{S}^{(p,q)}$ (a left coset representative of $\mathfrak{S}_p \times \mathfrak{S}_q$ in $\mathfrak{S}_{p+q}$), consider the map corresponding to the *right coset* of ζ^{-1}.

$$\rho_\zeta \;:\; \mathfrak{S}_p \times \mathfrak{S}_q \;\to\; \mathfrak{S}_{p+q}\,, \qquad \rho_\zeta(u,v) \;:=\; (u \times v) \cdot \zeta^{-1} \,.$$

This order-preserving map is injective and its image is an interval in $\mathfrak{S}_{p+q}$. For $u \in \mathfrak{S}_p$, $v \in \mathfrak{S}_q$ and $w \in \mathfrak{S}_{p+q}$, define $A_{u,v}^w \subseteq \mathfrak{S}^{(p,q)}$ to be

$$A_{u,v}^w \;=\; \left\{\zeta \in \mathfrak{S}^{(p,q)} \mid (u,v) = \max \rho_\zeta^{-1}[1,w]\right\}, \tag{2.3}$$

where $[w, w'] := \{w'' \mid w \le w'' \le w'\}$ denotes the interval between w and w'. This set has another description as the set of those $\zeta \in \mathfrak{S}^{(p,q)}$ satisfying

$$\begin{array}{ll} (i) & (u \times v) \cdot \zeta^{-1} \le w, \text{ and} \\ (ii) & \text{if } u \le u' \text{ and } v \le v' \text{ satisfy } (u' \times v') \cdot \zeta^{-1} \le w, \\ & \text{then } u = u' \text{ and } v = v'. \end{array} \tag{2.4}$$

Set $\alpha_{u,v}^w := \# A_{u,v}^w$.

Theorem 2.2. *For any* $u \in \mathfrak{S}_p$ *and* $v \in \mathfrak{S}_q$, *we have*

$$\mathcal{M}_u \cdot \mathcal{M}_v \;=\; \sum_{w \in \mathfrak{S}_{p+q}} \alpha_{u,v}^w \, \mathcal{M}_w \,. \tag{2.5}$$

For instance, in (2.2) the coefficient of $\mathcal{M}_{2431}$ in $\mathcal{M}_{12} \cdot \mathcal{M}_{21}$ is 2 because among the six permutations in $\mathfrak{S}^{(2,2)}$,

$$1234,\ 1324,\ 1423,\ 2314,\ 2413,\ 3412\,,$$

only the first two satisfy conditions (i) and (ii) of (2.4).

The structure constants $\alpha_{u,v}^{w}$ admit a geometric-combinatorial description in terms of the permutahedron. The vertices of the $(n-1)$-dimensional permutahedron are indexed by the elements of $\mathfrak{S}_n$ so that the 1-skeleton is the Hasse diagram of the weak order (see Figure 1). Facets of the permutahedron are products of two lower dimensional permutahedra, and the image of ρ_ζ is the set of vertices in a facet. Moreover, every facet arises in this way for a unique triple (p, q, ζ) with $p + q = n$ and $\zeta \in \mathfrak{S}^{(p,q)}$ (see [5, Exer. 2.9]). Such such a facet has *type* (p, q). Figure 3 shows the image of ρ_{1324}, a facet of the 3-permutahedron of type $(2, 2)$, and the permutation 2431.

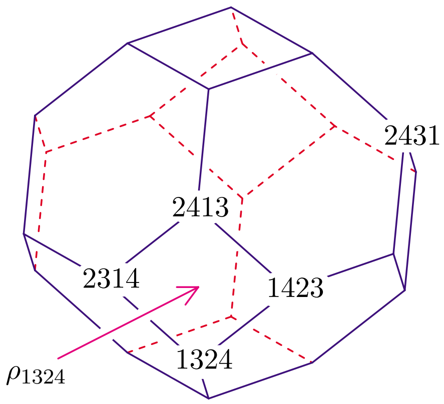


FIGURE 3. The facet ρ_{1324} of type $(2, 2)$ and $w = 2431$.

The description (2.3) of $A_{u,v}^{w}$ (and hence of $\alpha_{u,v}^{w}$) can be interpreted as follows: Given $u \in \mathfrak{S}_p$, $v \in \mathfrak{S}_q$, and $w \in \mathfrak{S}_{p+q}$, the structure constant $\alpha_{u,v}^{w}$ counts the number of facets of type (p, q) of the $(p+q-1)$-permutahedron such that the vertex $\rho_\zeta(u, v)$ is below w and it is the maximum vertex in that facet below w.

For instance, the facet ρ_{1324} contributes to the structure constant $\alpha_{12,21}^{2431}$ because the vertex $\rho_{1324}(12, 21) = 1423$ satisfies the required properties in relation to the vertex $w = 2431$, as shown in Figure 3.

2.3. **The antipode of $\mathfrak{S}Sym$.** Malvenuto left open the problem of an explicit formula for the antipode of $\mathfrak{S}Sym$ [15, pp. 59-60]. We identify the coefficients of the antipode in terms of both bases in explicit combinatorial terms. These are based upon a general formula for the antipode of a connected Hopf algebra due to Milnor and Moore [17].

For any subset $\mathsf{S} = \{p_1 < p_2 < \cdots < p_k\} \subseteq [n-1]$ and $v \in \mathfrak{S}_n$ set

$$v_{\mathsf{S}} \ :=\ \mathrm{st}(v_1, \ldots, v_{p_1}) \times \mathrm{st}(v_{p_1+1}, \ldots, v_{p_2}) \times \cdots \times \mathrm{st}(v_{p_k+1}, \ldots, v_n) \ \in\ \mathfrak{S}_{\mathsf{S}}\,.$$

Theorem 2.3. *For $v, w \in \mathfrak{S}_n$ set*

$$\begin{aligned}\lambda(v, w) \ :=\ & \#\{\mathsf{S} \subseteq [n-1] \mid \mathrm{Des}(w^{-1}v_{\mathsf{S}}) \subseteq \mathsf{S} \textit{ and } \#\mathsf{S} \textit{ is odd}\} \\ & - \#\{\mathsf{S} \subseteq [n-1] \mid \mathrm{Des}(w^{-1}v_{\mathsf{S}}) \subseteq \mathsf{S} \textit{ and } \#\mathsf{S} \textit{ is even}\}.\end{aligned}$$

Then

$$S(\mathcal{F}_v) \;=\; \sum_{w\in\mathfrak{S}_n} \lambda(v,w)\,\mathcal{F}_w\,. \tag{2.6}$$

These coefficients of the antipode may indeed be positive or negative. For instance,

$$S(\mathcal{F}_{231}) \;=\; \mathcal{F}_{132} - \mathcal{F}_{213} - 2\mathcal{F}_{231} + \mathcal{F}_{312}\,.$$

The coefficient of $\mathcal{F}_{312}$ is 1 because $\{1\}$, $\{2\}$, and $\{1,2\}$ are the subsets S of $\{1,2\}$ which satisfy $\mathrm{Des}\big((312)^{-1}(231)_{\mathsf{S}}\big) \subseteq \mathsf{S}$.

Our description of these coefficients is semi-combinatorial, in the sense that it involves a difference of cardinalities of sets. On the monomial basis the situation is different. The sign of the coefficients of $S(\mathcal{M}_v)$ only depends on the number of global descents of v. We provide a fully combinatorial description of these coefficients. Let $v,w \in \mathfrak{S}_n$ and suppose $\mathsf{S} \subseteq \mathrm{GDes}(v)$. Define $C_{\mathsf{S}}(v,w) \subseteq \mathfrak{S}^{\mathsf{S}}$ to be those $\zeta \in \mathfrak{S}^{\mathsf{S}}$ satisfying

$$\begin{array}{rl} (i) & v_{\mathsf{S}}\zeta^{-1} \le w, \\ (ii) & \text{if } v \le v' \text{ and } v'_{\mathsf{S}}\zeta^{-1} \le w \text{ then } v = v', \text{ and} \\ (iii) & \text{if } \mathrm{Des}(\zeta) \subseteq \mathsf{R} \subseteq \mathsf{S} \text{ and } v_{\mathsf{R}}\zeta^{-1} \le w \text{ then } \mathsf{R} = \mathsf{S}. \end{array} \tag{2.7}$$

Set $\kappa(v,w) := \#C_{\mathrm{GDes}(v)}(v,w)$.

Theorem 2.4. *For $v,w \in \mathfrak{S}_n$, we have*

$$S(\mathcal{M}_v) \;=\; (-1)^{\#\mathrm{GDes}(v)+1} \sum_{w\in\mathfrak{S}_n} \kappa(v,w)\,\mathcal{M}_w\,. \tag{2.8}$$

For instance,

$$\begin{aligned} S(\mathcal{M}_{3412}) \;=\; & \mathcal{M}_{1234} + 2\mathcal{M}_{1324} + \mathcal{M}_{1342} + \mathcal{M}_{1423} \\ & + \mathcal{M}_{2314} + \mathcal{M}_{2413} + \mathcal{M}_{3124} + \mathcal{M}_{3142} + \mathcal{M}_{3412}\,. \end{aligned}$$

Consider the coefficient of $\mathcal{M}_{3412}$. In this case, $\mathsf{S} = \mathrm{GDes}(3412) = \{2\}$, so

$$\mathfrak{S}^{\mathsf{S}} \;=\; \{1234,\ 1324,\ 1423,\ 2314,\ 2413,\ 3412\}\,.$$

We invite the reader to verify that 3412 is the only element of $\mathfrak{S}^{\{2\}}$ that satisfies all three conditions of (2.7). Therefore $C_{\mathsf{S}}(3412,3412) = \{3412\}$ and the coefficient is $\kappa(3412,3412) = 1$.

Remark 2.5. The antipode of $\mathfrak{S}\mathit{Sym}$ has infinite order. A computation gives that

$$S^{2m}(\mathcal{M}_{231}) \;=\; \mathcal{M}_{231} + 2m(\mathcal{M}_{213} - \mathcal{M}_{132}) \qquad \forall\, m \in \mathbb{Z}\,.$$

3. Hopf-algebraic structure of $\mathfrak{S}\mathit{Sym}$

3.1. **Cofreeness of $\mathfrak{S}\mathit{Sym}$ and the coradical filtration.** The basis $\{\mathcal{M}_u\}$ reveals the existence of a second coalgebra grading on $\mathfrak{S}\mathit{Sym}$, given by the number of global descents. With respect to this grading, $\mathfrak{S}\mathit{Sym}$ is a cofree graded coalgebra. We deduce an elegant description of the coradical filtration: it corresponds to a filtration of the symmetric groups by certain lower order ideals determined by the number of global descents. In particular, the space of primitive elements is spanned by those $\mathcal{M}_u$ where u has no global descents.

Let V be a vector space and consider the graded vector space

$$Q(V) \;:=\; \bigoplus_{k\ge 0} V^{\otimes k}\,.$$

The space $Q(V)$ is a graded connected coalgebra under the *deconcatenation* coproduct

$$\Delta(v_1 \otimes \cdots \otimes v_k) \; = \; \sum_{i=0}^{k} (v_1 \otimes \cdots \otimes v_i) \otimes (v_{i+1} \otimes \cdots \otimes v_k)\,,$$

and counit $\epsilon(v_1 \otimes \cdots \otimes v_k) = 0$ if $k \geq 1$.

The following universal property is satisfied for the canonical projection $\pi : Q(V) \to V$. Given a graded coalgebra $C = \oplus_{k \geq 0} C^k$ and a linear map $\varphi : C \to V$ such that $\varphi(C^k) = 0$ for $k \neq 1$, there is a unique morphism of graded coalgebras $\hat{\varphi} : C \to Q(V)$ such that the following diagram commutes

$$\begin{tikzcd} C \arrow[rr, dashed, "\hat{\varphi}"] \arrow[dr, "\varphi"'] & & Q(V) \arrow[dl, "\pi"] \\ & V & \end{tikzcd}$$

Explicitly, $\hat{\varphi}$ is defined by

$$\hat{\varphi}|_{C^k} = \varphi^{\otimes k} \Delta^{(k-1)}\,. \tag{3.1}$$

In particular, $\hat{\varphi}|_{C^0} = \epsilon$, $\hat{\varphi}|_{C^1} = \varphi$ and $\hat{\varphi}|_{C^2} = (\varphi \otimes \varphi)\Delta$.

To establish the cofreeness of $\mathfrak{S}Sym$, we first define a second coalgebra grading. Let $\mathfrak{S}^0 := \mathfrak{S}_0$, and for $k \geq 1$, let

$$\begin{aligned} \mathfrak{S}_n^k \; &:= \; \{u \in \mathfrak{S}_n \mid u \text{ has exactly } k{-}1 \text{ global descents}\}, \quad \text{and} \\ \mathfrak{S}^k \; &:= \; \coprod_{n \geq 0} \mathfrak{S}_n^k\,. \end{aligned}$$

For instance,

$$\begin{aligned} \mathfrak{S}^1 \; = \; \{1\} \; \cup \; \{12\} \; \cup \; \{123,\, 213,\, 132\} \; \cup \; \{&1234,\, 2134,\, 1324,\, 1243,\, 3124, \\ &2314,\, 2143,\, 1423,\, 1342,\, 3214,\, 3142,\, 2413,\, 1432\} \; \cup \; \cdots \end{aligned}$$

Let $(\mathfrak{S}Sym)^k$ be the vector subspace of $\mathfrak{S}Sym$ spanned by $\{\mathcal{M}_u \mid u \in \mathfrak{S}^k\}$.

Theorem 3.1. *The decomposition $\mathfrak{S}Sym = \oplus_{k \geq 0} (\mathfrak{S}Sym)^k$ is a coalgebra grading. Moreover, endowed with this grading, $\mathfrak{S}Sym$ is a cofree graded coalgebra.*

The coradical $C^{(0)}$ of a graded connected coalgebra C is the 1-dimensional component in degree 0 (identified with the base field via the counit). The primitive elements are

$$\mathrm{Prim}(C) \; := \; \{x \in C \mid \Delta(x) = x \otimes 1 + 1 \otimes x\}\,.$$

Set $C^{(1)} := C^{(0)} \oplus \mathrm{Prim}(C)$, the first level of the coradical filtration. More generally, the k-th level of the coradical filtration is

$$C^{(k)} \; := \; \big(\Delta^{(k)}\big)^{-1} \Big(\sum_{i+j=k} C^{\otimes i} \otimes C^{(0)} \otimes C^{\otimes j} \Big)\,.$$

We have $C^{(0)} \subseteq C^{(1)} \subseteq C^{(2)} \subseteq \cdots \subseteq C = \bigcup_{k \geq 0} C^{(k)}$, and

$$\Delta(C^{(k)}) \; \subseteq \; \sum_{i+j=k} C^{(i)} \otimes C^{(j)}\,.$$

Thus, the coradical filtration measures the complexity of iterated coproducts.

For a cofree graded coalgebra $Q(V)$, the coradical filtration is easy to describe. The space of primitive elements is just V, and the k-th level of the coradical filtration is $\oplus_{i=0}^{k} V^{\otimes i}$. These are immediate from the definition of the deconcatenation coproduct.

Define

$$\mathfrak{S}_n^{(k)} := \coprod_{i=0}^{k} \mathfrak{S}_n^k \quad\text{and}\quad \mathfrak{S}^{(k)} := \coprod_{i=0}^{k} \mathfrak{S}^k .$$

In other words, $\mathfrak{S}^{(0)} = \mathfrak{S}_0$ and for $k \geq 1$,

$$\mathfrak{S}_n^{(k)} = \{u \in \mathfrak{S}_n \mid u \text{ has at most } k{-}1 \text{ global descents}\} .$$

Proposition 1.2 asserts that GDes: $\mathfrak{S}_n \to \mathcal{Q}_n$ is order-preserving. Since $\mathcal{Q}_n$ is ranked by the cardinality of a subset, it follows that $\mathfrak{S}_n^{(k)}$ is a lower order ideal of $\mathfrak{S}_n$, with $\mathfrak{S}_n^{(k)} \subseteq \mathfrak{S}_n^{(k+1)}$. The coradical filtration corresponds precisely to this filtration of the symmetric groups by lower ideals.

Corollary 3.2. *A linear basis for the k-th level of the coradical filtration of $\mathfrak{S}Sym$ is*

$$\{\mathcal{M}_u \mid u \in \mathfrak{S}^{(k)}\} .$$

In particular, a linear basis for the space of primitive elements is

$$\{\mathcal{M}_u \mid u \text{ has no global descents}\} .$$

3.2. **The descent map to quasi-symmetric functions.** We study the effect of the morphism of Hopf algebras (1.11)

$$\mathcal{D} : \mathfrak{S}Sym \to \mathcal{Q}Sym, \qquad \text{defined by} \qquad \mathcal{F}_u \mapsto F_{\mathrm{Des}(u)}$$

on the monomial basis. Here, we use subsets S of $[n{-}1]$ to index monomial quasi-symmetric functions of degree n.

Definition 3.3. A permutation $u \in \mathfrak{S}_n$ is *closed* if we have $u = Z(\mathsf{T})$ for some $\mathsf{T} \in \mathcal{Q}_n$. Equivalently, u is closed if and only if $\mathrm{Des}(u) = \mathrm{GDes}(u)$.

Theorem 3.4. *Let $u \in \mathfrak{S}_n$. Then*

$$\mathcal{D}(\mathcal{M}_u) = \begin{cases} M_{\mathrm{GDes}(u)} & \textit{if } u \textit{ is closed,} \\ 0 & \textit{if not.} \end{cases}$$

3.3. $\mathfrak{S}Sym$ **as a crossed product over** $\mathcal{Q}Sym$**.** We describe the *algebra* structure of $\mathfrak{S}Sym$ as a crossed product over the Hopf algebra $\mathcal{Q}Sym$. See [18, §7] for a review of this construction in the general Hopf algebraic setting. Let us only say that the crossed product of a Hopf algebra K with an algebra A with respect to a Hopf cocycle $\sigma : K{\otimes}K \to A$ is a certain algebra structure on the space $A \otimes K$, denoted by $A\#_\sigma K$.

Proposition 3.5. *The map $M_{\mathsf{S}} \mapsto \mathcal{M}_{Z(\mathsf{S})}$ induces a morphism of coalgebras $\mathcal{Z}\colon \mathcal{Q}Sym \to \mathfrak{S}Sym$ that is a right inverse to the morphism of Hopf algebras $\mathcal{D}\colon \mathfrak{S}Sym \to \mathcal{Q}Sym$.*

In this situation, an important theorem of Blattner, Cohen and Montgomery [6] applies. Namely, suppose $\pi : H \to K$ is a morphism of Hopf algebras that admits a coalgebra splitting $\gamma : K \to H$. Then there is a *crossed product* decomposition

$$H \cong A\#_\sigma K$$

where A is the *left Hopf kernel* of π:

$$A = \{h \in H \mid \sum h_1{\otimes}\pi(h_2) = h{\otimes}1\}$$

and the *Hopf cocycle* $\sigma : K{\otimes}K \to A$ is

$$\sigma(k, k') = \sum \gamma(k_1)\gamma(k'_1)S\gamma(k_2k'_2) .$$

Note that if π and γ preserve gradings, then so does the rest of the structure.

Let A be the left Hopf kernel of $\mathcal{D}\colon \mathfrak{S}Sym \to \mathcal{Q}Sym$ and A_n its n-th homogeneous component. Once again the monomial basis of $\mathfrak{S}Sym$ proves useful in describing A.

Proposition 3.6. *A basis for A_n is the set $\{\mathcal{M}_u\}$ where u runs over all permutations of n that are not of the form*

$$* \ldots * \, 12 \ldots n{-}k$$

for any $k = 0, \ldots, n-1$. In particular,

$$\dim A_n \;=\; n! - \sum_{k=0}^{n-1} k!\,.$$

References

[1] Marcelo Aguiar, *Infinitesimal Hopf algebras and the cd-index of polytopes*, Discrete Comput. Geom. **27** (2002), no. 1, 3–28, Geometric combinatorics (San Francisco,CA/Davis, CA, 2000).

[2] Marcelo Aguiar and Frank Sottile, *The Malvenuto-Reutenauer Hopf algebra*, 2002.

[3] Nantel Bergeron, Stephan Mykytiuk, Frank Sottile, and Stephanie van Willigenberg, *Non-commutative pieri operations on posets*, J. Combin. Th. Ser. A **91** (2000), no. 1/2, 84–110, Special issue in memory of Gian-Carlo Rota.

[4] Louis J. Billera and Niandong Liu, *Noncommutative enumeration in graded posets*, J. Algebraic Combin. **12** (2000), no. 1, 7–24. MR **2001h:**05009

[5] Anders Björner, Michel Las Vergnas, Bernd Sturmfels, Neil White, and Günter M. Ziegler, *Oriented matroids*, second ed., Cambridge University Press, Cambridge, 1999. MR **2000j:**52016

[6] Robert J. Blattner, Miriam Cohen, and Susan Montgomery, *Crossed products and inner actions of Hopf algebras*, Trans. Amer. Math. Soc. **298** (1986), no. 2, 671–711. MR **87k:**16012

[7] Gérard Duchamp, Florent Hivert, and Jean-Yves Thibon, *Noncommutative symmetric functions VI: free quasi-symmetric functions and related algebras*, 2001.

[8] ———, *Some generalizations of quasi-symmetric functions and noncommutative symmetric functions*, Formal power series and algebraic combinatorics (Moscow, 2000), Springer, Berlin, 2000, pp. 170–178. MR **2001k:**05202

[9] Richard Ehrenborg, *On posets and Hopf algebras*, Adv. Math. **119** (1996), no. 1, 1–25. MR **97e:**16079

[10] Israel M. Gelfand, Daniel Krob, Alain Lascoux, Bernard Leclerc, Vladimir S. Retakh, and Jean-Yves Thibon, *Noncommutative symmetric functions*, Adv. Math. **112** (1995), no. 2, 218–348. MR **96e:**05175

[11] Ira M. Gessel, *Multipartite P-partitions and inner products of skew Schur functions*, Combinatorics and algebra (Boulder, Colo., 1983) (Providence, RI), Amer. Math. Soc., 1984, pp. 289–317. MR **86k:**05007

[12] Jean-Louis Loday and María O. Ronco, *Hopf algebra of the planar binary trees*, Adv. Math. **139** (1998), no. 2, 293–309. MR **99m:**16063

[13] ———, *Order structure on the algebra of permutations and of planar binary trees*, 2001, J. Alg. Combinatorics, to appear.

[14] Ian Macdonald, *Symmetric functions and Hall polynomials*, second ed., The Clarendon Press Oxford University Press, New York, 1995, With contributions by A. Zelevinsky, Oxford Science Publications. MR **96h:**05207

[15] Claudia Malvenuto, *Produits et coproduits des fonctions quasi-symétriques et de l'algèbre des descents*, no. 16, Laboratoire de combinatoire et d'informatique mathématique (LACIM), Univ. du Québec à Montréal, Montréal, 1994.

[16] Claudia Malvenuto and Christophe Reutenauer, *Duality between quasi-symmetric functions and the Solomon descent algebra*, J. Algebra **177** (1995), no. 3, 967–982. MR **97d:**05277

[17] John W. Milnor and John C. Moore, *On the structure of Hopf algebras*, Ann. of Math. (2) **81** (1965), 211–264. MR 30 #4259

[18] Susan Montgomery, *Hopf algebras and their actions on rings*, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1993. MR **94i:**16019

[19] Frédéric Patras and Christophe Reutenauer, *Lie representations and an algebra containing Solomon's*, 2001.

[20] Stéphane Poirier and Christophe Reutenauer, *Algèbres de Hopf de tableaux*, Ann. Sci. Math. Québec **19** (1995), no. 1, 79–90. MR **96g:**05146

[21] Christophe Reutenauer, *Free Lie algebras*, The Clarendon Press Oxford University Press, New York, 1993, Oxford Science Publications. MR **94j:**17002

[22] Richard P. Stanley, *Enumerative combinatorics. Vol. 2*, Cambridge University Press, Cambridge, 1999, With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin. MR **2000k:**05026

[23] John R. Stembridge, *Enriched P-partitions*, Trans. Amer. Math. Soc. **349** (1997), no. 2, 763–788. MR **97f:**06006

Department of Mathematics, Texas A&M University, College Station, Texas 77843 USA
E-mail address: `maguiar@math.tamu.edu`
URL: `http://www.math.tamu.edu/~maguiar`

THE TUTTE POLYNOMIAL OF A HYPERPLANE ARRANGEMENT (EXTENDED ABSTRACT)

FEDERICO ARDILA

Abstract. We define the Tutte polynomial of a hyperplane arrangement and study its properties. We show that it is the universal Tutte-Grothendieck invariant for the class of hyperplane arrangements. We also show that its coefficients are nonnegative. We introduce a new finite field method for computing Tutte polynomials, which generalizes several known results. We apply our method to several specific arrangements, thus relating the computation of Tutte polynomials to some problems in classical enumerative combinatorics. As a consequence, we obtain new formulas for the generating functions enumerating alternating trees, labeled trees, semiorders and Dyck paths.

Résumé. Nous définissons le polynôme de Tutte d'un arrangement d'hyperplans et étudions ses propriétés. Nous montrons que ce polynôme est l'invariant universel de Tutte-Grothendieck pour la classe des arrangements d'hyperplans, et aussi que ses coefficients sont non-négatifs. Une nouvelle méthode utilisant des corps finis est introduite, qui étend plusieurs résultats connus. Nous l'appliquons à plusieurs arrangements classiques, ce qui relie le calcul des polynômes de Tutte à divers problèmes classiques de combinatoire énumérative. Nous obtenons ainsi de nouvelles formules pour les séries génératrices dénombrant les arbres alternants, les arbres étiquetés, les semi-ordres et les chemins de Dyck.

1. Introduction

The aim of this paper is to define the Tutte polynomial of a hyperplane arrangement, and find out what we can say about it. Central arrangements inherit Tutte polynomial properties from their associated matroids; we want to know whether such properties hold for affine arrangements as well.

In Section 2 we introduce the basic notions that we will need in the paper. In Section 3 we define the Tutte polynomial of a hyperplane arrangement, and show that it is the universal Tutte-Grothendieck invariant on the class of hyperplane arrangements. Our first main result is that the coefficients of the Tutte polynomial of a hyperplane arrangement are nonnegative. In Section 4 we obtain our second main result, a finite field method for computing Tutte polynomials of hyperplane arrangements. This is done in terms of the coboundary polynomial, a simple transformation of the Tutte polynomial. We derive some consequences of this method. Finally, in Section 5, we compute the Tutte polynomials of several families of arrangements. In particular, for deformations of the braid arrangement, we relate the computation of Tutte polynomials to some enumeration problems in classical combinatorics. As a consequence, we obtain new formulas for the generating functions enumerating alternating trees, labeled trees, semiorders and Dyck paths.

2. Hyperplane arrangements

Given a field $\Bbbk$ and a positive integer n, an *affine hyperplane* in $\Bbbk^n$ is an $(n-1)$-dimensional affine subspace of $\Bbbk^n$. If we put a system of coordinates $x_1, \ldots, x_n$ on $\Bbbk^n$, a hyperplane can be seen as the set of points that satisfy a certain equation $c_1x_1 + \cdots + c_nx_n = c$, where $c_1, \ldots, c_n, c$ are constants in $\Bbbk$ with not all c_i's equal to 0. A *hyperplane arrangement* $\mathcal{A}$ in $\Bbbk^n$ is a finite collection of affine hyperplanes of $\Bbbk^n$. We will refer to hyperplane arrangements simply as *arrangements*. We will always assume that $\Bbbk = \mathbb{R}$ unless explicitly stated, although most of our results extend immediately to any field of characteristic zero.

We will say that an arrangement $\mathcal{A}$ is *central* if the hyperplanes in $\mathcal{A}$ have a nonempty intersection.[1] Similarly, we will say that a subset (or *subarrangement*) $\mathcal{B} \subseteq \mathcal{A}$ of hyperplanes is *central* if the hyperplanes in $\mathcal{B}$ have a nonempty intersection. The *rank function* $r_{\mathcal{A}}$ is defined for each central subset $\mathcal{B}$ by the equation $r_{\mathcal{A}}(\mathcal{B}) = n - \dim \cap \mathcal{B}$. The rank of a noncentral subset $\mathcal{B}$ is defined to be the largest rank of a central subset of $\mathcal{B}$. The *rank* of $\mathcal{A}$ is $r_{\mathcal{A}}(\mathcal{A})$, and it is denoted $r_{\mathcal{A}}$. We will usually omit the subscripts when the underlying arrangement is clear, and simply write $r(\mathcal{B})$ and r respectively.

To each hyperplane arrangement $\mathcal{A}$ we assign a partially ordered set called the *intersection poset* of $\mathcal{A}$, and denoted $L_{\mathcal{A}}$. It consists of the nonempty intersections $H_{i_1} \cap \cdots \cap H_{i_k}$, ordered by reverse inclusion. This poset is graded, with rank function $r(H_{i_1} \cap \cdots \cap H_{i_k}) = r_{\mathcal{A}}(\{H_{i_1}, \ldots, H_{i_k}\})$, and a unique minimal element $\hat{0} = \mathbb{R}^n$. We say that two hyperplane arrangements are *isomorphic* if their intersection posets are isomorphic.

We define the *characteristic polynomial* of $\mathcal{A}$ to be

$$\chi_{\mathcal{A}}(q) = \sum_{x \in L_{\mathcal{A}}} \mu(\hat{0}, x) q^{n-r(x)}.$$

where μ denotes the Möbius function [17, Section 3.7] of $L_{\mathcal{A}}$. The following theorem will be important for us in Section 5.

Theorem 2.1. *(Zaslavsky,* [25]*) Let $\mathcal{A}$ be a hyperplane arrangement in $\mathbb{R}^n$. The number of regions into which $\mathcal{A}$ dissects $\mathbb{R}^n$ is equal to $(-1)^n \chi_{\mathcal{A}}(-1)$.*

Even though it is not $(n-1)$-dimensional, we need to allow $\mathbb{R}^n$ as a possible "hyperplane" in $\mathcal{A}$. We will also allow $\mathcal{A}$ to contain repeated hyperplanes. We will say that a central subset $\mathcal{B}$ is *independent* if $r(\mathcal{B}) = |\mathcal{B}|$, and *dependent* otherwise. It is a *base* of $\mathcal{A}$ if $r(\mathcal{B}) = |\mathcal{B}| = r(\mathcal{A})$. A hyperplane $H \in \mathcal{A}$ is called a *loop* if it is $\mathbb{R}^n$, and an *isthmus* if $r(\mathcal{A} - H) = r(\mathcal{A}) - 1$. Here we are slightly abusing notation, writing $\mathcal{A} - H$ instead of $\mathcal{A} - \{H\}$. We will often do this for simplicity.

3. The Tutte polynomial

The *Tutte polynomial* of a hyperplane arrangement $\mathcal{A}$ is defined by

$$T_{\mathcal{A}}(q,t) = \sum_{\substack{\mathcal{B} \subseteq \mathcal{A} \\ \text{central}}} (q-1)^{r-r(\mathcal{B})} (t-1)^{|\mathcal{B}|-r(\mathcal{B})}, \tag{3.1}$$

where the sum is over all central subsets $\mathcal{B} \subseteq \mathcal{A}$.

The characteristic polynomial is essentially an evaluation of the Tutte polynomial: we have $\chi_{\mathcal{A}}(q) = (-1)^r q^{n-r} T_{\mathcal{A}}(1-q, 0)$. This is a consequence of the following result.

Theorem 3.1. *(Whitney* [24]*, Postnikov and Stanley* [13]*) For any hyperplane arrangement $\mathcal{A}$,*

$$\chi_{\mathcal{A}}(q) = \sum_{\substack{\mathcal{B} \subseteq \mathcal{A} \\ \text{central}}} (-1)^{|\mathcal{B}|} q^{n-r(\mathcal{B})}.$$

It is worth remarking that the same procedure can be used to define the Tutte polynomial of an arbitrary subspace arrangement, where the subspaces do not have to be hyperplanes. We can still define the rank function of $\mathcal{A}$ by $r_{\mathcal{A}}(\mathcal{B}) = n - \dim(\cap \mathcal{B})$ for each subarrangement $\mathcal{B} \subseteq \mathcal{A}$, and then use equation (3.1) to define $T_{\mathcal{A}}(q,t)$. In this paper we shall focus on hyperplane arrangements. However, we briefly mention that our second main result,

[1] Sometimes we will call an arrangement *affine* to emphasize that it does not need to be central.

Theorem 4.1 also holds for subspace arrangements; our proof extends to this case without difficulty. The other main result of this paper, Theorem 3.5, does not hold. The Tutte polynomial of a subspace arrangement does not necessarily have nonnegative coefficients.

3.1. **Deletion-contraction.** Let $\mathcal{A}$ be an arrangement and let H be a hyperplane in $\mathcal{A}$. The arrangement $\mathcal{A} - H$ is called the *deletion* of H in $\mathcal{A}$. It is an arrangement in $\mathbb{R}^n$. The arrangement $\mathcal{A}^H = \{H_0 \cap H \mid H_0 \in \mathcal{A} - H, H_0 \cap H \neq \emptyset\}$ is called the *contraction* (or *restriction*) of $\mathcal{A}$ to H. It is an arrangement in H.

A function f on the class of arrangements is called an *invariant* if $f(\mathcal{A}_1) = f(\mathcal{A}_2)$ whenever $\mathcal{A}_1$ and A_2 are isomorphic arrangements. There is a very important type of invariant, known as a *Tutte-Grothendieck invariant*. The Tutte polynomial is the universal *Tutte-Grothendieck invariant* on the class of arrangements. The following theorem shows that any other *generalized Tutte-Grothendieck invariant*, that is, an invariant satisfying the conditions of Theorem 3.2, is an evaluation of the Tutte polynomial. Analogous results are essentially known for matroids [4], [12], and for framed configurations [23]. Framed configurations include hyperplane arrangements as particular cases, but their generality does not allow for a statement as explicit as this one.

Theorem 3.2. *Let $\mathbb{A}$ be the set of isomorphism classes of arrangements in real vector spaces. Let $\Bbbk$ be a field and let $a, b \in \Bbbk$. Let R be a commutative ring containing $\Bbbk$, and let $f : \mathbb{A} \to R$ be a function satisfying the following conditions.*

(i) *If a hyperplane H in an arrangement $\mathcal{A}$ is neither an isthmus nor a loop, then $f(\mathcal{A}) = af(\mathcal{A} - H) + bf(\mathcal{A}^H)$.*
(ii) *If H is an isthmus in $\mathcal{A}$, then $f(\mathcal{A}) = f(I)f(\mathcal{A} - H)$.*
(iii) *If H is a loop in $\mathcal{A}$, then $f(\mathcal{A}) = f(L)f(\mathcal{A} - H)$.*

Then the function f is given by $f(\mathcal{A}) = a^{|\mathcal{A}|-r(\mathcal{A})}b^{r(\mathcal{A})}T_{\mathcal{A}}(f(I)/b, f(L)/a)$. Here I denotes the arrangement consisting of a single isthmus, and L denotes the arrangement consisting of a single loop.

3.2. **Base activity.** We now show that the Tutte polynomial of a hyperplane arrangement has nonnegative coefficients, by giving a combinatorial interpretation of them. Crapo interpreted the coefficients of the Tutte polynomial of a matroid as enumerators of bases with a given internal and external activity [6]. Our interpretation for hyperplane arrangements is analogous.

Define a *circuit* of an arrangement to be a minimal set of hyperplanes which is central and dependent. Define a *bond* to be a minimal set of hyperplanes, the removal of which makes the rank of the arrangement decrease.

Lemma 3.3. *Let B be a base of $\mathcal{A}$, and let e be a hyperplane not in B such that $B \cup e$ is central. Then $B \cup e$ contains a unique circuit.*

Lemma 3.4. *Let B be a base of $\mathcal{A}$, and let i be a hyperplane in B. Then $\mathcal{A} - B \cup i$ contains a unique bond.*

From now on, we will fix a linear order on $\mathcal{A}$. Now each k-subset of $\mathcal{A}$ corresponds to a strictly increasing sequence of k integers between 1 and $|\mathcal{A}|$. For each $0 \leq k \leq |\mathcal{A}|$, order the k-subsets of $\mathcal{A}$ using the lexicographic order on these sequences.

Given a base B, we will say that a hyperplane e not in B is an *external activity hyperplane for B* if $B \cup e$ is central, and e is the smallest hyperplane[2] of the unique circuit in $B \cup e$. Let $E(B)$ be the set of external activity hyperplanes for B, and let $e(B) = |E(B)|$. We call $e(B)$ the *external activity* of B.

[2]according to the fixed linear order

We will say that a hyperplane i in B is an *internal activity hyperplane for* B if i is the smallest hyperplane of the unique bond in $\mathcal{A} - B \cup i$. Let $I(B)$ be the set of internal activity hyperplanes for B, and let $i(B) = |I(B)|$. We call $i(B)$ the *internal activity* of B.

Now we are in a position to state the main theorem of this section.

Theorem 3.5. *For any arrangement* $\mathcal{A}$,

$$T_{\mathcal{A}}(q,t) = \sum_{\substack{B \text{ base} \\ \text{of } \mathcal{A}}} q^{i(B)} t^{e(B)}.$$

Theorem 3.5 shows that the coefficients of the Tutte polynomial are nonnegative integers. The coefficient of $q^i t^e$ is equal to the number of bases of $\mathcal{A}$ with internal activity equal to i and external activity equal to e.

A useful ingredient in the proof of Theorem 3.5 is the following characterization of internal and external activity hyperplanes. Given a subarrangement $\mathcal{B} \subseteq \mathcal{A}$ and a hyperplane H, let $\mathcal{B}_{>H} = \{H_0 \in \mathcal{B} \mid H_0 > H\}$. Define $\mathcal{B}_{<H}$ analogously.

Lemma 3.6. *Let B be a base and e be a hyperplane not in B such that $B \cup e$ is central. Then e is an external activity hyperplane for B if and only if $r(\mathcal{B}_{>e} \cup e) = r(\mathcal{B}_{>e})$.*

Lemma 3.7. *Let B be a base and i be a hyperplane in B. Then i is an internal activity hyperplane for B if and only if $r(B - i \cup \mathcal{A}_{<i}) < r$.*[3]

Now we wish to present a different description of the central subarrangements of $\mathcal{A}$. To do it, we need two definitions. For each subarrangement $\mathcal{B}$, let $d\mathcal{B}$ be the lexicographically largest base of $\mathcal{B}$. For each independent central subarrangement $\mathcal{B}$, let $u\mathcal{B}$ be the lexicographically smallest base of $\mathcal{A}$ which contains $\mathcal{B}$. Notice that, for any subarrangement $\mathcal{B}$, $ud\mathcal{B}$ is a base of $\mathcal{A}$.

Let S_1 be the set of triples (B, I, E) such that B is a base of $\mathcal{A}$, $I \subseteq I(B)$ is a set of internal activity hyperplanes for B, and $E \subseteq E(B)$ is a set of external activity hyperplanes for $\mathcal{B}$. Let S_2 be the set of central subarrangements of $\mathcal{A}$. We want to establish a bijection between S_1 and S_2. Define two maps ϕ_1 and ϕ_2 as follows. Given $(B, I, E) \in S_1$, let $\phi_1(B, I, E) = B - I \cup E$. Given $\mathcal{C} \in S_2$, let $\phi_2(\mathcal{C}) = (ud\mathcal{C}, ud\mathcal{C} - d\mathcal{C}, \mathcal{C} - d\mathcal{C})$. The maps ϕ_1 and ϕ_2 give the desired bijection: every central subarrangement $\mathcal{B}$ of $\mathcal{A}$ can be written uniquely in the form $\mathcal{B} = B - I \cup E$ where B is a base of $\mathcal{A}$, $I \subseteq I(B)$ and $E \subseteq E(B)$.

Lemma 3.8. *The map ϕ_1 maps S_1 to S_2.*

Lemma 3.9. *The map ϕ_2 maps S_2 to S_1.*

Proposition 3.10. *The map ϕ_1 is a bijection from S_1 to S_2, and the map ϕ_2 is its inverse.*

Proposition 3.10 is the key ingredient of our proof of Theorem 3.5. It is a consequence of the following lemmas.

Lemma 3.11. *For all $(B, I, E) \in S_1$, we have $r(B - I \cup E) = r - |I|$.*

Lemma 3.12. *For all $(B, I, E) \in S_1$, we have $d(B - I \cup E) = B - I$.*

Lemma 3.13. *For all $(B, I, E) \in S_1$, we have $ud(B - I \cup E) = B$.*

Proving these lemmas requires some work; once we have done that, Theorem 3.5 follows easily from Proposition 3.10 and Lemma 3.11.

[3] In fact, this is true if and only if $r(B - i \cup \mathcal{A}_{<i}) = r - 1$.

4. A finite field method

In [14], Reiner asked whether it is possible to define the Tutte polynomial of a subspace arrangement. We have shown how to do this in the introduction to Section 3. Reiner also asked whether it is possible to use [14, Corollary 3] to compute explicitly the Tutte polynomials of some nontrivial families of arrangements. Compared to all the work that has been done on computing characteristic polynomials explicitly, very little is known about computing Tutte polynomials.

In this section, we introduce a new method for computing Tutte polynomials of hyperplane arrangements. This method also works for arbitrary subspace arrangements. Our approach does not use Reiner's result; it is closer to Athanasiadis's finite field method for computing characteristic polynomials. In fact, Athanasiadis's result [1, Theorem 2.2] can be obtained as a special case of the main result of this section, Theorem 4.1, by setting $t = 0$.

Using Crapo's terminology [6], define the *coboundary polynomial* $\overline{\chi}_{\mathcal{A}}(q,t)$ of an arrangement $\mathcal{A}$ by

$$\overline{\chi}_{\mathcal{A}}(q,t) = \sum_{\substack{\mathcal{B}\subseteq\mathcal{A} \\ \text{central}}} q^{r-r(\mathcal{B})}(t-1)^{|\mathcal{B}|}. \tag{4.1}$$

It is easy to see that the coboundary and Tutte polynomials are simple transformations of each other, and computing the coboundary polynomial of an arrangement is essentially equivalent to computing its Tutte polynomial. Our results can be presented more elegantly in terms of the coboundary polynomial.

Let $\mathcal{A}$ be a $\mathbb{Z}$-arrangement in $\mathbb{R}^n$; that is, an arrangement where the defining equations have integer coefficients. Let q be a prime power. The arrangement $\mathcal{A}$ induces an arrangement $\mathcal{A}_q$ in the vector space $\mathbb{F}_q^n$. If we consider the equations defining the hyperplanes of $\mathcal{A}$ and regard them as equations over $\mathbb{F}_q$, they define the hyperplanes of $\mathcal{A}_q$.

Theorem 4.1. *Let $\mathcal{A}$ be a $\mathbb{Z}-$arrangement in $\mathbb{R}^n$. Let q be a power of a large enough prime, and let $\mathcal{A}_q$ be the induced arrangement in $\mathbb{F}_q^n$. Then*

$$q^{n-r}\,\overline{\chi}(q,t) \;=\; \sum_{p\,\in\,\mathbb{F}_q^n} t^{h(p)}, \tag{4.2}$$

where $h(p)$ denotes the number of hyperplanes of $\mathcal{A}_q$ that p lies on.

We remark that Theorem 4.1 was also discovered independently by Welsh and Whittle [23, Theorem 7.4].

Theorem 4.1 generalizes two classical enumerative results. The first result concerns vertex colorings of graphs. Given a graph G on $[n]$, we can associate to it an arrangement $\mathcal{A}_G$ in $\mathbb{R}^n$. It consists of the hyperplanes $x_i = x_j$, for all $1 \le i < j \le n$ such that ij is an edge in the graph G. When we apply Theorem 4.1 to this arrangement, we obtain the following result.

Theorem 4.2. ([5, Proposition 6.3.26]) *Let G be a graph with n vertices and c connected components. Then*

$$q^c\,\overline{\chi}_{\mathcal{A}_G}(q,t) = \sum_{\substack{q-\text{colorings} \\ \kappa\,\text{of}\,G}} t^{\text{mono}(\kappa)},$$

where $\text{mono}(\kappa)$ *is the number of monochromatic edges in κ.*

The second result concerns linear codes. An $[n,r]$ *linear code* C over $\mathbb{F}_q$ is an r-dimensional subspace of $\mathbb{F}_q^n$. A *generator matrix* for C is an $r \times n$ matrix U over $\mathbb{F}_q$, the rows of which form a basis for C. It is not difficult to see that the isomorphism class of the matroid on the columns of U depends only on C. We shall denote the corresponding matroid M_C.

The elements of C are called *codewords*. The *weight* $w(v)$ of a codeword is the cardinality of its support; that is, the number of nonzero coordinates of v.

The translation of Theorem 4.1 to this setting is the following.

Theorem 4.3. *(Greene, [7]) For any linear code C over $\mathbb{F}_q$,*

$$\sum_{v\in C} t^{w(v)} = t^n\,\overline{\chi}_{M_C}\left(q,\frac{1}{t}\right).$$

We conclude this section by presenting two new results, which are also relatively simple consequences of Theorem 4.1.

Theorem 4.4. *Let $\mathcal{A}$ be an arrangement and let $0\le t\le 1$ be a real number. Let $\mathcal{B}$ be a random subarrangement of $\mathcal{A}$, obtained by independently removing each hyperplane from $\mathcal{A}$ with probability t. Then the expected characteristic polynomial $\chi_{\mathcal{B}}(q)$ of $\mathcal{B}$ is $q^{n-r}\overline{\chi}_{\mathcal{A}}(q,t)$.*

Theorem 4.5. *For an arrangement $\mathcal{A}$ and an affine subspace x in the intersection poset $L_{\mathcal{A}}$, let $h(x)$ be the number of hyperplanes of A containing x. Then*

$$\overline{\chi}_{\mathcal{A}}(q,t) = \sum_{\substack{x\le y\\ \text{in } L_{\mathcal{A}}}} \mu(x,y)\,q^{r-r(y)}t^{h(x)}.$$

5. Computation of Tutte polynomials

In this section we use Theorem 4.1 to compute the coboundary polynomials of several families of arrangements. As remarked in Section 4, this is essentially the same as computing their Tutte polynomials.

5.1. Coxeter arrangements.

Let Φ be an irreducible crystallographic root system in $\mathbb{R}^n$, with the standard inner product, and let W be its associated Weyl group. The Coxeter arrangement of type W consists of the hyperplanes $(\alpha,x)=0$ for each $\alpha\in\Phi^+$. See [8] for an introduction to root systems and Weyl groups, and [11, Chapter 6] or [3, Section 2.3] for more information on Coxeter arrangements.

In this section we present the coboundary polynomials of the Coxeter arrangements of type A_n, B_n and D_n. (The arrangement of type C_n is the same as the arrangement of type B_n.) The best way to state our results is to compute the exponential generating function for the coboundary polynomials of each family.

Theorem 5.1. *Let $\mathcal{A}_n$ be the Coxeter arrangement of type A_{n-1} in $\mathbb{R}^n$, consisting of the hyperplanes $x_i=x_j$ for $1\le i<j\le n$.[4] We have*

$$1+q\sum_{n\ge1}\overline{\chi}_{\mathcal{A}_n}(q,t)\frac{x^n}{n!} = \left(\sum_{n\ge0} t^{\binom{n}{2}}\frac{x^n}{n!}\right)^q.$$

Theorem 5.2. *Let $\mathcal{B}_n$ be the Coxeter arrangement of type B_n in $\mathbb{R}^n$, consisting of the hyperplanes $x_i=x_j$ and $x_i+x_j=0$ for $1\le i<j\le n$, and the hyperplanes $x_i=0$ for*

[4] This arrangement is also known as the *braid arrangement.*

$1 \le i \le n$. *We have*

$$\sum_{n\ge 0} \overline{\chi}_{\mathcal{B}_n}(q,t)\frac{x^n}{n!} = \left(\sum_{n\ge 0} 2^n t^{\binom{n}{2}} \frac{x^n}{n!}\right)^{\frac{q-1}{2}} \left(\sum_{n\ge 0} t^{n^2} \frac{x^n}{n!}\right).$$

Theorem 5.3. *Let $\mathcal{D}_n$ be the Coxeter arrangement of type D_n in $\mathbb{R}^n$, consisting of the hyperplanes $x_i = x_j$ and $x_i + x_j = 0$ for $1 \le i < j \le n$. We have*

$$\sum_{n\ge 0} \overline{\chi}_{\mathcal{D}_n}(q,t)\frac{x^n}{n!} = \left(\sum_{n\ge 0} 2^n t^{\binom{n}{2}} \frac{x^n}{n!}\right)^{\frac{q-1}{2}} \left(\sum_{n\ge 0} t^{n(n-1)} \frac{x^n}{n!}\right).$$

These results follow fairly easily from Theorem 4.1. Theorem 5.1 is equivalent to a result of Tutte [20], who computed the coboundary polynomial of the complete graph. It is also an immediate consequence of a more general result of Stanley [19, equation (15)]. Theorems 5.2 and 5.3 have never been stated explicitly in the literature, but they are implicit in the work of Zaslavsky [26].

Setting $t = 0$ in Theorems 5.1, 5.2 and 5.3, it is easy to recover the well-known formulas for the characteristic polynomials of the above arrangements:

$$\begin{aligned} \chi_{\mathcal{A}_n}(q) &= q(q-1)(q-2)\cdots(q-n+1), \\ \chi_{\mathcal{B}_n}(q) &= (q-1)(q-3)\cdots(q-2n+1), \\ \chi_{\mathcal{D}_n}(q) &= (q-1)(q-3)\cdots(q-2n+3)(q-n+1). \end{aligned}$$

5.2. **Two more examples.**

Theorem 5.4. *Let $\mathcal{A}_n^{\#}$ be a generic deformation of the arrangement $\mathcal{A}_n$, consisting of the hyperplanes $x_i - x_j = a_{ij}$ $(1 \le i < j \le n)$, where the a_{ij} are generic real numbers* [5]*. For $n \ge 1$,*

$$q\,\overline{\chi}_{\mathcal{A}_n^{\#}}(q,t) = \sum_F q^{n-e(F)}(t-1)^{e(F)}$$

where the sum is over all forests F on $[n]$, and $e(F)$ denotes the number of edges of F. Also,

$$1 + q\sum_{n\ge 1} \overline{\chi}_{\mathcal{A}_n^{\#}}(q,t)\frac{x^n}{n!} = \left(\sum_{n\ge 0} f(n)\frac{x^n(t-1)^n}{n!}\right)^{\frac{q}{t-1}},$$

where $f(n)$ is the number of forests on $[n]$.

Theorem 5.5. *The threshold arrangement $\mathcal{T}_n$ in $\mathbb{R}^n$ consists of the hyperplanes $x_i + x_j = 0$, for $1 \le i < j \le n$. For all $n \ge 0$ we have*

$$\overline{\chi}_{\mathcal{T}_n}(q,t) = \sum_G q^{\mathrm{bc}(G)}(t-1)^{e(G)},$$

where the sum is over all graphs G on $[n]$. Here $\mathrm{bc}(G)$ is the number of connected components of G which are bipartite, and $e(G)$ is the number of edges of G. Also,

$$\sum_{n\ge 0} \overline{\chi}_{\mathcal{T}_n}(q,t)\frac{x^n}{n!} = \left(\sum_{n\ge 0}\sum_{k=0}^{n} \binom{n}{k} t^{k(n-k)} \frac{x^n}{n!}\right)^{\frac{q-1}{2}} \left(\sum_{n\ge 0} t^{\binom{n}{2}} \frac{x^n}{n!}\right).$$

[5] The a_{ij} are "generic" if no n of the hyperplanes have a nonempty intersection, and any nonempty intersection of k hyperplanes has rank k. This can be achieved, for example, by requiring that the a_{ij}'s are linearly independent over the rational numbers. Almost all choices of a_{ij}'s are generic.

5.3. **Deformations of the braid arrangement.** A deformation of the braid arrangement is an arrangement in $\mathbb{R}^n$ consisting of the hyperplanes $x_i - x_j = a_{ij}^{(1)}, \ldots, a_{ij}^{(k_{ij})}$ for $1 \leq i < j \leq n$, where the k_{ij} are nonnegative integers, and the $a_{ij}^{(r)}$ are real numbers. Such arrangements have been studied extensively by Athanasiadis [2] and Postnikov and Stanley [13]. In this section we study their coboundary polynomials.

The most natural deformations of the braid arrangement are the following. Fix a set A of k distinct integers $a_1 < \ldots < a_k$. Let $\mathcal{E}_n$ be the arrangement in $\mathbb{R}^n$ consisting of the hyperplanes

$$x_i - x_j = a_1, \ldots, a_k, \qquad 1 \leq i < j \leq n. \tag{5.1}$$

A *graded graph* is a triple $G = (V, E, h)$, where V is a linearly ordered set of vertices (usually $V = [n]$), E is a set of (nonoriented) edges, and h_G is a function $h_G : V \to \mathbb{N}$, called a *grading*. We will drop the subscript when the underlying graded graph is clear, and we will refer to $h(v)$ as the *height* of v. The *height* of G, denoted $h(G)$, is the largest height of a vertex of G. The vertices v such that $h(v) = r$ form the r-th *level* of G, for each $r \geq 0$. If $u < v$ are connected by edge e, the *slope* of e is $s(e) = h(u) - h(v)$. A graded graph is an *A-graph* if the slopes of all edges of G are in A. A graded graph is *planted* if each connected component has a vertex on the 0-th level.

Recall that, for a graph G, we let $e(G)$ be the number of edges and $c(G)$ be the number of connected components of G. We also let $v(G)$ be the number of vertices of G.

Proposition 5.6. *Let $\mathcal{E}_n$ be the arrangement (5.1). Then, for $n \geq 1$,*

$$q\,\overline{\chi}_{\mathcal{E}_n}(q,t) = \sum_G q^{c(G)}(t-1)^{e(G)},$$

where the sum is over all planted graded A-graphs on $[n]$.

Proof. We associate to each planted graded A-graph $G = (V, E, h)$ on $[n]$ an arrangement $\mathcal{A}_G$ in $\mathbb{R}^n$. It consists of the hyperplanes $x_i - x_j = h(i) - h(j)$, for each $i < j$ such that ij is an edge in G. This is a subarrangement of $\mathcal{E}_n$ because $h(i) - h(j)$, the slope of edge ij, is in A. It is central because the point $(h(1), \ldots, h(n)) \in \mathbb{R}^n$ belongs to all these hyperplanes.

This is in fact a bijection between planted graded A-graphs on $[n]$ and central subarrangements of $\mathcal{E}_n$. To see this, take a central subarrangement $\mathcal{A}$. We will recover the planted graded A-graph G that it came from. For each pair (i,j) with $1 \leq i < j \leq n$, $\mathcal{A}$ can have at most one hyperplane of the form $x_i - x_j = a_t$. If this hyperplane is in $\mathcal{A}$, we must put edge ij in G, and demand that the heights $h(i)$ and $h(j)$ satisfy $h(i) - h(j) = a_t$. When we do this for all the hyperplanes in $\mathcal{A}$, the height requirements that we introduce are consistent, because $\mathcal{A}$ is central. However, these requirements do not fully determine the heights of the vertices; they only determine the relative heights within each connected component of G. Since we want G to be planted, we demand that the vertices with the lowest height in each connected component of G should have height 0. This does determine G completely, and clearly $\mathcal{A} = \mathcal{A}_G$.

With this bijection in hand, and keeping equation (4.1) in mind, we just need that $r(\mathcal{A}_G) = n - c(G)$ and $|\mathcal{A}_G| = e(G)$. Both of these claims are easy. $\square$

Theorem 5.7. *Let $\mathcal{E}_n$ be the arrangement (5.1), and let*

$$A_r(t,x) = \sum_{n \geq 0} \left(\sum_{f:[n] \to [r]} t^{a(f)} \right) \frac{x^n}{n!}, \tag{5.2}$$

where $a(f)$ denotes the number of pairs (i,j) with $1 \le i < j \le n$ such that $f(i) - f(j) \in A$. Then

$$1 + q \sum_{n \ge 1} \overline{\chi}_{\mathcal{E}_n}(q,t) \frac{x^n}{n!} = \left(\lim_{r \to \infty} \frac{A_r(t,x)}{A_{r-1}(t,x)} \right)^q. \tag{5.3}$$

Remark. The limit in (5.3) is a limit in the sense of convergence of formal power series. For more information on this notion of convergence, see [17, Section 1.1] or [9].

Proof of Theorem 5.7. First we prove that

$$A_r(t,x) = \sum_G (t-1)^{e(G)} \frac{x^{v(G)}}{v(G)!} \tag{5.4}$$

where the sum is over all graded A-graphs G of height less than r. The coefficient of $\frac{x^n}{n!}$ in the right-hand side of (5.4) is $\sum_G (t-1)^{e(G)}$, summing over all graded A-graphs G on $[n]$ with height less than r. We have

$$\begin{aligned} \sum_G (t-1)^{e(G)} &= \sum_{h:[n]\to[0,r-1]} \sum_{\substack{G \text{ such that} \\ h_G = h}} (t-1)^{e(G)} \\ &= \sum_{h:[n]\to[0,r-1]} (1 + (t-1))^{a(h)} \\ &= \sum_{f:[n]\to[r]} t^{a(f)} \end{aligned}$$

The only tricky step here is the second: if we want all graded A-graphs G on $[n]$ with a specified grading h, we need to consider the possible choices of edges of the graph. Any edge ij can belong to the graph, as long as $h(i) - h(j) \in A$, so there are $a(h)$ possible edges.

Equation (5.4) suggests the following definitions. Let

$$B_r(t,x) = \sum_G t^{e(G)} \frac{x^{v(G)}}{v(G)!}$$

where the sum is over all *planted* graded A-graphs G of height less than r, and let

$$B(t,x) = \sum_G t^{e(G)} \frac{x^{v(G)}}{v(G)!}$$

where the sum is over all *planted* graded A-graphs G.

The equation

$$1 + q \sum_{n \ge 1} \overline{\chi}_{\mathcal{E}_n}(q,t) \frac{x^n}{n!} = B(t-1,x)^q, \tag{5.5}$$

follows from Proposition 5.6, using the compositional formula for exponential generating functions.

It is not difficult to see that $B(t-1,x) = \lim_{r\to\infty} B_r(t-1,x)$, so it suffices to show that

$$B_r(t-1,x) = A_r(t,x)/A_{r-1}(t,x) \tag{5.6}$$

or, equivalently, that $A_r(t,x) = B_r(t-1,x)A_{r-1}(t,x)$. We need to show that the ways of putting the structure of a graded A-graph G with $h(G) < r$ on $[n]$ can be put in correspondence with the ways of doing the following: first splitting $[n]$ into two disjoint sets S_1 and S_2, then putting the structure of a *planted* graded A-graph G_1 with $h(G_1) < r$ on

S_1, and then putting the structure of a graded A-graph G_2 with $h(G_2) < r-1$ on S_2. We also need that, in that correspondence, $(t-1)^{e(G)} = (t-1)^{e(G_1)}(t-1)^{e(G_2)}$.

We do this as follows. Let G be a graded A-graph G with $h(G) < r$. Let G_1 be the union of the connected components of G which contain a vertex on the 0-th level. Put a grading on G_1 by defining $h_{G_1}(v) = h_G(v)$ for $v \in G_1$. Let $G_2 = G - G_1$. Since $h_G(v) \geq 1$ for all $v \in G_2$, we can put a grading on G_2 by defining $h_{G_2}(v) = h_G(v) - 1$ for $v \in G_2$. G_1 is a planted graded A-graph with $h(G_1) < r$, and G_2 is a graded A-graph with $h(G_2) < r-1$.

It is clear that our map from G to a pair (G_1, G_2) is a one-to-one correspondence and that $(t-1)^{e(G)} = (t-1)^{e(G_1)}(t-1)^{e(G_2)}$. This completes the proof of (5.6), and Theorem 5.7 follows. □

The *Catalan arrangement* C_n in $\mathbb{R}^n$ consists of the hyperplanes

$$x_i - x_j = -1, 0, 1, \qquad 1 \leq i < j \leq n. \tag{5.7}$$

When the arrangement in Theorem 5.7 is a subarrangement of the Catalan arrangement, we can say more about the power series A_r of equation (5.2). Let

$$A(t,x,y) = \sum_r A_r(t,x)y^r = \sum_{n\geq 0}\sum_{r\geq 0}\left(\sum_{f:[n]\to[r]} t^{a(f)}\right)\frac{x^n}{n!}y^r \tag{5.8}$$

be the generating function for the power series in Theorem 5.7, and let

$$S(t,x,y) = \sum_{n\geq 0}\sum_{r\geq 0}\left(\sum_{f:[n]\to[r]} t^{a(f)}\right)\frac{x^n}{n!}y^r \tag{5.9}$$

where the inner sum is over all *surjective* functions $f : [n] \to [r]$. The following proposition reduces the computation of $A(t,x,y)$ to the computation of $S(t,x,y)$, which is easier in practice.

Proposition 5.8. *If $A \subseteq \{-1, 0, 1\}$ in the notation of Theorem 5.7, we have*

$$A(t,x,y) = \frac{S(t,x,y)}{1 - yS(t,x,y)}.$$

Considering the different subsets of $\{-1,0,1\}$, we get six nonisomorphic subarrangements of the Catalan arrangement. They come from the subsets $\emptyset$, $\{0\}$, $\{1\}$, $\{0,1\}$, $\{-1,1\}$ and $\{-1,0,1\}$. The corresponding subarrangements are the empty arrangement, the braid arrangement, the *Linial arrangement*, the *Shi arrangement*, the *interval arrangement* and the Catalan arrangement, respectively. The empty arrangement is trivial, and the braid arrangement was already treated in detail in Section 5.1. We now have a technique that lets us talk about the remaining four arrangements under the same framework. We will do this in the remainder of this section.

5.3.1. *The Linial arrangement.* The Linial arrangement $\mathcal{L}_n$ consists of the hyperplanes $x_i - x_j = 1$ for $1 \leq i < j \leq n$. This arrangement was first considered by Linial and Ravid. It was later studied by Athanasiadis [1] and Postnikov and Stanley [13], who independently computed its characteristic polynomial:

$$\chi_{\mathcal{L}_n}(q) = \frac{q}{2^n}\sum_{k=0}^{n}\binom{n}{k}(q-k)^{n-1}.$$

They also put the regions of $\mathcal{L}_n$ in bijection with several different sets of combinatorial objects. Perhaps the simplest such set is the set of *alternating trees* on $[n+1]$: the set of trees such that every vertex is either larger or smaller than all its neighbors.

Now we present the consequences of Proposition 5.6, Theorem 5.7 and Proposition 5.8 for the Linial arrangement. Recall that a poset P on $[n]$ is *naturally labeled* if $i < j$ in P implies $i < j$ in $\mathbb{Z}^+$.

Proposition 5.9. *For all $n \geq 1$ we have*

$$q\,\overline{\chi}_{\mathcal{L}_n}(q,t) = \sum_P q^{c(P)}(t-1)^{e(P)}$$

where the sum is over all naturally labeled, graded posets P on $[n]$. Here $c(P)$ and $e(P)$ denote the number of components and edges of the Hasse diagram of P, respectively.

Theorem 5.10. *Let*

$$\frac{1+ye^{x(1+y)}}{1-y^2e^{x(1+y)}} = \sum_{r\geq 0} A_r(x)y^r. \tag{5.10}$$

Then we have

$$\sum_{n\geq 0} \chi_{\mathcal{L}_n}(q)\frac{x^n}{n!} = \left(\lim_{r\to\infty}\frac{A_r(x)}{A_{r-1}(x)}\right)^q.$$

In particular, if f_n is the number of alternating trees on $[n+1]$,

$$\sum_{n\geq 0}(-1)^n f_n\frac{x^n}{n!} = \lim_{r\to\infty}\frac{A_{r-1}(x)}{A_r(x)}.$$

Proof. In view of Theorem 5.7 and Proposition 5.8, the exponential generating functions that we are after are determined by $S(0,x,y)$. From equation (5.9), the coefficient of $\frac{x^n}{n!}y^r$ in $S(0,x,y)$ is equal to the number of surjective functions $f:[n]\to[r]$ which never have $f(i)-f(j)=1$ for $i<j$. These are just the nondecreasing surjective functions $f:[n]\to[r]$. For $n\geq 1$ there are $\binom{n-1}{r-1}$ such functions. For $n=0$ and $r\geq 1$ there are no such functions, and for $n=r=0$ there is one such function. Therefore

$$\begin{aligned}
S(0,x,y) &= 1+\sum_{n\geq 1}\sum_{r\geq 1}\binom{n-1}{r-1}\frac{x^n}{n!}y^r \\
&= 1+\sum_{n\geq 1}\frac{x^n}{n!}y(1+y)^{n-1} \\
&= \frac{1+ye^{x(1+y)}}{1+y}.
\end{aligned}$$

Proposition 5.8 then implies that

$$A(0,x,y) = \frac{1+ye^{x(1+y)}}{1-y^2e^{x(1+y)}},$$

in agreement with equation (5.10). The theorem then follows since $\chi_{\mathcal{L}_n}(q) = q\overline{\chi}_{\mathcal{L}_n}(q,0)$, and the number of regions of $\mathcal{L}_n$ is $f_n = (-1)^n\chi_{\mathcal{L}_n}(-1)$ by Theorem 2.1. $\square$

Now we apply our machinery to the Shi arrangement, the interval arrangement and the Catalan arrangement.

The Shi arrangement $\mathcal{S}_n$ consists of the hyperplanes $x_i - x_j = 0, 1$ for $1 \leq i < j \leq n$. Its number of regions is $(n+1)^{n-1}$, the number of labeled trees on $n+1$ vertices; its characteristic polynomial is $\chi_{\mathcal{S}_n}(q) = q(q-n)^{n-1}$ [15], [16].

Theorem 5.11. *Let*

$$A_r(x) = \sum_{n=0}^{r} (r-n)^n \frac{x^n}{n!}.$$

Then we have

$$\sum_{n\geq 0} \chi_{\mathcal{S}_n}(q) \frac{x^n}{n!} = \left(\lim_{r\to\infty} \frac{A_r(x)}{A_{r-1}(x)} \right)^q.$$

In particular,

$$\sum_{n\geq 0} (-1)^n (n+1)^{n-1} \frac{x^n}{n!} = \lim_{r\to\infty} \frac{A_{r-1}(x)}{A_r(x)}.$$

A *semiorder on* $[n]$ is a poset P on $[n]$ for which there exist n unit intervals $I_1, \ldots, I_n$ of $\mathbb{R}$, such that $i < j$ in P if and only if I_i is disjoint from I_j and to the left of it. Let i_n be the number of semiorders on $[n]$.

The interval arrangement $\mathcal{I}_n$ consists of the hyperplanes $x_i - x_j = -1, 1$ for $1 \leq i < j \leq n$. Its number of regions is i_n [18],[13].

Theorem 5.12. *Let*

$$\frac{1-y+ye^x}{1-y+y^2-y^2e^x} = \sum_{r\geq 0} A_r(x) y^r.$$

Then we have

$$\sum_{n\geq 0} \chi_{\mathcal{I}_n}(q) \frac{x^n}{n!} = \left(\lim_{r\to\infty} \frac{A_r(x)}{A_{r-1}(x)} \right)^q.$$

In particular,

$$\sum_{n\geq 0} (-1)^n i_n \frac{x^n}{n!} = \lim_{r\to\infty} \frac{A_{r-1}(x)}{A_r(x)}.$$

The Catalan arrangement $\mathcal{C}_n$ consists of the hyperplanes $x_i - x_j = -1, 0, 1$ for $1 \leq i < j \leq n$. Its number of regions is $n!\, C_n$, where $C_n = \frac{1}{n+1}\binom{2n}{n}$ is the n-th Catalan number [18].

Theorem 5.13. *Let*

$$A_r(x) = \sum_{n=0}^{\lfloor \frac{r+1}{2} \rfloor} \binom{r-n+1}{n} x^n.$$

Then we have

$$\sum_{n\geq 0} \chi_{\mathcal{C}_n}(q) \frac{x^n}{n!} = \left(\lim_{r\to\infty} \frac{A_r(x)}{A_{r-1}(x)} \right)^q.$$

In particular,

$$\frac{\sqrt{1+4x}-1}{2x} = \sum_{n\geq 0} (-1)^n C_n x^n = \lim_{r\to\infty} \frac{A_{r-1}(x)}{A_r(x)}. \tag{5.11}$$

References

[1] C. A. Athanasiadis. Characteristic polynomials of subspace arrangements and finite fields, *Adv. Math.* **122** (1996), 193-233.

[2] C. A. Athanasiadis. Deformations of Coxeter hyperplane arrangements and their characteristic polynomials, in M. Falk and H. Terao (eds.), *Arrangements - Tokyo 1998*, Advanced Studies in Pure Mathematics, in press.

[3] A. Björner, M. Las Vergnas, B. Sturmfels, N. White and G. M. Ziegler. *Oriented matroids*, Encyclopedia of Mathematics and Its Applications 46, Cambridge University Press, Cambridge, 1993.

[4] T. Brylawski. A decomposition for combinatorial geometries, *Trans. Amer. Math. Soc.* **171** (1972), 235-82.

[5] T. Brylawski and J. Oxley. The Tutte polynomial and its applications, in N. White (ed.), *Matroid Applications*, Encyclopedia of Mathematics and Its Applications, 40, Cambridge Univ. Press, Cambridge, 1992, 123-225.

[6] H. H. Crapo. The Tutte polynomial, *Aequationes Math.* **3** (1969), 211-229.

[7] C. Greene. Weight enumeration and the geometry of linear codes, *Stud. Appl. Math.* **55** (1976), 119-128.

[8] J. Humphreys. *Reflection groups and Coxeter groups*, Cambridge Studies in Advanced Mathematics 29, Cambridge University Press, Cambridge, 1990.

[9] I. Niven. Formal power series, *Amer. Math. Monthly*, **76** (1969), 871-889.

[10] P. Orlik and L. Solomon. Combinatorics and topology of complements of hyperplanes, *Invent. Math.* **56** (1980), 167-189.

[11] P. Orlik and H. Terao. *Arrangements of hyperplanes*, Springer-Verlag, Berlin/Heidelberg/New York, 1992.

[12] J. G. Oxley and D. J. A. Welsh. The Tutte polynomial and percolation, in J.A. Bondy and U.S.R. Murty (eds.), *Graph theory and related topics*, Academic Press, New York, 1979, 329-339.

[13] A. Postnikov, R. Stanley. Deformations of Coxeter hyperplane arrangements, *J. Combin. Theory, Ser. A.* **91** (2000), 544-597.

[14] V. Reiner. An interpretation for the Tutte polynomial, *European J. Combin.* **20** (1999), 149-161.

[15] J.-Y. Shi. *The Kazhdan-Lusztig cells in certain affine Weyl groups*, Lecture Notes in Mathematics, no. 1179, Springer, Berlin/Heidelberg/New York, 1986.

[16] J.-Y. Shi. Sign types corresponding to an affine Weyl group, *J. London Math. Soc.* **35** (1987), 56-74.

[17] R. P. Stanley. *Enumerative Combinatorics, vol. 1*, Wadsworth and Brooks - Cole, Belmont, CA, 1986; reprinted by Cambridge University Press, Cambridge, 1997.

[18] R. P. Stanley. Hyperplane arrangements, interval orders, and trees. *Proc. Nat. Acad. Sci. U.S.A.* **93** (1996), 2620-2625.

[19] R. P. Stanley. Graph colorings and related symmetric functions: ideas and applications, *Discrete Mathematics* **193** (1998), 267-286.

[20] W. T. Tutte. A contribution to the theory of chromatic polynomials, *Canad. J. of Math.* **6** (1953), 80-91.

[21] M. L. Wachs and J. W. Walker. On geometric semilattices, *Order.* **2** (1986), 367-385.

[22] D. J. A. Welsh. *Matroid theory*, Academic Press, New York, 1976.

[23] D.J.A. Welsh and G.P. Whittle. Arrangements, channel assignments and associated polynomials, *Advances in Applied Mathematics* **23** (1999), 375-406.

[24] H. Whitney. A logical expansion in mathematics, *Bull. Amer. Math. Soc.* **38** (1932), 572-579.

[25] T. Zaslavsky. Facing up to arrangements: face-count formulas for partitions of space by hyperplanes, *Mem. Amer. Math. Soc.*, vol. 1, no. 154, 1975.

[26] T. Zaslavsky. Biased graphs III. Chromatic and dichromatic invariants, *J. Combin. Theory, Ser. B.* **64** (1995), 17-88.

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA
E-mail address: fardila@math.mit.edu

GENERALIZED DYCK EQUATIONS AND MULTICOLOR TREES

DIDIER ARQUÈS AND AND ANNE MICHELI

Abstract. New topological operations are introduced in order to recover in another way the generalized Dyck equation presented by D. Arquès and J.F. Béraud, for the generating function of maps, by decomposing maps topologically and bijectively. Applying repeatedly the operations which allowed to reveal the generalized Dyck equation to the successive transformed maps, a one-to-one correspondence is obtained between maps on any surface and trees whose vertices can be colored with several colors, following rules. This bijection provides us with a coding of these maps.

Résumé. De nouvelles opérations topologiques sont introduites afin de nous permettre de retrouver l'équation de Dyck généralisée aux cartes de genre quelconque donnée par D. Arquès et J.F. Béraud, par des méthodes topologiques et bijectives de décomposition des cartes. En appliquant plusieurs fois les opérations qui nous ont permis de retrouver l'équation de Dyck généralisée aux cartes successives obtenues, on obtient une bijection entre cartes de genre quelconque et des arborescences où les sommets peuvent être coloriés de plusieurs couleurs suivant des règles que nous définirons. Cette bijection nous fournit un codage de ces cartes.

1. Introduction

The enumerative study of maps starts in 1962 with W. Tutte [17, 18], who enumerates rooted planar maps with n edges. Maps can also be described as combinatorics objects [13]. In 1975, R. Cori [8] studies planar maps in this perspective and extends these results with A. Machi [9] to orientable maps. In 1987, D. Arquès [1] determines functional relations satisfied by generating functions of rooted maps on the torus and obtains closed formulas to enumerate these maps by vertices and faces. Several studies follow on maps of greater genus, orientable or not, as for example the papers of E. Bender and E. Canfield [7] and also D. Arquès and A. Giorgetti [4].

The study of rooted maps independently of their genus begins with T.R.S. Walsh and A. Lehman [19]. They give a recursive relation on the number of rooted maps with respect to the number of edges, which does not lead to an explicit enumeration formula of these maps. In 1990, D.M. Jackson and I.T. Visentin [12] use an algebraic approach and obtain a closed formula for the generating functions of orientable rooted maps with respect to the number of edges and vertices.

More recently, D. Arquès and J.F. Béraud [2, 3] determine a functional equation satisfied by the generating functions of rooted maps with respect to the number of edges and vertices, that generalizes the Dyck equation on trees, and express the solution in a continued fraction form. This continued fraction reveals an interesting bijection, since it also enumerates connected fixed-point free involutions. P. Ossona de Mendez and P. Rosenstiehl [16] describe this bijection. From the combinatorial structure they give for rooted maps, they deduce a code for each map with a connected fixed-point free involution.

Topological operations applied to a map such as the removal or the addition of an edge, the fusion of two vertices, modify sometimes the genus of the map. These operations can not therefore be carried out in a systematic way when one works with fixed genus. However these elementary operations make it possible to find new functional equations on maps studied independently of genus and to establish bijections between families of maps.

In Section 2, we recall general definitions on maps. New topological operations are introduced in Section 3, in order to establish in Section 4, a bijection between maps of indifferent genus, and maps of indifferent genus with a root bridge edge, in which a subset of their vertices has been selected. This bijection provides us with a new proof of the generalized Dyck equation on orientable rooted maps given by D. Arquès and J.F. Béraud. They obtain this equation by an analytic resolution of a differential equation satisfied by the generating function of rooted maps. We here present a new proof of this equation, without any transformation on the generating function, but only by transcription of the given bijection. P. Flajolet [11] moreover showed that many continued fractions having integer coefficients can be explained in a purely combinatorial way, and here is another instance of this assertion.
We then give in Section 5, a bijection between orientable rooted maps and a family of trees whose vertices can be colored by several colors according to certain rules, which is deduced from the one presented in Section 4 by successive applications of this bijection. A generalization of the bijection between planar maps and well labelled trees [10], to maps of genus g and well labelled g-trees [14], allowed M. Marcus and B. Vauquelin to obtain a code for maps of genus g by words product of a shuffle of Dyck words with constraints and of a sequence of integers. The bijection enables us to determine a new language coding the maps of indifferent genus.

2. Definitions

Let us recall some definitions used in the sequel (for further details, see for example [8, 9]).

A *topological map* C in an orientable surface Σ of $\mathbb{R}^3$ is a partition of Σ in three finite sets of cells:

1. The set of vertices of C, which is a finite set of dots ;
2. The set of edges of C, which is a finite set of open Jordan arcs, pairwise disjoint, whose extremities are vertices ;
3. The set of faces of C. Each face is simply connected and its border is the union of vertices and edges.

The *genus* of the map C is the genus of Σ. A cell is *incident* to another cell if one is contained in the boundary of the other. A *bridge* is an edge incident on both sides to the same face. We call half-edge an oriented edge of the map.

Let B be the set of half-edges of the map. With each half-edge, one can associate its initial vertex, its final vertex and its underlying edge. α (resp. σ) is the permutation in B associating to each half-edge b its opposite half-edge (resp. the first half-edge met when turning round the initial vertex of b in the positive way of the surface). The cycles of α (resp. σ) represent the edges (resp. the vertices) of the map. The cycles of $\bar{\sigma} = \sigma \circ \alpha$ are the oriented borders of the faces of the map. (B, σ, α) is the *combinatorial definition* of the topological orientable map associated C.

A map $C = (B, \sigma, \alpha)$ is *rooted* if a half-edge $\tilde{b}$ is distinguished. The half-edge $\tilde{b}$ is called the *root half-edge* of C, and its initial vertex is the *root vertex*. C is then defined as the triplet $(\sigma, \alpha, \tilde{b})$. Face $\bar{\sigma}^*(\tilde{b})$ is called the *exterior face* of C. By convention, the one vertex map (one vertex, no edge) is said to be rooted.

Two orientable maps of the same genus are isomorphic if there is a homeomorphism of the surfaces, preserving its orientation, mapping vertices, edges and faces of one map onto vertices, edges and faces respectively of the other map. An isomorphic class of orientable rooted maps of genus g will simply be called an orientable rooted map.

Let us denote by $\{p\}$ the one vertex map, $\mathcal{M}$ the set of orientable rooted maps, $\mathcal{I}$ the subset of $\mathcal{M}$ of maps with a bridge root edge, and for any map $I \in \mathcal{I}$, $Right(I)$ (resp. $Left(I)$) the maximal submap of I incident to the root vertex (resp. the final vertex of $\tilde{b}$) such that the root half-edge $\tilde{b}$ (resp. $\alpha(\tilde{b})$) of I does not belong to $Right(I)$ (resp. $Left(I)$) (see Figure 3).

3. Preliminaries

In Section 3.1, we describe two algorithms of half-edges and vertices numbering of a map. Numbering induces an order relation on half-edges and vertices that allows us to define in Section 3.2, new topological operations on maps. These operations will be useful to prove Theorem 1. These two operations are reciprocal, and they are interesting since the derivation allows to gather in one vertex a subset of vertices of a map, and the integration allows to get back to this subset of vertices.

3.1. **Order relations in a rooted map.** Order relations on half-edges and vertices of a map are introduced in this Section. We show an algorithm that explains how to traverse a map along its half-edges: they are numbered beginning with the root half-edge and in their order of appearance in the oriented circuit given by the algorithm (see map C in Figure 1. Each number appears near the initial vertex of the half-edge). Half-edges are then naturally ordered by their number.

The root half-edge $\tilde{b}$ gets number 0, then the other half-edges of its face, $\bar{\sigma}^*(\tilde{b})$, are numbered. Afterwards while there still are numberless half-edges:

- Among numbered half-edges, the smallest half-edge b is chosen with a numberless opposite half-edge.
- Along the face $\bar{\sigma}^*(\alpha(b))$, beginning with $\alpha(b)$, half-edges are numbered.

Definition 1. Order relation on vertices. Let C be a rooted map and s_1, s_2 two vertices of C. The vertex s_1 is smaller than s_2 if the smallest half-edge of s_1 is smaller than the smallest half-edge of s_2.

Vertices are numbered by this order relation. Number 1 is affected to the root vertex and other vertices are numbered in an ascending order such that if vertex v_1 is encountered in the traversal of the map earlier than vertex v_2, its number must be smaller than the number of v_2 (see numbers in bold on map C of Figure 1).

A map is *ordered* when its half-edges and vertices are numbered by the algorithms given above.

Definition 2. Path and subpath of a map. The *path* of an ordered map C corresponds to the increasing ordered sequence of the half-edges of C, starting from its root half-edge. A *subpath* of C is defined as an increasing subsequence of ordered and successive half-edges of C.

Property 1. On the smallest half-edges of a face and of a vertex of an ordered map. *The smallest half-edge b_s of a vertex s different from the root vertex, of an ordered map $C = (\sigma, \alpha, \tilde{b})$, is not the smallest half-edge of its face $\bar{\sigma}^*(b_s)$.*
The smallest half-edge b_f of a face f different from the exterior face, of an ordered map $C = (\sigma, \alpha, \tilde{b})$, is not the smallest half-edge of its initial vertex.

Proof of property 1. If b_s belongs to the exterior face of C, as s is different from the root vertex, we have $\tilde{b} < b_s$ and b_s cannot be the smallest half-edge of its face.
If b_s does not belong to the exterior face of C, half-edges of face $\bar{\sigma}^*(\alpha(b_s))$ have been

numbered before b_s (see the algorithm above). Thus $\alpha(b_s)$ is smaller than b_s. Then $\bar{\sigma}(\alpha(b_s)) = \sigma(b_s)$, which belongs to vertex s, is smaller than b_s.

Since b_f is the smallest half-edge of face f, the half-edge $\alpha(b_f)$ is smaller than b_f and $\bar{\sigma}(\alpha(b_f)) = \sigma(b_f)$, which belongs to the initial vertex of b_f, is smaller than b_f $\diamond$

3.2. **Topological and bijective operations on maps.** In Section 3.2.1, we define the *derivation* operation, that gathers a subset of vertices of a map and the root vertex of a second map, in one vertex. These vertices can be recovered by applying the inverse operation, called *integration* operation and defined in Section 3.2.2, which uses the order properties on a map to get back all the gathered vertices. These operations are the main tools used in the proof of Theorem 1.

Let us denote by $\mathcal{M}_2$ the subset of maps of $\mathcal{M}$ that have at least two distinct vertices.

3.2.1. *Derivation of maps.* In this Section, we define a derived map of a pair of maps (C, R) of $\mathcal{M}_2 \times \mathcal{M}$ with respect to certain vertices of C. To derive a pair of maps with respect to vertices $s_1, \ldots, s_m$ of C means to collect these vertices in one vertex while respecting an order and afterwards to glue this vertex to the root vertex of R, as described in definition below.

Definition 3. Derived map. Let $C = (\sigma, \alpha, \tilde{b})$ be a map of $\mathcal{M}_2$, with root vertex $\tilde{s}_C$ and $R = (\sigma_R, \alpha)$ be a map of $\mathcal{M}$, with root vertex $\tilde{s}_R$ and if $R \neq \{p\}$, let $(b_{\tilde{s}_R,1}, b_{\tilde{s}_R,2}, \ldots, b_{\tilde{s}_R,l_{\tilde{s}_R}})$ be the half-edges of $\tilde{s}_R$ and $b_{\tilde{s}_R,1}$ be the root half-edge of R. Let $\mathscr{S} = \{s_1, \ldots, s_m\}$ be a set of distinct vertices of C such that $\tilde{s}_C < s_1 < s_2 < \ldots < s_m$. Forall i in $[1, m]$, let $(b_{s_i,1}, \ldots, b_{s_i,l_{s_i}}) = \sigma^*(b_{s_i,1})$, be the half-edges of initial vertex s_i, in which $b_{s_i,1}$ is the smallest half-edge of s_i.

The *derived map* $C'_{\mathscr{S},R} = (\sigma', \alpha, \tilde{b})$ of $(C, \mathscr{S}, R)$ is then the map obtained from C and R after the gathering in a unique vertex s, of the vertices of $\mathscr{S} \bigcup \{\tilde{s}_R\}$ in the following way (see Figure 1):

$$s = (\underbrace{b_{s_1,1}, \ldots, b_{s_1,l_{s_1}}}_{s_1}, \underbrace{b_{s_2,1}, \ldots, b_{s_2,l_{s_2}}}_{s_2}, \ldots, \underbrace{b_{s_m,1}, \ldots, b_{s_m,l_{s_m}}}_{s_m}, \underbrace{b_{\tilde{s}_R,1}, \ldots, b_{\tilde{s}_R,l_{\tilde{s}_R}}}_{\tilde{s}_R}) = \sigma'^*(b_{s_1,1}).$$

In terms of permutation, it means: $\sigma' = \tau_{1R}\tau_{1m} \ldots \tau_{12}\sigma = \gamma\sigma$ with $\tau_{1i} = (b_{s_1,1} b_{s_i,1})$, $\tau_{1R} = (b_{s_1,1} b_{\tilde{s}_R,1})$ and $\gamma = (b_{s_1,1} \ldots b_{s_m,1} b_{\tilde{s}_R,1})$.

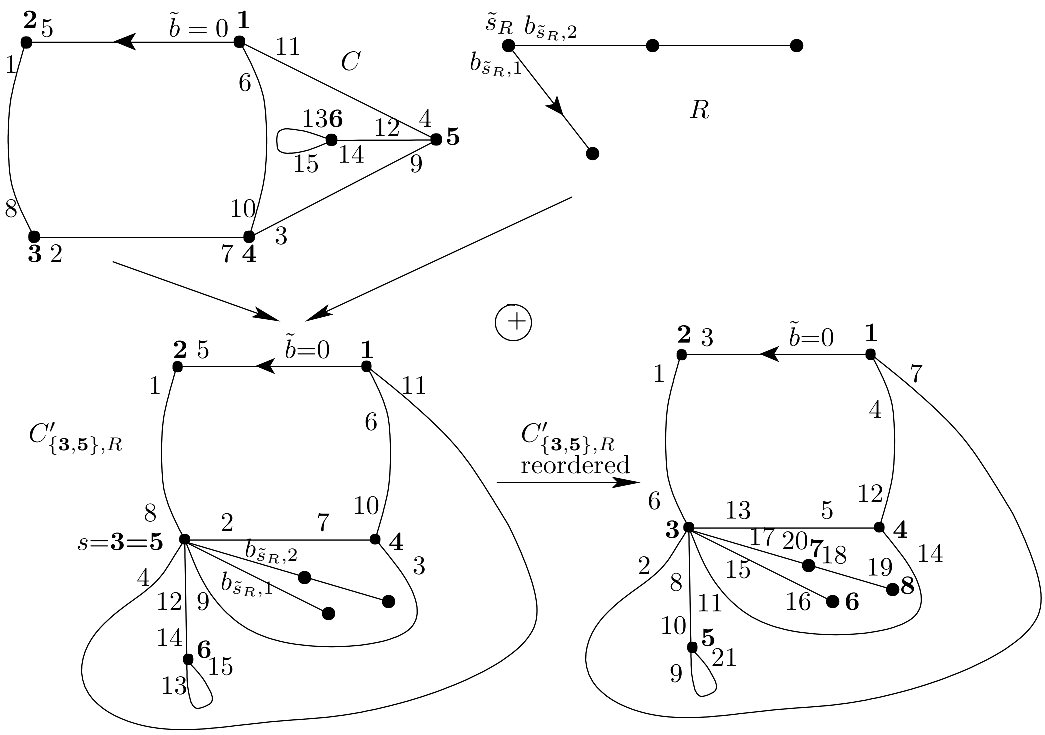
Property 2. Orders of $C'_{\mathscr{S},R}$, of C and of R

(1) *In $C'_{\mathscr{S},R}$, if $R \neq \{p\}$, $b_{\tilde{s}_R,1}$ is the smallest half-edge among the half-edges of R (see Figure 1 in which $b_{\tilde{s}_R,1} = 15$).*
(2) *The half-edges smaller than or equal to $\alpha(b_{s_1,l_{s_1}})$ have the same numbering in the ordered maps $C'_{\mathscr{S},R}$ and C.*

Proof of property 2. (1) By construction, R is recovered if in $C'_{\mathscr{S},R}$, the subset of half-edges belonging also to R, i.e. $\{b_{\tilde{s}_R,1}, \ldots, b_{\tilde{s}_R,l_{\tilde{s}_R}}\}$, is unglued from vertex s. Thus in the traversal of $C'_{\mathscr{S},R}$, starting from its root half-edge, $\tilde{b}$, to reach any half-edge of R, one has to cross s. It implies that there exists i, $1 \leq i \leq l_{\tilde{s}_R}$ such that $b_{\tilde{s}_R,i}$ is the smallest half-edge of the half-edges of R in $C'_{\mathscr{S},R}$. If $l_{\tilde{s}_R} > 1$, let us prove that $b_{\tilde{s}_R,1}$ is the smallest half-edge of the half-edges of R in $C'_{\mathscr{S},R}$.

$b_{\tilde{s}_R,i}$ cannot be the smallest half-edge of its face, $\overline{\sigma'}^*(b_{\tilde{s}_R,i})$, otherwise $\alpha(b_{\tilde{s}_R,i})$, which belongs to R and has been previously numbered to the face $\overline{\sigma'}^*(b_{\tilde{s}_R,i})$, is smaller than $b_{\tilde{s}_R,i}$.

If $i > 1$, $b_{\tilde{s}_R,i} = \sigma'(b_{\tilde{s}_R,i-1}) = \overline{\sigma'}(\alpha(b_{\tilde{s}_R,i-1}))$, so that $\alpha(b_{\tilde{s}_R,i-1})$, which belongs to



$$
\begin{array}{lllllll}
 & (b_{s_2,1} & , b_{s_2,2} & , b_{s_2,3}=b_{s_2,l_{s_2}} & , b_{s_1,1} & , b_{s_1,2}=b_{s_1,l_{s_1}} &) \\
= & (4 & , 12 & , 9 & , 2 & , 8 &)
\end{array}
$$

FIGURE 1. Derived map with respect to vertices **3** and **5** of a pair of maps.

R, is smaller than $b_{\tilde{s}_R,i}$ (as $b_{\tilde{s}_R,i}$ is not the smallest half-edge of its face), which contradicts definition of $b_{\tilde{s}_R,i}$. Thus $i=1$.

(2) In C, $s_1 < s_2 < \ldots < s_m$ implies that $b_{s_1,1} < b_{s_2,1} < \ldots < b_{s_m,1}$.
Furthermore, forall i in $[1,m]$, $\bar{\sigma}(\alpha(b_{s_i,l_{s_i}})) = b_{s_i,1}$ and $b_{s_i,1}$ is not the smallest half-edge of its face (see Property 1), so that $\alpha(b_{s_i,l_{s_i}})$ precedes $b_{s_i,1}$ in the ordered map C.
One then has in C, $\tilde{b} < \alpha(b_{s_1,l_{s_1}}) < b_{s_1,1} < \alpha(b_{s_2,l_{s_2}}) < b_{s_2,1} < \ldots < \alpha(b_{s_m,l_{s_m}}) < b_{s_m,1}$.
Thus in C, the subpath from $\tilde{b}$ to $\alpha(b_{s_1,l_{s_1}})$ does not cross any half-edge $\alpha(b_{s_i,l_{s_i}})$.
If one proves that in $C'_{\mathscr{S},R}$, the subpath from $\tilde{b}$ to $\alpha(b_{s_1,l_{s_1}})$ does not cross $\alpha(b_{\tilde{s}_R,l_{\tilde{s}_R}})$, then one will conclude from what precedes that in $C'_{\mathscr{S},R}$, the subpath from $\tilde{b}$ to $\alpha(b_{s_1,l_{s_1}})$ does not cross any of the half-edges $\alpha(b_{s_i,l_{s_i}})$. As

$$
\overline{\sigma'}(a) = \begin{cases} b_{s_{i+1},1} & \text{if } a=\alpha(b_{s_i,l_{s_i}}) \ \ \forall 1 \le i < m \\ b_{\tilde{s}_R,1} & \text{if } a=\alpha(b_{s_m,l_{s_m}}) \\ b_{s_1,1} & \text{if } a=\alpha(b_{\tilde{s}_R,l_{\tilde{s}_R}}) \\ \bar{\sigma}(a) & \text{if } a \in C, a \neq b_{s_i,l_{s_i}} \ \ \forall 1 \le i \le m \\ \overline{\sigma_R}(a) & \text{if } a \in R, a \neq b_{\tilde{s}_R,l_{\tilde{s}_R}} \end{cases},
$$

it means that the subpath from $\tilde{b}$ to $\alpha(b_{s_1,l_{s_1}})$ in $C'_{\mathscr{S},R}$ is unchanged.
Let us then prove that the subpath of $C'_{\mathscr{S},R}$ from $\tilde{b}$ to $\alpha(b_{s_1,l_{s_1}})$ does not cross the half-edge $\alpha(b_{\tilde{s}_R,l_{\tilde{s}_R}})$.

Since $\overline{\sigma'}(\alpha(b_{s_m,l_{s_m}})) = b_{\tilde{s}_R,1}$ and $b_{\tilde{s}_R,1}$ is not the smallest half-edge of its face (see item 1 of this proof), $\alpha(b_{s_m,l_{s_m}})$ precedes $b_{\tilde{s}_R,1}$ in the path of $C'_{\mathscr{S},R}$.
Furthermore, from Property 2.1, $b_{\tilde{s}_R,1} < \alpha(b_{\tilde{s}_R,l_{\tilde{s}_R}})$ as $\alpha(b_{\tilde{s}_R,l_{\tilde{s}_R}}) \in R$ and $b_{\tilde{s}_R,1}$ is the smallest half-edge of R in $C'_{\mathscr{S},R}$. Thus in $C'_{\mathscr{S},R}$, $\alpha(b_{s_1,l_{s_1}}) < \alpha(b_{\tilde{s}_R,l_{\tilde{s}_R}})$ $\diamond$

The following technical lemma gives us the way to recover vertices $s_1, \ldots, s_m, \tilde{s}_R$, which compose vertex s, as will be showed in Lemma 2.
Notations of Definition 3 are used here.

Lemma 1. *In $C'_{\mathscr{S},R}$, $\sigma'(b_{s_1,l_{s_1}}) = \begin{cases} b_{s_2,1} & \text{if } m > 1 \\ b_{\tilde{s}_R,1} & \text{if } R \neq \{p\} \text{ and } m = 1 \\ b_{s_1,1} & \text{if } R = \{p\} \text{ and } m = 1 \end{cases}$ is the smallest half-edge among half-edges of vertex s.*

Proof of lemma 1. (1) If $R = \{p\}$ and $m = 1$ then $C = C'_{\mathscr{S},R}$, $s = s_1$ and thus, $\sigma'(b_{s_1,l_{s_1}}) = b_{s_1,1}$ is the smallest half-edge among the half-edges of s.

(2) Let us assume that $R \neq \{p\}$ or $m \neq 1$. One has

$$\overline{\sigma'}(a) = \begin{cases} b_{s_{i+1},1} & \text{if } a = \alpha(b_{s_i,l_{s_i}}) \;\; \forall 1 \leq i < m \\ b_{\tilde{s}_R,1} & \text{if } a = \alpha(b_{s_m,l_{s_m}}) \\ b_{s_1,1} & \text{if } a = \alpha(b_{\tilde{s}_R,l_{\tilde{s}_R}}) \\ \bar{\sigma}(a) & \text{if } a \in C, a \neq \alpha(b_{s_i,l_{s_i}}) \;\; \forall 1 \leq i \leq m \\ \overline{\sigma_R}(a) & \text{if } a \in R, a \neq \alpha(b_{\tilde{s}_R,l_{\tilde{s}_R}}) \end{cases}.$$

Let $\hat{b}$ be the smallest half-edge of face $\overline{\sigma}^*(b_{s_1,1})$ in C.

(a) In C, $b_{s_1,1}$ is the smallest half-edge of vertex s_1. From Property 1, as $s_1 \neq \tilde{s}_C$, $b_{s_1,1}$ is not the smallest half-edge of its face. It implies that there exists $j > 0$ such that $\bar{\sigma}^j(\hat{b}) = b_{s_1,1}$.

(b) Let us prove at last Lemma 1, that is: $\sigma'(b_{s_1,l_{s_1}})$ is the smallest half-edge of s in $C'_{\mathscr{S},R}$.
From Property 2.2, one knows that the subpath from $\tilde{b}$ to $\alpha(b_{s_1,l_{s_1}})$ in $C'_{\mathscr{S},R}$ is identical to the one in C. Thus $\alpha(b_{s_1,l_{s_1}}) = \bar{\sigma}^{j-1}(\hat{b}) = \overline{\sigma'}^{j-1}(\hat{b})$.
Furthermore, in C, the subpath from $\tilde{b}$ to $\alpha(b_{s_1,l_{s_1}})$ does not cross s as $b_{s_1,1}$ is the smallest half-edge of the half-edges of s in C and $\alpha(b_{s_1,l_{s_1}})$ is smaller than $b_{s_1,1}$ in C (see the proof of Property 2.2). It is the same in $C'_{\mathscr{S},R}$.
Thus $\sigma'(b_{s_1,l_{s_1}}) = \overline{\sigma'}(\alpha(b_{s_1,l_{s_1}}))$ is the smallest half-edge of s in $C'_{\mathscr{S},R}$ $\diamond$

$b_{s_1,1}$ is the smallest half-edge of $\mathscr{S}$ in C. Its predecessor in the path of C, is the half-edge $\alpha(b_{s_1,l_{s_1}})$ as $b_{s_1,1}$ is not the smallest half-edge of its face (see Property 1). In map $C'_{\mathscr{S},R}$, built from C and R by gluing together vertices of C and the root vertex of R in one vertex s, the successor of $\alpha(b_{s_1,l_{s_1}})$ becomes $b_{s_2,1}$, which then is the smallest half-edge of s in $C'_{\mathscr{S},R}$ reordered. If $b_{s_1,1}$ has been marked, one gets back thus vertex s_1 which is detached from s, then recursively vertices $s_2, \ldots, s_m$. Thus the pair of initial maps can be recovered from its derived map. A formal definition of this inverse operation, which will be called integration, is given in the next Section.

3.2.2. *Integration of a map.* A topological operation of opening of a vertex into two vertices is introduced in order to define the integration of a map, which consists in the splitting of a vertex into several vertices. It will then be seen that to recover a pair of maps (C, R) and the subset of vertices of C if its derived map is known, one has to integrate this last map.

Definition 4. Topological operation of opening of a map with respect to a half-edge. Let $C = (\sigma, \alpha, \tilde{b})$ be a map and b a half-edge of C. Let b_s be the smallest half-edge of a vertex $s = \sigma^*(b)$. The *opening* of C with respect to b consists in the splitting of the vertex s into two vertices s_1 and s_2 in the following way:

$$s = (b, \ldots, \sigma^{-1}(b_s), b_s, \ldots, \sigma^{-1}(b)) \quad \rightarrow \quad s_1 = (b, \ldots, \sigma^{-1}(b_s)) \text{ and } s_2 = (b_s, \ldots, \sigma^{-1}(b)).$$

It means that the following permutation $\widehat{\sigma_b}$ is applied to the half-edges of C: $\widehat{\sigma_b} = \tau\sigma$ with $\tau = (bb_s)$.

The result of the opening of C with respect to b is a map or a pair of maps:

(i) If $b_s \neq b$ and if the group generated by $(\widehat{\sigma_b}, \alpha, \tilde{b})$ acts transitively on the set of half-edges of C (i.e. $(\widehat{\sigma_b}, \alpha, \tilde{b})$ generates a map and not two disconnected maps), then a new map $\widehat{C_b} = (\widehat{\sigma_b}, \alpha, \tilde{b})$ is defined.

(ii) Otherwise, a pair of maps $(\widehat{C_b}, D)$, $\widehat{C_b} = (\widehat{\sigma_b}, \alpha, \tilde{b})$, $D = (\widehat{\sigma_b}, \alpha, b_s)$, is obtained, D being the map $\{p\}$ if $b_s = b$.

Remark 1. If $s \neq \tilde{s}$, $\widehat{C_b} \in \mathcal{M}_2$.

The next definition explains that in order to integrate a map C with respect to a given half-edge b, one has to recursively apply this topological operation of opening of C until getting a pair of maps.

Definition 5. Integration of a map. Let $C = (\sigma, \alpha, \tilde{b})$ be a map of $\mathcal{M}_2$, of root vertex $\tilde{s}$. Let $s \neq \tilde{s}$ be a vertex of C and $b \in s$. Let $\mathscr{S} = \emptyset$.

It will be said that a map C *is integrated* with respect to an half-edge b, when the operation of the opening of C is recursively applied until case (ii) of Definition 4 is reached, that is:

- Let us denote by b_s the smallest half-edge of $\sigma^*(b)$, then C is opened with respect to b (see Definition 4).
- If this operation gives a map $\widehat{C_b}$ (see Figure 2, drawing $\boxed{2}$), the vertex obtained after the opening, incident to b (the other obtained vertex is incident to b_s), is added to $\mathscr{S}$ and the opening operation starts again with $C \leftarrow \widehat{C_b}$ and $b \leftarrow b_s$.
- Otherwise, a pair of maps of $\mathcal{M}_2 \times \mathcal{M}$, $(\widehat{C_b}, D)$ is obtained (see Figure 2, drawing $\boxed{3}$), and also a set of vertices of $\widehat{C_b}$, $\mathscr{S}$ with the added vertex of $\widehat{C_b}$ which was split from the root vertex of D (vertex of $\widehat{C_b}$ to which b belongs).

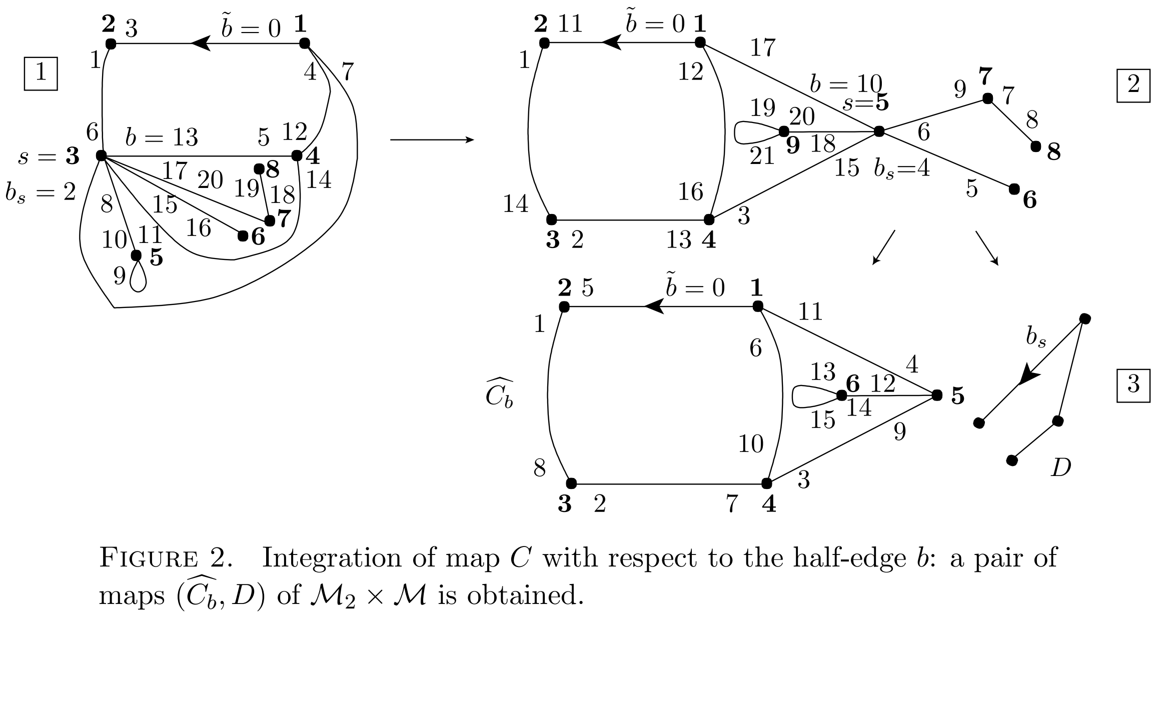


FIGURE 2. Integration of map C with respect to the half-edge b: a pair of maps $(\widehat{C_b}, D)$ of $\mathcal{M}_2 \times \mathcal{M}$ is obtained.

Lemma 2. *Let $C'_{\mathscr{S},R}$, be the derived map of a pair of maps (C,R) of $\mathcal{M}_2 \times \mathcal{M}$ with respect to a set of vertices $\mathscr{S}$ of C. Let us denote by b ($= b_{s_1,1}$ of Definition 3) the smallest half-edge of $\mathscr{S}$ in C. Integration of $C'_{\mathscr{S},R}$ with respect to b gives $(C, \mathscr{S}, R)$.*

Proof of lemma 2. With notations of Definitions 3 and 5, the map $C'_{\mathscr{S},R} = (\sigma', \alpha, \tilde{b})$ is integrated with respect to the half-edge $b_{s_1,1}$: $b = b_{s_1,1}$ and $b_s = b_{s_2,1}$ (from Lemma 1). The opening operation of vertex s unglues vertex s_1 from s, and one gets the map $(\widehat{C'_{\mathscr{S},R}})_b = (\widehat{\sigma'}_b, \alpha, \tilde{b})$:

$$s = (\underset{\substack{\uparrow \\ b}}{b_{s_1,1}}, \ldots, b_{s_1,l_{s_1}}, \underset{\substack{\uparrow \\ b_s}}{b_{s_2,1}}, \ldots, b_{s_2,l_{s_2}}, \ldots, b_{s_m,1}, \ldots, b_{s_m,l_{s_m}}, b_{\tilde{s}_R,1}, \ldots, b_{\tilde{s}_R,l_{\tilde{s}_R}})$$

Two vertices are obtained, a vertex $s_1 = (b_{s_1,1}, \ldots, b_{s_1,l_{s_1}})$ and a vertex $s = (b_{s_2,1}, \ldots, b_{s_2,l_{s_2}}, \ldots, b_{s_m,1}, \ldots, b_{s_m,l_{s_m}}, b_{\tilde{s}_R,1}, \ldots, b_{\tilde{s}_R,l_{\tilde{s}_R}})$. One has: $\widehat{\sigma'}_b = \tau_{12}\sigma'$.
Thus, $(\sigma_2 = \widehat{\sigma'}_b, \alpha, \tilde{b}) = C'_{\{s_2,\ldots,s_m\},R}$ and $\mathscr{S} = \{s_1\}$. One successively obtains maps $C'_{\{s_i,\ldots,s_m\},R} = (\sigma_i = \tau_{i-1i}\sigma_{i-1}, \alpha, \tilde{b})$ for $\tau_{i-1i} = (b_{s_{i-1},1} b_{s_i,1})$, and $\mathscr{S} = \{s_1, \ldots, s_{i-1}\}$, with $3 \le i \le m$. Applying for the last time to $C'_{\{s_m\},R}$ the topological operation of opening of $s = (b_{s_m,1}, \ldots, b_{s_m,l_{s_m}}, b_{\tilde{s}_R,1}, \ldots, b_{\tilde{s}_R,l_{\tilde{s}_R}})$, two disconnected maps, $C = (\sigma, \alpha, \tilde{b})$ and $R = (\sigma, \alpha, b_{\tilde{s}_R,1})$, are recovered and also $\mathscr{S} = \{s_1, \ldots, s_m\}$. One has: $\sigma = \tau_{Rm}\tau_{mm-1}\ldots\tau_{12}\sigma' = \delta\sigma'$ with $\delta = \gamma^{-1}$ (see Definition 3) $\diamond$

4. Generalized Dyck equation on maps of indifferent genus

The well-known Dyck equation on trees, is based on a one-to-one correspondence between rooted planar trees $\mathcal{A}$ deprived of the one vertex tree, and $\mathcal{A}^2$. In Section 4.1, an equation generalizing the Dyck equation to rooted maps studied independently of genus, is given. This equation is equivalent to an equation on sets which is determined. A proof of the equation on sets is given in Section 4.2. Topological operations introduced in Section 3.2 will be used for this demonstration.

4.1. Generalized Dyck equations.

The equation on sets is given as a bijection between the set of rooted maps of indifferent genus, $\mathcal{M}$, and the set of pairs of maps of $\mathcal{M}$, where in one of these maps a subset (possibly empty) of its vertices is selected. Equation (2) is then a translation with generating functions of this bijection.

For any map of $\mathcal{M}$, let us denote by $\mathcal{V}_M$ the set of vertices of M and $\mathcal{P}(\mathcal{V}_M)$ the set of all subsets of $\mathcal{V}_M$.

Theorem 1. Equation on sets

$$\mathcal{M} \leftrightarrow \{p\} \bigcup \left[\bigcup_{M \in \mathcal{M}} M \times \mathcal{P}(\mathcal{V}_M) \right] \times \mathcal{M} \tag{1}$$

The proof of this theorem is given in Section 4.2.

The translation of this bijection with generating functions provides a generalized Dyck equation generalizing the Dyck equation on trees.

Let us denote by y the variable which exponent enumerates the vertices of a map of $\mathcal{M}$, and by z the variable which exponent enumerates the edges of a map of $\mathcal{M}$ and $M(y,z)$ the generating function of rooted maps of indifferent genus.

One gets the following corollary:

Corollary 1. Generalized Dyck equation

$$M(y,z) = y + zM(y,z)M(y+1,z) \tag{2}$$

4.2. **Proof of Theorem 1.** A bijection between maps of $\mathcal{M}$, different from the one vertex map and $\left(\bigcup_{M\in\mathcal{M}} M \times \mathcal{P}(\mathcal{V}_M)\right) \times \mathcal{M}$ is described, which means between maps of $\mathcal{M}$ and maps of $\mathcal{I}$ in which for each map I of $\mathcal{I}$, a set $\mathscr{S}$ of vertices of the submap incident to the final vertex of the root half-edge, $Left(I)$, has been selected. As a matter of fact, $\mathcal{I}$ is in one-to-one correspondence with $\mathcal{M}^2$, as to each map I of $\mathcal{I}$, a pair of maps of $\mathcal{M}^2$, $(Left(I), Right(I))$, can be associated. Furthermore the set of pairs $(Left(I), \mathscr{S})$ is the set $\bigcup_{M\in\mathcal{M}} M \times \mathcal{P}(\mathcal{V}_M)$.

Lemma 3. Bijection of theorem 1. *There is a one-to-one correspondence between $\mathcal{M}$ and the set of pairs $(I, \mathscr{S})$, in which I is a map of $\mathcal{I}$ and $\mathscr{S}$ a set of vertices of $Left(I)$, possibly empty.*

Proof of lemma 3. Integration of a map with respect to an half-edge b allows to recover a pair of maps as well as a set of vertices of one of the obtained maps. Thus when a derived map I' is obtained, to have the possibility of going back, one has to memorize the half-edge b. To do this, if the root vertex of I' is only incident to the root half-edge, then it is sufficient to glue the root half-edge just before b in order to obtain a map M of $\mathcal{M}$.

Starting with a map I of $\mathcal{I}$ in which a set $\mathscr{S}$ of vertices of $Left(I)$ has been selected, we will first see how to obtain a map M of $\mathcal{M}$, and then how to recover map I and its set of vertices $\mathscr{S}$ from M.

Let $I = (\sigma, \alpha, \tilde{b})$ be a map of $\mathcal{I}$ of root vertex $\tilde{s}_I$ (see Figure 3). Let us denote by I_L, the map I deprived of $Right(I)$, with the same root half-edge than I and $\tilde{s}_R = \tilde{s}_I \setminus \{\tilde{b}\}$ the root vertex of $Right(I)$.

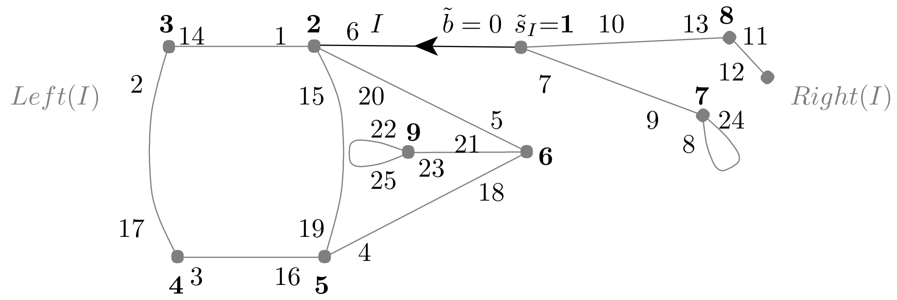


FIGURE 3. Map of $\mathcal{I}$

Stage 1: Derivation of $(I_L, \mathscr{S}, Right(I))$. Let $\mathscr{S}$ be a subset of vertices of $Left(I_L) = Left(I)$.

If $\mathscr{S}$ is not empty, let $\{s_1, \ldots, s_m\}$ be m distinct vertices of $\mathscr{S}$ such that $s_1 < \ldots < s_m$. Forall i in $[1, m]$, let $(b_{s_i,1}, \ldots, b_{s_i,l_{s_i}}) = \sigma^*(b_{s_i,1})$, be the half-edges of initial vertex s_i, in which $b_{s_i,1}$ is the smallest half-edge of s_i. Let $(b_{\tilde{s}_R,1}, \ldots, b_{\tilde{s}_R,l_{\tilde{s}_R}}) = \tilde{s}_R$, with $b_{\tilde{s}_R,1} = \sigma(\tilde{b})$. Let $I' = (I_L)'_{\mathscr{S},Right(I)} = (\sigma', \alpha, \tilde{b})$ be the derived map of $(I_L, Right(I))$ with respect to $\mathscr{S}$. Let us recall that the vertices of $\mathscr{S} \bigcup \{\tilde{s}_R\}$ are joined into one vertex s_d in the following way (see Figure 4):

$$s_d = (\underbrace{b_{s_1,1}, \ldots, b_{s_1,l_{s_1}}}_{s_1}, \underbrace{b_{s_2,1}, \ldots, b_{s_2,l_{s_2}}}_{s_2}, \ldots, \underbrace{b_{s_m,1}, \ldots, b_{s_m,l_{s_m}}}_{s_m}, \underbrace{b_{\tilde{s}_R,1}, \ldots, b_{\tilde{s}_R,l_{\tilde{s}_R}}}_{\tilde{s}_R}) = \sigma'^*(b_{s_1,1}).$$

If $\mathscr{S}$ is empty, i.e. $m = 0$, then $I' = I$.

Stage 2 : Labelization of $b_{s_1,1}$, getting of a map of $\mathcal{M}$.

- If $\mathscr{S} = \emptyset$, $I' = I$ ($(I_L, Right(I))$ has been derived with respect to no vertex) and $M = I' = I$.
- Otherwise a map $M = (\sigma_M, \alpha, \tilde{b})$ is built (see Figure 4), gluing the root vertex of $\tilde{b}$ to the vertex s_d in the following way:

  $(\underbrace{b_{s_1,1}, \ldots, b_{s_1,l_{s_1}}}_{s_1}, \underbrace{b_{s_2,1}, \ldots, b_{s_2,l_{s_2}}}_{s_2}, \ldots, \underbrace{b_{s_m,1}, \ldots, b_{s_m,l_{s_m}}}_{s_m}, \underbrace{b_{\tilde{s}_R,1}, \ldots, b_{\tilde{s}_R,l_{\tilde{s}_R}}}_{\tilde{s}_R}, \widetilde{\mathbf{b}})$
  $= {\sigma_M}^*(\tilde{b})$

  The following permutation is then applied to the half-edges of I': $\sigma_M = (\tilde{b} b_{s_1,1})\sigma'$.

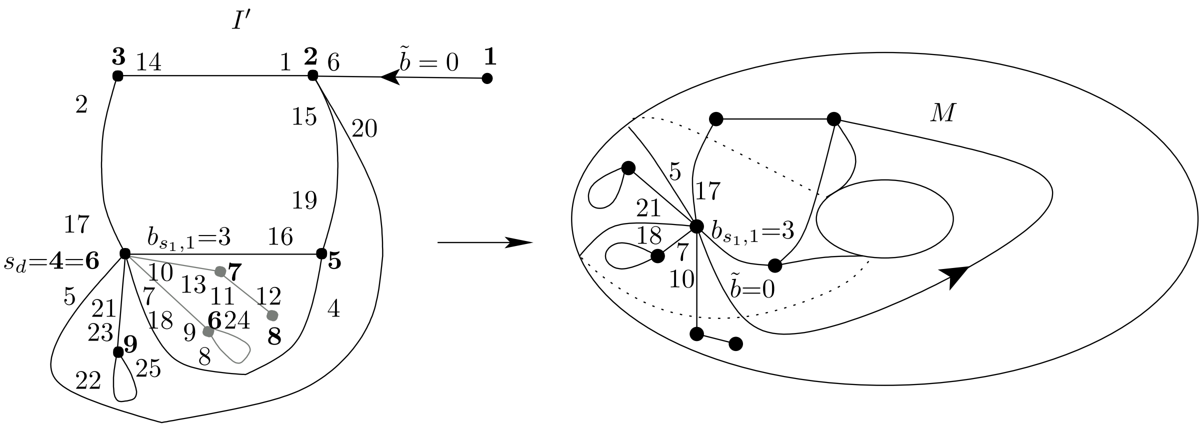


FIGURE 4. Getting of map $I' = (I_L)'_{\{\mathbf{4},\mathbf{6}\},Right(I)}$ and of a map M of $\mathcal{M}$ from map I' and its half-edge $b_{s_1,1}$ (I' and M have not been reordered)

Remark 2. If $\mathscr{S} = \emptyset$, $M = I$ and if I is a tree, then M is a tree.

Recovering of $(I, \mathscr{S})$ from M. If the map M obtained belongs to $\mathcal{I}$, it means that $\mathscr{S}$ was empty and then $M = I$. Thus to get back I from M, nothing has to be done. Let us remark that, thanks to Remark 2, when we restrict ourselves to the case of trees, one recovers the decomposition induced by the Dyck equation on trees.

Let us assume that M does not belong to $\mathcal{I}$. Then $\sigma_M(\tilde{b}) = b_{s_1,1}$. In map I, $b_{s_1,1}$ is the smallest half-edge among the half-edges incident to the vertices of $\mathscr{S}$. Conditions of Lemma 2 are satisfied, and one can apply this lemma. Thus, to recover $(I, \mathscr{S})$ from M, one has to:

- unglue $\tilde{b}$ from the root vertex,
- integrate this new map M_1 with respect to $b_{s_1,1} = \sigma_M(\tilde{b})$. From Lemma 2, one gets back I_L and $Right(I)$, respectively rooted in $\tilde{b}$ and $b_{\tilde{s}_R,1}$, and $\mathscr{S}$.
- Then $Right(I)$ is glued to the root vertex of I_L, such that $\sigma(\tilde{b}) = b_{\tilde{s}_R,1}$ $\diamond$

5. Bijection between maps of indifferent genus and multicolor trees

The operation that allowed to prove Theorem 1 transforms a map of $\mathcal{M}$ into a map with a bridge root edge in which a subset of its vertices has been selected. If this operation is iterated on the successive submaps incident to the two vertices incident to the bridge half-edge, and if the subset of vertices associated with each map is colored (one distinct color for

each subset), the initial map is transformed into a tree whose vertices can be colored with several colors, following repartition rules. One then obtains what we will call a *multicolor tree*.
In Section 5.1, we give the definition of a multicolor tree and in Section 5.2, we prove the one-to-one correspondence between maps of $\mathcal{M}$ and multicolor trees. This bijection leads to a coding of maps by words of language, as shown in Section 5.3.

5.1. **Multicolor trees.** We give the definition of a multicolor tree, called hereafter simply multicolor tree. We then define a one-to-one correspondence in Section 5.2 between multicolor trees and maps of indifferent genus. These multicolor trees are trees whose vertices can be colored with several colors, following repartition rules that will be defined. Order relations given in Section 3.1 are applied to multicolor trees. An order on half-edges and vertices is thus established in a classical in-depth descent of the tree. Let us notice that the smallest half-edge of a vertex is also its left son in the tree structure, since a tree has only one face.

Definition 6. Multicolor tree. Let $T = (\sigma, \alpha, \tilde{b})$ be a rooted tree. Let $\mathcal{W} = \{w_1, \dots, w_n\}$ be a set of n distinct colors, eventually empty ($n \geq 0$). Each vertex of T can be colored by 0 to n colors. forall i in $[1, n]$, let us denote by s_i the smallest vertex of T of color w_i.
T is a multicolor tree (see T in Figure 5) if T complies with the following rules:

(1) each color of W is assigned to at least two distinct vertices from T;
(2) let $(b_{s_i,1}, \dots, b_{s_i,l_{s_i}}) = \sigma^*(b_{s_i,1})$ be the half-edges of initial vertex s_i, where $b_{s_i,1}$ is the smallest half-edge of $\sigma^*(b_{s_i,1})$, i.e. the left son of s_i. The half-edges $b_{s_i,j}, 1 \leq j < l_{s_i}$ are the half-edges, sons of s_i, and $b_{s_i,l_{s_i}}$ is the half-edge which goes up towards the father of s_i. Let $T_{s_i,j}$, be the subtree of T incident to the final vertex of $b_{s_i,j}$, rooted in $\bar{\sigma}(b_{s_i,j})$ and $\overline{T}_{s_i,j}$, the tree composed of $T_{s_i,j}$ and of the half-edge $b_{s_i,j}$ which is its root half-edge. Then:
   (a) there is a single j_i such that in T, w_i colors s_i and exclusively vertices of T_{s_i,j_i}. Let us denote this subtree by $T_{s_i,w_i} = T_{s_i,j_i}$, $\tilde{b}_{s_i} = b_{s_i,j_i}$ its root half-edge and $\overline{T}_{s_i,w_i} = \overline{T}_{s_i,j_i}$;
   (b) forall k in $[1, n]$, $k \neq i$, if $s_i = s_k$ then $\overline{T}_{s_i,w_i} \cap \overline{T}_{s_k,w_k} = \emptyset$.
(3) For all distinct colors w_i and w_j, if there is a vertex s of colors w_i and w_j where s_i is smaller than s_j, then $s = s_j$ and s is the only vertex of color w_j which is also of color w_i.

We will say that two multicolor trees are isomorphic if one can be obtained from the other by a permutation on its colors. A class of isomorphism of multicolor trees will simply be called *multicolor tree*.

Let $\mathcal{T}$ be the set of multicolor trees.

Remark 3. If T is a multicolor tree with m vertices and n distinct colors, then $n < m$.

5.2. **Bijection between $\mathcal{M}$ and $\mathcal{T}$.**

Theorem 2. *The set of rooted maps of indifferent genus with n edges is in bijection with the family of multicolor trees with n edges.*

Proof of theorem 2. In order to simplify our notations, a map whose vertices can be colored (by several colors) will also be called a map.

Let M be a map of $\mathcal{M}$ not reduced to the one vertex map and w_1 be a color. Let $w = w_1$. A map T of $\mathcal{T}$ is obtained from map M (see Figure 5), while proceeding in the following way:

(1) (a) If M deprived of its colors does not belong to $\mathcal{I}$, one applies to M the decomposition induced by Lemma 3, which transforms bijectively a map of $\mathcal{M}$ into

a pair $(I, \mathscr{S})$, where $I \in \mathcal{I}$ and $\mathscr{S}$ is a set of vertices of $Left(I)$. Then one assigns color w to the vertices resulting from the partition of the root vertex, $\tilde{s}_M$, of M, i.e. to the vertices of $\mathscr{S} \bigcup \{\tilde{s}_I\}$, where $\tilde{s}_I$ is the root vertex of the obtained map I. If colors are not taken into account, $I \in \mathcal{I}$. Let $\mathcal{W}_{\tilde{s}_M}$ be the set of colors which color the root vertex $\tilde{s}_M$ of M. Then in I, the set $\mathcal{W}_{\tilde{s}_I}$ of colors assigned to $\tilde{s}_I$ is equal to $\mathcal{W}_{\tilde{s}_M} \bigcup \{w\}$. Colors of $\mathcal{W}_{\tilde{s}_M}$ are not deferred to the other vertices resulting from $\tilde{s}_M$.

(b) Otherwise M is renamed I.

(2) If $I \notin \mathcal{T}$, let w_{left} and w_{right} two distinct colors, also distinct from all the colors already coloring I. One begins again at stage 1a with $M = Left(I)$, $w = w_{left}$ and $M = Right(I)$, $w = w_{right}$.

From Lemma 3, one gets that each stage of the transformation of a map of $\mathcal{M}$ into a multicolor tree is bijective.

By construction, T follows all the rules of Definition 6, and $T \in \mathcal{T}$, as:

- *1* is checked since a color w is assigned to the vertices resulting from the same vertex s.
- *2a* is checked since if M deprived of its colors does not belong to $\mathcal{I}$, the root half-edge of M is unglued from the root vertex and a map M_1 is obtained. Thus all the vertices to be colored belong to $Left(M_1)$, the root vertex excluded.
- *2b* is checked since after application of the transformation induced by Lemma 3 to a map of $\mathcal{M}$, its root half-edge becomes a bridge.
- *3* is checked according to item 1a above $\diamond$

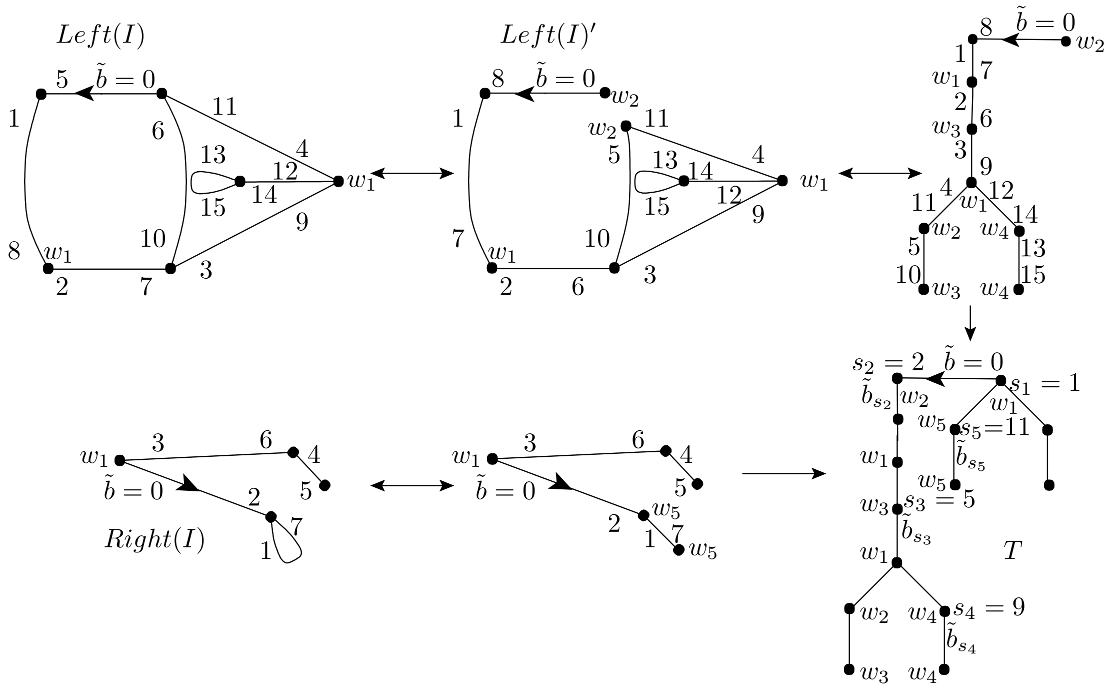


FIGURE 5. The multicolor tree associated with the map M of Figure 4

5.3. **Application: a language coding maps of indifferent genus.** In this Section we present a language coding rooted maps. The equation defining this language is a generalization of the well-known equation on Dyck words. In fact this language codes multicolor trees and thus by bijection rooted maps.

In order to clarify the signification of each letter of the alphabet of the language that we present, we need to give a definition.

Definition 7. Twin colors. Two colors w and w' of a tree of $\mathcal{T}$ are *twin* if there is a vertex of T colored by these two colors or if there is a subsequence of colors of T, $w_1 = w, w_2, \ldots, w_n = w'$ such that forall i in $[1, n]$, w_i and w_{i+1} color the same vertex. One thus defines classes of equivalence of colors, where two colors are in the same class if they are twin.

Let us denote by c the variable coding a half-edge, whose opposite half-edge is not coded, $\bar{c}$ the variable coding an half-edge, whose opposite half-edge is coded, y the variable coding a vertex in case of a map or in case of a multicolor tree, a vertex not colored or the smallest vertex among the vertices having the same or a twin color, and y_i, $i \geq 1$, the variable coding a vertex of color w_i (with $w_i \neq w_j$ if $i \neq j$) of a multicolor tree. In a rooted map, y_i, codes the half-edges belonging to a subset of the set of half-edges of initial vertex s_i, for a given vertex s_i of arity strictly superior to 1 (s_i can be equal to s_j if $i \neq j$).

Theorem 3. *The set of rooted maps is coded by the language $L_\infty = \lim_{n\to\infty} L_n$, where L_n represents the language coding rooted maps with at most n edges and is defined in the following way:*

$$L_n(y, y_1, \ldots, y_n, c, \bar{c}) = y + c\ L_{n-1}(y + y_n, y_1, \ldots, y_{n-1}, c, \bar{c}) \qquad (3)$$
$$\bar{c}\ L_{n-1}(y, y_1, \ldots, y_{n-1}, c, \bar{c})\ (1 - \epsilon_n + y_n\epsilon_n)\delta_{c,n}$$
$$L_0(y, c, \bar{c}) = y \qquad (4)$$

where for every word m_1 of $L_{n-1}(y + y_n, y_1, \ldots, y_{n-1}, c, \bar{c})$:

- $\epsilon_n = \begin{cases} 1 & \text{if } y_n \in m_1 \\ 0 & \text{otherwise} \end{cases}$
- *for every word m_2 of $L_{n-1}(y, y_1, \ldots, y_{n-1}, c, \bar{c})$:*
  $$\delta_{c,n} = \begin{cases} 1 & \text{if the number of occurrences of } c \text{ in } c\ m_1\ \bar{c}\ m_2 \leq n \\ & \text{and } \nexists 1 \leq k \leq n / y_k \in m_1 \text{ and } y_k \in m_2 \\ 0 & \text{otherwise} \end{cases}$$

Conclusion

The bijection determined in Section 4 which leads to a generalized Dyck equation on maps can easily be extended to n-colored orientable rooted maps [15] and gives us a direct bijective proof of the generalized Dyck equation on n-colored maps precedently determined in an analytic way [5, 6].

The one-to-one correspondence between maps and multicolor trees raises many questions. Can new equations on generating functions of families of maps be determined? Could it lead to new enumeration formulas of these families? This bijection can be specialized to planar maps [15] and it would be interesting to see if it can also be done to maps of genus g, $g > 0$. It would also be interesting to see what kind of informations we can get from the obtained coding of maps. It is straightforward to deduce the number of vertices and edges of a map from its associated word but we have not yet searched for other information.

References

[1] D. Arquès, Relations fonctionnelles et dénombrement des cartes pointées sur le tore, J. Combin. Theory Ser. B 43 (1987) 253-274.

[2] Arquès, D. and Béraud, J.F., "Rooted maps on orientable surfaces, Riccati's equation and continued fractions." *Discrete Math.* **215**(2000), 1–12.

[3] D. Arquès et J.-F. Béraud, "Rooted maps and hypermaps on surfaces", Proceedings of FPSAC'99, Barcelona (1999), 18-29.

[4] D. Arquès and A. Giorgetti, "Enumération des cartes pointées de genre quelcquonque en fonction des nombres de sommets et de faces", J. Combin. Theory Ser. B 77 (1999) 1-24.

[5] Arquès, D. and Micheli, A., "Enumeration of multi-colored rooted maps", *Actes du Lacim 2000* **27**(2000), 31-42.

[6] Arquès, D. and Micheli, A., "Enumeration of multi-colored rooted maps", *Discrete Math.*, to be published.

[7] E.A. Bender and E.R. Canfield, "The number of rooted maps on an orientable surface", J. Combin. Theory Ser. B, 53 (1991), 293-299.

[8] Cori, R., "Un code pour les graphes planaires et ses applications", *Astérisque* **27**(1975).

[9] Cori, R. and A. Machi, "Maps, hypermaps and their automorphisms: a survey I, II, III", Expo. Math. **10** (1992), 403-427, 429-447, 449-467.

[10] R. Cori et B. Vauquelin, "Planar maps are well labelled trees", Canad. J. Math. **33 (5)** (1981), 1023-1042.

[11] P. Flajolet, "Combinatorial aspects of continued fractions", Discrete Math. **32** (1980) 125-161.

[12] D.M. Jackson et T.I. Visentin, "A character theoretic approach to embeddings of rooted maps in an orientable surface of given genus", Trans. Amer. Math. Soc. **322** (1990), 343-363.

[13] A. Jacques, "Sur le genre d'une paire de substitutions", C. R. Acad. Sci. Paris **267** (1968), 625-627.

[14] M. Marcus et B. Vauquelin, "Un codage des cartes de genre quelconque", Actes de FPSAC'95, Marne-La-Vallée (1995), 399-415.

[15] A. Micheli, "Combinatoire des cartes de genre quelconque et arborescences multicouleurs", Thèse de Doctorat, Université de Marne-La-Vallée (2001).

[16] P. Ossona de Mendez et P. Rosenstiehl, "Connected permutations and hypermaps", les cahiers du CAMS 183, available at http://www.ehess.fr/centres/cams/papers.

[17] W.T. Tutte, "A census of planar maps", Canad. J. Math. 15 (1963) 249-271.

[18] W.T. Tutte, "On the enumeration of planar maps", Bull. Amer. Math. Soc. 74 (1968) 64-74.

[19] T.R.S. Walsh et A.B. Lehman, "Counting rooted maps by genus. I", J. Combin. Theory Ser. B **13** (1972), 192-218.

Équipe de synthèse d'images, Institut Gaspard Monge, Unité CNRS UPRES-A 8049, Université de Marne la Vallée, 5 Boulevard Descartes, Champs sur Marne, 77454 Marne la Vallée Cedex 2, France

E-mail address: `{arques,micheli}@univ-mlv.fr`

EXACT KAZHDAN CONSTANTS OF SOME COXETER GROUPS AND WREATH PRODUCTS (EXTENDED ABSTRACT)

ELI BAGNO

ABSTRACT. We compute exact Kazhdan constants for the Coxeter groups of types B and D, and for certain wreath products. The computations rely on the combinatorics of Young tableaux and on suitable decomposition of representation spaces.

RÉSUMÉ. Nous calculons les constantes exactes de Kazhdan des groupes de Coxeter de type B ainsi que de certains autres produits qui leur sont etrangers. Les calculs s'appuient sur la theorie combinatoire des tableaux de Young et sur une decomposition adequate des espaces de representation.

1. Introduction

The parameters now known as Kazhdan constants were introduced by Kazhdan [KA] in the study of semi-simple Lie groups to serve as a corner stone in the representation theory of discrete groups. This notion was applied to algebraic graph theory and computer science and, playing a key role in the construction of expanding graphs [MA] [LU] . Much research had been devoted to the calculation of exact values of Kazhdan constants for various groups. For the group $SL(3, \mathbb{Z})$, explicit Kazhdan constants for a certain family of representations have been computed by Burger in [BU]. More recently, new remarkable examples of groups with Kazhdan property (T) were discovered by Zuk. Explicit Kazhdan constants for them were computed in [ZU1].The problem of calculating Kazhdan constants for semi-simple groups and their lattices was solved by Shalom [SH]. Kazhdan constants of the symmetric groups with respect to their Coxeter generators were computed by Bacher and de la Harpe [BD].

In this paper we generalize the work of Bacher and de la Harpe on the symmetric groups and compute the Kazhdan constant of three infinite families of groups, namely the Coxeter groups of types B and D as well as $C_r \wr S_n$, the wreath product of the cyclic group C_r by the symmetric group S_n. The proof makes an extensive use of the combinatorial representation theory of the above groups, explicitly the analysis of Young tableaux.

2. Background

2.1. The Kazhdan constant.

Definition 2.1. *Let G be a discrete group, generated by a finite set S. We say that G has the* property (T) of Kazhdan *if there exists an $\epsilon > 0$ such that for every nontrivial irreducible unitary representation ρ of G on a Hilbert space H, and for every vector $v \in H$ with $\|v\| = 1$, there exists $s \in S$ such that $||\rho(s)(v) - v|| > \epsilon$.*

(Equivalent definitions may be found in [LU],Chapter 4).

Intuitively, this amounts to saying that for every nontrivial representation and every vector there is some generator which moves the vector by at least ϵ. Kazhdan property

Key words and phrases. Kazhdan constant, representation theory, Young tableaux.

(T) can also be expressed topologically, using the Fell Topology on the space of all unitary representations of G. In this setting, property (T) amounts to saying that the trivial representation is an isolated point of this space.

Definition 2.2. *Let G be a group with a finite generating set S and let π be an irreducible representation of G with representation space V_π. Define the* Kazhdan constant

$$K_G(S,\pi) := inf_{\xi \in S(V_\pi)} max_{s\in S} \|\pi(s)\xi - \xi\|$$

where $S(V_\pi) = \{\xi \in V_\pi : \|\xi\| = 1\}$.

We define also:

$$K_G(S) := inf\{K_G(S,\pi) : \pi \textit{ is irreducible and non trivial}\}$$

We cite here the main result of Bacher and de la Harpe [BD] concerning Kazhdan constants for the symmetric groups:

Theorem 2.3. *For the symmetric group S_n with the Coxeter system*

$$S = \{(1,2),(2,3),...,(n-1,n)\}$$

we have:

$$K_{S_n}(S) = \sqrt{\frac{24}{n^3-n}}.$$

2.2. **Coxeter groups of type B.**

Definition 2.4. *The* Coxeter group *B_n is the group of signed permutations of $\{1,...,n\}$. Namely, B_n consists of all permutations π of $\{-n,...,-1,1,...,n\}$ such that $\pi(-k) = -\pi(k)$ for all $1 \le k \le n$.*

We represent elements of B_n in cycle notation as permutations of $\{-n,...,-1,1,...,n\}$. The elements

$$s_0 = (1,-1)$$

and

$$s_i = (i-1,i)(-(i-1),-i) \quad (1 \le i \le n-1)$$

generate B_n and satisfy the relations:

$$s_i^2 = 1 \quad (\forall i)$$

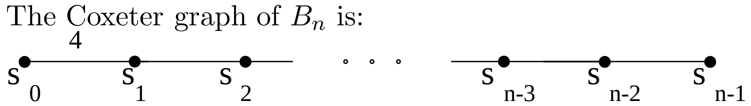
$$s_is_j = s_js_i \quad (|i-j|>1),$$

$$(s_is_{i+1})^3 = 1 \quad (1 \le i \le n-1),$$

$$(s_0s_1)^4 = 1.$$

They are called *Coxeter generators.* Denote $S_{B_n} = \{s_0, s_1, ..., s_{n-1}\}$.

The Coxeter graph of B_n is:



2.2.1. *Representations of the groups of type B.* We start with some definitions:

Definition 2.5. *Let n be an integer. A partition of n of length l is a sequence $\alpha = (a_1, ..., a_l)$ of nonnegative integers such that $a_1 \geq a_1 \geq ... \geq a_l$ and $a_1 + ... + a_l = n$. The size of α is defined by: $|\alpha| = a_1 + ... + a_l (= n)$.*

A partition $\alpha = (a_1, ..., a_l)$ of n can be represented by an array of n boxes in l rows with row i containing a_i boxes, $(1 \leq i \leq l)$. This is called the *Young diagram* of the partition α.

Definition 2.6. *A* double partition*, $\lambda = (\lambda_1, \lambda_2)$ of size n is an ordered pair of partitions λ_1 and λ_2 such that $|\lambda_1| + |\lambda_2| = n$.*

Every double partition $\lambda = (\lambda^1, \lambda^2)$ of size n is equipped with a *double Young diagram* which is a pair of Young tableaux, one for each λ^i. We also use the term *shape* for a double Young diagram.

The irreducible representations of the groups B_n are indexed by the shapes $\lambda = (\lambda^1, \lambda^2)$ such that $|\lambda| = |\lambda^1| + |\lambda^2| = n$.

A standard Young tableau $L = (L^{\lambda^1}, L^{\lambda^2})$ is a filling of the Young diagram of λ with the numbers $1, ..., n$ such that in each of L^{λ^1} and L^{λ^2} separately, numbers are increasing along rows and along columns. For example,

1	3	4	9
6	8		

2	5	7

is a standard tableau of the shape $((4,2),(3))$

2.3. **The groups of type D.** The Coxeter group D_n is the group of signed permutations of $\{1, 2, ..., n\}$ with an even number of negative signs. More precisely, D_n consists of all permutations π of $\{-n, ..., -1, 1, ..., n\}$ such that $\pi(-k) = -\pi(k)$ for all $1 \leq k \leq n$ and an even number of the numbers of $\pi(1), \pi(2), ..., \pi(n)$ are negative. We represent elements of D_n in cycle notation as permutations of $\{-n, ..., -1, 1, ..., n\}$. The element

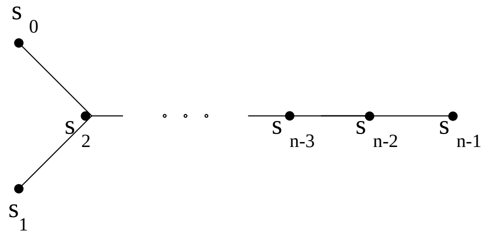
$$s_0 = (1, -2)(2, -1)$$

together with $s_i = (i-1, i)(-(i-1), -i), (\quad 1 \leq i \leq n-1)$ generate D_n and satisfy the relations:

$$\begin{gathered} s_i s_j = s_j s_i, \quad (|i-j| > 1, \quad i, j > 0), \\ s_0 s_j = s_j s_0 \quad (j \neq 2) \\ s_0 {s_2}^3 = 1, \\ s_i {s_{i+1}}^3 = 1, \quad (1 \leq i \leq n-1), \\ {s_i}^2 = 1 \quad (1 \leq i \leq n-1). \end{gathered}$$

The Coxeter group D_n can be realized as a normal subgroup of the Coxeter group B_n of index 2.

The Coxeter graph of The groups of type D is:



2.3.1. *Representations of D_n.* Being a normal subgroup of B_n of index 2, D_n essentialy inherits its irreducible representations from B_n. By Clifford theory, the restriction to D_n of an irreducible representation corresponding to a shape $\lambda = (\lambda_1, \lambda_2)$ with $(\lambda_1 \neq \lambda_2)$ is an irreducible representation of D_n. On the other hand, if $\lambda = (\lambda_1, \lambda_2)$ with $\lambda_1 = \lambda_2$ then the restriction to D_n of the corresponding B_n-representation splits into a sum of two non-isomorphic irreducible representations of D_n. All irreducible representations of D_n are obtained in this fasion.

2.4. **The groups of type G(r,n).** Let $G(r,n) = C_r \wr S_n$ be the wreath product of C_r and S_n . $G(r,n)$ is a unitary reflection group consisting of all monomial matrices (i.e., products of diagonal and permutation matrices) of order $n \times n$ whose non-zero entries are complex r-th roots of unity.

Abstractly, the group $G(r,n)$ can be presented by a set of generators $S_W = \{s_0, s_1, ..., s_{n-1}\}$ with the following set of relations:

$$s_0^r = 1$$
$$s_i^2 = 1 \quad (i = 1, ..., n-1)$$
$$s_0 s_1 s_0 s_1 = s_1 s_0 s_1 s_0$$
$$s_i s_j = s_j s_i, (|i-j| \geq 2)$$
$$s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1} \quad (i = 1, ..., n-1)$$

Note that for $r = 2$, $G(r,n) = B_n$.

2.4.1. *Representations of G(r,n).* The representation theory of the groups $G(r,n)$ is a generelization of the representation theory of B_n. The only difference is that one works with $n-$ tuples of partitions and Young tableaux instead of double partitions and double Young tableaux. For more details see e.g. [HR].

3. Exact Kazhdan constants

Our first main result is an exact computation of the Kazhdan constant for the Coxeter groups of type B.

Theorem 3.1. *The Kazhdan constant of the group B_n with respect to the set of Coxeter generators S_{B_n} is:*

$$K_{B_n}(S_{B_n}) = \sqrt{\frac{4}{\sum_{j=1}^{n} \left(1 + \sqrt{2}(j-1)\right)^2}}$$

Sketch of the proof: The Kazhdan constant is achieved by the following procedure. We begin by computing an upper bound for the Kazhdan constant of a specific representation of the group B_n, namely the natural representation which reflects the action of the Coxeter generators on the Euclidean space $\mathbb{R}^n$. The vector of $\mathbb{R}^n$ on which the upper bound is attained is moved the same amount by all of the Coxeter generators and thus taken from the central chamber of the natural action of B_n on $\mathbb{R}^n$. This vector can be computed by solving some linear equations.

The next step is to prove that that the value for the natural representation serves also as a lower bound for every nontrivial irreducibe representation of B_n. Since the natural representation of B_n is irreducible, we conclude that this value is the Kazhdan constant for the whole set of representations of B_n.

The proof that the upper bound is also a lower bound is combinatorial in nature. The idea is to divide the representation space of every irreducible representation which is composed of standard Young tableaux of some given shape into n subspaces chosen according to the digit located in top left box of the second tableau. The subspaces are chosen in such a way that almost all of the Coxeter generators act invariantly on every single subspace.

We note that that a similar idea appears first in [BD], but their choice of the box which splits the representation space into subspaces was inadequate to our case. Moreover, due to the structure of the representation theory of B_n, our choice yields a more elegant proof.

We deal next with the family of Coxeter groups of type D. Here, since some of the irreducible representations of D_n split into two irreducible representations the computation is much more complicated. The case of non-splitting representations is very similar to the case of the groups of type B while the case of splitting representations requires a new parameterization of the basis by tableaux. The Kazhdan constant for the Coxeter groups of type D is given in the following:

Theorem 3.2. *The Kazhdan constant of the Coxeter groups of type D with respect to the set S_{D_n} of Coxeter generators is:*

$$K_{D_n}(S_{D_n}) = \sqrt{\frac{2}{\sum_{j=2}^{n}(j-1)^2}} = 2\sqrt{3}\sqrt{\frac{1}{n(2n^2-3n+1)}}$$

The following result generalizes theorem 3.1

Theorem 3.3. *The Kazhdan constants for the groups $G(r,n)$ with respect to the set S_W of generators is:*

$$K_{G(r,n)}(S_W) = \sqrt{\frac{|\rho_r - 1|^2}{\sum_{j=1}^{n}\left(1+\frac{|\rho_r-1|}{\sqrt{2}}(j-1)\right)^2}}.$$

where $\rho_r = e^{\frac{2\pi i}{r}}$.

Naturally, the next family of groups whose Kazhdan constant is interesting is the complex reflection groups $G(r,n,p)$. Work in this direction is in progress.

Acknowledgments

This work is a part of a Ph.D. thesis at Bar-Ilan University carried out under the supervision of Ron Adin and Yuval Roichman. The author would like to thank Alex Lubotzki, Pierre de la Harpe and Andre Zuk for useful discussions.

References

[BA] R. Bacher, Valeur propre minimale du laplacien de Coxeter pour le group symetŕique, *J. of Algebra*, 167, (1994), 460-472.

[BD] R. Bacher and P. de la Harpe, Exact values of Kazhdan constants for some finite groups, *J. of Algebra*, 163 , 1994, 495–515

[BU] M. Burger, Kazhdan constants for $SL(3,\mathbb{Z})$, *J. Reine Angew. Math.*, 413, (1991), 36–37.

[HR] T. Halverson and A. Ram, Murnaghan-Nakayama rules for characters of Iwahori-Hecke algebras of the complex reflection groups $G(r,p,n)$, *Can. J. Mat* Vol.50 (1), 1997,167-192

[KA] D.Kazhdan, Connection of the dual space of a group with the structure of its closed subgroups., *Funct. Anal. Appl* 1(1967), 63-65.

[LU] A. Lubotzky, Discrete Groups, Expanding Graphs and Invariant Measures, Birkha´user, Germany, 1994.

[MA] G. A. Margulis, Explicit constructions of expanders, *Problems Inform. Transmission 9*, No. 4, (1973), 325–332.

[SH] Y. Shalom, Explicit Kazhdan constants for representations of semisimple and arithmetic groups, *Ann. Inst. Fourier*, to appear.

[ZU1] A. Zuk, La propriete (T) de Kazhdan pour les groupes agissant sur les polyedres, *C.R. Acad. Sci. Paris*, Ser I 323 no. 5, (1996), 453–458.

[ZU2] A. Zuk, Property (T) and Kazhdan constants for discrete groups, preprint (1999).

Department of Mathematics and Computer Sciences, Bar-Ilan University, Ramat-Gan, Israel 52900

E-mail address: bagnoe@macs.biu.ac.il

LATTICE PATHS WITH AN INFINITE SET OF JUMPS

CYRIL BANDERIER AND DONATELLA MERLINI

Abstract. Whereas walks on $\mathbb{N}$ with a finite set of jumps were the subject of numerous studies, walks with an infinite number of jumps remain quite rarely studied. Even for relatively well structured models, the classical approach with context-free grammars fails as we deal with rewriting rules over an infinite alphabet. However, several classes of such walks offer a surprising structure: we make here explicit the associated bivariate functions, and give several theorems on their nature (rational, algebraic) via the kernel method or Riordan arrays theory. We give some examples of recent problems in combinatorics or theoretical computer science which lead to such rules.

Résumé. Tandis que les propriétés énumératives et asymptotiques des marches sur $\mathbb{N}$ avec un nombre fini de sauts ont fait l'objet de nombreuses études, les marches avec un nombre infini de sauts demeurent assez peu étudiées. Même pour des modèles relativement structurés, on ne peut utiliser les approches classiques par grammaires algébriques, puisqu'il s'agit de règles de récriture sur un alphabet infini. Toutefois, diverses classes de telles marches offrent une surprenante structure : nous explicitons ici la nature (algébrique, rationnelle) de la série génératrice bivariée associée (via la méthode du noyau ou la théorie des tableaux de Riordan). Nous illustrons l'intêret de telles marches en combinatoire et informatique théorique par quelques exemples.

Introduction

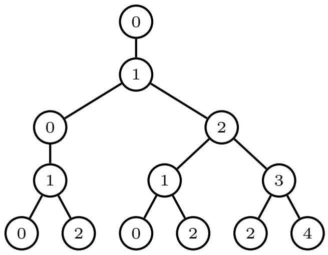
A considerable number of problems from computer science deals with a sum of independent identical distributed random variables $\Sigma_n = X_1 + X_2 + \cdots + X_n$ (where each of the X_i's assumes integer values). We will consider here the following model of random walks: the walk starts (at time 0) from a point Σ_0 of $\mathbb{Z}$ and at time n, one makes a jump $X_n \in \mathbb{Z}$; so the new position is given by the recurrence $\Sigma_n = \Sigma_{n-1} + X_n$ where, when $\Sigma_{n-1} = k$, the X_n's are constrained to belong to a fixed set $\mathcal{P}_k$ (that is, the possible jumps depend on the position of the walk).

These "walks on $\mathbb{Z}$" are homogeneous in time (that is to say, the set of jumps when one is at position k is independent from the time). When the positions Σ_n's are constrained to be nonnegative, we talk about "walks on $\mathbb{N}$". The probabilistic model under consideration here is the uniform distribution on all paths of length n.

When the sets $\mathcal{P}_k$'s are equal to a fixed set $\mathcal{P}$ (the simplest interesting case being $\mathcal{P} = \{-1, +1\}$), the corresponding walks have been deeply studied both in combinatorics and in probability theory. We refer to [3] for asymptotic properties of such "walks on $\mathbb{N}$ with a finite set of jumps". When the sets $\mathcal{P}_k$'s are unbounded, both enumeration and asymptotics become cumbersome: contrary to the previous case, the walks are not space-homogeneous (the set of available jumps depends on the position) and it is not possible to generate them by context-free grammars. However, if the sets $\mathcal{P}_k$'s have a "combinatorial" shape, it is reasonable to hope that the generating function associated to the corresponding walk would have some nice properties. We show here that this hope is legitimate and we present several classes of such walks, for which we are able to give the nature of their generating function.

Our results have potential impacts on the theory of generating trees (generation of combinatorial objects), the enumeration of general classes of lattice paths, and on the study of rewriting rules on an infinite alphabet.

A definition of the generating function associated to the walk is given in Section 1. In this first section, we also present the generating tree and Riordan array viewpoints. In Section 2, we give several theorems related to the nature of the generating functions associated to some walks and then we give some asymptotic results. In Section 3, we give some examples of problems in which some of the new classes of walks that we study in this article appeared.



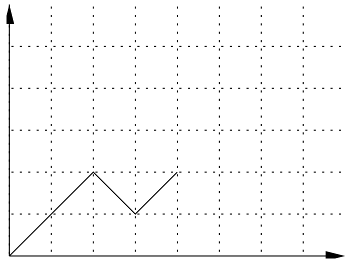


FIGURE 1. The generating tree of the walk on $\mathbb{N}$ with jumps $\mathcal{P} = \{+1, -1\}$ starting in 0 (and up to length $n = 4$). Each branch corresponds to a path. The branch $(0, 1, 2, 1, 2)$ corresponds to the path drawn on the lattice.

1. LATTICE PATHS AND GENERATING TREES

In combinatorics, it is classical to represent a particular walk as a path in a two dimensional lattice. Thus the drawing corresponds to the walk of length n linking the points $\big((0, \Sigma_0), (1, \Sigma_1), \ldots, (n, \Sigma_n)\big)$. It is also convenient to represent all the walks of length $\leq n$ as a tree of height n, where the root (at level 0 by convention) is labeled with the starting point of the walks and where the label of each node at level n encodes a possible position of the walk (see Figure 1).

We note $w_{n,k}$ the number of walks on $\mathbb{N}$ of length n going from the starting point to k (or, equivalently, the number of nodes with label k at level n in the tree) and we want to find the bivariate generating function

$$W(z, u) = \sum_{n \geq 0} w_n(u) z^n = \sum_{k \in \mathbb{Z}} W_k(z) u^k = \sum_{k \in \mathbb{Z}, n \geq 0} w_{n,k} u^k z^n \,,$$

where u encodes the final altitude of the walk (the label in the tree), z the length of the walk (the level in the tree), and where $w_n(u)$ is a Laurent polynomial (that is, a polynomial with finitely many monomials of negative and positive degree). When the walk is constrained to remain nonnegative (or equivalently when negative labels in the tree are not allowed), we consider similarly the bivariate generating function

$$F(z, u) = \sum_{n \geq 0} f_n(u) z^n = \sum_{k \in \mathbb{N}} F_k(z) u^k = \sum_{k \in \mathbb{N}, n \geq 0} f_{n,k} u^k z^n \,. \tag{1}$$

Generating trees and rewriting rules. The concept of generating trees has been used from various points of view and was introduced in the literature by Chung, Graham, Hoggatt and Kleiman [6] to examine the reduced Baxter permutations. This technique has been successively applied to other classes of permutations. A generating tree is a rooted labeled tree with the property that if v_1 and v_2 are any two nodes with the same label then, for each label ℓ, v_1 and v_2 have exactly the same number of children with label ℓ. To specify a generating tree it therefore suffices to specify: 1) the label of the root; 2) a set of rules explaining how to derive from the label of a parent the labels of all of its children. Points 1) and 2) define what we call a *rewriting rule*. For example, Figure 1 illustrates the upper part of the generating tree which corresponds to the rewriting rule $[(0), \{(k) \rightsquigarrow (k-1)(k+1)\}]$.

Riordan arrays We introduce now the concept of *matrix associated to a generating tree*: this is an infinite matrix $\{d_{n,k}\}_{n,k \in \mathbb{N}}$ where $d_{n,k}$ is the number of nodes at level n with label $k + r$, r being the label of the root. For example, the matrix associated to the generating tree of the Figure 1 is the following:

n/k	0	1	2	3	4
0	1				
1	0	1			
2	1	0	1		
3	0	2	0	1	
4	2	0	3	0	1

Many such matrices can be studied from a *Riordan array* viewpoint. In fact, the concept of a Riordan array provides a remarkable characterization of many lower triangular arrays that arise in combinatorics and algorithm analysis. The theory has been introduced in the literature in 1991 by Shapiro, Getu, Woan and Woodson [11]. Riordan arrays are a powerful tool in the study of many counting problems [7].

A Riordan array is an infinite lower triangular array $\{d_{n,k}\}_{n,k\in\mathbb{N}}$, defined by a pair of formal power series $D = (d(z), h(z))$, such that the k-th column is given by $d(z)(zh(z))^k$, i.e.:

$$d_{n,k} = [z^n]d(z)(zh(z))^k, \qquad n, k \geq 0.$$

From this definition we have $d_{n,k} = 0$ for $k > n$. The bivariate generating function for D is:

$$\sum_{n,k\geq 0} d_{n,k} u^k z^n = \frac{d(z)}{1 - uzh(z)}.$$

In what follows, we always assume that $d(0) \neq 0$; if we also have $h(0) \neq 0$ then the Riordan array is said to be *proper*; in the proper-case the diagonal elements $d_{n,n}$ are different from zero for all $n \in \mathbb{N}$. The most simple example is the Pascal triangle for which we have

$$\binom{n}{k} = [z^n]\frac{1}{1-z}\left(\frac{z}{1-z}\right)^k,$$

where we recognize the proper Riordan array with $d(z) = h(z) = 1/(1-z)$. Proper Riordan arrays are characterized by the existence of a sequence $A = \{a_i\}_{i\in\mathbb{N}}$ with $a_0 \neq 0$, called the *A-sequence*, such that every element $d_{n+1,k+1}$ can be expressed as a linear combination, with coefficients in A, of the elements in the preceding row, starting from the preceding column:

$$d_{n+1,k+1} = a_0 d_{n,k} + a_1 d_{n,k+1} + a_2 d_{n,k+2} + \cdots$$

It can be proved that $h(z) = A(zh(z))$, $A(z)$ being the generating function for A. For example, for the Pascal triangle we have: $A(z) = 1 + z$ and the previous relation reduces to the well-known recurrence relation for binomial coefficients. The A-sequence doesn't characterize completely $(d(z), h(z))$ because $d(z)$ is independent of $A(z)$. But it can be proved that there exists a unique sequence $Z = \{z_0, z_1, z_2, \ldots\}$, such that every element in column 0 can be expressed as a linear combination of all the elements of the preceding row:

$$d_{n+1,0} = z_0 d_{n,0} + z_1 d_{n,1} + z_2 d_{n,2} + \cdots$$

This property has been recently studied in [7], where it is proved that $d(z) = d(0)/(1 - zZ(zh(z)))$, $Z(z)$ being the generating function for Z. Thus the triple $(d(0), Z(z), A(z))$ characterizes every proper Riordan array.

2. Walks on $\mathbb{Z}$ with an infinite set of negative jumps

2.1. **Lattice paths and generating trees.** Consider a sequence $(e_i(k))_{i\geq -a}$ (for a given integer $a > 0$) of polynomials assuming nonnegative integers values then the walk with an infinite set of jumps under consideration here are of the following kind:

$$[(r), \{(k) \rightsquigarrow (0)^{e_k(k)}(1)^{e_{k-1}(k)}(2)^{e_{k-2}(k)} \ldots (k-1)^{e_1(k)}(k)^{e_0(k)} \ldots (k+a)^{e_{-a}(k)}\}], \tag{2}$$

where the exponent $e_i(k)$ is the multiplicity of the jumps $-i$ when one is at position k and where r is the starting position of the walk (or equivalently, the root of the associated generating tree).

If the sequence of polynomials $(e_i(k))_{i\geq -a}$ is ultimately $e_i(k) = 0$, then the situation covers the case of walks with a finite set of jumps. If the sequence is ultimately $e_i(k) = 1$, then this covers the case of "factorial rules" which are of great interests for the generation of combinatorial objects [4] and for which it was proven in [2] that the associated generating functions are algebraic.

We still note $f_{n,k}$ the number of walks on $\mathbb{N}$ of length n going from the starting point to k and we want to find the bivariate generating function $F(z,u) = \sum_{n,k\geq 0} f_{n,k} u^k z^n$. These random walks on $\mathbb{N}$ can equivalently be seen as lattice paths, generating trees and also as Riordan arrays (when $a = 1$).

Rule	EIS approximate description	Generating Function $F(z,u)$
	Rational OGF	**OGF**
$(0), \{(k) \rightsquigarrow (0)^k (k+1)\}$	F_0, $F(z,1)$: powers of 2	$\frac{1-2z-z^2}{1-(u+2)z-2uz^2}$
$(0), \{(k) \rightsquigarrow (0)^{2k} (k+1)\}$	$F(z,1)$:A001333 continued fraction convergents to $\sqrt{2}$ F_0: A052542 (ECS)	$\frac{1-2z+z^2}{1-(u+2)z+(2u-1)z^2+uz^3}$
$(0), \{(k) \rightsquigarrow (0)^{3k} (k+1)\}$	$F(z,1)$: A026150 (ECS)	$\frac{1-2z+z^2}{1-(u+2)z+(2u-2)z^2+2uz^3}$
$(0), \{(k) \rightsquigarrow (0)^{4k} (k+1)\}$	$F(z,1)$: A046717 half of 3^n	$\frac{1-2z+z^2}{1-(u+2)z+(2u-3)z^2+3uz^3}$
$(0), \{(k) \rightsquigarrow (0)^k (k+1)(k+2)\}$	$F(z,1)$: A001075 and F_0: A005320 Pell's equation	$\frac{1-4z+4z^2}{1-(4+u+u^2)z+(4u^2+u-1)z^2-\dots}$
$(1), \{(k) \rightsquigarrow (0)(1)^2(k)(k+2)^2(k+3)^5\}$	6^n and A003464 $(6^n-1)/5$	$\frac{(4u-1)z-u}{(1-6z)((2u^2+1)z-1)}$
$(0), \{(k) \rightsquigarrow (0)^{k^2}(2)^{3k-1}(3)(k)(k+1)^2(k+3)^5\}$		see Theorem 1
	Algebraic OGF	**OGF**
$(1), \{(k) \rightsquigarrow (1)\dots(k+s-2)(k+s-1)\}$	$F(z,1)$: s-ary trees	
$(1), \{(k) \rightsquigarrow (1)^2\dots(k)^2(k+1)\}$	$F(z,1)$: A001003 Schröder's second problem	$\frac{u}{2}\frac{1-(2u+1)z-\sqrt{1-6z-z^2}}{(1-u)z+(u^2+u)z^2}$
$(0), \{(k) \rightsquigarrow (0)^{k^2}(2)^{3k-1}(3)(k)(k+1)^2(k+3)^5\}$		
$(0), \{(k) \rightsquigarrow (0)^k(1)^{k-1}\dots(k-1)^1(k)^0(k+1)\}$	A036765 $F(z,1)$: rooted trees with a degree constraint	equation of degree 3
$(0), \{(k) \rightsquigarrow (0)^{k+2}(1)^{k+1}\dots(k-1)^3(k)^2(k+1)\}$	F_0: A006013 A046648 noncrossing trees on a circle $F(z,1)$: A001764 ternary trees	equation of degree 3
$(0), \{(k) \rightsquigarrow (0)^{k+3}\dots(k-1)^4(k)^3(k+1)^2(k+2)\}$	$F(z,1)$: A066357 planar trees with root parity constraint	equation of degree 4
$(0), \{(k) \rightsquigarrow (0)^{C_k}\dots(k-1)^{C_1}(k)^{C_0}(k+1)\}$ (where C_k is the k-th Catalan number)	F_0: A006318 large Schröder numbers	$\frac{1}{2}\frac{3-(4u+1)z-\sqrt{1-6z-z^2}}{1-3uz+(2u^2+u)z^2}$
$(0), \{(k) \rightsquigarrow (0)^{C_k}\dots(k-1)^{C_1}(k+1)\}$	F_0: A052705 (ECS)	$\frac{1}{2}\frac{3-(4u+2)z-\sqrt{1-4z-4z^2}}{1-(3u+2)z+(2u^2-2u+1)z^2}$
$(0), \{(k) \rightsquigarrow (0)^{T_k}\dots(k-1)^{T_1}(k)^{T_0}(k+1)\}$ (where T_k is the k-th tri-Catalan number)	F_0: A054727 noncrossing forests of rooted trees	equation of degree 3

TABLE 1. Some rewriting rules with simple combinatorial patterns. The ordinary generating functions $F(z,1)$ and $F_0(z)$ are defined as in Equation 1.

In Table 1, we give a list of rewriting rules with simple combinatorial patterns, the reference to famous numbers or combinatorial problems they refer to, the generating function $F(z,1)$, and the numbers identifying the corresponding sequences in the On-Line Encyclopedia of Integer Sequences `http://www.research.att.com/~njas/sequences/`; ECS stands for the Encyclopedia of Combinatorial Structures `http://algo.inria.fr/encyclopedia/`.

2.2. Rationality and algebraicity of classes of rewriting rules.

Theorem 1. *For a constant $B \geq 0$, the rule*

$$[(r), \{(k) \rightsquigarrow (0)^{e_k(k)} \dots (B)^{e_{k-B}(k)} \quad (k)^{e_0} \dots (k+a)^{e_{-a}}\}]$$

(where $e_k(k), \dots, e_{k-B}(k)$ are polynomial in k, $e_i(k) = 0$ for $0 < i < k-B$ and $e_i(k) = e_i$, some fixed constants, for $i \leq 0$) has a rational generating function $F(z,u)$.

Proof. First, we illustrate the general case by the following example:

$$[(0), \{(k) \rightsquigarrow (0)^{k^2}(2)^{3k-1}(3)(k)(k+1)^2(k+3)^5\}],$$

for which $B = 3$, the polynomials in k are $e_k(k) = k^2, e_{k-1} = 0, e_{k-2} = 3k-1, e_{k-3} = 1$, and the fixed constants are $e_0 = 1$, $e_{-1} = 2$, $e_{-2} = 0$, $e_{-3} = 5$.

The part $(k) \rightsquigarrow (0)^{k^2}$ implies a transformation $u^k \rightsquigarrow k^2 u^0$. The part $(k) \rightsquigarrow (2)^{3k-1}$ implies a transformation $u^k \rightsquigarrow (3k-1)u^2$. The part $(k) \rightsquigarrow (3)$ implies a transformation $u^k \rightsquigarrow u^3$. It is possible to perform all these transformations using the derivation, evaluation in $u = 1$ and multiplication by a monomial: in the first case, the multiplicity k^2 is obtained by $\partial(u\partial(u^k))$ and then evaluating in $u = 1$; for the second case, the multiplicity $3k-1$ is obtained by taking $\partial_u(u^{3k})/u$ and then evaluating in $u = 1$; for the third case simply evaluate in $u = 1$ and multiply by u^3. The part $(k) \rightsquigarrow (k)(k+1)^2(k+3)^5$ gives $u^k \rightsquigarrow P(u)u^k$ where $P(u) = 1 + 2u + 5u^3$. All these transformations are in fact linear, so to act on u^k or a polynomial in u (like $f_n(u)$) is the same.

Finally, evaluating $\partial(u\partial f_n(u))$ in $u=1$ gives $f_n''(1)+f_n'(1)$ and evaluating $u^2\partial_u f_n(u^3)/u$ in $u=1$ gives $u^2(3f_n'(1)-f_n(1))$, so these trivial simplifications gives the following recurrence:

$$f_{n+1}(u) = P(u)f_n(u) + u^0(f_n''(1)+f_n'(1)) + u^2(3f_n'(1)-f_n(1)) + u^3 f_n(1)\,.$$

Multiplying by z^{n+1} and summing for $n \geq 0$ leads to the functional equation

$$(1-zP(u))F(z,u) = 1 + z(u^3-1)F(z,1) + z(3u^2+1)F'(z,1) + zF''(z,1)\,.$$

Taking the first 2 derivatives and instantiating in $u=1$ gives a rational system of full rank, hence $F(z,u)$ is rational:

$$F(z,u) = \frac{u^3(22z^2-112z^3-z)+u^2(480z^3-60z^2)+528z^3-250z^2+31z-1}{(1-zP(u))(872z^3-212z^2+30z-1)}\,.$$

For the general case, one has the following functional equation

$$(1-zP(u))F(z,u) = u^r + z\sum_{i=0}^{d} t_i(u)\partial_u^i F(z,1)$$

(d is the largest degree of the polymonials $e_i(k)$, and the t_i's are some Laurent polynomials which can be made explicit). Taking the first d derivatives and instantiating in $u=1$ gives a system (for $m=0,\dots,d$):

$$\begin{aligned}\partial_u^m u^r \quad &+ \quad \left(\sum_{i=0}^{m-1}\left(z\partial_u^m t_i(1) + z\binom{m}{i}\partial_u^{m-i}P(1)\right)\partial_u^i F(z,1)\right)\\ &+ \quad (z\partial_u^m t_i(1) - (1-zP(1)))\,\partial_u^m F(z,1) + z\sum_{i=m+1}^{d}\partial_u^m t_i(1)\partial_u^i F(z,1) = 0\,.\end{aligned}$$

This gives a matricial equation $M.\overrightarrow{F} = \overrightarrow{v}$ where $\overrightarrow{F} = (\partial_u^0 F(z,1),\dots,\partial_u^d F(z,1))^T$ and $\overrightarrow{v} = (u^r,0,\dots,0)^T$. The coefficients of the main diagonal of M are $-1+z\dots$ (as they are the coefficients of the $\partial_u^m F(z,1)$ summand) and all the other coefficient of M are monomials in z of degree 1. Thus, one has $[z^0]\det M = \pm 1$ and then $\det M \neq 0$. Consequently, this system is of full rank. Solving it gives rational expressions for the $\partial_u^i F(z,1)$ and for $F(z,u)$. □

We now give a generalization of a result of [2] which was giving the algebraicity of "factorial rules": we allow here initial multiplicities which are not space-homogeneous.

Theorem 2. *For a constant $B \geq 0$, the rule*

$$[(r),\{(k) \rightsquigarrow (0)^{e_k(k)}\dots(B)^{e_{k-B}(k)}(B+1)\dots(k-b-1)(k-b)^{e_b}\dots(k+a)^{e_{-a}}\}]$$

(where $e_k(k),\dots,e_{k-B}(k)$ are polynomial in k, $e_i(k)=1$ for $b<i<k-B$ and $e_i(k)=e_i$, some fixed constants, for $i \leq b$) has an algebraic generating function $F(z,u)$.

Proof. We illustrate the general case by the following example:

$$[(0),\{(k) \rightsquigarrow (0)^{k^2}(2)^{3k^5-2}(6)(7)\dots(k-5)(k-4)^2(k-2)^3(k)(k+3)^2(k+23)\}]$$

for which $B=5, b=4, a=23$, the polynomials in k are $e_k(k)=k^2$, $e_{k-2}(k)=3k^5-2$, $e_{k-1}(k)=e_{k-3}(k)=e_{k-4}(k)=e_{k-5}(k)=0$ and the fixed constants are $e_4=2$, $e_2=3$, $e_0=1$, $e_{-3}=2$, $e_{-23}=1$. One sets $P(u)=2u^{-4}+3u^{-2}+1+2u^3+u^{23}$, the recurrence is

$$f_{n+1}(u) = P(u)f_n(u) - \{u^{<0}\}P(u)f_n(u) + \sum_{i=0}^{5} t_i(u)\partial_u^i f_n(1)\,,$$

where $\{u^{<0}\}$ stands for the sum of the monomials in u with a negative degree. Multiplying by z^{n+1} and summing for $n \geq 0$ leads to the functional equation

$$(1-zP(u))F(z,u) = 1 - z\sum_{k=0}^{4-1} r_k(u)F_k(z) + z\sum_{i=0}^{5} t_i(u)\partial_u^i F(z,1)\,, \tag{3}$$

where $r_k(u) := \{u^{<0}\}P(u)u^k$ and $t_i(u)$ are (Laurent) polynomials which can be made explicit.

One can use the kernel method (we refer to [3, 5] for recent applications of this method) to solve this equation. We call $1 - zP(u)$ the *kernel* of the equation. Solving $1 - zP(u) = 0$ with respect to u gives 4 roots $u_1(z)$, $u_2(z)$, $u_3(z)$ and $u_4(z)$ which are Puiseux series in $z^{1/4}$ and which tend to zero in 0. There are also 23 others roots which behave like $z^{-1/23}$ around 0, so we call $u_1, \dots, u_4$ the *small* roots of the kernel. Plugging the 4 small roots of the kernel in Equation 3 and considering the 6 other equations obtained by taking the first 5 derivatives of Equation 3 (and then setting $u = 1$) gives a system of full rank with 10 equations with 10 unknown univariate generating functions, which are thus all algebraic, and then one has a formula for $F(z, u)$, involving the u_i, which implies its algebraicity. For the general case, simply replace 4 by b and 5 by d in Equation 3 and then one can argue as in Theorem 1 above, with a new matricial equation $M.\overrightarrow{F} = \overrightarrow{v}$; looking at the valuation in z of each entries in M (some of them involves the small roots u_i's, but at most a product of b of them) gives $\det M \neq 0$ and thus a system of full rank, so $F(z, u)$ can be expressed as a rational function in z, u and the small roots u_i's. As these roots are algebraic, $F(z, u)$ is algebraic. □

Consider now the case where, for each i, the exponent $e_i(k)$ of the rule (2) is a constant (that is, the polynomial in $e_i(k)$ does not depend on k, so one simply writes e_i). How far can we relate the behavior of the walk

$$[(0), \{(k) \rightsquigarrow (0)^{e_k}(1)^{e_{k-1}} \dots (k-2)^{e_2}(k-1)^{e_1}(k)^{e_0}(k+1)^{e_{-1}} \dots (k+a)^{e_{-a}}\}] \tag{4}$$

to the generating function of the exponents $E(u) = \sum_{i \geq -a} e_i u^i$? We give here a first element of answer:

Theorem 3. *Consider the rule*

$$[(0), \{(k) \rightsquigarrow (0)^{e_k}(1)^{e_{k-1}} \dots (k-1)^{e_1}(k)^{e_0} \dots (k+a)^{e_{-a}}\}]\,. \tag{5}$$

If the generating function of the exponents $E(u)$ is algebraic then the bivariate generating function of the walk $F(z, u)$ is algebraic. For $a = 1$, one has

$$F(z, u) = \frac{F_0(z)}{1 - u\, e_{-1} z\, F_0(z)} \qquad \textit{with} \qquad F_0(z) = \frac{1}{e_{-1} z}\, E^{<-1>}(\frac{1}{z})$$

where $E^{<-1>}$ is the compositional inverse of $E(u)$ and where e_{-1} is the multiplicity of the $+1$ jump. More generally, for $a \geq 1$, the generating function $F(z, u)$ is expressed in terms of the a solutions $u_1(z), \dots, u_a(z)$ of $1 - zE(u) = 0$ which satisfy $u_i(z) \sim 0$ for $z \sim 0$:

$$F(z, u) = \sum_{k \geq 0} F_0(z) \left(\sum_{i_1 + \dots + i_a = k} u_1^{i_1} \dots u_a^{i_a} \right) u^k \qquad \textit{with} \qquad F_0(z) = \frac{(-1)^{a+1}}{z e_{-a}} \prod_{i=1}^{a} u_i(z)\,.$$

Writing σ_i for the homogeneous symmetrical polynomial of total degree i and of degree 1 in each variable and where each coefficient equals 1, one has

$$F(z, 1) = \frac{(-1)^{a+1}}{z e_{-a}} \frac{\prod_{i=1}^{a} u_i(z)}{1 + \sum_{i=1}^{a} (-1)^i \sigma_i(u_1, \dots, u_a)}\,.$$

Proof. For $a = 1$, the first identity reflects the combinatorial decomposition (one to one correspondence, in fact) "a walk from 0 to $k+1$" is "a walk from 0 to k" then followed by a jump $+1$ then followed by "a walk from $k+1$ to $k+1$ never going below $k+1$". The generating function of these last walks is clearly $F_0(z)$, thus one has $F_{k+1}(z) = F_k(z) e_{-1} z F_0(z) = F_0(z)(z e_{-1} F_0(z))^{k+1}$.

For the walks corresponding to the rule (5), the set of jumps is given by $E(1/u)$; if one reverses the time direction, one gets a new walk where the set of available jumps is given by $E(u)$. Define $\widetilde{F}(z, u)$ as the corresponding generating function (one starts at altitude 0), one has:

$$\widetilde{f}_{n+1}(u) = \{u^{\geq 0}\} E(u) \widetilde{f}_n(u), \qquad \widetilde{f}_0(u) = 1$$

where $\{u^{\geq 0}\}$ stands for the sum of all monomials in u with a nonnegative degree. Multiplying by z^{n+1} and summing for $n \geq 0$ gives

$$\widetilde{F}(z, u) = \widetilde{f}_0(u) + zE(u)\widetilde{F}(z, u) - z\{u^{-1}\} \frac{e_{-1}}{u} \widetilde{F}(z, u)\,,$$

that one rewrites as the following functional equation

$$(1 - zE(u))\widetilde{F}(z,u) = 1 - z\frac{e_{-1}}{u}\widetilde{F}_0(z)\,.$$

Then solving the "kernel" $1 - zE(u) = 0$ with respect to u gives a series $u_1(z) = E^{<-1>}(1/z)$, which is algebraic as the compositional inverse of an invertible algebraic function is algebraic (simply plug the inverse in the polynomial equation $\Phi(E(u), u) = 0$ satisfied by $E(u)$ to check this fact).

If one then evaluates the above functional equation at $u = u_1(z)$, one gets $0 = 1 - z\frac{e_{-1}}{u_1}\widetilde{F}_0(z)$ and thus $\widetilde{F}_0(z) = \frac{u_1}{e_{-1}z}$. As one has $\widetilde{F}_0(z) = F_0(z)$ (a walk from 0 to 0 from left to right is still a walk from 0 to 0 from right to left), one gets the result from the theorem. Note that if one sets $\widetilde{f}_0(u) = \frac{1}{1-u}$, $\widetilde{F}_0$ enumerates walks from anywhere to 0, so $\widetilde{F}_0(z) = \frac{u_1/(ze_{-1})}{1-u_1} = F(z,1)$, which is coherent with the theorem (case $a = 1$).

For $a \geq 1$, one sets $P(u) := \sum_{i=-a}^{-1} e_i u^i$; one has

$$(1 - zE(u))\widetilde{F}(z,u) = \widetilde{f}_0(u) - z\{u^{<0}\}P(u)\widetilde{F}(z,u)\,.$$

This is rewritten as

$$(1 - zE(u))\widetilde{F}(z,u) = \widetilde{f}_0(u) - z\sum_{k=0}^{a-1} r_k(u)\widetilde{F}_k(z)\,. \tag{6}$$

where $r_k(u) := \{u^{<0}\}P(u)u^k$ is a Laurent polynomial with monomials of degree going from -1 down to $k - a$.

The kernel equation $1 - zE(u) = 0$ has a roots $u_1(z), \ldots, u_a(z)$ which are Puiseux series in $z^{1/a}$ and which tend to 0 when z tends to 0. When $\widetilde{f}_0(u) = 1$, plugging these roots in the functional equation shows that they correspond to the a roots of the polynomial $u^a - zu^a\sum_{k=0}^{a-1} r_k(u)F_k(z)$, whose leading term is z^a and whose constant term is so $-ze_{-a}\widetilde{F}_0(z)$. This gives $\widetilde{F}_0(z) = \frac{-\prod_{i=1}^a u_i}{-ze_{-a}}$. When $\widetilde{f}_0(u) = \frac{1}{1-u}$ this gives a system of a equations for a unknowns (the $\widetilde{F}_k$'s). Solving it for $\widetilde{F}_0$ gives $F(z,1)$. Solving the $\widetilde{F}_0$ for $\widetilde{f}_0(u) = u^k$ gives the $F_k(z)$. The last identity for $F(z,1)$ follows from $\big(\sum_{k\geq 0}\sum_{i_1+\cdots+i_a=k} u_1^{i_1}\ldots u_a^{i_a}\big)\big(1 + \sum_{i=1}^{a}(-1)^i\sigma_i(u_1,\ldots,u_a)\big) = 1$.

For $a = 1$, the Riordan arrays approach that we presented in Section 1 also gives the algebraicity of $F(z,u)$. In fact, a theorem from [10] gives $F(z,u) = \frac{d(z)}{1-uzh(z)}$ where $h(z) = A(zh(z))$ and $d(z) = 1/(1 - zZ(zh(z)))$ for the rule $[(0), \{(k) \rightsquigarrow (0)^{z_k}(1)^{a_k}(2)^{a_{k-1}}\ldots(k)^{a_1}(k+1)^{a_0}\}]$. For $a > 1$, the matrix associated (see Section 1) to the rule (4) is called a *horizontally stretched Riordan array*. With this concept, it can be shown, like with the kernel method, that the algebraicity of the corresponding generating function $F(z,u)$ depends on the algebraicity of $A(z) = \sum_{k\geq 0} a_k z^k$ and $F_0(z), \ldots, F_{a-1}(z)$ (the generating functions of the first a columns of the matrix). While the theory of Riordan arrays has been intensively studied, the theory of stretched Riordan arrays, from a generating function point of view, is still in progress. □

Remark: as D-finite functions are not necessarily closed under compositional inverse, it is not true that if $E(u)$ is D-finite, then $F(z,1)$ or $F_0(z)$ (and a fortiori $F(z,u)$) are D-finite, even in the case $a = 1$.

We end with a last application of the kernel method.

Theorem 4. *Consider the rewriting rule (4) when the e_i's are ultimately constants (say, equal to a constant C): $[(0), \{(k) \rightsquigarrow (0)^C \ldots (k-b-1)^C(k-b)^{e_b}\ldots(k)^{e_0}\ldots(k+a)^{e_{-a}}\}]$. Then $F(z,u)$ is algebraic and satisfies*

$$F(z,u) = \frac{\prod_{i=0}^{b} u - u_i(z)}{K(z,u)}\,,$$

where the u_i's and K are defined as below.

Proof. One has the recurrence $f_{n+1}(u) = C\frac{f_n(u)-f_n(1)}{u-1} + P(u)f_n(u)$ this leads to the functional equation

$$\left(1 - zP(u) - z\frac{C}{u-1}\right)F(z,u) = 1 + \frac{C}{u-1}F(z,1) - z\sum_{k=0}^{b-1}\{u^{<0}\}P(u)u^k F_k(z) \tag{7}$$

where $P(u) = \sum_{i=1}^{b}(e_i - C)\frac{1}{u^i} + \sum_{i=0}^{a} e_{-i}u^i$. Define the kernel K as $K(u,z) = u^b(1-u)(1-zP(u) - \frac{zC}{u-1})$. It has b roots $u_1(z), \ldots, u_b(z)$ which are Puiseux series in $z^{1/b}$ and which tend to 0 in 0 and one root $u_0(z)$ which tends to 1 in 0. These are exactly the $b+1$ roots of the right hand part of (7) (once multiplied by $(1-u)u^b$). So $F(z,u) = \frac{\prod_{i=0}^{b} u-u_i(z)}{K(z,u)}$, where the u_i's are the $b+1$ small roots of the kernel. □

2.3. **Asymptotics.** Given a peculiar rule for Theorem 1, 2, 3 or 4, it is possible to find an asymptotic expansion for the number of walks. It is not really possible to merge all these results in a single one, as the rules are too unconstrained. However, for the algebraic case, a kind of universality holds for the behavior of the roots of the kernel. This leads to following theorem, which has to be adapted case by case for rules of Theorems 2 and 3 (and is easily applied to rules of Theorem 4).

Theorem 5. *The number of walks of length n for the "factorial" rule*

$$[(0), \{(k) \rightsquigarrow (0)(1)\ldots(k-b-1)(k-b)^{e_b}\ldots(k)^{e_0}\ldots(k+a)^{e_{-a}}\}]$$

(where $e_i(k) = 1$ for $b < i \leq k$ and $e_i(k) = e_i$, some fixed constants, for $i \leq b$) has the following asymptotics $A\frac{\rho^{-n}}{\sqrt{2\pi n^3}}$, where A and ρ are algebraic constants depending on the finite set of jumps $\mathcal{P}$.

Proof. See [1] for a proof and applications to the limit laws of final altitude and number of factors. The approach is similar to the one used for walks with a finite number of jumps but there are some complications due to the fact that the kernel is now of the kind $1 - z\phi(u)$ where $\phi(u)$ is not unimodal. One can however establish that the real positive root u_0 now dominates and has a square-root behavior. □

3. Examples

We now give a series of examples from combinatorics or computer science in which rewriting rules studied in Section 2 appear.

Example 1. *Two families of rules leading to an algebraic generating function.*
For the rule $[(0), \{(k) \rightsquigarrow (0)^{e_k}(1)^{e_{k-1}}\ldots(k-1)^{e_1}(k)^{e_0}(k+1)\}]$, where e_k for $i \geq 0$ is the number of t-ary trees with k nodes, $F(z,u)$ satisfies a algebraic equation of degree t. E.g., for $t = 3$, one has: $1 - \big(3 + (4-3u)z\big)F(z,u) - \big(-3 + (6u-7)z + (-3u^2+8u-3)z^2\big)F(z,u)^2 - \big(1 + (3-3u)z + (3u^2 - 7u+3)z^2 + (-u^3+4u^2-3u+1)z^3\big)F(z,u)^3 = 0$.

For the rule $[(0), \{(k) \rightsquigarrow (0)^{c+k}(1)^{c+k-1}\ldots(k-2)^{c+2}(k-1)^{c+1}(k)^c(k+1)\}]$, $F(z,u)$ satisfies an algebraic equation of degree 3:
$\big((1-2u)z^2 + (c-(c+1)+2u^2)\big)F^3 + \big((u-2)z + (-c-2+4u-2u^2)z^2\big)F^2 + \big(1+(2-2u)z\big)F = 1$.
□

Example 2. *Tennis ball problem.* Let $s \geq 2$ be an integer and consider the following problem known as *the s-tennis ball problem.* At the first turn one is given balls numbered one through s. One throws one of them out of the window onto the lawn. At the second turn balls numbered $s+1$ through $2s$ are brought in and now one throws out on the lawn any of the $2s-1$ remained. Then balls $2s+1$ through $3s$ are brought in and one throws out one of the $3s-2$ available balls. The game continues for n turns. At this point, one picks up the n balls in the lawn and consider the ordered sequence $B = (b_1, b_2, \ldots, b_n)$ with $b_1 < b_2 < \cdots < b_n$. This sequence will be called a *tennis ball s-sequence* and the first question is: how many tennis ball s-sequences of length n exist? The second question is: what is the sum of all the balls in all the possible s-sequences of length n ? Obviously, if we answer to both these questions, we also know the average sum of the balls in an s-sequence of length n. The general case $s \geq 1$ has been studied in [8] from a generating function viewpoint.

In fact, the authors consider an infinite tree with root 0 and with s children. Each $(n+1)$-length path in this tree corresponds to an s-sequence of length n. This infinite tree is isomorphic to the generating tree with specification $[(1), \{(k) \rightsquigarrow (1)\dots(k+s-2)(k+s-1)\}]$.

By using this result the authors find that the number of tennis ball s-sequences of length n are counted by T_{n+1}, where $T_n = \frac{1}{1+(s-1)n}\binom{sn}{n}$ (the number of s-ary trees with n-nodes) and the cumulative sum of all the balls thrown onto the lawn in n turn is

$$\Sigma_n = \frac{1}{2}(sn^2 + (3s-1)n + 2s)T_{n+1} - \frac{1}{2}\sum_{k=0}^{n+1}\binom{sk}{k}\binom{s(n+1-k)}{n+1-k}.$$

□

Example 3. A new rewriting rule for $(4,2)$-tennis ball problem.

The problem of balls on the lawn admits many other variants. For example, one could be supplied with s balls at each turn but now throw out t balls at a time with $t < s$. The general (s,t) case is an open problem while the $(4,2)$ case has been treated in [8], where the authors study the problem by introducing a bilabeled generating tree technique. Anyway, recently Merlini and Sprugnoli found that the problem can be expressed by the rule (4) with $e_i = i+3$ and $a = 2$, namely:

$$[(0), \{(k) \rightsquigarrow (0)^{k+3}(1)^{k+2}(2)^{k+1}\dots(k+2)\}] \tag{8}$$

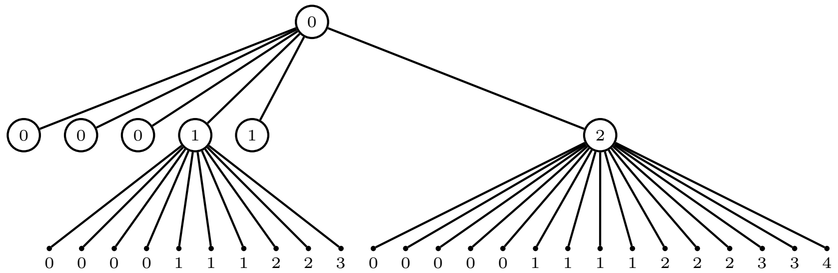


Figure 2. The partial generating tree for the specification (8)

In fact, if we don't care of the order of the balls thrown away, so that the configuration $(1,4)$, $(5,8)$, $(2,10)$ is considered to be the same as $(1,2)$, $(4,5)$, $(8,10)$, it can be proved that the number of $(4,2)$-sequences of length $2n$ in which the last-but-one element is $2n+k-1$ corresponds to the number of nodes with label k at level n in the generating tree of Figure 2 (for example, the possible sequences of length 2 are $(1,2)$, $(1,3)$, $(1,4)$, $(2,3)$, $(2,4)$ and $(3,4)$). □

Example 4. *Printers.*

In [9] the authors present a combinatorial model for studying the characteristics of job scheduling in a slow device, for example a printer in a local network. The policy usually adopted by spooling systems is called *First Come First Served* (FCFS) and can be realized by queuing the processes according to their arrival time and by using a FIFO algorithm. A job (printing a file) consists in a finite number of *actions* (printing-out a single page). Each action takes constant time to be performed (a *time slot*). If we fix n time slots, and suppose that at the end of the period the queue becomes empty, while it was never empty before, the successive states of the jobs queue can be described by a combinatorial structure called *labeled 1-histograms*. A *1-histogram* of length n is a histogram whose last column only contains 1 cell and, whenever a column is composed by k cells, then the next column contains at least $k-1$ cells. It is at all obvious that a 1-histogram corresponds to a path in the generating tree produced by the specification $[(1), (k) \rightsquigarrow (1)\dots(k+1)]$. A *labeled 1-histograms* of length n is a 1-histogram in which we label each cell according to some rules (see [9] for the details). Figure 3 illustrates the possible schedules for two particular 1-histograms of length 3: the first one, for example, corresponds to i) a first job which consists in printing two pages and a second job, which starts at time slot 2, and corresponds to printing a page at time slot 3, and ii) three different jobs which consists in printing a single page, the first at time slot 1, the second

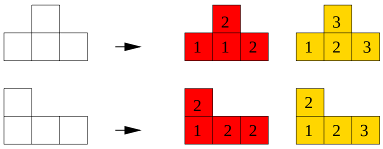


FIGURE 3. The schedules corresponding to two particular 1-histograms.

at time slot 2 and the third at time slot 3, after queuing at time slot 2. It can be proved that the number of schedules of length n with k jobs request at the first time slot corresponds to the number of nodes at level n having label $k+1$ in the generating tree with specification:

$$[(1), \{(k) \rightsquigarrow (1)^2 \ldots (k)^2 (k+1)\}] \,.$$

This gives that the number S_n of possible schedules corresponds to the n^{th} small Schröder number, that is, the generating function for S_n is $(1 - 3z - \sqrt{1 - 6z + z^2})/(4z)$.

$\square$

Acknowledgements. This work was partially supported by the Future and Emerging Technologies programme of the EU under contract number IST-1999-14186 (ALCOM-FT), by the INRIA postdoctoral programme and by the Max-Planck Institut. The first author also benefited of pasta and of an invitation in Florence in June and November 2001. Finally, the authors are grateful to an anonymous referee for her constructive comments.

References

[1] Cyril Banderier. Limit laws for basic parameters of lattice paths with unbounded jumps. *Submitted*, 2002.
[2] Cyril Banderier, Mireille Bousquet-Mélou, Alain Denise, Philippe Flajolet, Danièle Gardy, and Dominique Gouyou-Beauchamps. Generating functions for generating trees. *Discrete Mathematics*, 246(1-3):29–55, 2002.
[3] Cyril Banderier and Philippe Flajolet. Basic analytic combinatorics of directed lattice paths. *To appear in Theoretical Computer Science*, June 2001.
[4] Elena Barcucci, Alberto Del Lungo, Elisa Pergola, and Renzo Pinzani. From Motzkin to Catalan permutations. *Discrete Mathematics*, 217(1-3):33–49, 2000. Formal power series and algebraic combinatorics (Vienna, 1997).
[5] Mireille Bousquet-Mélou. Walks on the slit plane: other approaches. *Applied Math.*, 27(2-3):243–288, 2001.
[6] F. R. K. Chung, R. L. Graham, V. E. Hoggatt, and M. Kleiman. The number of Baxter permutations. *Journal of Combinatorial Theory, Series A*, 24:382–394, 1978.
[7] D. Merlini, D. G. Rogers, R. Sprugnoli, and M. C. Verri. On some alternative characterizations of Riordan arrays. *Canadian Journal of Mathematics*, 49(2):301–320, 1997.
[8] D. Merlini, D. G. Rogers, R. Sprugnoli, and M. C. Verri. The tennis ball problem, 2002. To appear in Journal of Combinatorial Theory-A.
[9] D. Merlini, R. Sprugnoli, and M. C. Verri. Waiting patterns for a printer. In *FUN with algorithms 2, E. Lodi, L. Pagli, N. Santoro, Editors*, Carleton Scientific, pages 183–198, 2001.
[10] D. Merlini and M. C. Verri. Generating trees and proper Riordan Arrays. *Discrete Mathematics*, 218:167–183, 2000.
[11] L. W. Shapiro, S. Getu, W.-J. Woan, and L. Woodson. The Riordan group. *Discrete Applied Mathematics*, 34:229–239, 1991.

INRIA-Rocquencourt, 78150 Le Chesnay, France, and Max-Planck-Institut, 66123 Saarbrücken, Germany
E-mail address: `Cyril.Banderier@inria.fr`
URL: `http://algo.inria.fr/banderier`

Dipartimento di Sistemi e Informatica, Universit degli Studi di Firenze, Via Lombroso 6/17, 50134 Firenze, Italy
E-mail address: `merlini@dsi.unif.it`
URL: `http://www.dsi.unifi.it/~merlini/`

DIRECTED-CONVEX POLYOMINOES: ECO METHOD AND BIJECTIVE RESULTS

ELENA BARCUCCI, ANDREA FROSINI, AND SIMONE RINALDI

Abstract. In this paper we provide bijective proofs for the number of directed-convex polyominoes having a fixed number of rows and columns, both by means of the ECO method, and of a direct mapping into the set of 2-colored Grand Motzkin paths.

Résumé. Dans cet article, nous donnons des preuves bijectives pour le nombre de polyominos dirigés convexes ayant un nombre fixé de lignes et de colonnes, en utilisant la méthodo;ogie ECO ainsi qu'une application bijective dans l'ensemble des grands chemins de Motzkin bi-colorés.

1. ECO method and directed-convex polyominoes

A *polyomino* is a finite union of elementary cells of the lattice $Z \times Z$, whose interior is connected. A polyomino is said to be *vertically convex* [*horizontally convex*] when its intersection with any vertical [horizontal] line is convex. A polyomino is *convex* if it is both vertically and horizontally convex. A polyomino is said to be *directed* when each of its cells can be reached from a distinguished cell, called the root, by a path which is contained in the polyomino and uses only north and east unitary steps. A polyomino is *directed-convex* if it is both directed and convex (see Fig. 1 (a)).

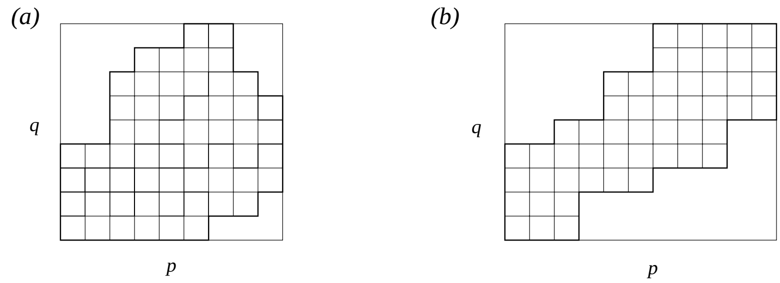


Figure 1. (a) A directed convex polyomino; (b) a parallelogram polyomino.

A *parallelogram polyomino* is a polyomino whose boundary consists of two lattice paths that intersect only initially and finally. The boundary paths, which we call upper and lower path, use the positively directed unit steps, $(1,0)$ and $(0,1)$ (see Fig. 1, (b)). Chang and Lin [3] used analytic methods to prove that the number of directed-convex polyominoes and the number of parallelogram polyominoes having q rows and p columns are equal to

$$\binom{p+q-2}{p-1}\binom{p+q-2}{q-1}, \tag{1}$$

and

$$\frac{1}{p+q-1}\binom{p+q-1}{p-1}\binom{p+q-1}{q-1}, \tag{2}$$

respectively (the second formula is originally due to Narayana, [7]). For polyominoes having $n+1$ rows and $n+1$ columns, these formulas reduce to

$$\binom{2n}{n}^2, \tag{3}$$

and

$$\frac{1}{2n+1}\binom{2n+1}{n}^2, \tag{4}$$

respectively. Furthermore, from (1) and (2) it arises that the number of directed-convex polyominoes having semiperimeter $p+q=n+2$ is equal to

$$\sum_{\substack{p,q\geq 1\\ p+q=n+2}} \binom{p+q-2}{p-1}^2 = \sum_{k=0}^{n}\binom{n}{k}^2 = \binom{2n}{n}, \tag{5}$$

that is, the well known central binomial coefficient (sequence M1645 in [11]), and, analogously, the number of parallelogram polyominoes having semiperimeter $n+2$ is equal to

$$\frac{1}{n+1}\binom{2n}{n}, \tag{6}$$

the nth Catalan number (sequence M1459 in [11]).

In this paper we consider the classes:

(1) $\mathcal{D}_n$ the class of directed-convex polyominoes having semi-perimeter n;
(2) $\mathcal{P}_n$ the class of parallelogram polyominoes having semi-perimeter n;
(3) $\mathcal{D}_{p,q}$ the class of directed-convex polyominoes having p columns and q rows;
(4) $\mathcal{P}_{p,q}$ the class of parallelogram polyominoes having p columns and q rows.

In this section we use ECO method to provide a new bijective proof that the numbers of directed-convex polyominoes and the number of parallelogram polyominoes having semiperimeter $n+2$ are equal to (5) and (6), respectively, and then we obtain some statistics on directed-convex polyominoes.

1.1. **An ECO operator to construct directed-convex polyominoes.** ECO (Enumerating Combinatorial Objects) [1] is a method for the enumeration and the recursive construction of a class of combinatorial objects, $\mathcal{O}$, by means of an operator ϑ which performs "local expansions" on the objects of $\mathcal{O}$. More precisely, let p be a parameter on $\mathcal{O}$, such that $|\mathcal{O}_n| = |\{O \in \mathcal{O} : p(O) = n\}|$ is finite. An operator ϑ on the class $\mathcal{O}$ is a function from $\mathcal{O}_n$ to $2^{\mathcal{O}_{n+1}}$, where $2^{\mathcal{O}_{n+1}}$ is the power set of $\mathcal{O}_{n+1}$.

Proposition 1. *Let ϑ be an operator on $\mathcal{O}$. If ϑ satisfies the following conditions:*

1.: *for each $O' \in \mathcal{O}_{n+1}$, there exists $O \in \mathcal{O}_n$ such that $O' \in \vartheta(O)$,*
2.: *for each $O, O' \in \mathcal{O}_n$ such that $O \neq O'$, then $\vartheta(O) \cap \vartheta(O') = \emptyset$,*

then the family of sets $\mathcal{F}_{n+1} = \{\vartheta(O) : O \in \mathcal{O}_n\}$ is a partition of $\mathcal{O}_{n+1}$.

We refer to [1] for further details, proofs and definitions. The recursive construction determined by ϑ can be described by a *generating tree* [4], whose vertices are objects of $\mathcal{O}$. The objects having the same value of the parameter p lie at the same level, and the sons of an object are the objects it produces through ϑ. Thus a generating tree defines a non-decreasing sequence $(f_n)_{n\geq 0}$ of positive integers, f_n being the number of nodes at level n in the tree.

Let us now consider the operator:

$$\vartheta : \mathcal{D}_n \to 2^{\mathcal{D}_{n+1}},$$

working as follows on any given $P \in \mathcal{D}_n$, such that the length of its rightmost column is equal to k (see Fig. 2):

i): ϑ glues a unitary cell to the right of each cell constituting the rightmost column of P;

ii): ϑ glues a column of length h, $2 \leq h \leq k$ to the rightmost column of P;

iii): ϑ glues a cell onto the top of the rightmost cell of the topmost row of P, one cell onto the top of the rightmost column of P, and two cells onto the top of each column between the two inserted ones (if there is any). If P is a parallelogram polyomino, then the operation **iii)** reduces to adding a cell on the top of the rightmost column of P.

The reader can check that ϑ satisfies conditions 1. and 2. of Proposition 1, meaning that ϑ constructs each polyomino $P' \in \mathcal{D}_{n+1}$ from the polyominoes in $\mathcal{DC}_n$, and each polyomino $P' \in \mathcal{DC}_{n+1}$ is obtained from one and only one $P \in \mathcal{D}_n$.

It should be clear that $|\vartheta(P)| = 2k$, then the generating tree of ϑ can be described by means of a *succession rule* of the form:

$$(7) \qquad \begin{cases} (2) \\ (2) \rightsquigarrow (2)(4) \\ (2k) \rightsquigarrow (2)^k(4)(6)\dots(2k)(2k+2), \end{cases}$$

where the power notation stands for repetitions. The expression in (7) means that the root object has 2 sons, and the $2k$ objects $O'_1, \dots, O'_{2k}$, produced by an object O are such that $|\vartheta(O'_i)| = 2$, $1 \leq i \leq k$, and $\left|\vartheta(O'_{k+j})\right| = 2(j+1)$, $1 \leq j \leq k$.

In the next paragraph we prove (5) and (6) in a bijective way, instead of determining the generating function of the rule in (7).

1.2. **A construction for Grand Dyck paths.** In the discrete plane, a Grand Dyck path is a sequence of rise $(1,1)$, and fall $(1,-1)$ steps running from $(0,0)$ to $(2n,0)$. In a Grand Dyck path, we call peak (valley, resp.) each pair of consecutive rise and fall steps (fall and rise steps, resp.). If the path ends with a fall step, we call last descent the last sequence of fall steps of the path. The number of Grand Dyck paths having semi-length n is well-known to be equal to $\binom{2n}{n}$. We define an operator ϑ' which constructs the class of Grand Dyck paths according to the succession rule (7). Let G be a Grand Dyck path of semi-length n:

: - if the last step of G is a rise step, then $\vartheta'(G)$ is obtained by inserting a peak or a valley into the last point of G (Figure 3, (a)); in this case $|\vartheta'(G)| = 2$;

: - otherwise, $\vartheta'(G)$ is obtained by (Figure 3, (b)):

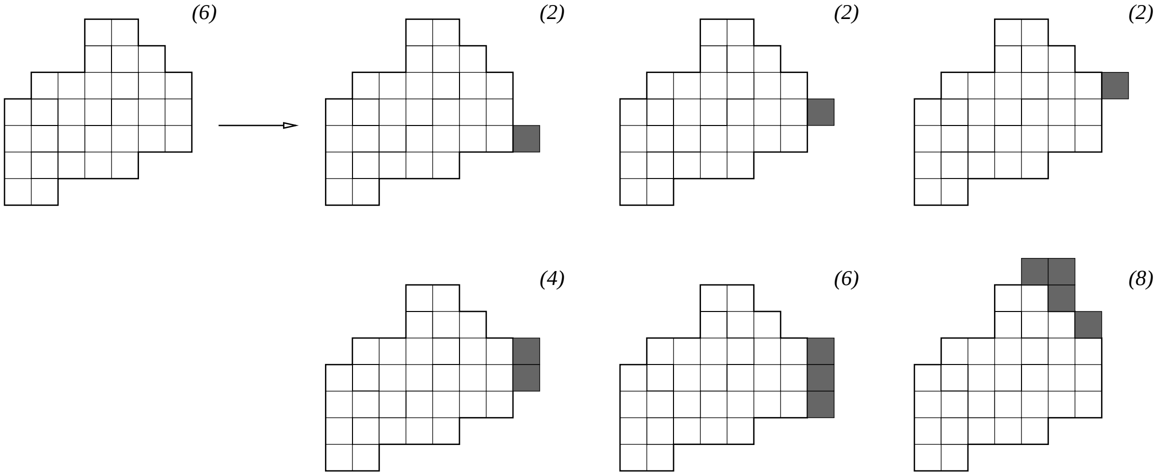


FIGURE 2. The ECO operator on directed-convex polyominoes.

A: inserting a peak into any point of the last descent of G,
B: reflecting all the paths obtained trough A (excepted the one obtained adding a peak at the end of G),
C: inserting a valley into the last point of G.

In this case, if k is the number of points in the last descent of G, then $|\vartheta'(G)| = k + (k-1) + 1 = 2k$.

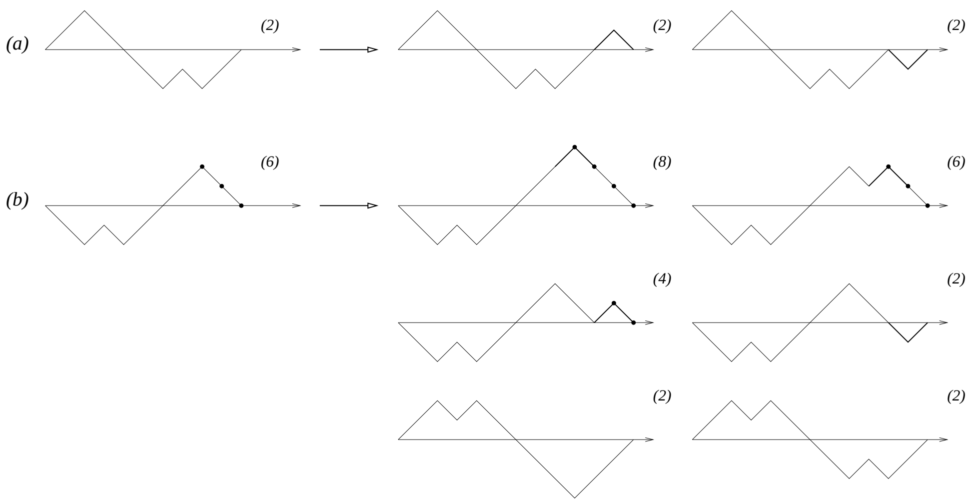


FIGURE 3. The application of the operator ϑ' to a Grand Dyck path.

It is not difficult to verify that ϑ' satisfies the conditions of Proposition 1. This fact gives a proof of (5) and establishes a bijection between the class of directed-convex polyominoes having semi-perimeter $n+2$ and the class of Grand Dyck paths having semi-length n.

1.3. **Further results.** Let $a_{n,k}$, $n \geq 0$, $k \geq 1$, be the number of labels $(2k)$ at level n of the generating tree of the succession rule (7) and consider the infinite lower triangular matrix $(a_{n,k})_{n,k\geq 0}$; in particular, for $n \geq 1$:

$$
(8) \qquad \begin{aligned} a_{n+1,1} &= a_{n,1} + 2a_{n,2} + \ldots + (n+1)a_{n,n+1}, \\ a_{n+1,k} &= a_{n,k-1} + a_{n,k} + \ldots + a_{n,n+1}, \qquad k > 1 \end{aligned}
$$

n	$\binom{2n}{n}$	2	4	6	8	10	12
0	1	1	0	0	0	0	0
1	2	1	1	0	0	0	0
2	6	3	2	1	0	0	0
3	20	10	6	3	1	0	0
4	70	35	20	10	4	1	0
5	252	126	70	35	15	5	1

TABLE 1. The matrix filled by the numbers $a_{n,k}$, and the row sums.

Moreover, the numbers $a_{n,k}$ have a nice closed form:

$$
a_{n,k} = \binom{2n-k}{n-1}.
$$

As a neat consequence of the constructions in Sections 1.1 and 1.2, we have that:

: i) There is a bijection between directed-convex polyominoes having semiperimeter $n+2$ where the right-most column is made of a single cell and Grand Dyck paths having semi-length n and ending with a rise step. The cardinality of these sets (sequence M2848 in [11]) is:

$$
a_{n,1} = \frac{1}{2}\binom{2n}{n}, \quad n \geq 1.
$$

: ii) There is a bijection between directed-convex polyominoes having semiperimeter $n+2$ where the right-most column is made of k cells, $k \geq 2$, and Grand Dyck paths having semi-length n where the last descent is made of $k-1$ fall steps. The cardinality of these sets is:

$$
a_{n,k} = \binom{2n-k}{n-1}, \quad n \geq 1.
$$

Remark 1. Let ϑ_P be the restriction of ϑ to the set of parallelogram polyominoes. Then

$$
\vartheta_P : \mathcal{P}_n \to 2^{\mathcal{P}_{n+1}},
$$

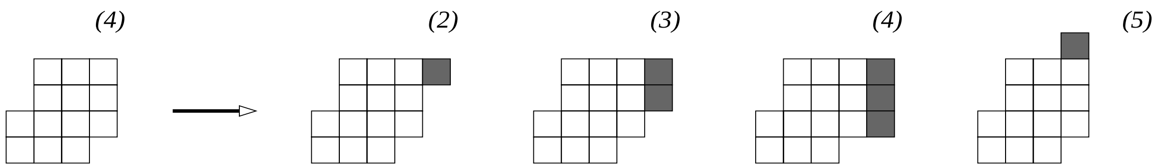


FIGURE 4. The operator ϑ_P for parallelogram polyominoes.

works as follows on a parallelogram polyomino P whose rightmost column is made of k cells (see Fig. 4):

i): it glues a column of length h, $1 \le h \le k$, to the rightmost column of P;
ii): it glues a cell onto the top of the rightmost column of P.

The operator $\vartheta_{\mathcal{P}}$ coincides with the classical ECO operator for parallelogram polyominoes [1], and gives rise to the well-known succession rule:

$$\begin{cases} (2) \\ (k) \rightsquigarrow (2)(3)(4)\dots(k)(k+1), \end{cases} \tag{9}$$

defining Catalan numbers.
Remark 2. As it is known [8], also the following rule defines central binomial coefficients,

$$\begin{cases} (2) \\ (2) \rightsquigarrow (3)(3) \\ (k) \rightsquigarrow (3)(3)(4)\dots(k)(k+1) \end{cases} \tag{10}$$

and an ECO operator exists which describes the recursive growth of Grand Dyck paths according to (10). The authors wish to point out that to find an ECO operator describing the growth of directed-convex polyominoes according to (10) is still an open problem.

2. A bijection between 2-colored Grand Motzkin paths and directed convex polyominoes

In this section we prove that:
- the class $\mathcal{DC}_{p,q}$ is enumerated by (1),
- the class $\mathcal{PP}_{p,q}$ is enumerated by (2),
by establishing a direct bijection between the class $\mathcal{D}_{p,q}$ (with $p, q \ge 1$), and a special subclass of 2-colored Grand Motzkin paths. Then we naturally extend our bijection to the class of directed-convex polyominoes with semiperimeter $p+q$.
The 2-colored Grand Motzkin paths are paths in the $Z \times Z$ plane which use four different steps: the rise step $(1,1)$, the fall step $(1,-1)$, the α-colored horizontal step $(1,0)$ and β-colored horizontal step $(1,0)$. They start from $(0,0)$ and end in $(n,0)$ (see Fig. 5). Let us define the class of 2-colored Motzkin paths as the class of 2-colored Grand Motzkin paths which remain weakly above the x-axis.
For any $p, q \ge 1$, let $M_{p,q}$ denote the class of 2-colored Grand Motzkin paths of length $p+q-2$ where the difference between the number of α-colored and β-colored steps is $p-q$ and $H_{p,q}$ the class prefixes of 2-colored Motzkin paths of length $p+q-2$ where the difference between the number of α-colored and β-colored steps is $p-q$ and whose final point ordinate is a positive even number.

2.1. The bijection between $\mathcal{D}_{p,q}$ and $M_{p,q}$.

We first easily prove that the classes $\mathcal{D}_{p,q}$ and $M_{p,q}$ have the same cardinality.
The cardinality of $M_{p,q}$. Let us take into consideration a generic path $P \in M_{p,q}$ and code each step with a 2×1 matrix as follows:

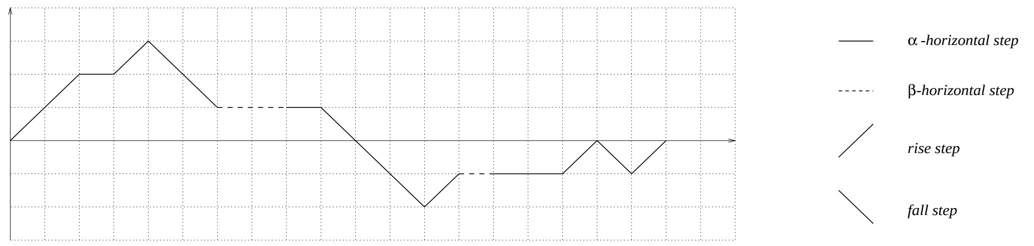


FIGURE 5. A 2-colored Grand Motzkin path.

$$
\begin{array}{llll}
\binom{1}{0} & \text{for a rise step,} & \binom{0}{1} & \text{for a fall step,} \\
\binom{1}{1} & \text{for a } \alpha\text{-horizontal step,} & \binom{0}{0} & \text{for a } \beta\text{-horizontal step.}
\end{array}
\tag{11}
$$

P can be univocally represented by the $2 \times (p+q-2)$ matrix M obtained concatenating the coding of its steps as shown in Fig. 6. Let u and l be the upper and the lower row of M.

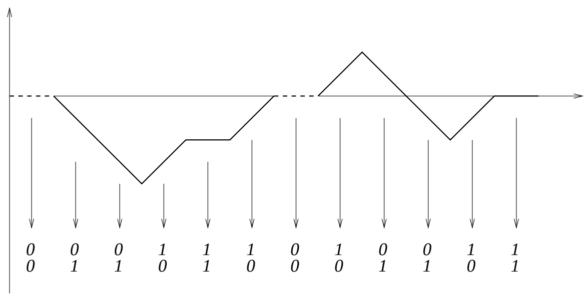


FIGURE 6. Encoding of a 2-colored Grand Motzkin path with a binary matrix.

Let j be a binary sequence, we indicate with $|j|_0$ and $|j|_1$ the number of 0 and 1 in j. We observe that:

: - $|u|_0 = |l|_0 = q - 1$;
: - $|u|_1 = |l|_1 = p - 1$.

From the previous two equations naturally follows:

$$|M_{p,q}| = \binom{p+q-2}{p-1}\binom{p+q-2}{q-1}.$$

The bijection between $\mathcal{D}_{p,q}$ and $H_{p,q}$. Let Q be a polyomino in $\mathcal{D}_{p,q}$ and $p \geq q$ (we can assume that with no loss of generality). Let $(p-h, q-h)$ be the last intersection point between Q and the line running from $(p-q, 0)$ to (p, q). Figure 7 shows in a simple graphical fashion that the polyomino is uniquely determined by two internal paths, both running from the initial cell to the cell identified by the points $(p-h, q-h)$, and $(p-h-1, q-h-1)$. These two paths are made up of $p+q-2$ steps: the upper, say u, is made of $(0,1)$, $(1,0)$ and $(0,-1)$ steps, and the lower, say l, is made of $(1,0)$, $(0,-1)$ and $(-1,0)$ steps.
Both u and l can be coded by a $p+q-2$ length binary vector, where 1 represents both the steps $(0,1)$ and $(0,-1)$ and 0 represents both the steps $(1,0)$ and $(-1,0)$. Then a $2 \times (p+q-2)$ binary matrix M is associated to the Q, where the first row is the coding

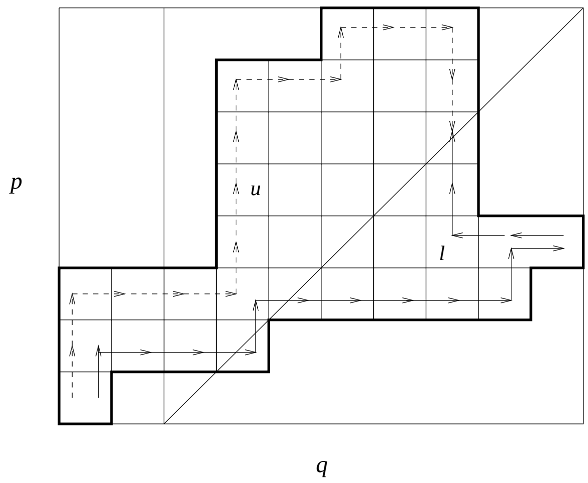


FIGURE 7. The upper and lower paths for a directed convex polyomino.

of u and the second row is the coding of l. For example, the following matrix encodes the polyomino in Fig. 7:

$$M = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Let us point out three properties of u and l:

(1) for every prefix s of u and every prefix v of l, having the same length, we have $|s|_1 \geq |v|_1$;
(2) the difference between the number of columns $\binom{0}{0}$ and $\binom{1}{1}$ of M is $p - q$;
(3) $|u|_1 - |l|_1 = 2h$.

Besides, the matrix M can be viewed as an array of $p + q - 2$ vectors 2×1 where each column represents a unitary step, using the code in (11). The properties 1., 2., and 3. guarantee that the obtained path is an element of $H_{p,q}$. In Figure 8 the 2-colored Motzkin prefix corresponding to the polyomino in Fig. 7 is depicted.

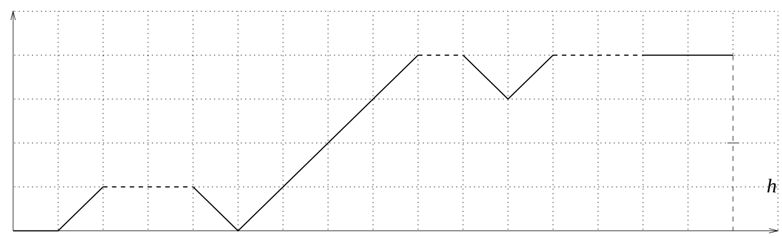


FIGURE 8. The 2-colored Motzkin prefix associated with the polyomino in Fig. 7.

To complete our bijection let us consider a generic $P' \in H_{p,q}$ and evaluate the parameters p and q. Using again the coding (11) we obtain a binary $2 \times (p + q - 2)$ matrix where the first row represents the upper paths u and the second row the lower path l of an element $Q \in \mathcal{D}_{p,q}$.

In the particular case that Q is a parallelogram polyomino we have $|u|_1 - |l|_1 = 0$ and the corresponding $P \in H_{p,q}$ is a 2-colored Motzkin path. We wish to point out that the classical bijection between parallelogram polyominoes and 2-colored Motzkin paths [5] arises naturally as a special case of our bijection.

The bijection between $H_{p,q}$ and $M_{p,q}$. Let P' be a path in $H_{p,q}$ and let $(p+q-2, 2h)$ be its final point coordinates. If $h = 0$, then P' is already a 2-colored Motzkin path. Otherwise, for every $i = 0, \ldots, h-1$ we consider the vertical side of unitary length from the point $(p+q-2, i)$ to the point $(p+q-2, i+1)$. We then draw a horizontal ray to the left from the center of this side. There are h such rays. Each ray hits for the first time a rise step in P'.

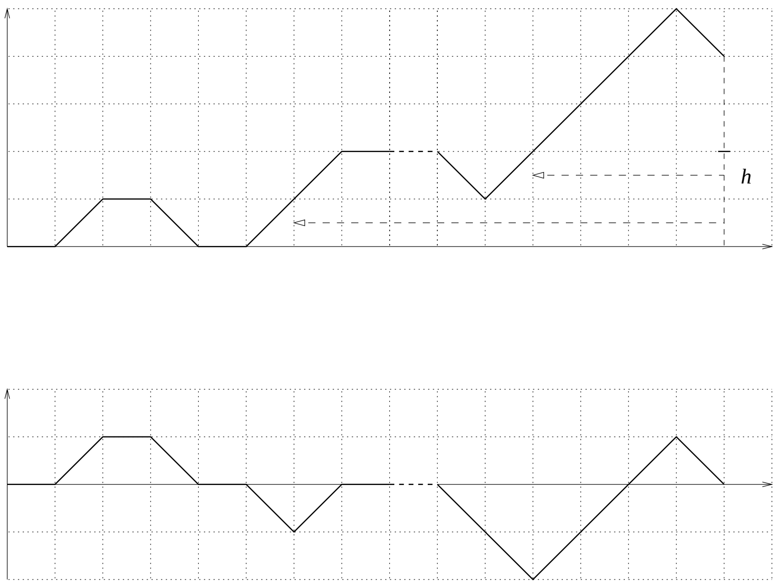


FIGURE 9. The mapping of a prefix of $H_{p,q}$ into a path of $M_{p,q}$.

We modify P' by changing the steps that are hit to fall steps. In this modified path the number of rise step is equal to the number of fall steps, while the difference between the numbers of α-horizontal and β-horizontal steps is the same as in P'. The obtained path is that corresponding to P' (see Fig. 9).

This mapping is inverted as follows (see Fig. 10). Let Q be a generic path in $M_{p,q}$, and let $-h$, $h > 0$ be the ordinate of the lowest point of Q. From each of the points $(0, -\frac{1}{2})$, $(0, -1-\frac{1}{2})$, …, $(0, -h+1-\frac{1}{2})$, we draw a ray to the right until it hits Q, necessarily at a fall step. Let Q' be the path obtained from Q in which each hit step is changed to a rise step. The path $Q' \in H_{p,q}$, and its final point ordinate is equal to $2h$.

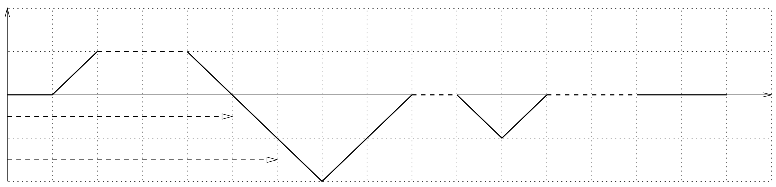


FIGURE 10. The inverse mapping of a $M_{p,q}$ path into the prefix of Fig. 8.

Remark 3. As a consequence of the previously defined bijection we have that:

- **-:** the height of a directed convex polyomino P is equal to the sum of the numbers of rise and α-colored horizontal steps plus one, in the path corresponding to P;

-: the width of a directed convex polyomino P is equal to the sum of the numbers of rise and β-colored horizontal steps plus one, in the path corresponding to P.

Remark 4. The reader should be convinced that the previously defined bijection defines also a bijection between the set of directed-convex polyominoes having $n+1$ rows and $n+1$ columns and the set of Grand Motzkin paths of length $2n$ and having the same number of α and β-colored horizontal steps. The latter class is trivially enumerated by the numbers

$$\binom{2n}{n}^2,$$

and this proves (3). The result (4) is then a neat application of the cycle lemma.

Acknowledgements. The authors wish to thank Emeric Deutsch for carefully reading the paper and for his enthusiastic help in the study of succession rules.

References

[1] E. Barcucci, A. Del Lungo, E. Pergola, R. Pinzani, ECO: a methodology for the Enumeration of Combinatorial Objects, *Journal of Difference Equations and Applications,* Vol.5 (1999) 435-490.

[2] W. Breckenridge, H. Gastineau-Hills, A. Nelson, P. Bos, G. Calvert, and K. Wehrhahn, Lattice paths and Catalan numbers, *Bull. Inst. Comb. and its App.*, 1 (1991) 41-55.

[3] S. J. Chang, and K. Y. Lin, Rigorous results for the number of convex polygons on the square and honeycomb lattices, *J. Phys. A: Math. Gen.*, 21 (1988) 2635-2642.

[4] F. R. K. Chung, R. L. Graham, V. E. Hoggatt, M. Kleimann, The number of Baxter permutations, *Journal of Combinatorial Theory Ser. A,* 24 (1978) 382-394.

[5] M. Delest, and X. Viennot, Algebraic languages and polyominoes enumeration, *Theor. Comp. Sci.*, 34 (1984) 169-206.

[6] R. K. Guy, C. Krattenthaler, and B. Sagan, Lattice paths, reflections and dimension-changing bijections, *Ars Combinatorica*, 34 (1992) 3-15.

[7] T. V. Narayana, Sur les treillis formés par les partitions d'un entier; leurs applications à la théorie des probabilités, *Comp. Rend. Acad. Sci. Paris*, 240 (1955) 1188-9.

[8] E. Pergola, R. Pinzani, S. Rinaldi Approximating algebraic functions by means of rational ones, *Theoretical Computer Science,* (to appear).

[9] G. Pólya, On the number of certain lattice polygons, *Journal Combinatorial Theory,* 6 (1969) 102-105.

[10] L. Shapiro, S. Getu, W-J. Woan, L. C. Woodson, The Riordan group, *Discrete Applied Mathemathics,* 34 (1991) 229-239.

[11] N. J. A. Sloane, and S. Plouffe, *The Encyclopedia of Integer Sequences,* Academic Press, New York (1995).

[12] R. P. Stanley, *Enumerative Combinatorics,* Vol.2, Cambridge University Press, Cambridge (1999).

Università di Firenze, Dipartimento di Sistemi e Informatica, via Lombroso, 6/17, 50134, Firenze, Italy
E-mail address: `barcucci@dsi.unifi.it.`

Università di Siena, Dipartimento di Matematica, via del Capitano, 15, 53100, Siena, Italy
E-mail address: `{frosini, rinaldi}@unisi.it.`

DESCENT NUMBER AND MAJOR INDICES FOR THE EVEN-SIGNED PERMUTATION GROUP

RICCARDO BIAGIOLI

ABSTRACT. We introduce and study three new statistics on the even-signed permutation group D_n. We show that two of these are Mahonian, i.e. are equidistributed with length, and that a pair of them gives a generalization of Carlitz's identity on the Euler-Mahonian distribution of the descent number and major index over S_n.

RÉSUMÉ. Nous presentons et étudions trois nouvelles statistiques sur le groupe de Coxeter de type D_n. Nous demontrons que parmi ces trois, deux sont "Mahonian", c'est-à-dire equidistribuées avec la longueur, et que deux autres portent à une généralisation de l'identité de Carlitz sur la distribution de Euler-Mahonian du nombre de descentes et du major index sur S_n.

1. INTRODUCTION

A well known classical result due to MacMahon (see [15]) asserts that the inversion number and the major index are equidistributed on the symmetric group. The joint distribution of major index and descent number was studied by Carlitz [7] and others. Several results of this nature have been generalized to the hyperoctahedral group B_n (see, e.g., [6],[13]) and many candidates for a major index for B_n have been suggested (see, e.g., [8],[9],[10],[12],[18]), but no generalizations of MacMahon's result have been found until the discovery of the flag major index in the recent paper [1]. After that, Foata posed the problem of finding a "descent statistic" that, together with the flag major index, allows the generalization to B_n of the well known Carlitz's identity on the Euler-Mahonian distribution of descent number and major index over S_n. In [2] Adin, Brenti and Roichman give two answers to Foata's question. Now it's natural to wonder if some of these statistics and results can be generalized to the even-signed permutation group D_n.

The goal of this paper is to show that this is the case. More precisely, we introduce and study three new statistics on D_n; the D-negative descent number ($ddes$), the D-negative major index ($dmaj$) and the D-flag major index ($fmaj_D$). When restricted to S_n, $ddes$ reduces to descent number and $dmaj$ to the major index. The two major indices on D_n are equidistributed with length, and the pair $(ddes, dmaj)$ gives a generalization of Carlitz's identity to D_n.

The organization of this extended abstract is as follows. In the next section we collect some definitions, notation and results that are needed in the rest of the work. In §3 we introduce a new "descent set" and hence in a very natural way new definitions of "descent number" and "major index" on D_n. It's shown that $dmaj$ is equidistributed with length and that $(ddes, dmaj)$ gives a generalization of Carlitz's identity. In §4 we define, in terms of Coxeter elements, the D-flag major index for D_n and we show that it's equidistributed with length. Furthermore, we describe a combinatorial algorithm to compute it.

2. Notation, Definitions and Preliminaries

In this section we give some definitions, notation and results that will be used in the rest of this work. We let $\mathbf{P} := \{1,2,3, \dots\}$, $\mathbf{N} := \mathbf{P} \cup \{0\}$, and $\mathbf{Z}$ be the set of integers; for $a \in \mathbf{N}$ we let $[a] := \{1, 2, \dots, a\}$ (where $[0] := \emptyset$). Given $n, m \in \mathbf{Z}$, $n \leq m$, we let $[n, m] := \{n, n+1, \dots, m\}$. The cardinality of a set A will be denoted by $|A|$ and we let $\binom{[n]}{2} := \{S \subseteq [n] : |S| = 2\}$. More generally, given a multiset $M = \{1^{a_1}, 2^{a_2}, \dots, r^{a_r}\}$ we denote by $|M|$ its cardinality, so $|M| = \sum_{i=1}^{r} a_i$. Given a variable q and a commutative ring R we denote by $R[q]$ (respectively, $R[[q]]$) the ring of polynomials (respectively, formal power series) in q with coefficient in R. For $i \in \mathbf{N}$ we let, as customary, $[i]_q := 1+q+q^2+\dots+q^{i-1}$ (so $[0]_q = 0$).
Given a sequence $\sigma = (a_1, \dots, a_n) \in \mathbf{Z}^n$ we say that $(i, j) \in [n] \times [n]$ is an *inversion* of σ if $i < j$ and $a_i > a_j$. We say that $i \in [n-1]$ is a *descent* of σ if $a_i > a_{i+1}$. We denote by $Inv(\sigma)$ and $Des(\sigma)$ the set of inversions and the set of descents of σ and with $inv(\sigma)$ and $des(\sigma)$ their cardinality, respectively. We also let

$$maj(\sigma) := \sum_{i \in Des(\sigma)} i \tag{1}$$

and call it the *major index* of σ.

Let S_n be the set of all bijections $\sigma : [n] \to [n]$. If $\sigma \in S_n$ then we write $\sigma = \sigma_1 \dots \sigma_n$ to mean that $\sigma(i) = \sigma_i$, for $i = 1, \dots, n$. If $\sigma \in S_n$ then we may also write σ in *disjoint cycle form* (see, e.g., [16, p.17]) and we will usually omit to write the 1-cycles of σ. For example, if $\sigma = 64175823$ then we also write $\sigma = (2, 4, 7)(1, 6, 8, 3)$. Given $\sigma, \tau \in S_n$ we let $\sigma\tau := \sigma \circ \tau$ (composition of functions) so that, for example, $(1, 2)(2, 3) = (1, 2, 3)$.

We denote by B_n the group of all bijections π of the set $[-n, n] \setminus \{0\}$ onto itself such that

$$\pi(-a) = -\pi(a)$$

for all $a \in [-n, n] \setminus \{0\}$, with composition as the group operation. This group is usually known as the group of *signed permutations* on $[n]$, or as the *hyperoctahedral group* of rank n. We identify S_n as a subgroup of B_n, and B_n as a subgroup of S_{2n}, in the natural ways.

If $\pi \in B_n$ then we write $\pi = [a_1, \dots, a_n]$ to mean that $\pi(i) = a_i$ for $i = 1, \dots, n$, we call this the *window* notation of w, and we let

$$\begin{array}{rcl} inv(\pi) & := & inv(a_1, \dots, a_n), \\ des(\pi) & := & des(a_1, \dots, a_n), \\ maj(\pi) & := & maj(a_1, \dots, a_n), \\ Neg(\pi) & := & \{i \in [n] : a_i < 0\}, \\ N_1(\pi) & := & |Neg(\pi)|, \end{array} \tag{2}$$

and

$$N_2(\pi) := |\{\{i, j\} \in \binom{[n]}{2} : a_i + a_j < 0\}|. \tag{3}$$

We denote by D_n the subgroup of B_n consisting of all the signed permutations having an even number of negative entries in their window notation, more precisely

$$D_n := \{\pi \in B_n : N_1(\pi) \equiv 0 \ (\ mod\ 2\ )\}.$$

Obviously, the definitions in (2) and (3) are still valid for $\pi \in D_n$.
It is well known (see, e.g., [5, §8.2]) that D_n is a Coxeter group with respect to the generating set $S := \{s_0, s_1, \dots, s_{n-1}\}$ where

$$s_0 := [-2, -1, 3, \dots, n]$$

and

$$s_i := [1, 2, \ldots, i-1, i+1, i, i+2, \ldots, n]$$

for $i = 1, \ldots, n-1$. This gives rise to another natural statistic on D_n the *length* (similarly definable for any Coxeter group), namely

$$l(\pi) := min\{r \in \mathbf{N} : \pi = s_{i_1} \ldots s_{i_r} \text{ for some } i_1, \ldots, i_r \in [0, n-1]\}.$$

There is a well known direct combinatorial way to compute this statistic for $\pi \in D_n$ (see, e.g., [5, §8.2]), namely

$$l(\pi) = inv(\pi) - \sum_{i \in Neg(\pi)} \pi(i) - N_1(\pi). \tag{4}$$

It's not hard to prove that for all $\pi \in B_n$ (and so also for $\pi \in D_n$),

$$N_1(\pi) + N_2(\pi) = - \sum_{i \in Neg(\pi)} \pi(i), \tag{5}$$

so equivalently (4) becomes

$$l(\pi) = inv(\pi) + N_2(\pi). \tag{6}$$

For example, if $\pi := [-4, 1, 3, -5, -2, -6] \in D_6$ then $inv(\pi) = 10$, $des(\pi) = 2$, $maj(\pi) = 8$, $N_1(\pi) = 4$, $N_2(\pi) = 13$ and $l(\pi) = 23$.

We follow [5] for general Coxeter group notation and terminology. In particular, let (W, S) be a Coxeter system, for $J \subseteq S$ we let W_J be the subgroup of W generated by J, and

$$W^J := \{w \in W : l(ws) > l(w) \text{ for all } s \in J\}.$$

We call W_J the *parabolic subgroup* generated by J and W^J the *set of minimal left coset representatives* of W_J or the *quotient*. The quotient W^J is a poset according to the Bruhat order (see, e.g., [5] or [14]).
The following is well known, (see, e.g., [5] or [14]).

Proposition 1. *Let $J \subseteq S$. Then:*

i) *Every $w \in W$ has a unique factorization $w = w^J w_J$ such that $w^J \in W^J$ and $w_J \in W_J$.*
ii) *For this factorization $l(w) = l(w^J) + l(w_J)$.*

Now we let

$$T := \{\pi \in D_n : des(\pi) = 0\}. \tag{7}$$

It is well known, and easy to see, that

$$D_n = \biguplus_{u \in S_n} \{\pi u : \pi \in T\}, \tag{8}$$

where $\biguplus$ denotes disjoint union. We will often use this decomposition in this paper. Note that (8) is one case of the multiplicative decomposition of a Coxeter group into a parabolic subgroup and its minimal coset representatives (see Proposition 1), more precisely T is the quotient corresponding to the maximal parabolic subgroup generated by $J := S \setminus \{s_0\}$. In the next section we will analyze this issue in more detail.
For $n \in \mathbf{P}$ we let

$$A_n(t, q) := \sum_{\sigma \in S_n} t^{des(\sigma)} q^{maj(\sigma)},$$

and $A_0(t,q) := 1$. For example, $A_3(t,q) = 1 + 2tq^2 + 2tq + t^2q^3$. The following result is due to Carlitz, and we refer the reader to [7] for its proof.

Theorem 2. *Let $n \in \mathbf{P}$. Then*

$$\sum_{r \geq 0} [r+1]_q^n t^r = \frac{A_n(t,q)}{\prod_{i=0}^{n}(1-tq^i)} \tag{9}$$

in $\mathbf{Z}[q][[t]]$.

3. The "Negative" Statistics

In this section we define a new "descent set" for elements of D_n. This gives rise, in a very natural way, to the definitions of "major index" and "descent number" for D_n. We then show that these two statistics give a generalization of Carlitz's identity to D_n, and that the former is equidistributed with length.

3.1. **The D-Negative Descent Multiset.** For $\pi \in D_n$ let

$$Des(\pi) := \{i \in [n-1] \; : \; \pi(i) > \pi(i+1)\},$$

we define the *D-negative descent multiset*

$$DDes(\pi) := Des(\pi) \biguplus \{-\pi(i) - 1 \; : \; i \in Neg(\pi)\} \setminus \{0\}. \tag{10}$$

For example, if $\pi = [-4, 1, 3, -5, -2, -6] \in D_6$ then $Des(\pi) = \{3, 5\}$ and $DDes(\pi) = \{1, 3^2, 4, 5^2\}$.

Note that if $\pi \in S_n$ then $DDes(\pi)$ is a set and coincides with the usual descent set of π. Also, note that $DDes(\pi)$ can be defined rather naturally also in purely Coxeter group theoretic terms. In fact, for $i \in [n-1]$ let $\xi_i \in D_n$ be defined by

$$\xi_i := [-1, 2, \ldots, i, -i-1, i+2, \ldots, n].$$

Then $\xi_1, \ldots, \xi_{n-1}$ are reflections (in the Coxeter group sense, see e.g., [5] or [14]) of D_n and it is clear from (4) that

$$DDes(\pi) := \{i \in [n-1] : l(\pi s_i) < l(\pi)\} \biguplus \{i \in [n-1] : l(\pi^{-1}\xi_i) < l(\pi^{-1})\}.$$

These considerations explain why it is natural to think of $DDes(\pi)$ as a "descent set", so the following definitions are natural.
For $\pi \in D_n$ we let

$$ddes(\pi) := |DDes(\pi)|$$

and

$$dmaj(\pi) := \sum_{i \in DDes(\pi)} i.$$

For example if $\pi = [-4, 1, 3, -5, -2, -6] \in D_6$ then $ddes(\pi) = 6$, and $dmaj(\pi) = 21$.
Note that from (10) there follows that

$$dmaj(\pi) = maj(\pi) - \sum_{i \in Neg(\pi)} \pi(i) - N_1(\pi) = maj(\pi) + N_2(\pi). \tag{11}$$

This formula is also one the motivations behind our definition of $dmaj(\pi)$, because of the corresponding formulas (4) and (6), (see also [2]) .
Also note that

$$ddes(\pi) = des(\pi) + N_1(\pi) + \epsilon(\pi), \tag{12}$$

where

$$\epsilon(\pi) := \begin{cases} -1 & \text{if } 1 \notin \pi([n]) \\ 0 & \text{if } 1 \in \pi([n]). \end{cases}$$

3.2. **Equidistribution.** Our first result shows that $dmaj$ and l are equidistribuited in D_n.

Proposition 3. [4, Proposition 3.1] *Let $n \in \mathbf{P}$. Then*

$$\sum_{\pi \in D_n} q^{dmaj(\pi)} = \sum_{\pi \in D_n} q^{l(\pi)}.$$

3.3. **Generalization of Carlitz's Identity.** We start with some notation and terminology concerning partitions (see [17, §7.2]). By an (integer) *strict partition* we mean a sequence of positive integers $\lambda = (\lambda_1, \ldots, \lambda_k)$ such that $\lambda_1 > \lambda_2 > \ldots > \lambda_k$. We denote by $|\lambda| := \sum_i \lambda_i$. We denote by $\tilde{P}_S$ the set of all (integer) strict partitions. Given $n \in \mathbf{P}$ we let

$$\tilde{P}_S(n) := \{\lambda \in \tilde{P}_S : \lambda \subseteq (n, n-1, \ldots, 2, 1)\}.$$

As before, let $T = \{\pi \in D_n : des(\pi) = 0\}$ so

$$T = \{\pi \in D_n : \pi(1) < \pi(2) < \ldots < \pi(n)\}.$$

Therefore, given $\pi \in T, \pi \neq e$, there is a unique $k \in [n]$ such that

$$\pi(k) < 0 < \pi(k+1).$$

Given $\pi \in T$ we associate to π the strict partition

$$\Lambda(\pi) := (-\pi(1) - 1, -\pi(2) - 1, \ldots, -\pi(k) - 1). \tag{13}$$

The following is known, (see, e.g., [3]).

Proposition 4. *The map Λ defined by (13) is a bijection between T and $\tilde{P}_S(n-1)$. Furthermore $\pi \leq \sigma$ in T if and only if $\Lambda(\pi) \subseteq \Lambda(\sigma)$ and $l(\pi) = |\Lambda(\pi)|$ for all $\pi, \sigma \in T$.*

We begin with the following lemma.

Lemma 1. [4, Lemma 3.3] *Let $n \in \mathbf{P}$. Then*

$$\sum_{\sigma \in T} t^{N_1(\sigma)+\epsilon(\sigma)} q^{N_2(\sigma)} = \sum_{S \subseteq [n-1]} t^{|S|} q^{\sum_{i \in S} i} = \prod_{i=1}^{n-1} (1 + tq^i).$$

We are now ready to state the main result of this work, namely that the pair of statistics $(ddes, dmaj)$ solves Foata's problem for the group of even-signed permutations D_n.

Theorem 5. [4, Theorem 3.5] *Let $n \in \mathbf{P}$. Then*

$$\sum_{r \geq 0} [r+1]_q^n t^r = \frac{\sum_{\pi \in D_n} t^{ddes(\pi)} q^{dmaj(\pi)}}{(1-t)(1-tq^n) \prod_{i=1}^{n-1} (1 - t^2 q^{2i})} \tag{14}$$

in $\mathbf{Z}[q][[t]]$.

Note that, as in (9) for S_n and in ([2, Theorem 3.2]) for B_n, the powers of q in the denominator of formula (14) are the Coxeter degrees of D_n (see [14, p.59]).

4. The Flag Major Index for D_n

In this section we introduce another new "major index" statistic for D_n. This is an analogue of the flag major index introduced in [1]. We show that this statistic is equidistributed with length and we give a combinatorial algorithm to compute it.

4.1. **The D-Flag Major Index.** For $i = 0, \ldots, n-1$ we define

$$t_i := s_i s_{i-1} \cdots s_0, \tag{15}$$

explicitly for all $i \in [n-1]$

$$t_i = [-1, -i-1, 2, 3, \ldots, i, i+2, \ldots, n], \tag{16}$$

and for $i = 0$

$$t_0 = [-2, -1, 3, \ldots, n] = s_0. \tag{17}$$

These are Coxeter elements (see e.g., [14, §3.16]), in a distinguished flag of parabolic subgroups

$$1 < G_1 < G_2 < \ldots < G_n = D_n$$

where $G_i \simeq D_i$ $(i \geq 2)$ is the parabolic subgroup of D_n generated by $s_0, s_1, \ldots, s_{i-1}$. The family $\{t_i\}_i$ is a new set of generators for D_n, and we have the following proposition.

Proposition 6. [4, Proposition 4.1] *For every $\pi \in D_n$ there exists a unique representation*

$$\pi = t_0^{h_{n-1}} t_{n-1}^{k_{n-1}} t_0^{h_{n-2}} t_{n-2}^{k_{n-2}} \cdots t_0^{h_1} t_1^{k_1} \tag{18}$$

with $0 \leq h_r \leq 1$, $0 \leq k_r \leq 2r-1$ and

$$k_r \in \{2r-1, r-1\} \ \text{ if } \ h_r = 1 \tag{19}$$

for all $r = 1, \ldots, n-1$.

Note that the representation (18) is not unique if we drop the condition (19). For example consider $\pi = [3, -2, 1, -4] \in D_4$, then π has two different representations of type (18), namely, $\pi = t_3^3 t_0 t_2^3 t_0$ and $\pi = t_0 t_3^3 t_2 t_0 t_1$. The representation of Proposition 6 is the first one.

Let $\pi \in D_n$, then we define the *D-flag major index* of π by

$$fmaj_D := \sum_{i=1}^{n-1} k_i + \sum_{i=1}^{n-1} h_i. \tag{20}$$

4.2. **Equidistribution.** For $0 \leq m \leq 2n-1$, $n \geq 2$, we define $r_{n,m} \in D_n$ as follows: for $n = 2$,

$$r_{2,m} := \begin{cases} e & \text{if } m = 0 \\ s_1 & \text{if } m = 1 \\ s_1 s_0 & \text{if } m = 2 \\ s_0 & \text{if } m = 3, \end{cases}$$

and for $n > 2$,

$$r_{n,m} := \begin{cases} e & \text{if } m = 0 \\ s_{n-m} s_{n-m+1} \cdots s_{n-1} & \text{if } 0 < m < n \\ s_{m-n+1} s_{m-n} \cdots s_0 s_2 s_3 \cdots s_{n-1} & \text{if } n \leq m < 2n-1 \\ s_0 s_2 s_3 \cdots s_{n-1} & \text{if } m = 2n-1. \end{cases}$$

The set $\{r_{n,m} \ : \ 0 \leq m < 2n\}$ forms a complete set of representatives of minimal length for the left cosets of D_{n-1} in D_n. Moreover this is still valid for every $i \in [2, n]$, namely, $r_{i,m} \in D_i^{J_i}$ for all $m \in [0, 2i-1]$, where $J_i := S \setminus \{s_{n-1}, \ldots, s_{i-1}\}$. Hence we have the following decomposition

$$D_n = D_n^{J_n} D_{n-1}^{J_{n-1}} \cdots D_2.$$

Note that the length of $r_{i,m}$ is $\bar{m}$, where

$$\bar{m} := \begin{cases} m & \text{if } 0 \leq m \leq 2i-2 \\ i-1 & \text{if } m = 2i-1. \end{cases}$$

From i) of Proposition 1 we know that each element $\pi \in D_n$ has a unique representation as a product

$$\pi = \prod_{k=1}^{n-1} r_{n+1-k, m_{n+1-k}} \tag{21}$$

where $0 \leq m_j < 2j$ for all j. From ii) of Proposition 1 there follows that

$$l(\pi) = \sum_{j=2}^{n} \bar{m}_j. \tag{22}$$

Thanks to the unique representation (21) we define a map $\phi : D_n \to D_n$ in the following way:

$$\phi(\prod_{k=1}^{n-1} r_{n+1-k, m_{n+1-k}}) := \prod_{k=1}^{n-1} \phi(r_{n+1-k, m_{n+1-k}}),$$

where for $i \neq 2$,

$$\phi(r_{i,m}) := \begin{cases} t_{i-1}^{m} & \text{if } m < 2i-2 \\ t_0 t_{i-1}^{\bar{m}-1} & \text{if } 2i-2 \leq m \leq 2i-1, \end{cases}$$

and for $i = 2$,

$$\phi(r_{2,m}) := \begin{cases} e & \text{if } m = 0 \\ t_1 & \text{if } m = 1 \\ t_0 t_1 & \text{if } m = 2 \\ t_0 & \text{if } m = 3. \end{cases}$$

The definition of ϕ, together with Proposition 6 and (21), imply the following result.

Proposition 7. [4, Proposition 4.2] *The map $\phi : D_n \to D_n$ is a bijection.*

This implies the main result of this section, namely that the D-flag major index is equidistributed with the length in D_n.

Theorem 8. [4, Theorem 4.3] *Let $n \in \mathbf{P}$. Then*

$$\sum_{\pi \in D_n} q^{fmaj_D(\pi)} = \sum_{\pi \in D_n} q^{l(\pi)}.$$

Note that the B-flag major index ($fmaj$) defined on B_n (see [1]) does not work on D_n. Namely if we consider $\pi \in D_n$ as an element of B_n, then $fmaj(\pi)$ is not equidistributed with length on D_n. For example let $\pi = [-2, -1]$ then $fmaj_D(\pi) = 1$ while $fmaj(\pi) = 4$, and in D_2 there is no element of length 4.
Note also that $fmaj_D$ restricted to S_n is not the major index and it's not equidistributed with length. It seems to be a new statistic on S_n. It's easy to see that for each $\pi \in S_n$, $fmaj_D(\pi)$ is always even and that $fmaj_D(\pi) \geq maj(\pi)$. If we let $E_n(q) := \sum_{\pi \in S_n} q^{fmaj_D(\pi)}$, for $n \leq 4$ we have $E_1(q) = 1$, $E_2(q) = 1 + q^2$, $E_3(q) = 1 + 3q^2 + q^4 + q^6$ and $E_4(q) = 1 + 5q^2 + 6q^4 + 7q^6 + 3q^8 + q^{10} + q^{12}$.

4.3. **A combinatorial algorithm.** Let $\sigma = (a_1, \ldots, a_n) \in \mathbf{Z}^n$, we use this *split-notation*

$$\sigma = [a_1][a_2, \ldots, a_{i+1}][a_{i+2}, \ldots, a_n].$$

Sometimes it will be useful to denote the first part with A and the second with C_i where i represents the number of its elements.
We define the following operations on $\sigma \in \mathbf{Z}^n$:

$$\overset{\rightarrow_i^0}{\sigma} := [-a_2][-a_1, a_3, \ldots, a_{i+1}][a_{i+2}, \ldots, a_n],$$

and

$$\overset{\rightarrow_i^1}{\sigma} := [-a_1][-a_{i+1}, a_2, \ldots, a_i][a_{i+2}, \ldots, a_n].$$

In these cases we will write $\overset{\rightarrow_i^0}{\sigma} = (A^0, C_i^0, [a_{i+2}, \ldots, a_n])$ and $\overset{\rightarrow_i^1}{\sigma} = (\overset{\rightarrow 1}{A}, \overset{\rightarrow 1}{C_i}, [a_{i+2}, \ldots, a_n])$.
Moreover for all $n \in \mathbf{P}$ we define

$$\rightarrow_i^n := \rightarrow_i^1 \circ \cdots \circ \rightarrow_i^1 \quad n\text{-times}. \tag{23}$$

Note that for every $\sigma \in \mathbf{Z}^n$, $\overset{\rightarrow_i^{2i}}{\sigma} = \sigma$.
For example let $\pi \in D_4$, $\pi = [-2][1, 3, -4, 5] = (A, C_4)$, then

$$\overset{\rightarrow_4^0}{\pi} = [-1][2, 3, -4, 5] = (A^0, C_4^0),$$

$$\overset{\rightarrow_4^5}{\pi} = [2][5, -1, -3, 4] = (\overset{\rightarrow 5}{A}, \overset{\rightarrow 5}{C_4}),$$

and

$$\overset{\rightarrow_3^2}{\pi} = [-2][-3, 4, 1][5] = (\overset{\rightarrow 2}{A}, \overset{\rightarrow 2}{C_3}, [5]).$$

These are the two technical properties that we will use in the algorithm. Fix $i \in [n-1]$, let t_i be as in (16),

$$t_i = [-1][-i-1, 2, 3, \ldots, i][i+2, \ldots n].$$

It's easy to see that for all $i \in [n-1]$ we have

$$t_i^2 = t_i t_i = \overset{\rightarrow_i^1}{t_i}, \tag{24}$$

and by (23) that for $k \in \mathbf{P}$

$$t_i^k = \overset{\rightarrow_i^{k-1}}{t_i}. \tag{25}$$

Now consider $t_{i-1} = [-1][-i, 2, \ldots, i-1][i+1, \ldots, n]$. As before it is not hard to see that

$$t_i t_{i-1} = \overset{\rightarrow_{i-1}^1}{t_i}. \tag{26}$$

Now we are able to state the algorithm to compute the unique representation of π as in Proposition 6, namely

$$\pi = f_{n-1} \cdots f_1$$

where for all $r \in [n-1]$, $f_r = t_0^{h_r} t_r^{k_r}$ with $h_r \in [0, 1]$ and $k_r \in [0, 2r-1]$.
We construct a sequence $e_0, \ldots, e_{n-1}$ of elements of D_n such that

i) $e_0 = e$, $e_{n-1} = \pi$;

ii) $e_i = f_{n-1} \cdots f_{n-i}$, for all $i \in [1, n-1]$;

iii) $\pi(j) = e_i(j)$, for all $j > n-i$.

From $iii)$ there immediately follows that $e_{n-1} = \pi$.

We need to do $n-1$ steps. From now to avoid confusion we put on A an index corresponding to the number of steps. We begin with $e_0 = [1][2, \ldots, n]$. We describe the $(n-i+1)$-th step. Assume that e_{n-i} has been constructed. Then by $iii)$,

$$e_{n-i} = (A_{n-i}, C_{i-1}, [\pi(i+1), \ldots \pi(n)]).$$

For simplicity, we define $p(i)$ and $p(-i)$ to be the positions of $\pi(i)$ and $-\pi(i)$ in C_{i-1} or C^0_{i-1}respectively. There are four cases to consider.

1) $\pi(i) \in C_{i-1}$

Then we let $k_{i-1} = i-1-p(i)$ and $h_{i-1} = 0$. Hence $f_{i-1} = t_{i-1}^{i-1-p(i)}$.

2) $-\pi(i) \in C_{i-1}$

Then we let $k_{i-1} = 2i-2-p(-i)$ and $h_{i-1} = 0$. Hence $f_{i-1} = t_{i-1}^{2i-2-p(-i)}$.

3) $\pi(i) \in A_{n-i}$

Then $-\pi(i) \in C^0_{i-1}$ and in particular $p(-i) = 1$. We let $k_{i-1} = 2i-3$ and $h_{i-1} = 1$. Hence $f_{i-1} = t_0 t_{i-1}^{2i-3}$.

4) $-\pi(i) \in A_{n-i}$

Then $\pi(i) \in C^0_{i-1}$ and $p(i) = 1$. We let $k_{i-1} = i-2$ and $h_{i-1} = 1$. Hence $f_{i-1} = t_0 t_{i-1}^{i-2}$.

We have determined the factor f_{i-1}. From (23) and (25) there follows that $e_{n-i+1}(i) = \pi(i)$ and by (26) $iii)$ again holds.
Therefore

$$e_{n-i+1} = (A_{n-i+1}, C_{i-2}, [\pi(i), \ldots, \pi(n)]),$$

where in cases 1) and 2),

$$A_{n-i+1} := \overset{\rightarrow k_{i-1}}{A_{n-i}} \ , \ \ C_{i-2} := \overset{\rightarrow k_{i-1}}{C_{i-1}} \setminus [\pi(i)],$$

while in cases 3) and 4),

$$A_{n-i+1} := \overset{\rightarrow k_{i-1}}{A^0_{n-i}} \ , \ \ C_{i-2} := \overset{\rightarrow k_{i-1}}{C^0_{i-1}} \setminus [\pi(i)].$$

Observe that in the first step $p(n) = \pi(n) - 1$ and $p(-n) = -\pi(n) + 1$. These can be used for the computation of e_1.
We finish this section by illustrating the procedure with an example.
Let $\pi = [5, 3, -4, 1, -2] \in D_5$. We start from

$$e = e_0 = [1][2, 3, 4, 5] = (A_0, C_4).$$

$1^{st}-step)$ $\quad -\pi(5) = -2 \in C_4$

We are in case 2) and $p(-5) = 1$, so $k_4 = 7$, $h_4 = 0$ and $f_4 = t_4^7$.
It follows that $A_1 = \overset{\rightarrow 7}{A_0} = [-1]$ and $C_3 = \overset{\rightarrow 7}{C_4} \setminus [-2] = [3, 4, 5]$. Hence,

$$e_1 = [-1][3, 4, 5][-2].$$

$2^{nd}-step)$ $\quad -\pi(4) = -1 \in A_1$

We are in case 4) so $k_3 = 2$, $h_3 = 1$ and $f_3 = t_0 t_3^2$.
It follows that $A_2 = \overset{\rightarrow 2}{A^0_1} = [-3]$ and $C_2 = \overset{\rightarrow 2}{C^0_3} \setminus [1] = [-4, -5]$. Hence,

$$e_2 = [-3][-4, -5][1, -2].$$

$3^{rd}-step)$ $\quad \pi(3) = -4 \in C_2$

We are in case 1) and $p(3) = 1$, so $k_2 = 1$, $h_2 = 0$ and $f_2 = t_2$.

It follows that $A_3 = \overset{\rightarrow 1}{A_2} = [3]$ and $C_1 = \overset{\rightarrow 1}{C_2} \setminus [-4] = [5]$. Hence,

$$e_3 = [5][-4, 1, -2].$$

$4^{th} - step$) $\pi(2) = 3 \in A_3$

We are in case 3) so $k_1 = 1$, $h_1 = 1$ and $f_1 = t_0 t_1$.

It follows that $A_4 = \overset{\rightarrow 1}{A_3^0} = [5]$ and $C_0 = \emptyset$. Hence,

$$e_4 = [5][3, -4, 1, -2] = \pi,$$

and we are done. Finally $\pi = t_4^7 t_0 t_3^2 t_2 t_0 t_1$ and $fmaj_D(\pi) = 12$.

Acknowledgements. I'm indebted to Francesco Brenti who introduced me to the subject and presented me with these questions. I would like to thank Anne Schilling, Paolo Papi and Domenique Foata for some useful conversations, and finally I express my gratitude to the Massachusetts Institute of Technology for hospitality during the preparation of this work.

References

[1] R. M. Adin and Y. Roichman, *The Flag Major Index and Group Actions on Polynomial Rings*, Europ. J. Combinarorics, **22** (2001), 431-446.
[2] R. M. Adin, F. Brenti and Y. Roichman, *Descent Numbers and Major Indices for the Hyperoctahedral Group*, Adv. Applied Math., **27** (2001), 210-224.
[3] R. Biagioli, *Closed product formulas for Extensions of generalized Verma modules*, Trans. Amer. Math. Soc., to appear.
[4] R. Biagioli, *Major and Descent Statistics for the even-signed Permutation Group*, preprint, December 2001.
[5] A. Björner and F. Brenti, *Combinatorics of Coxeter Groups*, to appear on Grad. Texts in Math., Springer-Verlag, Berlin, 2001
[6] F. Brenti, *q-Eulerian polynomials arising from Coxeter groups*, Europ. J. Combinarorics, **15** (1994), 417-441.
[7] L. Carlitz, *A combinatorial property of q-Eulerian numbers*, Amer. Math. Montly, **82** (1975), 51-54.
[8] R. J. Clarke and D. Foata *Eulerian calculus. I. Univariable statistics*, Europ. J. Combinarorics, **15** (1994), 345-362.
[9] R. J. Clarke and D. Foata *Eulerian calculus. II. An extension of Han's fundamental trasformation*, Europ. J. Combinarorics, **16** (1995), 221-252.
[10] R. J. Clarke and D. Foata *Eulerian calculus. III. The ubiquitous Cauchy formula*, Europ. J. Combinarorics, **16** (1995), 329-355.
[11] D. Foata *On the Netto inversion number of a sequence*, Proc. Amer. Math. Soc., **19** (1968), 236-240.
[12] V. Reiner, *Signed permutation statistics*, Europ. J. Combinatorics, **14** (1993), 553-567.
[13] V. Reiner, *The distribution of descent and length in a Coxeter group*, Electron. J. Combinarorics, **2** (1995), R25.
[14] J. E. Humphreys, *Reflection Groups and Coxeter Groups*, Cambridge Stud. Adv. Math., no.29, Cambridge Univ. Press, Cambridge, 1990.
[15] P. A. MacMahon, *Combinatory Analysis*, vol.1, Cambridge Univ. Press, London, 1915.
[16] R. P. Stanley, *Enumerative Combinatorics* , vol.1, Wadsworth and Brooks/Cole, Monterey, CA, 1986.
[17] R. P. Stanley, *Enumerative Combinatorics* , vol.2, Cambridge Stud. Adv. Math., no.62, Cambridge Univ. Press, Cambridge, 1999.
[18] E. Steingrimsson, *Permutation statistics of indexed permutations*, Europ. J. Combinarorics, **15** (1994), 187-205.

Dipartimento di Matematica, Università di Roma "La Sapienza", P.le A. Moro n.5, 00185 Roma, Italy.

E-mail address: biagioli@mat.uniroma1.it

$\mathbb{Z}$-TILINGS OF POLYOMINOES AND STANDARD BASIS

OLIVIER BODINI

Abstract. In this paper, we prove that for every set F of *polyominoes* (for us, a polyomino is a finite union of unit squares of a square lattice), we can find a $\mathbb{Z}$-tiling (*signed tile*) of polyominoes by copies of elements of F in polynomial time. We use for this the theory of generalized standard basis. So, we can algorithmically extend results of Conway and Lagarias on $\mathbb{Z}$-tiling problems.

Résumé. Nous montrons que, pour toute famille F de *polyominos généraux* (union finie de cases d'une grille), le problème du $\mathbb{Z}$-pavage (*pavage signé*) des polyominos par des copies d'éléments de F peut être résolu en temps polynomial par l'usage de la théorie des bases de Grobner. De plus, nous pouvons ainsi retrouver et étendre de manière algorithmique des résultats obtenus par Conway et Lagarias sur les $\mathbb{Z}$-pavages.

1. Introduction

A *cell* $c(i,j)$ in the square lattice denotes the set :

$$c(i,j) := \{(x,y)\,; i \leq x < i+1, j \leq y < j+1\}.$$

So, cells are labelled by their lower left corner. For us, a *polyomino* is a finite -not necessary connected- union of cells. In this paper, we are interested in the study of a variant of the problem of tiling, called the $\mathbb{Z}$-*tiling problem*. Precisely, let P a polyomino and F a set of polyominoes (*the tiles*), a $\mathbb{Z}$-*tiling of* P *by* F consists of a finite number of translated tiles placed in the lattice (possibly with overlaps), with each tile assigned a sign of +1 or -1, such that for each cell $c(i,j)$ in $\mathbb{Z}^2$ the sum of the signs of the tiles covering $c(i,j)$ is +1 if $c(i,j) \in P$ and 0 if $c(i,j) \notin P$ (fig.1). Obviously, a polyomino which is tilable by a set of tiles is also $\mathbb{Z}$-tilable by this set. So, the study of $\mathbb{Z}$-tiling creates important necessary conditions of tilability. J.H. Conway and J.C. Lagarias [3] have previously studied this notion. They particularly obtained the following necessary and sufficient condition for a simply connected polyomino P :
P has a $\mathbb{Z}$-tiling of P by F if and only if the combinatorial boundary $[\partial P]$ is included in the tile boundary group $\mathbf{B}(F)$. For these definitions, we can refer to the paper of J.H. Conway and J.C. Lagarias [3].
Nevertheless, this group theoretic theorem presents some drawbacks : Firstly, it only applies to simply connected polyominoes. Secondly, the new criterion seems in general no easier to verify than to solve the original problem. Thirdly, it seems to be impossible to extend theoretic group arguments in higher dimension. In this paper, we propose another way of solving the problem. We associate for each polyomino P a polynomial in $\mathbb{Z}[X_1, X_2, Y_1, Y_2]$, called P-*polynomial.* We denote it by Q_P. We prove that, given a set F of polyominoes, a polyomino P is $\mathbb{Z}$-tilable by F if and only if $Q_P \in \langle Q_{P'} \textit{ with } P' \in E, X_1Y_1 - 1, X_2Y_2 - 1\rangle_{\mathbb{Z}}$ (i.e. the ideal of $\mathbb{Z}[X_1, ..., X_n]$ generated by the polynomials $Q_{P'}$ *where* $P' \in E$ and by $X_1Y_1 - 1, X_2Y_2 - 1$). This new formulation allows us to use commutative algebraic tools like

1991 *Mathematics Subject Classification.* Primary: 05A30; Secondary: 05A99.
Key words and phrases. Polyominoes, Tiling.

the standard basis algorithm to solve the problem. The reader can find a good introduction to standard basis (for ideals in $\mathbb{K}[X_1, ..., X_n]$ where $\mathbb{K}$ is a field) in [2] [4].

This leads us to reconsider coloring arguments. These tools are frequently found in the literature [5],[6],[7]). This notion gives in general important necessary conditions of tilability. We define in this paper the *generalized coloring* associated to a set of tiles F. This coloring groups all the generalized coloring arguments defined by Conway-Lagarias [3]. Moreover, the generalized coloring of a polyomino P is null if and only if P is $\mathbb{Z}$-tilable by the set F. Finally, we prove that it is possible to determine the generalized coloring of a classical polyomino when we only know the colors of the squares which are adjacent to the boundary of P.

So, if a polyomino P is in a sense "big" then we have a better algorithm to determine the tilability of P. Now, we are going to introduce the abstract notions that constitute the general framework of this paper. Given a subdivision S of $\mathbb{R}^d$ in cells and $\mathbb{A}$ a unitary ring, a $\mathbb{A}$-*weighted polycell* or simply a $\mathbb{A}$-polycell is a map P of S in $\mathbb{A}$ with a finite support. For each cell c, we call *weight of* P in c the number $P(c)$. The space $\mathbb{P}_{\mathbb{A}}$ of $\mathbb{A}$-weighted polycells has a natural structure of free $\mathbb{A}$-module. Clearly, the cells of weight 1 constitute a base of $\mathbb{P}_{\mathbb{A}}$. We can canonically embed the set of polycells in the $\mathbb{A}$-module of the $\mathbb{A}$-weighted polycells (in assigning 1 to cells covered by the polycell and 0 to the other cells). We say that a $\mathbb{A}$-weighted polycell P is $\mathbb{A}$-*tilable* by a set of $\mathbb{A}$-weighted tiles if and only if P is a $\mathbb{A}$-linear combination of translated elements of this set.

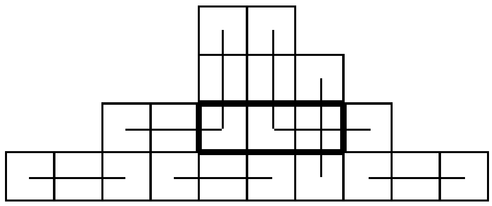


Fig.1. A (classical) polyomino P which is $\mathbb{Z}$-tilable by bars of length 3. In bold, a negative copy of a bar. The segments indicate positive copies of bars. These positive and negative bars constitute a $\mathbb{Z}$-tiling of P.

2. P-polynomials and $\mathbb{Z}$-tilings

In this section, in order to simplify, we only deal with $\mathbb{Z}$-polyominoes and $\mathbb{Z}$-polyhexes (sum of weighted hexagons of the regular hexagonal lattice). But all the theorems can be made in a more general framework. For instance, we have similar results for $\mathbb{Z}$-polycubes (in this case, the cells are the unit cubes of the regular cubical lattice of $\mathbb{R}^d$) or for $\mathbb{Z}$-polyamands (the cells are the triangles of the regular triangular lattice). The reader can find in [1] a more general presentation.

Firstly, for $a \in \mathbb{Z}^2$, we put by convention that

$$X^a = X_1^{\frac{a_1+|a_1|}{2}} X_2^{\frac{a_2+|a_2|}{2}} Y_1^{\frac{|a_1|-a_1}{2}} Y_2^{\frac{|a_2|-a_2}{2}}.$$

We encode the plan with 4 parameters to avoid to work with Laurent polynomials. Let us recall that we denote by $\langle P_1, ..., P_k \rangle_{\mathbb{Z}}$ the ideal of $\mathbb{Z}[X_1, ..., X_n]$ generated by the polynomials $P_1, ..., P_k$. For each $\mathbb{Z}$-polyomino P, we can define its *P-polynomial*

$$Q_P = \sum_{(a_1,a_2)\in\mathbb{Z}^2} P\left(c(a_1, a_2)\right) X^{(a_1,a_2)}$$

Lemma 2.1. *The space* $\mathbb{P}_{\mathbb{Z}}$ *is isomorphic to* $\mathbb{Z}[X_1, X_2, Y_1, Y_2] / \langle (X_1Y_1 - 1), (X_2Y_2 - 1) \rangle_{\mathbb{Z}}$.

Proof. There exists a unique linear map f from $\mathbb{Z}[X_1, X_2, Y_1, Y_2]$ to $\mathbb{P}_{\mathbb{Z}}$ such that : $f\left(X_1^{a_1} Y_1^{b_1} X_2^{a_2} Y_2^{b_2}\right)$ is the cell $(a_1 - b_1, a_2 - b_2)$ with weight 1. Now, we must prove that $\ker(f) = \langle (X_1Y_1 - 1), (X_2Y_2 - 1)\rangle_{\mathbb{Z}}$. We proceed by successive divisions by $(X_1Y_1 - 1)$ and $(X_2Y_2 - 1)$ in the successive rings

$$\mathbb{Z}[X_2, Y_1, Y_2][X_1] \text{ and } \mathbb{Z}[X_1, Y_1, Y_2][X_2].$$

So, we can write all polynomial Q as follows $Q = R + \sum_{i=1}^{2} Q_i (X_iY_i - 1)$ with R containing only monomials of the form X^a where $a \in \mathbb{Z}^2$ (i.e. without simultaneously X_i and Y_i). We have the following equivalence : $f(Q)$ is the empty polyomino, denoted by 0 (i.e. the polyomino P with 0 weight on all the squares), if and only if $f(R) = 0$ (because of $f(Q_i(X_iY_i - 1)) = 0$). Moreover, it is clear that $f(R) = 0 \Leftrightarrow R = 0$ and that $R = 0 \Leftrightarrow Q \in \langle (X_1Y_1 - 1), (X_2Y_2 - 1)\rangle_{\mathbb{Z}}$. So, $\ker(f) = \langle (X_1Y_1 - 1), (X_2Y_2 - 1)\rangle_{\mathbb{Z}}$. □

Theorem 2.2. *Let E be a set of $\mathbb{Z}$-polyominoes. A $\mathbb{Z}$-polyomino P is $\mathbb{Z}$-tilable by E if and only if*

$$Q_P \in \left\langle Q_{P'} \textit{ with } P' \in E, X_1Y_1 - 1, X_2Y_2 - 1\right\rangle_{\mathbb{Z}}.$$

Proof. By definition, a $\mathbb{Z}$-polyomino P is $\mathbb{Z}$-tilable by E if and only if there exists an integer t and for all i, $1 \le i \le t$, $\lambda_i \in \mathbb{Z}$, $P^i \in E$ and $a^i = \left(a_1^i, a_2^i\right) \in \mathbb{Z}^2$ such that $P = \sum_{i=1}^{t} \lambda_i P^i_{\left(a_1^i, a_2^i\right)}$ where $P^i_{\left(a_1^i, a_2^i\right)}$ denotes the translation of P^i by the vector $\left(a_1^i, a_2^i\right)$. So, in

$$\mathbb{Z}[X_1, X_2, Y_1, Y_2] / \langle (X_1Y_1 - 1), (X_2Y_2 - 1)\rangle_{\mathbb{Z}}$$

we have $Q_{P^i_{a^i}} = X^{a^i} Q_{P^i}$ and consequently, $Q_P = \sum_{i=1}^{t} \lambda_i X^{a_i} Q_{P^i}$. Finally, P is $\mathbb{Z}$-tilable by E if and only if

$$Q_P \in \left\langle Q_{P'} \textit{ with } P' \in E, X_1Y_1 - 1, X_2Y_2 - 1\right\rangle_{\mathbb{Z}}.$$

□

In fact, when we have a periodic tiling, we can always define the notion of polycell (finite union of cells of this tiling) and so, it is possible to translate, as we have done it, the problem of the $\mathbb{Z}$-tiling of a polycell by a set of polycells to algebraic problem of membership in an ideal of a polynomial ring on $\mathbb{Z}$. For instance, for the hexagonal lattice built in gluing copies of the hexagonal convex hull of the points $(0,0), (0,1), \left(\frac{-1}{2}, \frac{\sqrt{3}}{2}\right), \left(\frac{3}{2}, \frac{\sqrt{3}}{2}\right), \left(0, \sqrt{3}\right), \left(1, \sqrt{3}\right)$, a polycell is generally called a *polyhexe*. We denote by $[a_1, a_2]$ the hexagonal cell which the lower left corner is the point $\left(\frac{3}{2}(a_1 + a_2), \frac{\sqrt{3}}{2}(-a_1 + a_2)\right)$ where $(a_1, a_2) \in \mathbb{Z}^2$ (fig.2). For each $\mathbb{Z}$-polyhexe P, we can define the *P-polynomial*

$$Q_P = \sum_{(a_1, a_2) \in \mathbb{Z}^2} P([a_1, a_2]) X^{(a_1, a_2)}$$

Then, we have the following theorem :

Theorem 2.3. *Let E be a set of $\mathbb{Z}$-polyhexes. A $\mathbb{Z}$-polyhexe P is $\mathbb{Z}$-tilable by E if and only if*

$$Q_P \in \left\langle Q_{P'} \textit{ avec } P' \in E, X_1Y_1 - 1, X_2Y_2 - 1\right\rangle_{\mathbb{Z}}.$$

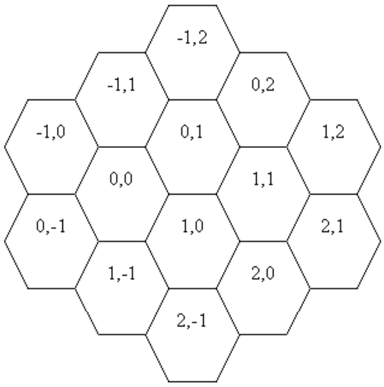


Fig.2. The hexagonal lattice with its coordinates.

3. STANDARD BASIS ON $\mathbb{Z}[X_1, ..., X_n]$.

In this section, we indicate briefly how to solve the problem of membership in an ideal of $\mathbb{Z}[X_1, ..., X_n]$. In fact, we use a non trivial extended version to $\mathbb{Z}[X_1, ..., X_n]$ of the Buchberger algorithm [1]. The original one only works for an ideal of $\mathbb{K}[X_1, ..., X_n]$ where $\mathbb{K}$ is a field and can be found in [4], [2]. First of all, we have to define a total order on the monomials of $\mathbb{Z}[X_1, ..., X_n]$. Let $\leq^*$ be the lexicographic order on the n-tuples and let $\alpha = (\alpha_1, ..., \alpha_n)$ be in $\mathbb{N}^n$, we denote by X^α the monomial $X_1^{\alpha_1}...X_n^{\alpha_n}$. Then, we put by definition that $X^\alpha \leq^* X^\beta$ if and only if $\alpha \leq^* \beta$. It is easy to verify that $\leq^*$ is a total order on the monomials of $\mathbb{Z}[X_1, ..., X_n]$ and that we have the following property :
For all $\gamma \in \mathbb{N}^n$, if $X^\alpha \leq^* X^\beta$, then $X^{\alpha+\gamma} \leq^* X^{\beta+\gamma}$. This is the *lexicographic order* induced by $X_1 > ... > X_n$. Now, we recall useful terminologies for multivariable polynomials. Let $P = \sum_{\alpha\in\mathbb{N}^n} a_\alpha X^\alpha$ be a non empty polynomial of $\mathbb{Z}[X_1, ..., X_n]$:
The *support* of P is $S(P) = \{\alpha \in \mathbb{N}^n \textit{ such that } a_\alpha \neq 0\}$. In particular $S(P)$ is always finite.
The *multidegree* of P is $m(P) = \max^*(\alpha \in S(P))$.
The *leading coefficient* of P is $LC(P) = a_{m(P)}$.
The *leading monomial* of P is $LM(P) = X^{m(P)}$.
The *leading term* of P is $LT(P) = LC(P)LM(P)$.

Theorem 3.1. *Let $F = (P_1, ..., P_s)$ be a s-tuple of polynomials of $\mathbb{Z}[X_1, ..., X_n]$. Then every polynomial P of $\mathbb{Z}[X_1, ..., X_n]$ can be written in the following non unique form $P = R + \sum_{k=1}^{s} Q_k P_k$ where :*
i) $Q_1, ..., Q_s, R \in \mathbb{Z}[X_1, ..., X_n]$.
ii) $R = \sum_{\alpha\in\mathbb{N}^n} c_\alpha X^\alpha$ *and* $\forall \alpha \in S(R), c_\alpha X^\alpha$ *is not divisible by any of* $LT(P_1), ..., LT(P_s)$.

Proof. This proof is an easy consequence of the following generalized division algorithm. □

Algorithm 3.2. *Generalized Division Algorithm.*
We denote by $trunc(s)$ *the integer part of* s.
Input : $(P_1, ..., P_s), P$
Output : $(a_1, ..., a_s), R$
$a_1 := 0, ..., a_s := 0,\ R := 0$
$Q := P$

While $Q \neq 0$ *Do*
  $i := 1$
  $division := false$
  While ($i \leq s$ *and division=false*) *Do*
    If $LM(P_i)$ *divides* $LM(Q)$ *and* $|LC(P_i)| \leq |LC(Q)|$ *Then*
      $a_i := a_i + trunc(LC(Q)/LC(P_i))LM(Q)/LM(P_i)$
      $Q := Q - (trunc(LC(Q)/LC(P_i))LM(Q)/LM(P_i))P_i$
      $division := true$
    Else
    $i := i + 1$
    EndIf
  EndWhile
  If $division = false$ *Then*
    $R := R + LT(Q)$
    $Q := Q - LT(Q)$
  EndIf
EndWhile
Return $(a_1, ..., a_s), R$

R is the *remainder* of P by $(P_1, ..., P_s)$. We denote it by $\bar{P}^{(P_1,...,P_s)}$.

Example 3.3. *If we have* $P = X_1X_2^2 + X_1X_2 + X_2^2$ *and* $(P_1 = X_2^2 - 1, P_2 = X_1X_2 - 1)$, *then we obtain* $P = P_1 \times (X_1 + 1) + P_2 + X_1 + 2$. *The remainder is* $X_1 + 2$.

Example 3.4. *The remainder of the division of* $P = X_1X_2^2 - X_2^2$ *by* $(P_1 = X_2^2 - 1, P_2 = X_1X_2 - 1)$ *is null. Nevertheless, the division of* $P = X_1X_2^2 - X_2^2$ *by* $(P_1 = X_1X_2 - 1, P_2 = X_2^2 - 1)$ *gives* $\bar{P}^{(P_1,P_2)} = -X_2^2 + 1$. *So, we point out that the division depends on the ordering in the s-tuple of the polynomials. Actually, the division does not allow us to determine if a polynomial belongs or not to an ideal* I *of* $\mathbb{Z}[X_1, ..., X_n]$.

We recall that in $\mathbb{R}[X]$ a polynomial $P \in I$ if and only if Q divides P where Q is the minimal polynomial of I. We have the following analogous version in $\mathbb{Z}[X_1, ..., X_n]$:

Theorem 3.5. *For every ideal* I *of* $\mathbb{Z}[X_1, ..., X_n]$ *other than* $\{0\}$, *there exists a s-tuple of polynomials* $(P_1, ..., P_s)$ *such that* P *belongs to* I *if and only if the remainder of* P *by* $(P_1, ..., P_s)$ *is null.*

Such a s-tuple is called a *standard basis* of I. We do not proof here this important theorem. The reader who wants to take this theorem further can find a constructive proof of this in the following report [1]. We continue this section by an application to a classical problem solved by Conway and Lagarias [3] by using group theoretic arguments. Let T_N denote the triangular array of cells in the hexagonal lattice having $N(N+1)/2$ cells (fig.3).

Theorem 3.6. *(Conway-Lagarias, Theo 1.3 and 1.4)*
a) The triangular region T_N *in the hexagonal lattice has a* $\mathbb{Z}$*-tiling by congruent copies of* T_2 *polyhexes if and only if* $N = 0$ *or* $2 \mod 3$.
b) The triangular region T_N *in the hexagonal lattice has a* $\mathbb{Z}$*-tiling by congruent copies of three-in-line polyhexes if and only if* $N = 0$ *or* $8 \mod 9$.

Proof.
a) Firstly, we put $Y_1 > Y_2 > X_1 > X_2$ and we compute a standard basis of

$$\langle X_1 + X_2 + 1, X_1X_2 + X_1 + X_2, Y_1X_1 - 1, Y_2X_2 - 1 \rangle_{\mathbb{Z}}.$$

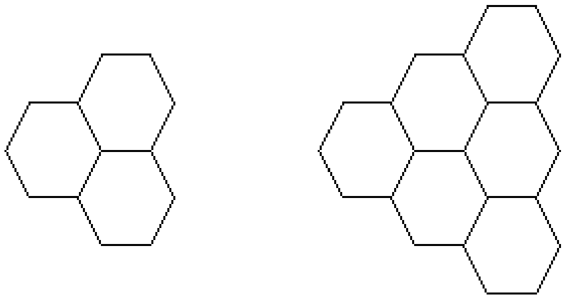


Fig.3. T_2 and T_3

We obtain $(X_1+X_2+1, X_2^2+X_2+1, Y_2+X_2+1, Y_1-X_2)$. Now, as $Q_{T_N} = \sum_{i=0}^{N} X_2^i \left(\sum_{j=0}^{N-i-1} X_1^j \right)$, we can easily compute that the remainder of Q_{T_N} by $(X_1+X_2+1, X_2^2+X_2+1, Y_2+X_2+1, Y_1-X_2)$ is equal to $\begin{cases} 0 & \text{if } N = 0 \text{ or } 2 \mod 3. \\ 1 & \text{if } N = 1 \mod 3. \end{cases}$

b) we compute a standard basis of

$$\langle X_1^2+X_1+1, X_2^2+X_2+1, X_1^2Y_2^2+X_1Y_2+1, Y_1X_1-1, Y_2X_2-1\rangle_{\mathbb{Z}}.$$

with $Y_1 > Y_2 > X_1 > X_2$. We obtain $B = (X_1^2+X_1+1, X_2^2+X_2+1, X_1+Y_1+1, X_2+Y_2+1, 3X_2+3X_1+3, X_2X_1-X_1-X_2-2)$.

The remainder of Q_{T_N} by B is equal to $\begin{cases} 0 & \text{if } N = 0 \text{ or } 8 \mod 9. \\ 1 & \text{if } N = 1 \mod 9. \\ X_1+X_2+1 & \text{if } N = 2 \text{ or } 3 \mod 9. \\ -2X_1-2X_2-1 & \text{if } N = 4 \mod 9. \\ 2X_1+2X_2+2 & \text{if } N = 5 \text{ or } 6 \mod 9. \\ -X_1-X_2 & \text{if } N = 7 \mod 9. \end{cases}$

□

4. General Coloring

Definition 4.1. *Let E be a set of $\mathbb{Z}$-polyominoes, the ideal*

$$I(E) = \langle Q_P; P \in E \text{ and } X_1Y_1-1, X_2Y_2-1\rangle_{\mathbb{Z}}$$

is the ideal associated to E. *If B is a standard basis of I then B is said to* be associated *to E.*

Definition 4.2. *A* general coloring χ_E *is the map from $\mathbb{Z}[X_1,X_2,Y_1,Y_2]$ into $\mathbb{Z}[X_1,X_2,Y_1,Y_2]$ such that $\chi_E(Q) = \bar{Q}^B$.*

Theorem 4.3. *A $\mathbb{Z}$-polyomino P is $\mathbb{Z}$-tilable by a set of $\mathbb{Z}$-polyominoes $E = \{P_1, ..., P_s\}$ if and only if $\chi_E(P) = 0$.*

Proof. By theorem 2.2, P is $\mathbb{Z}$-tilable by $E = \{P_1, ..., P_s\}$ if and only if Q_P belongs to $I(E)$, and by definition, Q_P belongs $I(E)$ if and only if $\chi_E(Q_P) = 0$. □

Remark 4.4. *This definition seems to be tautological. Indeed, this is very useful to have a geometric visualization. Let us observe the following explicit example.*

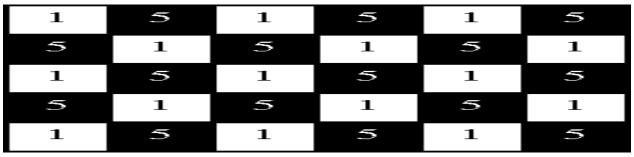


Fig.4. The T-tetraminoes

Example 4.5. *Consider that we want to have a chromatic characterization of the* $\mathbb{Z}$*-tilability by the set* E *of classical polyominoes described below (fig.4) called* T-tetraminoes.
We have

$$I(E) = \langle X_1^2 + X_1X_2 + X_1 + 1, X_1^2X_2 + X_1X_2 + X_1 + X_2, X_1X_2 + X_2^2 + X_2 + 1, X_1X_2^2 + X_1X_2 + X_1 + X_2, X_1Y_1 - 1, X_2Y_2 - 1\rangle.$$

We compute a standard basis for the order $Y_1 > Y_2 > X_1 > X_2$: $(X_1 + 3, X_2 + 3, 8, Y_1 + 3, Y_2 + 3)$. *So, we have* $\chi_E(Q_{c(i,j)}) = \chi_E(X^{(i,j)}) = \begin{cases} 1 & \text{if } i+j = 0 \mod 2 \\ 5 & \text{if } i+j = 1 \mod 2 \end{cases}$. *Moreover, we always have the following classical remainder property* $\chi_E(Q_P) = \chi_E(\sum_{c(i,j)\in P} Q_{c(i,j)}) = \chi_E\left(\sum_{c(i,j)\in P} \chi_E\left(Q_{c(i,j)}\right)\right)$. *Now, as* $A = \chi_E\left(\sum_{c(i,j)\in P} Q_{c(i,j)}\right)$ *is an integer,* $\chi_E(A) = A \mod 8$. *So, suppose that the squares of the plan have a chessboard-like coloration (fig.5), a polyomino* P *is* $\mathbb{Z}$*-tilable by* E *if and only if when assigning 5 on the white squares and 1 on the black ones in* P*, the sum of values on the squares is a multiple of 8.*

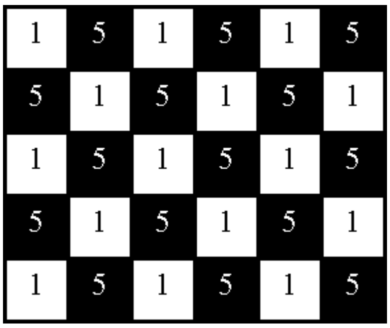


Fig.5. A general coloring for the T-tetraminoes. The lower left corner square is the square (0,0).

5. $\mathbb{Z}$-tilability and boundary conditions

In this section, we only deal with classical polyominoes and not with $\mathbb{Z}$-polyominoes. In the paper of Conway and Lagarias, it is possible to know if a polyomino P has a $\mathbb{Z}$-tiling in travelling the boundary of P. We prove that we have a similar situation with our characterization. We do not need to compute the remainder of Q_P, but only the remainder of a shorter polynomial associated to the boundary of P.

Theorem 5.1. *Let* P *be a polyomino and* χ_E *a general coloring associated to* E*. We suppose that*

$$\chi_E\left(\sum_{i=1}^{2}(X_i + Y_i)\right) - 4$$

is not a zero divisor of $\mathbb{Z}[X_1, X_2, Y_1, Y_2]/I(E)$*. In this case,*

$$\chi_E(Q_P) = 0 \text{ if and only if } \sum_{(c_1,c_2)\in S} (\chi_E(Q_{c_1}) - \chi_E(Q_{c_2})) = 0$$

where (c_1, c_2) *belongs to* S *if* c_1 *is a square in* P*,* c_2 *does not belong to* P *and* c_2 *has a common side with* c_1*.*

Proof. Let χ_E the general coloring, we denote by

$$v_{\chi_E}(X^a) = \sum_{i=1}^{2} (\chi_E(X^a) - \chi_E(X^a X_i)) + \sum_{i=1}^{2} (\chi_E(X^a) - \chi_E(X^a Y_i)).$$

We have $\sum_{(c_1,c_2)\in S} (\chi_E(Q_{c_1}) - \chi_E(Q_{c_2})) = \sum_{c\in P} v_{\chi_E}(Q_c)$ because, if the squares c and c' belong to P, the contributions of the couples (c, c') et (c', c) vanish themselves. We denote by T the set of (c, c') where c is a square in P, and c' has a common side with c_1. Moreover, we have $\sum_{(c_1,c_2)\in S} (\chi_E(Q_{c_1}) - \chi_E(Q_{c_2})) = \sum_{c\in P} v_{\chi_E}(Q_c) = \sum_{c\in P} \sum_{(c,c')\in T} (\chi_E(Q_c) - \chi_E(Q_{c'}))$. Now, if we consider that the image of χ_E is in $\mathbb{Z}[X_1, X_2, Y_1, Y_2]/I(E)$, it is obvious that χ_E is a morphism of algebra. So,

$$\sum_{c\in P} \sum_{(c,c')\in T} (\chi_E(Q_c) - \chi_E(Q_{c'})) = \chi_E(Q_P)\left(4 - \chi_E\left(\sum_{i=1}^{2} (X_i + Y_i)\right)\right) \sum_{c\in P} \chi_E(Q_c)$$

in $\mathbb{Z}[X_1, X_2, Y_1, Y_2]/I(E)$. As $\chi_E\left(\sum_{i=1}^{2} (X_i + Y_i)\right) - 4$ is not a zero divisor, $\chi_E(Q_P) = 0$ if and only if $\sum_{(c_1,c_2)\in S} (\chi_E(Q_{c_1}) - \chi_E(Q_{c_2})) = 0$. □

To conclude this section, we give an example related to the paper of Thurston [8]. Let us consider that we want to $\mathbb{Z}$-tile a polyomino with dominoes (union of two adjacent squares). The associated ideal is $I(E) = \langle X_1 + 1, X_2 + 1, X_1 Y_1 - 1, X_2 Y_2 - 1\rangle$. We obtain that $(1 + X_1, 1 + X_2, Y_1 + 1, Y_2 + 1)$ is a standard basis of $I(E)$ for the order $Y_1 > Y_2 > X_1 > X_2$. So, $\chi_E(X^{(i,j)}) = \begin{cases} 1 & \text{if } i + j = 0 \mod 2 \\ -1 & \text{if } i + j = 1 \mod 2 \end{cases}$. Hence, a polyomino is $\mathbb{Z}$-tilable by dominoes if and only if it has the same number of black (when $i + j = 1 \mod 2$) and white (when $i + j = 0 \mod 2$) squares $c_{(i,j)}$. In this case, the polyomino is said *balanced*. Independently, we have $\chi_E(\sum_{i=1}^{2} (X_i + Y_i)) - 4 = -8$ which is not a zero divisor. So, we can apply the theorem 5.1. The values of $\chi_E(Q_{c_1}) - \chi_E(Q_{c_2}) = \begin{cases} 2 & \text{if } c_1 = c_{(i,j)} \text{ and } i + j = 0 \mod 2 \\ -2 & \text{if } c_1 = c_{(i,j)} \text{ and } i + j = 1 \mod 2 \end{cases}$ where $(c_1, c_2) \in S$. Thus, P is balanced if and only if we have the same number of black and white edges on its boundary (a *black* (resp. *white*) *edge* is an edge which borders a black (resp. white) square of P).

References

[1] O. Bodini, Pavage des polyominos et Bases de Grobner, Rapport de recherche N^oRR2001-51, LIP, 2001.
[2] B. Buchberger, Introduction to Grobner basis, Logic of computation (Marktoberdorf 95) 35-66, NATO Adv. Sci. Inst. Ser. F Comput. Systems Sci.,157, Springer Berlin (1997).
[3] J.H Conway, J.C. Lagarias, Tiling with polyominoes and combinatorial group theory, J.C.T. Series A 53 (1990) 183-208.
[4] D. Cox, J. Little, D. O'Shea, Ideals,varieties and algorithms, 2nde edition, Undergraduate Text in Mathematics, Springer Verlag, New York (1997) XV 536 pp.
[5] S. W. Golomb, Tiling with polyominoes, J.C.T. Series A 1 (1966) 280-296.
[6] S. W. Golomb, Polyominoes which tile rectangles, J.C.T. Series A 51 (1989) 117-124.
[7] D.A. Klarner, Packing a rectangle with congruent n-ominoes, J.C.T. Series A 7 (1969) 107-115.
[8] W.P. Thurston, Conway's tiling groups, Amer. Math. Monthly, 97, num. 8 (Oct. 1990) 757-773.

LIP, École Normale Supérieure de Lyon, 46, Allée d'Italie, 69364 Lyon Cedex 05, France
E-mail address: `olivier.bodini@ens-lyon.fr`

A BIJECTION BETWEEN REALIZERS OF MAXIMAL PLANE GRAPHS AND PAIRS OF NON-CROSSING DYCK PATHS

NICOLAS BONICHON

Abstract. Schnyder trees, or realizers, of maximal plane graphs, are widely used in the graph drawing domain. In this paper, a bijection between realizers and pairs of non-crossing Dyck paths is proposed. The transformation of a realizer into a pair of non-crossing Dyck paths and the opposite operation can be done in linear time. Applying this bijection, we enumerate the number of realizers of size n and we can efficiently generate all of them.

Résumé. Les arbres de Schnyder, ou réaliseurs, d'un graphe maximal planaire, sont largement répandus dans le domaine du dessin de graphe. Nous proposons ici, une bijection entre les réaliseurs et les paires de chemins de Dyck qui ne se coupent pas. La transformation d'un réaliseur en une paire de chemin de Dyck et son inverse se font en temps linéaire. Utilisant cette bijection, nous pouvons énumérer les réaliseurs de taille n et nous pouvons les générer exhaustivement de manière efficace.

1. Introduction

Schnyder showed that every maximal plane graph admits a special decomposition of its interior edges into three trees (see Fig. 2), called a realizer [16, 17]. Such decomposition can be constructed in linear time [17]. Using realizers, it has been proved in [17] that every plane graph with $n \geq 3$ vertices has a planar straight-line drawing in a rectangular grid area $(n-2) \times (n-2)$.

Realizers are useful for many graph algorithms, of course for graph drawing [17, 4, 1, 13] but also for graph encoding [5]. They are linked to canonical orderings (or shelling orders) [9, 14], with 3-orientations [6], and with orderly spanning trees [4]. They can also be used to characterize planar graphs in terms of the order of their incidence, i.e., a graph G is planar iff the dimension of the incidence order of vertices and edges is at most 3 [16].

Realizers of the same graph have already been investigated [6, 3]. Suitable operations transforming a realizer of a graph to another realizer of the same graph have been introduced [3]. A particular normal form is also characterized. Moreover, the structure of the set of realizers of a given graph turns out to be a distributive lattice [6, 3]. Operations on realizers of same size have also been investigated. In [18] diagonal flip operations have been introduced. For all the maximal planar graphs in [2], colored diagonal flip operations on realizers have been proposed.

Here, we deal with realizers of size n, i.e. realizers of maximal plane graphs of size n. The main motivations are the following: how many realizers of size n are there and how can they be generated. To answer these two questions, a bijection between realizers of size n and pairs of non-crossing Dyck paths of size $2n-6$ is proposed.

A *Dyck path* of size $2n$ is a path in the discrete plane that starts from the point $(0,0)$ and ends at the point $(2n,0)$. It is composed of length $sqrt(2)$ elementary steps North-East and South-East such that it stays in the positive quarter of the plane. (f,g) is a *pair of non-crossing Dyck paths* if g never goes below f. Such paths have been studied by D. Gouyou-Beauchamps [10, 11]. Pairs of non-crossing Dyck paths are a particular case of *vicious walkers* [8, 7, 12, 15]. In [10], the number of pairs of non-crossing Dyck paths of length $2n$ is calculated: $|V_n| = C_{n+2}C_n - C_{n+1}^2$, where C_n is the Catalan number $\frac{(2n)!}{n!(n+1)!}$.

The principle of the bijection is the following. To each realizer R we can associate a particular realizer R_c, called *star realizer*. A star realizer is a realizer in which the third tree is a star i.e. all the inner vertices are children of the root. A realizer R is totally defined by its associated star realizer R_c and a particular sequence of flips, called a *prefix flip sequence*, which transforms R_c into R. The star realizer and the prefix flip sequence can be encoded by two non-crossing Dyck paths of size $2n-6$, where n is the size of the realizer. The star realizer is totally defined by its first tree T_0. T_0 is encoded by a Dyck path. The prefix flip sequence is encoded by a second Dyck path obtained from the first one by local transformations.

The rest of this paper is organized as follows. In Section 2, realizers are presented and some of their properties are given. Star realizers and the prefix flip sequence are introduced in the section 3. The bijection between realizers and non-crossing Dyck paths is explained in section 4.

2. Realizers

2.1. **Definitions.** We assume that the reader is familiar with graph theory. In this paper we deal with simple and undirected graphs. A drawing of a graph is a mapping of each vertex to a point of the plane and of each edge to the continuous curve joining the two ends of this edge. A planar drawing, or *plane graph* is a drawing without crossing edges except, eventually, on a common extremity. A graph that has a planar drawing is a planar graph. A plane graph splits the plane into topologically connected regions, called *face regions*. A *face* is the counter-clockwise walk of the boundary of a face region. One of the regions is unbounded and its associated face is named the *external face* of the plane graph. The vertices and edges of this face are called *external vertices* and *external edges*. The other vertices are called *inner* ones. The *adjacency list* of a vertex u is the list of neighbors of u. In plane graphs, the neighbors of u are ordered in the clockwise order in the adjacency list of u.

A planar graph G is *maximal* (or *triangulated*) if all the other graphs with a same number of vertices that contain it are not planar. The faces of a maximal plane graph are triangular. In this case, we denote v_0, v_1, v_2 the three vertices of the external face of this plane graph.

Definition 1. *(Schnyder* [16]*)*

A realizer of a maximal plane graph G is a partition of the interior edges of G in three sets T_0, T_1, T_2 of directed edges such that for each interior vertex u there holds:

1. *u has out-degree exactly one in each of T_0, T_1, T_2.*
2. *The counter-clockwise order of the edges incident on u is: leaving in T_0, entering in T_2, leaving in T_1, entering in T_0, leaving in T_2 and entering in T_1.*

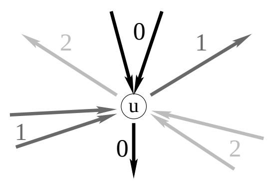


Figure 1. Edge coloration and orientation around a vertex

Schnyder show that, T_0, T_1 and T_2 were three *ordered rooted trees* where their edges are oriented toward their roots, which are the external vertices v_0, v_1, v_2. Each tree contains $n-2$ vertices.

An example of a graph, and a realizer of this graph are given in Figure 2.

FIGURE 2. An example of a realizer (a graph on the left side, and one of its realizers on the right side).

In the sequel, the edges of the tree T_i are colored with color i, where $i \in \{0,1,2\}$ such that the external edges $(v_i,\ v_{i+1})$ are of the color $i+1$.

$u_1 \underset{\longrightarrow}{i} u_2$ denotes the path colored i from u_1 to u_2. We write $u_1 >^i_{ccw} u_2$ (resp. $u_1 >^i_{cw} u_2$) if u_1 is after u_2 in the *counter-clockwise preordering* (resp. *clockwise preordering*) of the tree T_i. The parent of u in the tree T_i is denoted by $P_i(u)$. Let $Ch_i(u)$ be the list of children of u in clockwise order. $Ch_i(u,k)$ denotes the k^th child of the vertex u in T_i. We denote by $deg_i(u)$ the number of ingoing edges (number of children), of u in T_i. If u is not the element of $Ch_i(P_i(u))$, its predecessor u' in $Ch_i(P_i(u))$ is the *left brother* u. The *right branch* of a vertex u in a tree T, is the path in T that joins the right most leaf of the subtree of u to u. The *length* of a right branch is the number of edges of the right branch.

2.2. **Diagonal Flips on Realizers.** Let $\mathcal{R}_n$ be the set of realizers of graphs of size n.

In [18], R. Wagner proved that it is possible to obtain all maximal planar graphs of size n using a rewriting rule, called a diagonal flip. In this section, we extend this result to realizers using colored flips.

Definition 2. *Let G be an embedded graph. Let u_2, u_1, u_4 and u_3, u_4, u_1 be two adjacent faces where u_0 is not neighbor of u_3. A diagonal flip consists of removing the edge (u_1, u_4) and inserting the edge (u_2, u_3).*

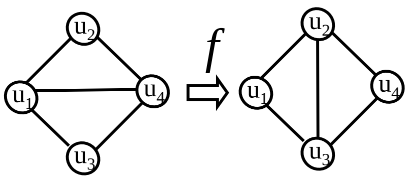


FIGURE 3. Diagonal flip operation

Theorem 1. *(Wagner* [18]*)*

Let G_1 and G_2 be two maximal planar graphs with n vertices. There exists a sequence of diagonal flips that transforms G_1 into G_2.

2.3. **Generalization to realizers.** As shown in Figure 4, we propose colored diagonal flips for realizers using two kinds of flips: f_1^i and f_2^i. It is easy to see that the application of a diagonal flip f_1^i or f_2^i on a realizer gives another realizer.

The choice between f_1^i and f_2^i depends on the quadrilateral configuration. Note that if, the edge (u_2, u_1) is colored $i-1$ and oriented towards u_1, and if the edge (u_3, u_1) is colored $i+1$ and oriented towards u_1, then $f_1^i(u_1)$ or $f_2^i(u_1)$ can be applied.

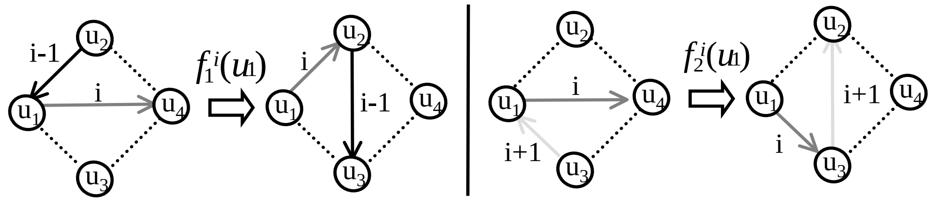


FIGURE 4. Flips on realizer

Unfortunately, these two operations are not always possible to apply. This occurs for the configuration of the quadrilateral of Figure 5.

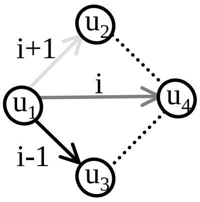


FIGURE 5. Configuration for which colored flip cannot be directly applied.

Theorem 2. [2] *There exists a sequence of colored flips that transforms any realizer R with n vertices into any other realizer R' with n vertices.*

3. Star realizers and prefix flip sequence

In this section, we present a particular class of realizers, called *star realizers*. Using these particular realizers, we construct in a unique way all realizers with a *prefix flip sequence*.

3.1. Star realizers.

Definition 3. *A star realizer $R_c = (T_0, T_1', C_{n-2})$ is a realizer where C_{n-2} is a star of size $n-2$ where all the edges are oriented toward the center of the star, i.e. C_{n-2} is a rooted tree of depth 1.*

In the first realizer of Figure 7, the vertex v_2 is a neighbor of all inner vertices of the graph. So this realizer is a star realizer.

Property 1. *Let T_0 be an ordered rooted tree of size $n-2$. There is a unique tree T_1' such that $R_c = (T_0, T_1', C_{n-2})$ is a star realizer.*

Proof. First, one can remark that there is only one way to connect T_0 and C_{n-2}: the clockwise prefix order in T_0 is the counter-clockwise order around v_2. Once T_0 and C_{n-2} are connected, we obtain a planar map. Let $F_k = (v_2, u_k, u_{k_1}, u_{k_2}, \ldots, u_{k_p}, u_{k+1})$ be a face of this planar map (see Figure 6). The parent in T_1' of the vertices $u_k, u_{k_1}, u_{k_2}, \ldots, u_{k_{p-1}}$ must be a vertex of this face. This is the only way to satisfy the second item of the definition 1 (see Figure 1). For the same reason the only vertex which can be the parent of $u_k, u_{k_1}, u_{k_2}, \dots, u_{k_{p-1}}$ is the vertex u_{k+}. For each vertex u_k, only one vertex can be the parent of u_k in T'_1. Hence, there is only one tree T'_1 such that $R_c = (T_0, T'_1, C_{n-2})$ is a realizer. $\sqcap$

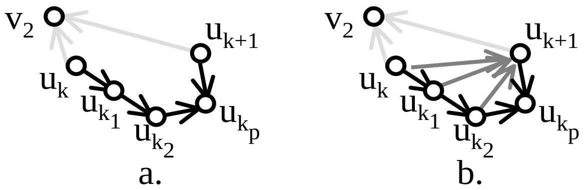


FIGURE 6. **a.** Face of the planar map obtained from the connection of T_0 and C_{n-2}. **b.** The same face, with the edges of T'_1 inside

From the construction of the tree T'_1 in the proof of property 1, the following property comes directly

Property 2. *Let $R_c = (T_0, T'_1, T_2)$ be a star realizer. Let G_c the maximal plane graph of R_c. Let $(u_1, ..., u_{n-3})$ be the inner vertices of G_c in the clockwise prefix order of T_0. The number of children of u_k in T'_1 is the length of the right branch of its left brother.*

3.2. **Prefix Flip sequence.** In this section, we will denote by $deg_1(u)$ the number of children of u in the tree T_1. Similarly, we will also denote by $deg'_1(u)$ the number of children of u in the tree T'_1.

Definition 4. *Let $R_c = (T_0, T'_1, C_{n-2})$ be a star realizer. The inner vertices u_k of G are numbered respecting the prefix order of T_0. A Prefix Flip Sequence, or PFS, is a sequence of flips $(f_1^2(u_{k_1}), f_1^2(u_{k_2}), \dots, f_1^2(u_{k_p}))$ that can be applied to R_c such that $i < j \Rightarrow k_i \leq k_j$.*

In the sequel, a PFS will be represented by a list of $n-2$ numbers, specifying the number of flips to apply on each inner vertex. For example, the PFS $(f_1^2(u_3), f_1^2(u_4), f_1^2(u_4))$ is represented by $(0, 0, 1, 2)$. $\#f(u_k)$ denotes the number of flips on u_k in the PFS. In the previous sequence, $\#f(u_4) = 2$. Figure 7 shows this PFS applied to a star realizer.

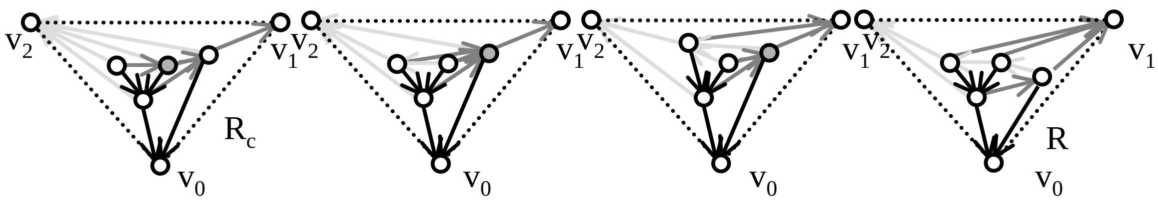


FIGURE 7. Example of prefix flip sequence $(0, 0, 1, 2)$

Remark 1. *A prefix flip sequence does not change the tree T_0 of the star realizer.*

Property 3. *Let $R_c = (T_0, T'_1, C_{n-2})$ be a star realizer. Let S be a prefix flip sequence and $R = (T_0, T_1, T_2)$ be the realizer obtained from R_c by S. For each inner vertex u_k, we have: $deg_1(u_k) = deg'_1(u_k) + \#f(u_{k-1}) - \#f(u_k)$*

Proof. The property can be reformulated in the following way: when a flip is applied on u_k in S, $deg_1(u_k)$ is decremented and $deg_1(u_{k+1})$ is incremented. Obviously, when a flip $f_1^2(u_k)$ is applied, $deg_1(u_k)$ is decremented. Let us show, that when a flip $f_1^2(u_k)$ is applied,

$deg_1(u_{k+1})$ is incremented. For this purpose, let us prove by induction on k that when a flip can be applied on u_k, u_{k+1} is just after $P_2(u_k)$ in the adjacency list of u_k.

First we can remark, that in a the star realizer R_c, for each k, u_{k+1} is just after $P_2(u_k)$ in the adjacency list of u_k.

Assume that after applying the flips of S on the $k-1$ first vertices, u_{i+1} is just after $P_2(u_i)$ in the adjacency list of u_i for all $i \geq k$. After applying $f_1^2(u_k)$, u_{k+1} is still just after $P_2(u_k)$ in the adjacency list of u_k (see Fig. 4). So for the $\#f(u_k)$ flips on u_k, u_{k+1} is just after $P_2(u_k)$ in the adjacency list of u_k.

Moreover, the modifications made by the flips $f_1^2(u_k)$ are enclosed in the region $(v_2, u_{k+1}, u_{k+1} \xrightarrow{0} v_0)$. Hence, $P_2(u_{k+1})$ in the adjacency list of u_{k+1} and for each $i > k+1$, the adjacency list of u_i is unchanged.

Hence, in a prefix flip sequence, each time we operate a flip $f_1^2(u_k)$, the number of children of u_k in T_1 is decremented and the number of children of u_{k+1} is incremented. □

Remark 2. *The property 3 can be also expressed:* $\#f(u_k) = deg_1'(u_k) + \#f(u_{k-1}) - deg_1(u_k)$.

Lemma 1. *Let* $R = (T_0, T_1, T_2)$ *be a realizer and* $R_c = (T_0, T_1', C_{n-2})$ *be its star realizer. There exists a unique prefix flip sequence* S_cw *that transforms* R_c *into* R.

Proof. Existence: Let R be a realizer. Let us consider the following algorithm:

for each vertex u_k in counter prefix order of T_0 **do**
  while u_k is not a neighbor of v_2 **do**
    Make the flip $f_2^1(P_2(u_k))$
  end while
end for

We cannot operate an infinite number of times the flip $f_2^1(P_2(u_k))$. Hence the algorithm terminates. When this algorithm ends, a star realizer is obtained, since all the inner vertices are neighbors of v_2. The reverse of the flip $f_2^1(P_2(u_k))$ is the flip $f_1^2(u_k)$ (see Figure 4). The reverse of the sequence of flips built by the previous algorithm is a prefix flip sequence. Hence, for every realizer R, there exists a flip sequence that transforms the star realizer R_c of R into R.

Unicity: Two realizers with two different star realizers cannot be identical since, they will have different trees T_0. Let R_c be a star realizer. Let S_{cw1} and S_{cw2} be two prefix flip sequences. Let R_1 (resp. R_2) be the realizer obtained from R_c by the flip sequence S_{cw1} (resp. S_{cw1}). Let k be the first index where the two sequences are different. The $deg_1(u_k)$ in R_1 is different from $deg_1(u_k)$ in R_2 (see property 3). Hence, a star realizer with two different sequences cannot produce the same realizer. □

Remark 3. *If the flips can be operated in any order, there are several ways to transform* R_c *into* R. *For example, the sequence* $(f_1^2(u_4), f_1^2(u_3))$ *transforms the star realizer* R_c *into the realizer* R *of Figure 7.*

4. Encoding and Decoding Realizers

In this section the bijection between realizers and pairs of non-crossing Dyck words is presented. This bijection is described as an encoding scheme. A first Dyck word is used to encode a star realizer. Precisely, this word encodes the tree T_0 of the star realizer. Then, a second Dyck word, which is used to encode the prefix flip sequence, is obtained from the first one by applying some permutations. In the decoding process, the first word is used to reconstruct the tree T_0 and the star realizer. The difference between the two words encodes the PFS.

4.1. **Non-crossing Dyck paths.** We use a finite set called *alphabet* where the elements are called *letters*. Here, we will use the alphabet $A = \{'(',')'\}$. A *word* is a finite sequence of letters denoted by $f = f_1 f_2 \dots f_n$. The set A^* of all words on the alphabet A is equipped with the concatenation. The length of a word, denoted by $|f|$, is the number of letters of f. For a letter x, $|f|_x$ denotes the number of letters x in the word f. A word f' is a left factor of f if there exists a word f'' such that $f = f'f''$. A morphism δ from A^* to $\mathbb{N}$ is defined by: $\delta('(') = 1$, $\delta(')') = -1$ and $\delta(f'f'') = \delta(f') + \delta(f'')$. The *Dyck language* is the following: $D = \{f \in A^* | \delta(f) = 0 \text{ and } \forall f' \text{ left factor of } f, \delta(f') \geq 0\}$. We denote by $D_n = D \cap A^{2n}$. We denote by $open(k, f)$ the position of the k^{th} opening parenthesis in f.

Dyck paths are paths coded by Dyck words. A step North-East is coded by $'('$ and a step South-East is coded by $')'$. These paths start from the point $(0, 0)$, never go bellow the x-axis and end on the x-axis. Dyck words of $2n - 2$ length are classically used to encode ordered rooted trees of size n. Figure 8 a. shows an ordered rooted tree and its coding with a Dyck word.

The pair (g, h) of $D_n \times D_n$ are *non-crossing Dyck words* if for all g' (resp. h') left factor of g (resp. h) such that $|g'| = |h'|$, $\delta(h') \geq \delta(g')$. V_n denotes the set of pairs of non-crossing words of $D_n \times D_n$. Obviously, a pair of non-crossing Dyck words encodes a pair of non-crossing Dyck paths. Figure 8 b. shows an example of non-crossing Dyck paths.

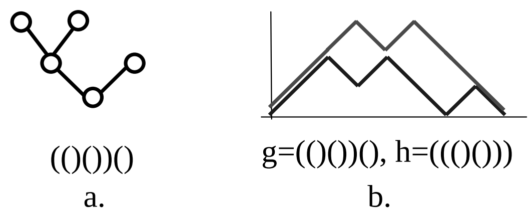


FIGURE 8. **a.** Encoding an ordered rooted tree with a Dyck word. **b.** Example of non-crossing Dyck paths

Theorem 3. *(Gouyou-Beauchamps [10])*
$|V_n| = C_{n+2}C_n - C_{n+1}^2$, *where* C_n *is the Catalan number* $\frac{(2n)!}{n!(n+1)!}$.

The first values of $|V_n|$ are $1, 1, 3, 14, 84, 594, 4719, \dots$

Algorithm 1 Encoding algorithm

```
Build the corresponding star realizer R_c of R
Code the tree T_0 by a Dyck word g.
h ← g
for each vertex u_k in the prefix order of T_0 do
   #f(u_k) ← deg'_1(u_k) − deg_1(u_k) + #f(u_{k−1})
   Move open(k, h) of #f(u_k) ranks to the left in h.
end for
```

4.2. **Encoding.**

Property 4. *In the algorithm 1, the number of flips on* u_k *is less than or equal to the number of consecutive closing brackets just before* $open(k, h)$ *in* h.

Proof. When no flips are previously made, $h = g$. The number of consecutive closing brackets just before $open(k, h)$ in h is exactly the length of the right branch of its left brother. $deg'_1(u_k)$ is equal to the length of the right branch of its left brother (see property 2). As $\#flips(u_k) \leq deg'_1(u_k)$, the property is satisfied in this case.

Suppose that the property is verified for $i \leq k-1$. The number of consecutive closing brackets just before $open(k,h)$ is $deg'_1(u_k) + \#flips(u_{k-1})$. As $\#flips(u_k) \leq deg'_1(u_k) + \#flips(u_{k-1})$ (see Remark 2) the property is still satisfied for $i = k$. □

Lemma 2. *The previous algorithm encodes a realizer R of size n with a pair of non-crossing Dyck words of lengths $2n-6$. This algorithm is linear time.*

Proof. First we can notice that g and h are non-crossing Dyck words.

Injectivity: Let $R = (T_0, T_1, T_2)$ and $R' = (T'_0, T'_1, T'_2)$ be two different realizers. Let (g,h) (resp. (g',h')) the pair of non-crossing Dyck words obtained by the previous algorithm from R (resp. R'). If $T_0 \neq T'_0$, then $g \neq g'$. Let S_f (resp. S'_f) be the PFS associated to R (resp. R'). Let k be the first index such that $\#f(u_k) \neq \#f'(u_k)$. After the k_{th} step in the loop, $open(k,h) \neq open(k,h')$. During the rest of the algorithm $open(k,h)$ and $open(k,h')$ are unchanged, so $h \neq h'$. Hence two different realizers are encoded with two different pairs of non-crossing words.

Complexity: To construct the star realizer, all vertices of T_0 are connected with an outgoing edge, colored 2, to v_2 and in-going edges, colored 1, to all the vertices of the right branch of its left brother. This construction can be done in linear time. The encoding of T_0 with the traditional algorithm is also done in linear time. The treatment of each vertex u_k is done in constant time. Hence the encoding algorithm computes in linear time. □

Example: to encode the realizer R of Figure 7, we can encode its star realizer R_c and the PFS $(0,0,1,2)$. The tree T_0 of R is the one of Figure 8 a. It can be encoded by $g = (()())()$. To encode the flip sequence, we need to move the third opening bracket of one step to the left and the fourth one to two steps to the left. So $h = ((()()))$. Hence, the realizer R of Figure 7 is encoded by the pair of non-crossing Dyck words (g,h).

4.3. **Decoding.** Let (g,h) be a pair of non-crossing words.

The function *Concat(L_1, L_2)* append the list L_2 at the end of the list L_1 and returns this new list. The function *Split(L, i)* removes the last i elements of L and returns a list which contains these i elements. The procedure *AddFirst(L, E)* added the element E at the begining of the list L. Naturally, *Del(E, i)* removes the i^{th} element of the list L.

Algorithm 2 Decoding algorithm

> Build the tree T_0 from g
> Build the star realizer $R_c = (T_0, T'_1, C_{n-2})$
> $R = (T_0, T_1, T_2) \leftarrow R_c$
> **for each** vertex u_k in the prefix order of T_0 **do**
> $\#f(u_k) \leftarrow open(g,k) - open(h,k)$
> $L \leftarrow Split(Ch_1(u_k), \#f(u_k))$
> $Ch_1(u_{k+1}) \leftarrow Concat(Ch_1(u_{k+1}), L)$
> $Del(Ch_2(P_2(u_k)), u_k)$
> $AddFirst(Ch_2(Ch_1(u_{k+1}, 0)), u_k)$
> **end for**

Lemma 3. *The algorithm 2 computes in linear time a realizer R of size n from a pair of non-crossing Dyck words of lengths $2n-6$.*

Proof. Validity: as $h \geq g$, $0 \leq \#f(u_k) \leq deg_1'(u_k) + \#flips(u_{k-1})$ encodes a star Realizer and a valid PFS. Moreover the algorithm 2 constructs the realizer encoded by the algorithm 1.

Complexity: as in the encoding algorithm, the construction of the star realizer can be operated in linear time. The algorithm uses chained lists to store the list of children of vertices in each tree. The split operation in the loop takes $O(deg_1(u_k))$ operations. Globally, it takes $O(m) = O(n)$ operations. The other operations in the loop take $O(1)$ operations. So globally it takes $O(n)$ operations. □

The following theorem comes directly from lemma 2 and lemma 3:

Theorem 4. *There is a bijection between realizers of size n and pairs of non-crossing Dyck paths of lengths* $2n - 6$.

Corollary 5. *The number of realizer of size* n *is* $|R_n| = |V_{n-3}| = C_{n-3}C_{n-1} - C_{n-2}^2$.

Acknowledgments

The author thanks Philippe Duchon, Mireille Bousquet-Mélou and Xavier Gérard Viennot for fruitful discussions and suggestions. Much thanks to Mohamed Mosbah for his help with writing the paper.

References

[1] G. Di Battista, R. Tamassia, and L.Vismara. Output-sensitive reporting of disjoint paths. In *Proc. 2nd Int. Conf. Computing and Combinatorics*, number 1090, pages 81–91. Springer-Verlag, 1996.

[2] N. Bonichon, B. Le Saëc, and M. Mosbah. Diagonal flip operations on realizers and their application to wagner's theorem on realizers. In *Applied Graph Transformation 2002 (AGT2002), Grenoble 12-13th April, France*, 2002.

[3] E. Brehm. *3-orientations and schnyder 3-Tree-Decompositions.* PhD thesis, FB Mathematik und Informatik, Freie Universität Berlin, 2000.

[4] Yi-Ting Chiang Ching-Chi and Hsueh-I Lu. Orderly spanning trees with applications to graph encoding and graph drawing. In *Proc. 12th Symp. Discrete Algorithms*, pages 506–515. ACM and SIAM, 2001.

[5] Richie Chih-Nan Chuang, Ashim Garg, Xin He, Ming-Yang Kao, and Hsueh-I Lu. Compact encodings of planar graphs via canonical ordering and multiple parentheses. In *Proc. 25th International Colloquium on Automata, Languages, and Programming (ICALP'98)*, volume 1443, pages 118–129, 1998.

[6] P. Ossona de Mendez. *Orientations bipolaires.* PhD thesis, Ecole des Hautes Etudes en Sciences Sociales, 1994.

[7] J.W. Essam and A.J. Guttmann. Vicious walkers and directed polymer networks in general dimension. *Phys. Rev. E 52*, pages 5849–5862, 1995.

[8] M. E. Fisher and M. P. Gelfand. The reunions of three dissimilar vicious walkers. *J. Stat. Phys 53*, pages 175–189, 1988.

[9] H. De Frayseix, J. Pach, and J. Pollack. How to draw a planar graph on a grid. *Combinatorica*, 10:41–51, 1990.

[10] D. Gouyou-Beauchamps. Chemins sous-diagonaux et tableaux de young. In *Colloque de Combinatoire Enumrative, Montral, UQAM 1985, Lecture Notes in Mathematics*, volume 1234, pages 112–125, 1986.

[11] D. Gouyou-Beauchamps. Standard young tableaux of height 4 and 5. *Europ. J. Combinatorics*, 10:69–82, 1989.

[12] A. J. Guttmann, A. L. Owczarek, and X. G. Viennot. Vicious walkers and young tableaux i: Without walls. *J. Phys. A; Math. Gen.*, 31:8123–8135, 1998.

[13] G. Kant. Drawing planar graphs using the lmc-ordering. In *Proc. 33th Ann. IEEE Symp. on Found. of Comp. Science*, pages 101–110, 1992.

[14] G. Kant. Drawing planar graphs using the canonical ordering. *Algorithmica*, 16:4–32, 1996.

[15] C. Krattenthaler, A. J. Guttmann, and X. G. Viennot. Vicious walkers, friendly walkers and young tableaux ii: with a walls. *J. Phys. A; Math. Gen.*, 33:8123–8135, 2000.

[16] W. Schnyder. Planar graphs and poset dimension. *Order*, 5:323–343, 1989.

[17] W. Schnyder. Embedding planar graphs on the grid. *Proc. 1st ACM-SIAM Symp. Discrete Algorithms*, pages 138–148, 1990.

[18] K. Wagner. Bemerkungen zum vierfarbenproblem. In *Jahresber. Deutsche Math. -Verein.*, volume 46, pages 26–32, 1936.

LaBRI-Université Bordeaux 1, 351 Cours de la Libération, 33405 Talence, France
E-mail address: `bonichon@labri.fr`

ENUMERATION OF SOLID 2-TREES

MICHEL BOUSQUET AND CEDRIC LAMATHE

Abstract. The goal of this paper is to enumerate solid 2-trees according to the number of edges (or triangles) and also according to the edge degree distribution. We first enumerate oriented solid 2-trees using the general methods of the theory of species. In order to obtain non oriented enumeration formulas we use quotient species which consists in a specialization of Pólya theory.

Résumé. Le but de cet article est d'obtenir l'énumération des 2-arbres solides selon le nombre d'arêtes (ou de triangles) ainsi que selon la distribution des degrés des arêtes. Nous obtenons d'abord le dénombrement des 2-arbres solides orientés en utilisant les méthodes de la théorie des espèces. Pour obtenir le dénombrement des 2-arbres solides non orientés, nous utilisons la notion d'espèce quotient qui provient d'une spécialisation de la théorie de Pólya.

1. Introduction

Definition 1. Let $\mathcal{E}$ be a non-empty finite set of n elements called *edges*. A *2-tree* is either a single edge (if $n = 1$) or a non-empty subset $\mathcal{T} \subseteq \mathcal{P}_3(\mathcal{E})$ whose elements are called *triangles*, satisfying the following conditions:

(1) For every pair $\{a, b\} = \{\{a_1, a_2, a_3\}, \{b_1, b_2, b_3\}\}$ of distinct elements of $\mathcal{T}$, we have $|a \cap b| \leq 1$, which means that two distinct triangles share at most one edge.
(2) For every ordered pair $(a, b) = (\{a_1, a_2, a_3\}, \{b_1, b_2, b_3\})$ of distinct elements of $\mathcal{T}$, there is a unique sequence $(t_0 = a, t_1, t_2, \ldots, t_k = b)$ such that for $i = 0, 1, \ldots, k-1$, we have $t_i \in \mathcal{T}$ and $|t_i \cap t_{i+1}| = 1$, which means that each pair of consecutive triangles in this sequence share exactly one edge.

An edge e and a triangle t are *incident* to each other if $e \in t$. The *degree* of an edge is the number of triangles which are incident to that edge. The *edge degree distribution* of a 2-tree is described by a vector $\vec{n} = (n_1, n_2, \ldots)$, where n_i is the number of edges of degree i. We denote by $\mathrm{Supp}(\vec{n})$, the *support* of $\vec{n}$ which is the set of indices i such that $n_i \neq 0$. Figure 1 shows a 2-tree having 11 edges, 5 triangles and edge degree distribution given by $\vec{n} = (8, 2, 1)$.

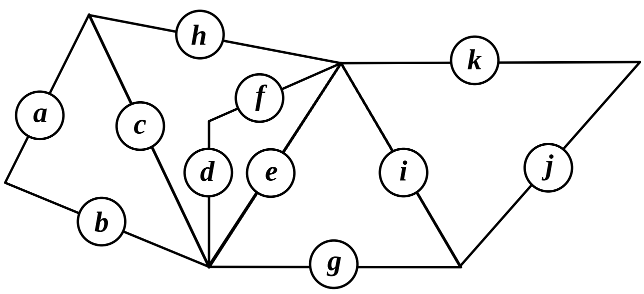


Figure 1. A 2-tree on $\mathcal{E} = \{a, b, c, d, e, f, g, h, i, j, k\}$.

Several classes of 2-trees have been studied before. Beineke and Pippert enumerate some k-dimensional trees in [1] labelled at vertices. In [7], Harary and Palmer count unlabelled 2-trees. For the enumeration of plane 2-trees see [10], and for a classification of plane and planar 2-trees see [8]. In [5, 6], Fowler et al. worked on general 2-trees and give asymptotical results. Here, we consider a new class of 2-trees, that is, *solid* 2-trees, *i.e.* 2-trees in which there is a cycle structure on the triangles around each edge.

Lemma 1. Let m, n be two nonnegative integers, and $\vec{n} = (n_1, n_2, \ldots)$, an infinite vector of nonnegative integers. Then

(1) There exists a 2-tree having m triangles and n edges if and only if $n = 2m + 1$.
(2) There exists a 2-tree having $\vec{n}$ as edge degree distribution if and only if

$$\sum_i n_i = n \quad \text{and} \quad \sum_i i n_i = 3m. \tag{1}$$

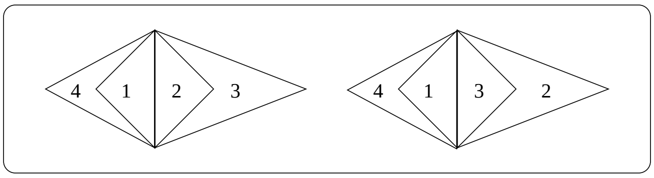


FIGURE 2. Two distinct solid 2-trees but the same 2-tree.

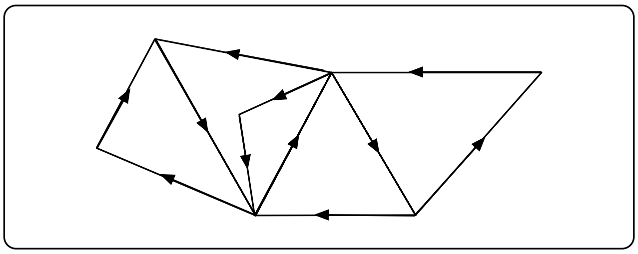


FIGURE 3. A well oriented 2-tree.

A *solid* 2-tree can be viewed topologically as a 2-tree in which the faces of the triangles cannot interpenetrate themselves. As a consequence, there is a cyclic configuration of triangles around each edge. Figure 2 shows an example of two different solid 2-trees which are in fact the same 2-tree. As we can see, in the case of a solid 2-tree, one has to take into account the cyclic order of the triangles around each edge. A *well oriented* solid 2-tree is obtained from a solid 2-tree in the following way: first, pick any triangle and give a cyclic orientation on its edges. Then each triangle adjacent to the first triangle inherits a cyclic orientation (see Figure 3). This process is repeated until all edges receive an orientation. By the arborescent nature of the structure, there will be no conflict (the orientation of each edge will always be well defined). Figure 3 shows an example of a well oriented 2-tree. The species of non-oriented and well oriented solid 2-trees will be denoted respectively by $\mathcal{A}$ and $\mathcal{A}_o$. In order to analyze these two species, the following auxiliary species will be used:

- The species of *triangles* X: a single triangle will be denoted by X.
- The species of *edges* Y: a single edge will be denoted by Y.
- The species L of *lists* or *linear orders*.

- The species C and C_3, respectively denoting the species of oriented cycles and of oriented cycles of length 3.
- The species $\mathcal{A}^{-}$ and $\mathcal{A}_o^{-}$, respectively denoting the species of non oriented and well oriented solid 2-trees *rooted at an edge.*
- The species $\mathcal{A}^{\triangle}$ and $\mathcal{A}_o^{\triangle}$, respectively denoting the species of non oriented and well oriented solid 2-trees *rooted at a triangle.*
- The species $\mathcal{A}^{\underline{\triangle}}$ and $\mathcal{A}_o^{\underline{\triangle}}$, respectively denoting the species of non oriented and well oriented solid 2-trees *rooted at a triangle having itself one of its edges distinguished.*
- Finally, the species $\mathcal{B}$ of *planted* oriented solid 2-trees which consists of an oriented root edge Y incident to a linear order (L-structure) of triangles X each of which having its two remaining sides being themselves $\mathcal{B}$-structures. Therefore, the species $\mathcal{B}$ satisfies the following combinatorial equation

$$\mathcal{B}(X,Y) = YL(X\mathcal{B}^2(X,Y)), \tag{2}$$

as illustrated by Figure 4.

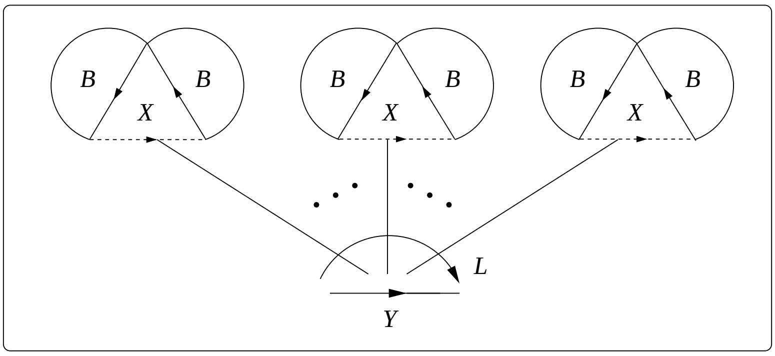


FIGURE 4. A $\mathcal{B}$-structure.

Note that $\mathcal{B}$ has been defined as a *two-sort* species where the sorts are X and Y. Since the numbers of edges n and of triangles m are linked by the relation $n = 2m+1$, equation (2) above can either be expressed as a one sort species in X alone by setting $Y := 1$, or in Y alone, by setting $X := 1$ respectively, giving the two following equations:

$$\mathcal{B}(X,1) = L(X\mathcal{B}^2(X,1)), \tag{3}$$

$$\mathcal{B}(1,Y) = YL(\mathcal{B}^2(1,Y)). \tag{4}$$

Recall that setting $X := 1$ in a two sort species $F(X,Y)$ essentially means unlabelling the elements of sort X. The second form in equation (4) is more suitable for the use of Lagrange inversion formula. Therefore the species Y of edges will be used as the base singleton species to make our computations and we will use the shorter form $\mathcal{B}(Y) = YL(\mathcal{B}^2(Y))$ for (4). Hence, the structures are labelled at edges. However, the results will be more elegant when expressed as a function of the number m of triangles.

• Lagrange Inversion Formula

In this paper we make an extensive use of Lagrange inversion formula (see [2]). Let A and R be formal series satisfying $A(Y) = YR(A)$ and $R(0) = 0$. If F is another series, then

$$[y^n]F(A(y)) = \frac{1}{n}[y^{n-1}]F'(t)R^n(t), \tag{5}$$

where $[y^n]F(A(y))$ denotes the coefficient of y^n in $F(A(y))$. Another main tool used in this paper is the following dissymmetry theorem which has been proved in [5, 6]. Note that in their paper, the authors made a proof for general 2-trees but obviously, the proof is also valid for both well oriented and non oriented solid 2-trees.

Theorem 1. The species $\mathcal{A}_o$ and $\mathcal{A}$, respectively of well oriented and (non oriented) solid 2-trees, satisfy the following relations:

$$\mathcal{A}_o^- + \mathcal{A}_o^{\Delta} = \mathcal{A}_o + \mathcal{A}_o^{\underline{\Delta}}, \tag{6}$$

and

$$\mathcal{A}^- + \mathcal{A}^{\Delta} = \mathcal{A} + \mathcal{A}^{\underline{\Delta}}. \tag{7}$$

2. Well oriented solid 2-trees

We begin this section by expressing the species appearing in the dissymmetry theorem (oriented case) in terms of the species $\mathcal{B}$.

Theorem 2. The species $\mathcal{A}_o^-$, $\mathcal{A}_o^{\Delta}$ and $\mathcal{A}_o^{\underline{\Delta}}$ satisfy the following isomorphisms of species :

$$\mathcal{A}_o^-(Y) = Y + YC(\mathcal{B}^2(Y)), \tag{8}$$

$$\mathcal{A}_o^{\Delta}(Y) = C_3(\mathcal{B}(Y)), \tag{9}$$

$$\mathcal{A}_o^{\underline{\Delta}}(Y) = \mathcal{B}(Y)^3, \tag{10}$$

where C and C_3 are the species of oriented cycles and of oriented cycles of length 3.

2.1. Enumeration according to the number of edges.

• Labelled case

Let $\mathcal{A}_o[n]$ be the set of edge labelled solid 2-trees over n edges. We similarly define $\mathcal{A}_o^-[n]$, $\mathcal{A}_o^{\Delta}[n]$ and $\mathcal{A}_o^{\underline{\Delta}}[n]$. Our first task is to determine $|\mathcal{A}_o^-[n]|$, the cardinality of the set $\mathcal{A}_o^-[n]$. By applying Lagrange inversion with $F(t) = C(t^2) = -\log(1-t^2)$ and $R(t) = L(t^2) = (1-t^2)^{-1}$, we find

$$\begin{aligned}
[y^n]\mathcal{A}_o^-(y) &= [y^{n-1}]C(\mathcal{B}^2(y)), \\
&= \frac{2}{n-1}[t^{n-3}]\sum_{j\geq 0}\frac{n(n+1)\cdots(n+j-1)}{j!}t^{2j} \\
&= \frac{2}{3(n-1)}\binom{3(n-1)/2}{n-1}.
\end{aligned}$$

Hence, we have

$$|\mathcal{A}_o^-[n]| = n![y^n]\mathcal{A}_o^-(y) = \frac{2}{3}n(n-2)!\binom{\frac{3(n-1)}{2}}{n-1}. \tag{11}$$

Note that when a solid 2-tree over n edges is labelled, we have n different choices for the root edge. Therefore

$$n|\mathcal{A}_o[n]| = |\mathcal{A}_o^-[n]|,$$

and the next proposition follows.

Proposition 1. The number $|\mathcal{A}_o[n]|$ of well oriented edge-labelled solid 2-trees over n edges is given by

$$|\mathcal{A}_o[n]| = \frac{2}{3}(n-2)!\binom{\frac{3(n-1)}{2}}{n-1}, \quad n > 1. \tag{12}$$

Note that if we express equation (12) as a function of m, the number of triangles, we obtain

$$|\mathcal{A}_{o,t}[m]| = \frac{m!}{3}\frac{1}{2m+3}\binom{3m+3}{m+1}, \quad m \geq 1, \tag{13}$$

where the index t in $|\mathcal{A}_{o,t}[m]|$ means that the strucures are labelled at triangles instead of edges.

• **Unlabelled case**

We first need to compute the generating series $\widetilde{\mathcal{A}_o^-}(y)$ of unlabelled $\mathcal{A}_o^-$-structures. In order to accomplish this, we use the following property: let F and G be two species, then we have

$$(F(G))^{\sim}(x) = Z_F(\widetilde{G}(x), \widetilde{G}(x^2), \widetilde{G}(x^3), \ldots), \tag{14}$$

where the *cycle index series* Z_F of a species is defined by

$$Z_F(x_1, x_2, \ldots) = \sum_{k\geq 0}\frac{1}{k!}\sum_{\sigma\in\mathcal{S}_k} \mathrm{fix}F[\sigma]x_1^{\sigma_1}x_2^{\sigma_2}x_3^{\sigma_3}\cdots, \tag{15}$$

where $\mathcal{S}_k$ is the symmetric group of order k, σ_i, the number of cycles of length i in σ and $\mathrm{fix}F[\sigma]$, the number of F-structures left fixed under the relabelling induced by σ. For example, if $F = C$, the species of oriented cycles, we have

$$Z_C(x_1, x_2, \ldots) = \sum_{k\geq 1}\frac{\phi(k)}{k}\ln\left(\frac{1}{1-x_k}\right), \tag{16}$$

where ϕ is the Euler function. Now, applying this to the species $\mathcal{A}_o^- - Y = YC(\mathcal{B}^2)$, we get

$$\begin{aligned}\widetilde{\mathcal{A}_o^-}(y) &= yZ_C(\widetilde{\mathcal{B}}^2(y), \widetilde{\mathcal{B}}^2(y^2), \widetilde{\mathcal{B}}^2(y^3), \ldots),\\ &= y\sum_{k\geq 1}\frac{\phi(k)}{k}\ln\left(\frac{1}{1-\widetilde{\mathcal{B}}^2(y^k)}\right).\end{aligned}$$

We note that since $\mathcal{B}$ is asymmetric (there are exactly $n!$ labelled structures for each unlabelled structures), we have $\widetilde{\mathcal{B}}(y) = \mathcal{B}(y)$, hence

$$\begin{aligned}|\widetilde{\mathcal{A}_o^-}[n]| &= [y^n]\widetilde{\mathcal{A}_o^-}(y),\\ &= [y^{n-1}]\sum_{k\geq 1}\frac{\phi(k)}{k}\ln\left(\frac{1}{1-\mathcal{B}^2(y^k)}\right).\end{aligned}$$

But

$$\begin{aligned}[y^{n-1}]\ln\left(\frac{1}{1-\mathcal{B}^2(y^k)}\right) &= \frac{2k}{n-1}[t^{\frac{n-1}{k}-2}](1-t^2)^{-\frac{n-1}{k}-1},\\ &= \frac{2k}{3(n-1)}\binom{3(n-1)/2k}{(n-1)/k}.\end{aligned}$$

Obviously, k must divide $n-1$ and $(n-1)/k$ must be even. Letting $d = (n-1)/k$, we finally get

$$|\widetilde{\mathcal{A}_o^-}[n]| = \frac{2}{3(n-1)}\sum_{d}\phi((n-1)/d)\binom{3d/2}{d}, \tag{17}$$

the sum being taken over all even divisors d of $n-1$. To compute $|\widetilde{\mathcal{A}}_o^{\triangle}[n]|$, we use equation (9) and the fact that

$$Z_{C_3}(y_1, y_2, \ldots) = \frac{1}{3}(y_1^3 + 2y_3).$$

We have

$$[y^n]\mathcal{B}^3(y) = \frac{1}{n}\binom{3(n-1)/2}{n-1},$$

and

$$[y^n]\mathcal{B}(y^3) = [y^{n/3}]\mathcal{B}(y) = \frac{3}{n}\binom{(n-3)/2}{n/3-1},$$

so that

$$|\widetilde{\mathcal{A}}_o^{\triangle}[n]| = \frac{1}{3n}\binom{\frac{3(n-1)}{2}}{n-1} + \frac{2}{n}\chi(3|n)\binom{\frac{(n-3)}{2}}{\frac{n}{3}-1}, \tag{18}$$

where $\chi(3|n) = 1$ if 3 divides n and 0 otherwise. It can be easily shown, by a very similar way that

$$|\widetilde{\mathcal{A}}_o^{\triangleq}[n]| = \frac{1}{n}\binom{\frac{3(n-1)}{2}}{n-1}. \tag{19}$$

And we get the following result:

Proposition 2. The number of unlabelled well oriented solid 2-trees over n edges is given by

$$|\widetilde{\mathcal{A}}_o[n]| = \frac{2}{3(n-1)}\sum_{d}\phi\left(\frac{n-1}{d}\right)\binom{3d/2}{d} + \chi(3|n)\frac{2}{n}\binom{\frac{n-3}{2}}{\frac{n}{3}-1} - \frac{2}{3n}\binom{\frac{3(n-1)}{2}}{n-1}, \tag{20}$$

the first sum being taken over all even divisors d of $n-1$.

We can also write $|\widetilde{\mathcal{A}}_{o,t}[m]|$, in function of the number m of triangles, as follows

$$|\widetilde{\mathcal{A}}_{o,t}[m]| = \frac{1}{3m}\sum_{d|m}\phi\left(\frac{m}{d}\right)\binom{3d}{d} + \chi(3|2m+1)\frac{2}{2m+1}\binom{m-1}{\frac{2m-2}{3}} - \frac{2}{3(2m+1)}\binom{3m}{m}.$$

Note that this expression is also the number of unlabelled 3-gonal cacti on m triangles (see [3]). There is an obvious bijection between these objects and solid 2-trees. The sequence of these numbers is known as sequence A054423 in the on-line encyclopedia of integers sequences ([11]).

2.2. **Enumeration according to edge degree distribution.**

Let $r = (r_0, r_1, r_2, \ldots)$ be an infinite set of formal variables. Recall that $\mathcal{A}[n]$ is the set of solid 2-trees over n edges. In order to keep track of the edge degree distribution, we introduce, for a given number n, the following weight function:

$$\begin{array}{rccc} w: & \mathcal{A}[n] & \longrightarrow & \mathbb{Q}[r_1, r_2, \ldots] \\ & s & \longmapsto & w(s) \end{array} \tag{21}$$

where $\mathbb{Q}[r_1, r_2, \ldots]$ is the ring of polynomials over the field of rational numbers $\mathbb{Q}$ in the variables $r_1, r_2, \ldots$ and where the weight of a given $\mathcal{A}$-structure s is defined by $w(s) = r_1^{n_1} r_2^{n_2} \cdots$, where n_i is the number of edges of degree i in s. Equations (2), (8), (9) and (10) have the following weighted versions:

$$\mathcal{B}_r = Y L_{r'}(\mathcal{B}_r^2), \tag{22}$$

and

$$\mathcal{A}_{o,w}^{-}(Y) = r_1 Y + Y C_r(\mathcal{B}_r^2), \tag{23}$$

$$\mathcal{A}_{o,w}^{\triangle}(Y) = C_3(\mathcal{B}_r), \tag{24}$$

$$\mathcal{A}_{o,w}^{\underline{\triangle}}(Y) = \mathcal{B}_r^3, \tag{25}$$

where C_r is the weighted species of cycles such that a cycle of length i has the weight r_i, and its derivative $L_{r'}$ which is the species of lists where a list of length i has the weight r_{i+1}. These species have the following generating series:

$$C_r(y) = r_1 y + \frac{r_2}{2} y^2 + \frac{r_3}{3} y^3 + \cdots,$$

and

$$L_{r'}(y) = (C_r(y))' = r_1 + r_2 y + r_3 y^2 + \cdots.$$

Let $\vec{n} = (n_1, n_2, n_3, \ldots)$ be a vector of nonnegative integers. Recall that there exists a 2-tree having a total of n edges and n_i edges of degree i if and only if the following relations are satisfied:

$$\sum_i n_i = n \quad \text{and} \quad \sum_i i n_i = 3\left(\frac{n-1}{2}\right). \tag{26}$$

• Labelled case

Let $\vec{n}$ be a vector satisfying (26). Then the number $|\mathcal{A}_o^-[\vec{n}]|$ of well oriented edge labelled solid 2-trees pointed at an edge, and having $\vec{n}$ as edge degree distribution, is given by

$$|\mathcal{A}_o^-[\vec{n}]| = n![r_1^{n_1} r_2^{n_2} \cdots][y^n]\mathcal{A}_{o,w}^-(y). \tag{27}$$

We have

$$\begin{aligned}
[y^n]\mathcal{A}_{o,w}^-(y) &= \frac{1}{n-1}[t^{n-2}]\frac{d}{dt}(C_r(t^2)) \cdot L_{r'}^{n-1}(t^2), \\
&= \frac{2}{n-1}[t^{n-3}](r_1 + r_2 t^2 + r_3 t^4 + \cdots)^n, \\
&= \frac{2}{n-1}[t^{n-3}] \sum_{\ell_1+\ell_2+\cdots=n} \binom{n}{\ell_1, \ell_2, \ldots} r_1^{\ell_1} r_2^{\ell_2} \cdots t^{2\ell_2+4\ell_3+6\ell_4+\cdots}.
\end{aligned}$$

Finally, we obtain

$$[y^n]\mathcal{A}_o^-(r, y) = \sum_{\ell_1, \ell_2, \ldots} \binom{n}{\ell_1, \ell_2, \ldots} r_1^{\ell_1} r_2^{\ell_2} \cdots,$$

the sum being taken over all vectors $(\ell_1, \ell_2, \ldots)$ satisfying

$$\sum_i \ell_i = n \quad \text{and} \quad \sum_i 2(i-1)\ell_i = n-3.$$

We note that this condition is the same as in (26). Hence using (27) we have

$$|\mathcal{A}_o^-[\vec{n}]| = 2n(n-2)!\binom{n}{n_1, n_2, \ldots}. \tag{28}$$

As in the unweighted case, we have

$$|\mathcal{A}_o^-[\vec{n}]| = n|\mathcal{A}_o[\vec{n}]|,$$

and therefore,

$$|\mathcal{A}_o[\vec{n}]| = 2(n-2)!\binom{n}{n_1, n_2, \ldots}. \tag{29}$$

• **Unlabelled case**

Let $\vec{n} = (n_1, n_2, \ldots)$ be a coherent edge degree distribution. In order to compute the number $|\widetilde{\mathcal{A}}_o^-[\vec{n}]|$ of unlabelled $\mathcal{A}_o^-$-structures having $\vec{n}$ as edge degree distribution, we use the fact that given two weighted species F_w and G_v, the generating series $\widetilde{H}(y)$ of unlabelled H-structures, where $H = F_w(G_v)$, is given by

$$\widetilde{H}(y) = Z_{F_w}(\widetilde{G}_v(y), \widetilde{G}_{v^2}(y^2), \widetilde{G}_{v^3}(y^3), \ldots). \tag{30}$$

In the present case, we have $\mathcal{A}_{o,w}^- = r_1 Y + Y C_r(\mathcal{B}_r^2)$, and since the species $\mathcal{B}$ is asymmetric, $\widetilde{\mathcal{B}}_r(y) = \mathcal{B}_r(y)$, hence

$$|\widetilde{\mathcal{A}}_o^-[\vec{n}]| = [r_1^{n_1} r_2^{n_2} \cdots][y^{n-1}] Z_{C_r}(\mathcal{B}_r^2(y), \mathcal{B}_{r^2}^2(y^2), \mathcal{B}_{r^3}^2(y^3), \ldots). \tag{31}$$

But $Z_{C_r}(y_1, y_2, \ldots)$ can be expressed as the following sum:

$$Z_{C_r}(y_1, y_2, \ldots) = \sum_{k \geq 1} \frac{r_k}{k} \sum_{d|k} \phi(d) y_d^{k/d}. \tag{32}$$

Combinatorially speaking, the integer k represents the degree of the root edge. Hence, k may only belong to $\mathrm{Supp}(\vec{n})$, the *support* of $\vec{n}$ which is the set of integers i such that $n_i \neq 0$. So, we have

$$|\widetilde{\mathcal{A}}_o^-[\vec{n}]| = [r_1^{n_1} r_2^{n_2} \cdots][y^{n-1}] \sum_{k \in \mathrm{Supp}(\vec{n})} \frac{r_k}{k} \sum_{d|k} \phi(d) \mathcal{B}_{r^d}^{2k/d}(y^d). \tag{33}$$

First, we compute

$$[y^{n-1}] \mathcal{B}_{r^d}^{2k/d}(y^d) = [y^{(n-1)/d}] \mathcal{B}_{r^d}^{2k/d}(y).$$

From Lagrange inversion, we have

$$\begin{aligned} [y^m] \mathcal{B}_{r^d}^{\ell}(y) &= \frac{1}{m} [t^{m-1}] \frac{d}{dt} \left(t^{\ell}\right) L_{r'^d}^m(t^2), \\ &= \frac{\ell}{m} \sum_{\ell_1, \ell_2, \ldots} \binom{m}{\ell_1, \ell_2, \ldots} r_1^{d\ell_1} r_2^{d\ell_2} \cdots, \end{aligned} \tag{34}$$

where the ℓ_i's satisfy $\sum_i \ell_i = m$ and $\sum_i 2(i-1)\ell_i = m - \ell$. Now, letting $m = (n-1)/d$ and $\ell = 2k/d$, we find

$$|\widetilde{\mathcal{A}}_o^-[\vec{n}]| = [r_1^{n_1} r_2^{n_2} \cdots] \frac{2}{n-1} \sum_{k \in \mathrm{Supp}(\vec{n})} \sum_{d|k} \phi(d) \sum_{\ell_1, \ell_2, \ldots} \binom{(n-1)/d}{\ell_1, \ell_2, \ldots} r_1^{d\ell_1} r_2^{d\ell_2} \cdots r_k^{d\ell_k + 1} \cdots. \tag{35}$$

Finally, we have

Proposition 3. Let $\vec{n}$ be a coherent edge degree distribution, then the number $|\widetilde{\mathcal{A}}_o^-[\vec{n}]|$ of unlabelled oriented solid 2-trees pointed at an edge and having $\vec{n}$ as edge degree distribution is given by

$$|\widetilde{\mathcal{A}}_o^-[\vec{n}]| = \frac{2}{n-1} \sum_{k \in \mathrm{Supp}(\vec{n})} \sum_{d | \{k, \vec{n} - \delta_k\}} \phi(d) \binom{\frac{n-1}{d}}{\frac{\vec{n} - \delta_k}{d}}, \tag{36}$$

where $\frac{\vec{n} - \delta_k}{d} = (\frac{n_1}{d}, \frac{n_2}{d}, \ldots, \frac{n_k - 1}{d}, \ldots)$, for $d \geq 1$ and

$$\binom{\frac{n-1}{d}}{\frac{\vec{n} - \delta_k}{d}} = \binom{\frac{n-1}{d}}{n_1/d, n_2/d, \ldots, (n_k - 1)/d, \ldots}.$$

Let $|\widetilde{\mathcal{A}}_o^{\triangle}[\vec{n}]|$ and $|\widetilde{\mathcal{A}}_o^{\underline{\triangle}}[\vec{n}]|$ be the numbers of unlabelled oriented solid 2-trees pointed respectively at a triangle and at a triangle pointed itself at one of its edge and having $\vec{n}$ as edge degree distribution. We have

Proposition 4. Let $\vec{n}$ be a coherent edge degree distribution, then the numbers $|\widetilde{\mathcal{A}}_o^{\triangle}[\vec{n}]|$ and $|\widetilde{\mathcal{A}}_o^{\underline{\triangle}}[\vec{n}]|$ are given by

$$|\widetilde{\mathcal{A}}_o^{\triangle}[\vec{n}]| = \frac{1}{n}\binom{n}{n_1, n_2, \ldots} + \frac{\chi(3|\vec{n})}{n}\binom{n/3}{n_1/3, n_2/3, \ldots}, \tag{37}$$

$$|\widetilde{\mathcal{A}}_o^{\underline{\triangle}}[\vec{n}]| = \frac{3}{n}\binom{n}{n_1, n_2, \ldots}, \tag{38}$$

where

$$\chi(3|\vec{n}) = \begin{cases} 1, & \text{if all components of } \vec{n} \text{ are multiples of 3,} \\ 0, & \text{otherwise.} \end{cases}$$

Proof. Let us start with $|\widetilde{\mathcal{A}}_o^{\triangle}[\vec{n}]|$. We have

$$\begin{aligned} |\widetilde{\mathcal{A}}_o^{\triangle}[\vec{n}]| &= [r_1^{n_1} r_2^{n_2} \cdots][y^n]\widetilde{\mathcal{A}}_{o,w}^{\triangle}(y), \\ &= [r_1^{n_1} r_2^{n_2} \cdots][y^n] Z_{C_3}(\widetilde{\mathcal{B}}_r(y), \widetilde{\mathcal{B}}_{r^2}(y^2), \ldots), \\ &= [r_1^{n_1} r_2^{n_2} \cdots][y^n] Z_{C_3}(\mathcal{B}_r(y), \mathcal{B}_{r^2}(y^2), \ldots). \end{aligned}$$

Since $Z_{C_3}(y_1, y_2, \ldots) = (y_1^3 + 2y_3)/3$,

$$|\widetilde{\mathcal{A}}_o^{\triangle}[\vec{n}]| = \frac{1}{3}[r_1^{n_1} r_2^{n_2} \cdots][y^n]\left(\mathcal{B}_r^3(y) + 2\mathcal{B}_{r^3}(y^3)\right). \tag{39}$$

From equation (34) letting $m = n$, $\ell = 3$ and $d = 1$, we get

$$[y^n]\mathcal{B}_r^3(y) = \frac{3}{n}\sum_{\ell_1, \ell_2, \ldots}\binom{n}{\ell_1, \ell_2, \ldots} r_1^{\ell_1} r_2^{\ell_2} \cdots, \tag{40}$$

where the ℓ_i's satisfy $\sum_i \ell_i = n$ and $\sum_i 2(i-1)\ell_i = n-3$. Now letting $m = n/3$, $\ell = 1$ and $d = 3$, we get

$$[y^n]\mathcal{B}_{r^3}(y^3) = [y^{n/3}]\mathcal{B}_{r^3}(y) = \frac{3}{n}\sum_{\ell_1, \ell_2, \ldots}\binom{n/3}{\ell_1, \ell_2, \ldots} r_1^{3\ell_1} r_2^{3\ell_2} \cdots, \tag{41}$$

where the ℓ_i's satisfy $\sum_i \ell_i = n$ and $\sum_i 2(i-1)\ell_i = n-1$. Now letting $\ell_i = n_i$ in (40) and $\ell_i = n_i/3$ in (41), we get equation (37). We obtain (38) in a very similar way. ■

Finally, using the dissymmetry theorem, we obtain the final result of this section:

Proposition 5. Let $\vec{n}$ be a coherent edge degree distribution, then the number $|\widetilde{\mathcal{A}}_o[\vec{n}]|$ of unlabelled oriented solid 2-trees having $\vec{n}$ as edge degree distribution is given by
(42)

$$|\widetilde{\mathcal{A}}_o[\vec{n}]| = \frac{2}{n-1}\sum_{k\in\mathrm{Supp}(\vec{n})}\sum_{d|\{k, \vec{n}-\delta_k\}}\phi(d)\binom{\frac{n-1}{d}}{\frac{\vec{n}-\delta_k}{d}} + \frac{\chi(3|\vec{n})}{n}\binom{\frac{n}{3}}{\frac{n_1}{3}, \frac{n_2}{3}, \ldots} - \frac{2}{3n}\binom{n}{n_1, n_2, \ldots},$$

where

$$\chi(3|\vec{n}) = \begin{cases} 1, & \text{if all components of } \vec{n} \text{ are multiples of 3,} \\ 0, & \text{otherwise,} \end{cases}$$

$$\frac{\vec{n}-\delta_k}{d} = \left(\frac{n_1}{d}, \frac{n_2}{d}, \ldots, \frac{n_k - 1}{d}, \ldots\right) \text{ for } d \geq 1,$$

and

$$\binom{\frac{n-1}{d}}{\frac{\vec{n}-\delta_k}{d}} = \binom{\frac{n-1}{d}}{n_1/d, n_2/d, \ldots, (n_k-1)/d, \ldots}.$$

3. Non-oriented solid 2-trees

In order to compute the numbers of labelled and unlabelled solid 2-trees, we use Burnside's Lemma with the group $\mathbb{Z}_2 = \{Id, \tau\}$, where the action of τ is to reverse the orientation of the structures.

3.1. **Enumeration according to the number of edges.**

• **Labelled case**

The labelled case is particularly simple since every labelled oriented 2-tree has exactly two possible orientations except the structure consisting of a single oriented edge. Hence, we have

Proposition 6. The number $|\mathcal{A}[n]|$ of edge labelled solid 2-trees over n edges is given by

$$|\mathcal{A}[n]| = \begin{cases} \frac{1}{2}|\mathcal{A}_o[n]| & \text{if } n > 1; \\ 1 & \text{if } n = 1. \end{cases} \tag{43}$$

Of course, the same argument will remain valid for all other pointed structures discussed in the previous section.

• **Unlabelled case**

In the unlabelled case, the action of τ is not so trivial. Figure 5 shows a structure which is left fixed under the action of τ. Let $\mathcal{A}^-$ be the species of (unoriented) solid 2-trees rooted at an edge. This species can be expressed as the following quotient species (see [4, 5, 6]):

$$\mathcal{A}^- = \frac{\mathcal{A}_o^-}{\mathbb{Z}_2} = \frac{Y + YC(\mathcal{B}^2(Y))}{\mathbb{Z}_2}, \tag{44}$$

where $\mathbb{Z}_2 = \{\mathrm{Id}, \tau\}$ is the two element group consisting of the identity and τ, whose action is to reverse the orientation of the edges. Hence, an unlabelled $\mathcal{A}^-$-structure is an orbit $\{a, \tau \cdot a\}$ under the action of $\mathbb{Z}_2$, where a is any (oriented) unlabelled $\mathcal{A}_o^-$-structure.

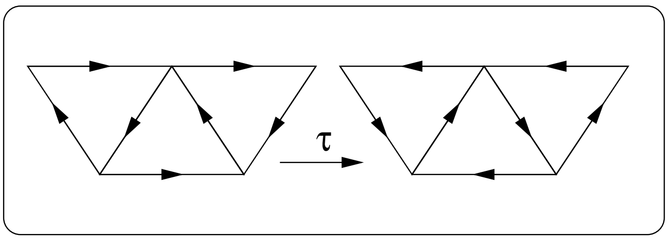


FIGURE 5. An unlabelled 2-tree invariant under the action of τ.

Let us introduce the auxiliary species $\mathcal{B}_{\mathrm{Sym}}$ of τ-symmetric $\mathcal{B}$-structures, *i.e* the species of $\mathcal{B}$-structures left fixed under the edge orientation inversion. Denote by $\widetilde{\mathcal{B}}_{\mathrm{Sym}}(y)$ its ordinary generating series. Recall the functional equation verified by the species $\mathcal{B}$:

$$\mathcal{B} = YL(\mathcal{B}^2).$$

In order to compute $\widetilde{\mathcal{B}}_{\mathrm{Sym}}(y)$, we have to distinguish two cases according to the parity of k, the length of the list of $\mathcal{B}^2$-structures attached to the rooted edge. First consider the case where k is odd (Figure 6 shows an example where $k = 5$). A τ-symmetric $\mathcal{B}$-structure must have a reflective symmetry plane. This plane contains the middle triangle of the list. When an inversion of the orientation of the rooted edge is applied, the two $\mathcal{B}$-structures glued on the two (non root) sides of the middle triangle (structures $\mathcal{B}_5$ and $\mathcal{B}_{5'}$ in Figure 6) are isomorphically exchange. The $k-1$ remaining triangles are exchanged pairwise carrying with them each of their attached $\mathcal{B}$-structures as shown in Figure 6. This gives a factor of $\mathcal{B}^k(y^2)$. We then have to sum the previous expression over all odd values of k. The case where k is even, is very similar except that the symmetry plane must passes between two triangles as shown in Figure 7 and we get the same expression summed over all even values of k. Therefore, we have

$$\widetilde{\mathcal{B}}_{\mathrm{Sym}}(y) = y \sum_{k \geq 0} \mathcal{B}^k(y^2) = \frac{y}{1 - \mathcal{B}(y^2)}. \tag{45}$$

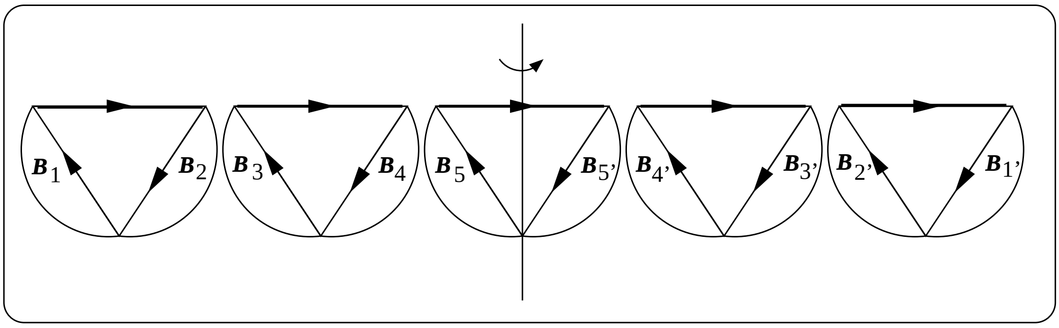


FIGURE 6. A $\mathcal{B}_{\mathrm{Sym}}$-structure, k odd.

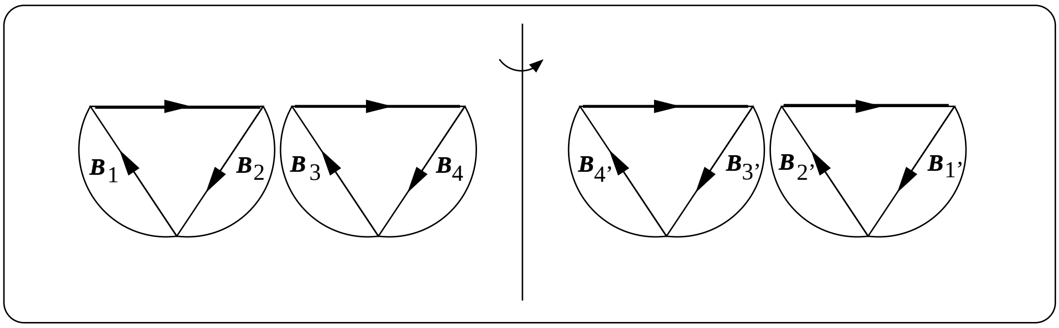


FIGURE 7. A $\mathcal{B}_{\mathrm{Sym}}$-structure, k even.

From expression (45) and another use of Lagrange inversion, we easily obtain the following result.

Proposition 7. The number $|\widetilde{\mathcal{B}}_{\mathrm{sym}}[m]|$ of τ-symmetric unlabelled oriented $\mathcal{B}$-structures is given by

$$|\widetilde{\mathcal{B}}_{\mathrm{sym}}[m]| = \begin{cases} \dfrac{1}{m+1}\dbinom{3m/2}{m} & \text{if } m \text{ is even,} \\[2ex] \dfrac{1}{m}\dbinom{(3m-1)/2}{m+1} + \dfrac{1}{3m}\dbinom{3(m+1)/2}{m+1} & \text{if } m \text{ is odd,} \end{cases} \tag{46}$$

where $m = (n-1)/2$ is the number of triangles and n, the number of edges.

We now give an expression for the generating function of unlabelled quotient structures, which will allow us to enumerate various kind of unlabelled solid 2-trees (see [4], proposition 2.2.4).

Proposition 8. Let F be any (weighted) species and G, a group acting on F. Then the ordinary generating series of the quotient species F/G is given by

$$(F/G)^{\sim}(y) = \frac{1}{|G|}\sum_{g\in G}\sum_{n\geq 0} |\mathrm{Fix}_{\widetilde{F}_n}(g)|_w y^n, \tag{47}$$

where $\mathrm{Fix}_{\widetilde{F}_n}(g)$ denotes the set of unlabelled F-structures left fixed under the action of $g \in G$ and $|\mathrm{Fix}_{\widetilde{F}_n}(g)|_w$ represents the total weight of this set.

Using an unweighted version of Proposition 8 with $F = \mathcal{A}_o^-$ and $G = \mathbb{Z}_2$, we obtain

$$\widetilde{\mathcal{A}}^-(y) = \frac{1}{2}\sum_{n\geq 0} |\mathrm{Fix}_{\widetilde{\mathcal{A}}^-_{o,n}}(\mathrm{Id})|y^n + \frac{1}{2}\sum_{n\geq 0} |\mathrm{Fix}_{\widetilde{\mathcal{A}}^-_{o,n}}(\tau)|y^n, \tag{48}$$

$$= \frac{1}{2}\widetilde{\mathcal{A}}_o^-(y) + \frac{1}{2}\widetilde{\mathcal{B}}_{\mathrm{sym}}(y), \tag{49}$$

since the oriented $\mathcal{A}^-$-structures left fixed under the action of τ have the same generating series as the $\mathcal{B}_{\mathrm{Sym}}$-structures. Then, it becomes easy to extract the coefficient of y^n in relation (49), and we get the number $|\mathcal{A}^-[n]|$ of edge pointed solid 2-trees over n edges

$$|\mathcal{A}^-[n]| = \frac{1}{2}|\widetilde{\mathcal{A}}_o^-[n]| + \frac{1}{2}|\widetilde{\mathcal{B}}_{\mathrm{Sym}}[n]|. \tag{50}$$

We now consider the species $\mathcal{A}^{\vartriangle}$ of triangle rooted solid 2-trees. Since $\mathcal{A}^{\vartriangle} = \mathcal{A}_o^{\vartriangle}/\mathbb{Z}_2$, by virtue of Proposition 8, we have

$$\widetilde{\mathcal{A}}^{\vartriangle}(y) = \frac{1}{2}\sum_{n\geq 0} |\mathrm{Fix}_{\widetilde{\mathcal{A}}^{\vartriangle}_{o,n}}(\mathrm{Id})|y^n + \frac{1}{2}\sum_{n\geq 0} |\mathrm{Fix}_{\widetilde{\mathcal{A}}^{\vartriangle}_{o,n}}(\tau)|y^n, \tag{51}$$

where $|\mathrm{Fix}_{\widetilde{\mathcal{A}}^{\vartriangle}_{o,n}}(\tau)|$, the number of τ-symmetric $\mathcal{A}^{\vartriangle}$-structures over n edges has to be determined. As shown in Figure 8, such a structure must have an axis of symmetry which coincides with one of the root triangle's medians. Since the structure is already considered up to rotation around the root triangle, the choice among the three possible axes is arbitrary. The base side of the triangle must be a $\mathcal{B}_{\mathrm{Sym}}$-structure while the two other sides must be isomorphic copies of the same $\mathcal{B}$-structure. Therefore,

$$\widetilde{\mathcal{A}}^{\vartriangle}(y) = \frac{1}{2}\widetilde{\mathcal{A}}_o^{\vartriangle}(y) + \frac{1}{2}\widetilde{\mathcal{B}}_{\mathrm{Sym}}(y)\mathcal{B}(y^2). \tag{52}$$

In a very similar way, since $\mathcal{A}^{\triangleq} = \mathcal{A}_o^{\triangleq}/\mathbb{Z}_2$, we obtain

$$\widetilde{\mathcal{A}}^{\triangleq}(y) = \frac{1}{2}\widetilde{\mathcal{A}}_o^{\triangleq}(y) + \frac{1}{2}\widetilde{\mathcal{B}}_{\mathrm{Sym}}(y)\mathcal{B}(y^2). \tag{53}$$

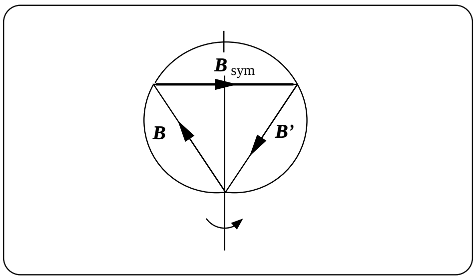


FIGURE 8. A τ-symmetric $\mathcal{A}_o^{\triangle}$-structure.

Finally, using equations (49), (52), (53) and the dissymmetry theorem, we get

Proposition 9. The ordinary generating function of unlabelled solid 2-trees is given by

$$\widetilde{\mathcal{A}}(y) = \frac{1}{2}(\widetilde{\mathcal{A}}_o(y) + \widetilde{\mathcal{B}}_{\mathrm{Sym}}(y)), \tag{54}$$

where $\widetilde{\mathcal{B}}_{\mathrm{Sym}}(y)$ is the ordinary generating series of τ-symmetric oriented $\mathcal{B}$-structures. Consequently, the number $|\widetilde{\mathcal{A}}_t[m]|$ of unoriented solid 2-trees over m triangles is given by

$$|\widetilde{\mathcal{A}}_t[m]| = \frac{1}{2}(|\widetilde{\mathcal{A}}_{o,t}[m]| + |\widetilde{\mathcal{B}}_{\mathrm{sym}}[m]|), \tag{55}$$

where

$$|\widetilde{\mathcal{A}}_{o,t}[m]| = \frac{1}{3m}\sum_{d|m}\phi\left(\frac{m}{d}\right)\binom{3d}{d} + \chi(3|2m+1)\frac{2}{2m+1}\binom{m-1}{\frac{2m-2}{3}} - \frac{2}{3(2m+1)}\binom{3m}{m}.$$

and

$$|\widetilde{\mathcal{B}}_{\mathrm{sym}}[m]| = \begin{cases} \dfrac{1}{m+1}\dbinom{3m/2}{m} & \text{if } m \text{ is even,} \\[2ex] \dfrac{1}{m}\dbinom{(3m-1)/2}{m+1} + \dfrac{1}{3m}\dbinom{3(m+1)/2}{m+1} & \text{if } m \text{ is odd.} \end{cases} \tag{56}$$

To express $|\widetilde{\mathcal{A}}_t[m]|$ in terms of n the number of edges, we only have to set $n := 2m+1$.

3.2. **Enumeration of solid 2-trees according to the edge degree distribution.**

We consider again the weight function defined by

$$\begin{array}{rccc} w: & \mathcal{A}[n] & \longrightarrow & \mathbb{Q}[r_1, r_2, \ldots] \\ & s & \longmapsto & w(s), \end{array} \tag{57}$$

where $r = (r_0, r_1, r_2, \ldots)$ is an infinite set of formal variables and n is any positive integer.

- **Labelled case**

Using the same argument as in the unweighted case, we have

$$|\mathcal{A}[\vec{n}]| = \begin{cases} \frac{1}{2}|\mathcal{A}_o[\vec{n}]| & \text{if } n > 1; \\ 1 & \text{if } n = 1, \end{cases} \tag{58}$$

where $\vec{n}$ is a valid edge degree distribution, n is the number of edges and $|\mathcal{A}[\vec{n}]| = [r_1^{n_1} r_2^{n_2} \cdots][y^n]\mathcal{A}_w(y)$.

- **Unlabelled case**

Using the weighted versions of equations (49), (52) and (53), we get

$$\widetilde{\mathcal{A}}_w^{-}(y) = \frac{1}{2}\widetilde{\mathcal{A}}_{o,w}^{-}(y) + \frac{1}{2}\widetilde{\mathcal{B}}_{\mathrm{sym},w}(y), \tag{59}$$

$$\widetilde{\mathcal{A}}_w^{\vartriangle}(y) = \frac{1}{2}\widetilde{\mathcal{A}}_{o,w}^{\vartriangle}(y) + \frac{1}{2}\widetilde{\mathcal{B}}_{\mathrm{sym},w}(y)\mathcal{B}_w(y^2), \tag{60}$$

$$\widetilde{\mathcal{A}}_w^{\underline{\vartriangle}}(y) = \frac{1}{2}\widetilde{\mathcal{A}}_{o,w}^{\underline{\vartriangle}}(y) + \frac{1}{2}\widetilde{\mathcal{B}}_{\mathrm{sym},w}(y)\mathcal{B}_w(y^2). \tag{61}$$

where $\widetilde{\mathcal{B}}_{\mathrm{sym},w}(y)$ is the ordinary generating series of unlabelled weighted τ-symmetric $\mathcal{B}$-structures. Now applying the dissymmetry theorem leads to

$$\widetilde{\mathcal{A}}(y) = \frac{1}{2}\widetilde{\mathcal{A}}_{o,w}(y) + \frac{1}{2}\widetilde{\mathcal{B}}_{\mathrm{sym},w}(y). \tag{62}$$

The only unknown term in the above equation is $\widetilde{\mathcal{B}}_{\mathrm{sym},w}(y)$. We first establish an additional condition on the vertex degree distribution for an edge rooted oriented solid 2-tree to be τ-symmetric. Since the root edge must remain fixed and all other edges are exchanged pairwise, the edge degree distribution vector $\vec{n}$ must have all its components even except one odd corresponding to the rooted edge.

For an edge degree distribution $\vec{n} = (n_1, n_2, \ldots)$ satisfying the previous condition, and using the fact that $\widetilde{\mathcal{B}}_{\mathrm{sym,w}}(y) = yr_k\mathcal{B}^k(y^2)$, we have

$$|\widetilde{\mathcal{B}}_{\mathrm{sym,w}}[\vec{n}]| = \frac{2k}{n-1}\binom{\frac{n-1}{2}}{\frac{\vec{n}-\delta_k}{2}}, \tag{63}$$

where k is the root edge degree. We now present the final result of this paper.

Proposition 10. Let $\vec{n}$ be a vector satisfying

$$\sum_i n_i = n \quad \text{and} \quad \sum_i in_i = 3m.$$

Then, the number $|\widetilde{\mathcal{A}}[\vec{n}]|$ of (non oriented) unlabelled solid 2-trees having $\vec{n}$ as edge degree distribution is given by

$$|\widetilde{\mathcal{A}}[\vec{n}]| = \frac{1}{2}|\widetilde{\mathcal{A}}_o[\vec{n}]| + \frac{1}{2}|\widetilde{\mathcal{B}}_{\mathrm{sym}}[\vec{n}]|, \tag{64}$$

where

$$|\widetilde{\mathcal{B}}_{\mathrm{sym}}[\vec{n}]| = \begin{cases} \dfrac{2k}{n-1}\dbinom{\frac{n-1}{2}}{\frac{\vec{n}-\delta_k}{2}}, & \text{if } \vec{n} \text{ has a unique odd component,} \\ 0, & \text{otherwise,} \end{cases}$$

δ_k being the vector having 1 at the k^{th} component and 0 everywhere else, and

$$|\widetilde{\mathcal{A}}_o[\vec{n}]| = \frac{2}{n-1}\sum_{k\in\mathrm{Supp}(\vec{n})}\sum_{d|\{k,\vec{n}-\delta_k\}}\phi(d)\binom{\frac{n-1}{d}}{\frac{\vec{n}-\delta_k}{d}} + \frac{\chi(3|\vec{n})}{n}\binom{n/3}{n_1/3, n_2/3, \ldots} - \frac{2}{3n}\binom{n}{n_1, n_2, \ldots}.$$

Appendix.

To conclude this paper, we give here two tables showing the numbers of unlabelled solid 2-trees oriented and unoriented as well as the number of unlabelled τ-symmetric $\mathcal{B}$-structures. The first table gives these numbers according to the number n of edges, and the second, according to edge degree distribution. We use the notation $1^{n_1}2^{n_2}\cdots$, where i^{n_i} means n_i edges of degree i.

n	$\vert\widetilde{\mathcal{A}}_o[n]\vert$	$\vert\widetilde{\mathcal{B}}_{\mathrm{sym}}[n]\vert$	$\vert\widetilde{\mathcal{A}}[n]\vert$
1	1	1	1
3	1	1	1
5	1	1	1
7	2	2	2
9	7	3	5
11	19	7	13
13	86	12	49
15	372	30	201
17	1825	55	940
19	9143	143	4643
21	47801	273	24037

$\vec{n}$	$\vert\widetilde{\mathcal{A}}_{o,w}[\vec{n}]\vert$	$\vert\widetilde{\mathcal{B}}_{\mathrm{sym},w}[\vec{n}]\vert$	$\vert\widetilde{\mathcal{A}}_w[\vec{n}]\vert$
$1^7 2^1 3^1$	2	0	1
$1^8 2^2 3^1$	9	3	6
$1^{12} 2^1 3^1 4^1$	46	0	23
$1^{10} 5^1$	3	1	2
$1^{15} 4^1 5^1$	2	0	1
$1^{16} 3^2 5^1$	17	5	11
$1^{15} 2^2 7^1$	34	0	17

References

[1] L. Beineke and R. Pippert, *The number of labeled k-dimensional trees*, Journal of Combinatorial Theory **6**, 200–205, (1969).

[2] F. Bergeron, G. Labelle, and P. Leroux, *Combinatorial Species and tree-like structures*, Encyclopedia of Mathematics and its Applications, vol. 67, Cambridge University Press, (1998).

[3] M. Bòna, M. Bousquet, G. Labelle and P. Leroux, *Enumeration of m-ary cacti*, Adv. in Appl. Math, **24**, 22–56, (2000).

[4] M. Bousquet, *Espèces de structures et applications au dénombrement de cartes et de cactus planaires*, Thèse de doctorat, UQÀM (1998), Publications du LaCIM, Vol. 24 (1999).

[5] T. Fowler, I. Gessel, G. Labelle, P. Leroux, *Specifying 2-trees*, Proceedings FPSAC'00, Moscow, 26-30 juin 2000, 202-213.

[6] T. Fowler, I. Gessel, G. Labelle, P. Leroux, *The Specification of 2-trees*, Advances in Applied Mathematics, to appear.

[7] F. Harary and E. Palmer, *Graphical Enumeration*, Academic Press, New York, (1973).

[8] G. Labelle, C. Lamathe and P. Leroux, *A classification of plane and planar 2-trees*, submitted to Theorical Computer Science.

[9] G. Labelle and P. Leroux, *Enumeration of (uni- or bicolored) plane trees according to their degree distribution*, Discrete Math. **157**, 227–240, (1996).

[10] E. Palmer and R. Read, *On the Number of Plane 2-trees*, J. London Mathematical Society **6**, 583-592, (1973).

[11] N. J. A. Sloane and S. Plouffe, *The Encyclopedia of Integer Sequences*, Academic Press, San Diego, (1995). `http://www.research.att.com/~njas/sequences`

LACIM, DÉPARTEMENT DE MATHÉMATIQUES, UNIVERSITÉ DU QUÉBEC À MONTRÉAL.
E-mail address: {`bousq2,lamathe`}`@math.uqam.ca`

WALKS IN THE QUARTER PLANE
A FUNCTIONAL EQUATION APPROACH

MIREILLE BOUSQUET-MÉLOU

Abstract. We study planar walks that start from a given point (i_0, j_0), take their steps in a finite set $\mathfrak{S}$, and are confined in the first quadrant $x \geq 0, y \geq 0$. Their enumeration can be attacked in a systematic way: the generating function $Q(x, y; t)$ that counts them by their length (variable t) and the coordinates of their endpoint (variables x, y) satisfies a linear functional equation encoding the step-by-step description of walks. For instance, for the square lattice walks starting from the origin, this equation reads

$$\left(xy - t(x + y + x^2y + xy^2)\right) Q(x, y; t) = xy - xtQ(x, 0; t) - ytQ(0, y; t).$$

The central question addressed in this paper is the *nature* of the series $Q(x, y; t)$. When is it algebraic? When is it D-finite (or holonomic)? Can these properties be derived from the functional equation itself?

Our first result is a new proof of an old theorem due to Kreweras, according to which one of these walk models has, for mysterious reasons, an algebraic generating function. Then, we provide a new proof of a holonomy criterion recently proved by M. Petkovšek and the author. In both cases, we work directly from the functional equation.

Résumé. Considérons les chemins du plan qui partent d'un point (i_0, j_0) donné, choisissent leurs pas dans un ensemble fini $\mathfrak{S}$, et restent confinés dans le premier quadrant $x \geq 0, y \geq 0$. On peut attaquer leur énumération de façon systématique : la série génératrice $Q(x, y; t)$ qui les énumère selon la longueur (variable t) et le point d'arrivée (variables x, y) est toujours solution d'une équation fonctionnelle linéaire qui code la construction pas à pas de ces chemins. Par exemple, l'équation régissant les chemins de la grille carrée qui, issus de l'origine, restent dans le premier quadrant, est donnée ci-dessus.

La question centrale qui motive cet article porte sur le *nature* de la série $Q(x, y; t)$. Quand est-elle algébrique ? holonome ? Comment déduire ces propriétés de l'équation fonctionnelle qu'elle satisfait ?

Notre premier résultat est une nouvelle preuve d'un vieux théorème dû à Kreweras, qui affirme qu'un de ces modèles de chemins a, pour des raisons mystérieuses, une série génératrice algébrique. Ensuite, nous donnons une nouvelle preuve d'un critère d'holonomie récemment démontré par M. Petkovšek et l'auteur. Dans les deux cas, le point de départ est l'équation fonctionnelle qui rgit la srie $Q(x, y; t)$.

1. Walks in the quarter plane

The enumeration of lattice walks is one of the most venerable topics in enumerative combinatorics, which has numerous applications in probabilities [15, 27, 36]. These walks take their steps in a finite subset $\mathfrak{S}$ of $\mathbb{Z}^d$, and might be constrained in various ways. One can only cite a small percentage of the relevant litterature, which dates back at least to the next-to-last century [1, 18, 24, 30, 31]. Many recent publications show that the topic is still active [4, 6, 11, 20, 22, 32, 33].

After the solution of many explicit problems, certain patterns have emerged, and a more recent trend consists in developping methods that are valid for generic sets of steps. A special attention is being paid to the *nature* of the generating function of the walks under consideration. For instance, the generating function for unconstrained walks on the line $\mathbb{Z}$ is rational, while the generating function for walks constrained to stay in the half-line $\mathbb{N}$ is always algebraic [3]. This result has often been described in terms of *partially directed*

2-dimensional walks confined in a quadrant (or *generalized Dyck walks* [13, 19, 25, 26]), but is, essentially, of a 1-dimensional nature.

Similar questions can be addressed for *real* 2-dimensional walks. Again, the generating function for unconstrained walks starting from a given point is clearly rational. Moreover, the argument used for 1-dimensional walks confined in $\mathbb{N}$ can be recycled to prove that the generating function for the walks that stay in the half-plane $x \geq 0$ is always algebraic. What about doubly-restricted walks, that is, walks that are confined in the quadrant $x \geq 0, y \geq 0$?

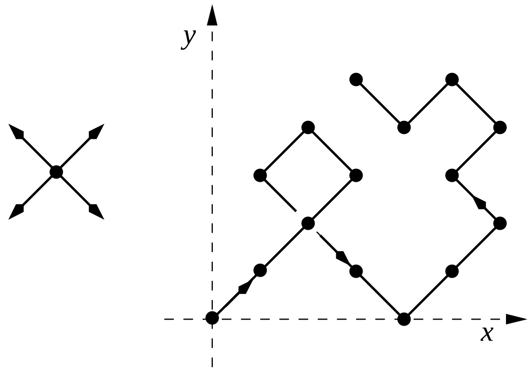


FIGURE 1. A walk on the diagonal square lattice confined in the first quadrant.

A rapid inspection of the most standard cases suggests that these walks might have always a D-finite generating function[1]. The simplest example is probably that of the diagonal square lattice, where the steps are North-East, South-East, North-West and South-West (Figure 1): by projecting the walks on the x- and y-axes, we obtain two decoupled prefixes of Dyck paths, so that the length generating function for walks that start from the origin and stay in the first quadrant is

$$\sum_{n\geq 0} \binom{n}{\lfloor n/2 \rfloor}^2 t^n,$$

a D-finite series. For the ordinary square lattice (with North, East, South and West steps), the generating function is

$$\sum_{m,n\geq 0} \binom{m+n}{m} \binom{m}{\lfloor m/2 \rfloor} \binom{n}{\lfloor n/2 \rfloor} t^{m+n} = \sum_{n\geq 0} \binom{n}{\lfloor n/2 \rfloor} \binom{n+1}{\lceil n/2 \rceil} t^n,$$

another D-finite series. The first expression comes from the fact that these walks are shuffles of two prefixes of Dyck walks, and the Chu-Vandermonde identity transforms it into the second simpler expression.

In both cases, the number of n-step walks grows asymptotically like $4^n/n$, which prevents the generating function from being algebraic (see [16] for the possible asymptotic behaviours of coefficients of algebraic series).

The two above results can be refined by taking into account the coordinates of the endpoint: if $a_{i,j}(n)$ denotes the number of n-step walks of length n ending at (i, j), then

[1]A series $F(t)$ is D-finite (or *holonomic*) if it satisfies a linear differential equation with polynomial coefficients in t. Any algebraic series is D-finite.

we have, for the diagonal square lattice:

$$\sum_{i,j,n\geq 0} a_{i,j}(n)x^iy^jt^n = \sum_{i,j,n\geq 0} \frac{(i+1)(j+1)}{(n+1)^2}\binom{n+1}{\frac{n-i}{2}}\binom{n+1}{\frac{n-j}{2}}x^iy^jt^n,$$

where the binomial coefficient $\binom{n}{(n-i)/2}$ is zero unless $0 \leq i \leq n$ and $i \equiv n \bmod 2$. Similarly, for the ordinary square lattice,

$$\sum_{i,j,n\geq 0} a_{i,j}(n)x^iy^jt^n = \sum_{i,j,n\geq 0} \frac{(i+1)(j+1)}{(n+1)(n+2)}\binom{n+2}{\frac{n+i-j+2}{2}}\binom{n+2}{\frac{n-i-j}{2}}x^iy^jt^n. \tag{1}$$

These two series can be seen to be D-finite in their three variables.

This holonomy, however, is not the rule: as proved in [10], walks that start from $(1,1)$, take their steps in $\mathfrak{S} = \{(2,-1),(-1,2)\}$ and always stay in the first quadrant have a non-D-finite length generating function. The same holds for the subclass of walks ending on the x-axis.

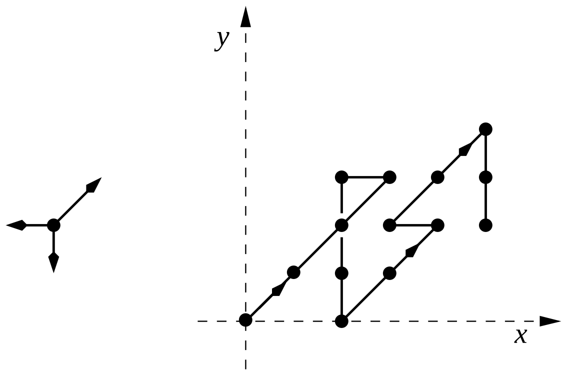


FIGURE 2. Kreweras' walks in a quadrant.

At the other end of the hierarchy, another walk model displays a mysteriously simple algebraic generating function: when the starting point is $(0,0)$, and the allowed steps South, West and North-East (Figure 2), the number of walks of length $3n+2i$ ending at the point $(i,0)$ is

$$\frac{4^n(2i+1)}{(n+i+1)(2n+2i+1)}\binom{2i}{i}\binom{3n+2i}{n}. \tag{2}$$

This result was first proved by Kreweras in 1965 [24, Chap. 3], and then rederived by Niederhausen [32] and Gessel [18]. It is, however, not well-understood, for two reasons:

– no direct proof of (2) is known, even when $i=0$. The number of walks ending at the origin is closely related to the number of non-separable planar maps, to the number of *cubic* non-separable maps [34, 35, 38, 39], and to the number of two-stack sortable permutations [5, 40, 41]. All available proofs of (2) are rather long and complicated. Moreover, in all of them, the result is *checked* rather than *derived*.

– most importantly, the three-variate generating function for these walks can be shown to be algebraic [18], but none of the proofs explain combinatorially this algebraicity.

All problems of walks confined in a quadrant can be attacked by writing a functional equation for their three-variate generating function, and it is this uniform approach that we discuss here. This functional equation simply encodes the step-by-step construction of the walks. For instance, for square lattice walks, we can write

$$\begin{aligned} Q(x,y;t) &:= \sum_{i,j,n\geq 0} a_{i,j}(n)x^i y^j t^n \\ &= 1 + t(x+y)Q(x,y;t) + t\,\frac{Q(x,y;t)-Q(0,y;t)}{x} + t\,\frac{Q(x,y;t)-Q(x,0;t)}{y}, \end{aligned}$$

that is,

$$\big(xy - t(x+y+x^2y+xy^2)\big)\, Q(x,y;t) = xy - xtQ(x,0;t) - ytQ(0,y;t), \tag{3}$$

and the solution of this equation, given by (1), is D-finite (but transcendental). Similarly, for the diagonal square lattice, we have

$$\big(xy - t(1+x^2)(1+y^2)\big)\, Q(x,y;t) = xy - t(1+x^2)Q(x,0;t) - t(1+y^2)Q(0,y;t) + tQ(0,0;t),$$

with again a D-finite transcendental solution, while for Kreweras' algebraic model, we obtain

$$\big(xy - t(x+y+x^2y^2)\big)\, Q(x,y;t) = xy - xtQ(x,0;t) - ytQ(0,y;t). \tag{4}$$

Finally, the equation that rules the non-holonomic model of [10] is

$$\big(xy - t(x^3+y^3)\big)\, Q(x,y;t) = x^2y^2 - tx^3Q(x,0;t) - ty^3Q(0,y;t).$$

The general theme of this paper is the following: the above equations completely solve, in some sense, the problem of enumerating the walks. But they are not the kind of solution one likes, especially if the numbers are simple, or if the generating function is actually algebraic! How can one derive these simple solutions from the functional equations? And what is the essential difference between, say, Eqs. (3) and (4), that makes one series transcendental, and the other algebraic?

We shall answer some of these questions. Our main result is a new proof of (2), which we believe to be simpler that the three previous ones. It has, at least, one nice feature: we *derive the algebraicity from the equation* without having to guess the formula first. Then, we give a new proof of a (refinement of) a holonomy criterion that was proved combinatorially in [10]: if the set of steps $\mathfrak{S}$ is symmetric with respect to the y-axis and satisfies a *small horizontal variations* condition, then the generating function for walks with steps in $\mathfrak{S}$, starting from any given point (i_0, j_0), is D-finite. This result covers the two above D-finite transcendental cases, but not Kreweras' model... We finally survey some perspectives of this work.

Let us conclude this section with a few more formal definitions on walks and power series.

Let $\mathfrak{S}$ be a finite subset of $\mathbb{Z}^2$. A walk with steps in $\mathfrak{S}$ is a finite sequence $w = (w_0, w_1, \ldots, w_n)$ of vertices of $\mathbb{Z}^2$ such that $w_i - w_{i-1} \in \mathfrak{S}$ for $1 \leq i \leq n$. The number of steps, n, is the *length* of w. The starting point of w is w_0, and its endpoint is w_n. The *complete generating function* for a set $\mathfrak{A}$ of walks starting from a given point $w_0 = (i_0, j_0)$ is the series

$$A(x,y;t) = \sum_{n\geq 0} t^n \sum_{i,j\in\mathbb{Z}} a_{i,j}(n) x^i y^j,$$

where $a_{i,j}(n)$ is the number of walks of $\mathfrak{A}$ that have length n and end at (i,j). This series is a formal power series in t whose coefficients are polynomials in $x, y, 1/x, 1/y$. We shall often denote $\bar{x} = 1/x$ and $\bar{y} = 1/y$.

Given a ring $\mathbb{L}$ and k indeterminates $x_1, \ldots, x_k$, we denote by $\mathbb{L}[x_1, \ldots, x_k]$ the ring of polynomials in $x_1, \ldots, x_k$ with coefficients in $\mathbb{L}$, and by $\mathbb{L}[[x_1, \ldots, x_k]]$ the ring of formal power series in $x_1, \ldots, x_k$ with coefficients in $\mathbb{L}$. If $\mathbb{L}$ is a field, we denote by $\mathbb{L}(x_1, \ldots, x_k)$ the field of rational functions in $x_1, \ldots, x_k$ with coefficients in $\mathbb{L}$.

Assume $\mathbb{L}$ is a field. A series F in $\mathbb{L}[[x_1, \ldots, x_k]]$ is *rational* if there exist polynomials P and Q in $\mathbb{L}[x_1, \ldots, x_k]$, with $Q \neq 0$, such that $QF = P$. It is *algebraic* (over the field $\mathbb{L}(x_1, \ldots, x_k)$) if there exists a non-trivial polynomial P with coefficients in $\mathbb{L}$ such that $P(F, x_1, \ldots, x_k) = 0$. The sum and product of algebraic series is algebraic.

The series F is *D-finite* (or *holonomic*) if the partial derivatives of F span a finite dimensional vector space over the field $\mathbb{L}(x_1, \ldots, x_k)$ (this vector space is a subspace of the fraction field of $\mathbb{L}[[x_1, \ldots, x_k]]$); see [37] for the one-variable case, and [28, 29] otherwise. In other words, for $1 \leq i \leq k$, the series F satisfies a non-trivial partial differential equation of the form

$$\sum_{\ell=0}^{d_i} P_{\ell,i} \frac{\partial^\ell F}{\partial x_i^\ell} = 0,$$

where $P_{\ell,i}$ is a polynomial in the x_j. Any algebraic series is holonomic. The sum and product of two holonomic series is still holonomic. The specializations of an holonomic series (obtained by giving values from $\mathbb{L}$ to some of the variables) are holonomic, if well-defined. Moreover, if F is an *algebraic* series and $G(t)$ is a holonomic series of one variable, then the substitution $G(F)$ (if well-defined) is holonomic [29, Prop. 2.3].

2. A new proof of Kreweras' result

Consider walks that start from $(0,0)$, are made of South, West and North-East steps, and always stay in the first quadrant (Figure 2). Let $a_{i,j}(n)$ be the number of n-step walks of this type ending at (i,j). We denote by $Q(x, y; t)$ the complete generating function of these walks:

$$Q(x, y; t) := \sum_{i,j,n \geq 0} a_{i,j}(n) x^i y^j t^n.$$

Constructing the walks step by step yields the following equation:

$$\big(xy - t(x + y + x^2y^2)\big)\, Q(x, y; t) = xy - xtQ(x, 0; t) - ytQ(0, y; t). \tag{5}$$

We shall often denote, for short, $Q(x, y; t)$ by $Q(x, y)$. Let us also denote the series $xtQ(x, 0; t)$ by $R(x; t)$ or even $R(x)$. Using the symmetry of the problem in x and y, the above equation becomes:

$$\big(xy - t(x + y + x^2y^2)\big)\, Q(x, y) = xy - R(x) - R(y). \tag{6}$$

This equation is equivalent to a recurrence relation defining the numbers $a_{i,j}(n)$ by induction on n. Hence, it defines completely the series $Q(x, y; t)$. Still, the characterization of this series we have in mind is of a different nature:

Theorem 1. *Let $X \equiv X(t)$ be the power series in t defined by*

$$X = t(2 + X^3).$$

Then the generating function for Kreweras' walks ending on the x-axis is

$$Q(x, 0; t) = \frac{1}{tx}\left(\frac{1}{2t} - \frac{1}{x} - \left(\frac{1}{X} - \frac{1}{x}\right)\sqrt{1 - xX^2}\right).$$

Consequently, the length generating function for walks ending at $(i, 0)$ is

$$[x^i]Q(x, 0; t) = \frac{X^{2i+1}}{2.4^i\ t}\left(C_i - \frac{C_{i+1}X^3}{4}\right),$$

where $C_i = \binom{2i}{i}/(i+1)$ is the i-th Catalan number. The Lagrange inversion formula gives the number of such walks of lenth $3n+2i$ as

$$a_{i,0}(3n+2i) = \frac{4^n(2i+1)}{(n+i+1)(2n+2i+1)}\binom{2i}{i}\binom{3n+2i}{n}.$$

The aim of this section is to derive Theorem 1 from the functional equation (5).

2.1. **The obstinate kernel method.** The kernel method is basically the only tool we have to attack Equation (6). This method had been around since, at least, the 70's, and is currently the subject of a certain rebirth (see the references in [2, 3, 9]). It consists in coupling the variables x and y so as to cancel the *kernel* $K(x,y) = xy - t(x+y+x^2y^2)$. This should give the "missing" information about the series $R(x)$.

As a polynomial in y, this kernel has two roots

$$\begin{aligned} Y_0(x) &= \frac{1-t\bar{x}-\sqrt{(1-t\bar{x})^2-4t^2x}}{2tx} = t+\bar{x}t^2+O(t^3), \\ Y_1(x) &= \frac{1-t\bar{x}+\sqrt{(1-t\bar{x})^2-4t^2x}}{2tx} = \frac{\bar{x}}{t}-\bar{x}^2 - t-\bar{x}t^2+O(t^3). \end{aligned}$$

The elementary symmetric functions of the Y_i are

$$Y_0+Y_1 = \frac{\bar{x}}{t}-\bar{x}^2 \quad \text{and} \quad Y_0Y_1 = \bar{x}. \tag{7}$$

The fact that they are polynomials in $\bar{x} = 1/x$ will play a very important role below.

Only the first root can be substituted for y in (6) (the term $Q(x, Y_1; t)$ is not a well-defined power series in t). We thus obtain a functional equation for $R(x)$:

$$R(x) + R(Y_0) = xY_0. \tag{8}$$

It can be shown that this equation – once restated in terms of $Q(x, 0)$ – defines uniquely $Q(x, 0; t)$ as a formal power series in t with polynomial coefficients in x. Equation (8) is the standard result of the kernel method.

Still, we want to apply here the *obstinate* kernel method. That is, we shall not content ourselves with Eq. (8), but we shall go on producing pairs (X, Y) that cancel the kernel and use the information they provide on the series $R(x)$.

Let $(X, Y) \neq (0, 0)$ be a pair of Laurent series in t with coefficients in a field $\mathbb{K}$ such that $K(X, Y) = 0$. We define $\Phi(X, Y) = (X', Y)$, where $X' = (XY)^{-1}$ is *the other solution* of $K(x, Y) = 0$, seen as a polynomial in x. Similarly, we define $\Psi(X, Y) = (X, Y')$, where $Y' = (XY)^{-1}$ is the other solution of $K(X, y) = 0$. Note that Φ and Ψ are involutions. Let us examine their action on the pair (x, Y_0). We obtain the following diagram[2]

All these pairs of power series cancel the kernel, and we have framed the ones that can be legally substituted[3] in the main functional equation (6). We thus obtain *two* equations for the unknown series $R(x)$:

$$R(x) + R(Y_0) = xY_0, \tag{9}$$

$$R(Y_0) + R(Y_1) = Y_0Y_1 = \bar{x}. \tag{10}$$

This obstinate kernel method is the first original ingredient of this proof.

Remark. Let p, q, r be three nonnegative numbers such that $p + q + r = 1$. Take $x = (pr)^{1/3}q^{-2/3}$, $y = (qr)^{1/3}p^{-2/3}$, and $t = (pqr)^{1/3}$. Then $K(x, y; t) = 0$, so that $R(x)+R(y) =$

[2]The transformations Φ and Ψ are related to, but distinct from, the transformations ξ and η of [14, Sect. 2.4]. The underlying idea – finding "all" pairs of roots of the kernel – is the same.

[3]For this choice of steps, the series $Q(Y_0, Y_1; t)$ and $Q(0, Y_1; t)$ *are* well-defined!

$$\begin{tikzcd}
 & \boxed{(x, Y_0)} \arrow[dl, "\Phi"'] \arrow[dr, "\Psi"] & \\
\boxed{(Y_1, Y_0)} \arrow[d, "\Psi"'] & & (x, Y_1) \arrow[d, "\Phi"] \\
(Y_1, x) \arrow[dr, "\Phi"'] & & \boxed{(Y_0, Y_1)} \arrow[dl, "\Psi"] \\
 & \boxed{(Y_0, x)} &
\end{tikzcd}$$

xy. This equation can be given a probabilistic interpretation by considering random walks that make a North-East step with (small) probability r and a South (resp. West) step with probability p (resp. q). This probabilistic argument, and the equation it implies, is the starting point of Gessel's solution of Kreweras problem [18, Eq. (21)].

2.2. **Symmetric functions of Y_0 and Y_1.** After the kernel method, the next tool of our approach is the extraction of the positive part of power series. More precisely, let $S(x;t)$ be a power series in t whose coefficients are Laurent polynomials in x:

$$S(x;t) = \sum_{n\geq 0} t^n \sum_{i\in\mathbb{Z}} s_i(n) x^i t^n,$$

where for each $n \geq 0$, only finitely many coefficients $s_i(n)$ are non-zero. We define the positive part of this series by

$$S^+(x;t) := \sum_{n\geq 0} t^n \sum_{i\in\mathbb{N}} s_i(n) x^i t^n.$$

This is where the values of the symmetric functions of Y_0 and Y_1 become crucial: the fact that they only involve negative powers of x (see (7)) will simplify the extraction of the positive part of certain equations. This observation is the second original ingredient of this proof.

Lemma 2. *Let $F(u, v; t)$ be a power series in t with coefficients in $\mathbb{C}[u, v]$, such that $F(u, v; t) = F(v, u; t)$. Then the series $F(Y_0, Y_1; t)$, if well-defined, is a power series in t with polynomial coefficients in $\bar{x}$. Moreover, the constant term of this series, taken with respect to $\bar{x}$, is $F(0, 0; t)$.*

Proof. All symmetric polynomials of u and v are polynomials in $u+v$ and uv with complex coefficients.

■

We now want to form a symmetric function of Y_0 and Y_1, starting from the equations (9–10). The first one reads

$$R(Y_0) - xY_0 = -R(x).$$

By combining both equations, we then obtain the companion expression:

$$R(Y_1) - xY_1 = R(x) + 2\bar{x} - 1/t.$$

Taking the product[4] of these two equations gives

$$(R(Y_0) - xY_0)(R(Y_1) - xY_1) = -R(x)(R(x) + 2\bar{x} - 1/t).$$

[4]An alternative derivation of Kreweras' result, obtained by considering the divided difference $(R(Y_0) - xY_0 - R(Y_1) + xY_1)/(Y_0 - Y_1)$, will be discussed on the complete version of this paper.

The extraction of the positive part of this identity is made possible by Lemma 2. Given that $R(x;t) = xtQ(x,0;t)$, one obtains:

$$x = -t^2x^2Q(x,0)^2 + (x-2t)Q(x,0) + 2tQ(0,0),$$

that is,

$$t^2x^2Q(x,0)^2 + (2t-x)Q(x,0) - 2tQ(0,0) + x = 0. \tag{11}$$

2.3. **The quadratic method.** Equation (11) – which begs for a combinatorial explanation – is typical of the equations obtained when enumerating planar maps, and the rest of the proof will be routine to all maps lovers. This equation can be solved using the so-called *quadratic method*, which was first invented by Brown [12]. The formulation we use here is different both from Brown's original presentation and from the one in Goulden and Jackson's book [21]. This new formulation is convenient for generalizing the method to equations of higher degree with more unknowns [8].

Equation (11) can be written as

$$P(Q(x), Q(0), t, x) = 0, \tag{12}$$

where $P(u,v,t,x) = t^2x^2u^2 + (2t-x)u - 2tv + x$, and $Q(x,0)$ has been abbreviated in $Q(x)$. Differentiating this equation with respect to x, we find

$$\frac{\partial P}{\partial u}(Q(x), Q(0), t, x)\frac{\partial Q}{\partial x}(x) + \frac{\partial P}{\partial x}(Q(x), Q(0), t, x) = 0.$$

Hence, if there exists a power series in t, denoted $X(t) \equiv X$, such that

$$\frac{\partial P}{\partial u}(Q(X), Q(0), t, X) = 0, \tag{13}$$

then one also has

$$\frac{\partial P}{\partial x}(Q(X), Q(0), t, X) = 0, \tag{14}$$

and we thus obtain a system of three polynomial equations, namely Eq. (12) written for $x = X$, Eqs. (13) and (14), that relate the three unknown series $Q(X)$, $Q(0)$ and X. This puts us in a good position to write an algebraic equation defining $Q(0) = Q(0,0;t)$.

Let us now work out the details of this program: Eq. (13) reads $X = 2t^2X^2Q(X) + 2t$, and since the right-hand side is a multiple of t, it should be clear that this equation defines a unique power series $X(t)$. The system of three equations reads

$$\begin{cases} t^2X^2Q(X)^2 + (2t-X)Q(X) - 2tQ(0) + X = 0, \\ 2t^2X^2Q(X) + 2t - X = 0, \\ 2t^2XQ(X)^2 - Q(X) + 1 = 0. \end{cases}$$

Eliminating $Q(X)$ between the last two equations yields $X = t(2 + X^3)$, so that the series X is the parameter introduced in Theorem 1. Going on with the elimination, we finally obtain

$$Q(0,0;t) = \frac{X}{2t}\left(1 - \frac{X^3}{4}\right),$$

and the expression of $Q(x,0;t)$ follows from (11).

∎

3. A holonomy criterion

Using functional equations, we can recover, and actually refine, a holonomy criterion that was recently proved combinatorially [10]. Let $\mathfrak{S}$ be a finite subset of $\mathbb{Z}^2$. We say that $\mathfrak{S}$ is symmetric with respect to the y-axis if

$$(i,j) \in \mathfrak{S} \Rightarrow (-i,j) \in \mathfrak{S}.$$

We say that $\mathfrak{S}$ has small horizontal variations if

$$(i,j) \in \mathfrak{S} \Rightarrow |i| \leq 1.$$

The usual square lattice steps satisfy these two conditions. So do the steps of the diagonal square lattice (Figure 1).

Theorem 3. *Let $\mathfrak{S}$ be a finite subset of $\mathbb{Z}^2$ that is symmetric with respect to the y-axis and has small horizontal variations. Let $(i_0, j_0) \in \mathbb{N}^2$. Then the complete generating function $Q(x,y;t)$ for walks that start from (i_0,j_0), take their steps in $\mathfrak{S}$ and stay in the first quadrant is D-finite.*

A combinatorial argument proving the holonomy of $Q(1,1;t)$ is presented in [10].

3.1. **Example.** Before we embark on the proof of this theorem, let us see the principle of the proof at work on a simple example: square lattice walks confined in a quadrant. The functional equation satisfied by their complete generating function is

$$\left(xy - t(x+y+x^2y+xy^2)\right) Q(x,y) = xy - xtQ(x,0) - ytQ(0,y) = xy - R(x) - R(y), \tag{15}$$

where, as in Kreweras' example, we denote by $R(x)$ the series $txQ(x,0)$. The kernel $K(x,y) = xy - t(x+y+x^2y+xy^2)$, considered as a polynomial in y, has two roots:

$$\begin{aligned}
Y_0(x) &= \frac{1-t(x+\bar{x}) - \sqrt{(1-t(x+\bar{x}))^2 - 4t^2}}{2t} &&= t + (x+\bar{x})t^2 + O(t^3),\\
Y_1(x) &= \frac{1-t(x+\bar{x}) + \sqrt{(1-t(x+\bar{x}))^2 - 4t^2}}{2t} &&= \frac{1}{t} - x - \bar{x} \quad - \quad t - (x+\bar{x})t^2 + O(t^3).
\end{aligned}$$

The elementary symmetric functions of the Y_i are

$$Y_0 + Y_1 = \frac{1}{t} - x - \bar{x} \quad \text{and} \quad Y_0 Y_1 = 1.$$

Observe that they are no longer polynomials in $\bar{x} = 1/x$.

If, as above, we apply to the pair (x, Y_0) the transformations Φ and Ψ, we obtain a very simple diagram:

$$\begin{tikzcd}
& \boxed{(x,Y_0)} \arrow[dl, "\Phi"'] \arrow[dr, "\Psi"] & \\
\boxed{(\bar{x},Y_0)} \arrow[dr, "\Psi"'] & & (x,Y_1) \arrow[dl, "\Phi"] \\
& (\bar{x},Y_1) &
\end{tikzcd}$$

From the two pairs that can be substituted for (x,y) in Equation (15), we derive the following system:

$$\begin{aligned}
R(x) + R(Y_0) &= xY_0,\\
R(\bar{x}) + R(Y_0) &= \bar{x}Y_0.
\end{aligned}$$

From here, the method has to diverge from what we did in Kreweras' case. Eliminating $R(Y_0)$ between the two equations gives

$$R(x) - R(\bar{x}) = (x - \bar{x})Y_0. \tag{16}$$

Since $R(0) = 0$, extracting the positive part of this identity gives $R(x)$ as *the positive part of an algebraic series*. It is known that the positive part of a D-finite series is always D-finite [28]. In particular, the series $R(x)$ is D-finite. The same holds for $Q(x, 0)$, and, by (15), for $Q(x, y)$.

This argument is enough for proving the holonomy of the series, but, given the simplicity of this model, we can proceed with explicit calculations. Given the polynomial equation defining Y_0,

$$Y_0 = t(1 + \bar{x}Y_0 + xY_0 + Y_0^2) = t(1 + \bar{x}Y_0)(1 + xY_0),$$

the Lagrange inversion formula yields the following expression for Y_0:

$$Y_0 = \sum_{m \ge 0} \sum_{i \in \mathbb{Z}} \frac{x^i \, t^{2m+|i|+1}}{2m + |i| + 1} \binom{2m + |i| + 1}{m + |i|} \binom{2m + |i| + 1}{m}.$$

Since $R(0) = 0$, extracting the positive part in the identity (16) now gives, after some reductions,

$$R(x) = txQ(x, 0) = \sum_{m \ge 0} \sum_{i \ge 0} \frac{x^{i+1} t^{2m+i+1} (i + 1)}{(2m + i + 1)(2m + i + 2)} \binom{2m + i + 2}{m + i + 1} \binom{2m + i + 2}{m}.$$

This naturally fits with the general expression (1).

3.2. **Proof of Theorem 3.** We define two Laurent polynomials in y by

$$P_0(y) := \sum_{(0,j) \in \mathfrak{S}} y^j \quad \text{and} \quad P_1(y) := \sum_{(1,j) \in \mathfrak{S}} y^j.$$

Let $-p$ be the largest down move; more precisely,

$$p = \max(0, \{-j : (i, j) \in \mathfrak{S} \text{ for some } i\}).$$

The functional equation obtained by constructing walks step-by-step reads:

$$K(x, y)Q(x, y) = x^{1+i_0} y^{p+j_0} - ty^p P_1(y)Q(0, y) - t \sum_{(i,-j) \in \mathfrak{S}} \sum_{m=0}^{j-1} \left(Q_m(x) - \delta_{i,1} Q_m(0)\right) x^{1-i} y^{p+m-j} \tag{17}$$

where

$$K(x, y) = xy^p \left(1 - tP_0(y) - t(x + \bar{x})P_1(y)\right)$$

is the kernel of the equation, and $Q_m(x)$ stand for the coefficient of y^m in $Q(x, y)$. All the series involved in this equation also depend on the variable t, but it is omitted for the sake of briefness. For instance, $K(x, y)$ stands for $K(x, y; t)$.

As above, we shall use the kernel method – plus another argument – to solve the above functional equation. The polynomial $K(x, y)$, seen as a polynomial in y, admits a number of roots, which are Puiseux series in t with coefficients in an algebraic closure of $\mathbb{Q}(x)$. Moreover, all these roots are distinct. As $K(x, y; 0) = xy^p$, exactly p of these roots, say $Y_1, \ldots, Y_p$, vanish at $t = 0$. This property guarantees that these p series can be substituted for y in (17), which yields

$$x^{1+i_0} Y^{p+j_0} = tY^p P_1(Y)Q(0, Y) + t \sum_{(i,-j) \in \mathfrak{S}} \sum_{m=0}^{j-1} \left(Q_m(x) - \delta_{i,1} Q_m(0)\right) x^{1-i} \, Y^{p+m-j} \tag{18}$$

for any $Y = Y_1, \ldots, Y_p$.

Given the symmetry of K in x and $\bar{x}$, each of the Y_i is invariant by the transformation $x \to 1/x$. Replacing x by $\bar{x}$ in the above equation gives, for any $Y = Y_1, \ldots, Y_p$,

$$\bar{x}^{1+i_0} Y^{p+j_0} = tY^p P_1(Y) Q(0,Y) + t \sum_{(i,-j)\in\mathfrak{S}} \sum_{m=0}^{j-1} \left(Q_m(\bar{x}) - \delta_{i,1} Q_m(0)\right) x^{i-1} Y^{p+m-j}. \tag{19}$$

We now combine (18) and (19) to eliminate $Q(0,Y)$:

$$(x^{1+i_0} - \bar{x}^{1+i_0}) Y^{p+j_0} = t \sum_{(i,-j)\in\mathfrak{S}} \sum_{m=0}^{j-1} \left(x^{1-i} Q_m(x) - x^{i-1} Q_m(\bar{x})\right) Y^{p+m-j}$$

for any $Y = Y_1, \ldots, Y_p$. This is the generalization of Eq. (16). The right-hand side of the above equation is a polynomial P in Y, of degree at most $p-1$. We know its value at p points, namely $Y_1, \ldots, Y_p$. The Lagrange interpolation formula implies that these p values completely determine the polynomial. As the left-hand side of the equation is algebraic, then each of the coefficients of P is also algebraic. That is,

$$t \sum_{(i,-j)\in\mathfrak{S}} \sum_{m=0}^{j-1} \left(x^{1-i} Q_m(x) - x^{i-1} Q_m(\bar{x})\right) y^{p+m-j} = \sum_{m=0}^{p-1} A_m(x) y^m,$$

where each of the A_m is an algebraic series. Let us extract the positive part of this identity. Given that i can only be 0, 1 or -1, we obtain

$$t \sum_{(i,-j)\in\mathfrak{S}} \sum_{m=0}^{j-1} \left(x^{1-i} Q_m(x) - \delta_{i,1} Q_m(0)\right) y^{p+m-j} = \sum_{m=0}^{p-1} H_m(x) y^m$$

where $H_m(x) := A_m^+(x)$ is the positive part of $A_m(x)$. Again, this series can be shown to be D-finite. Going back to the original functional equation (17), this gives

$$K(x,y)Q(x,y) = x^{1+i_0} y^{p+j_0} - ty^p P_1(y) Q(0,y) - \sum_{m=0}^{p-1} H_m(x) y^m.$$

Let us finally[5] consider the kernel $K(x,y)$ as a polynomial in x. One of its roots, denoted below X, is a formal power series in t that vanishes at $t = 0$. Replacing x by this root allows us to express $Q(0,y)$ as a D-finite series:

$$ty^p P_1(y) Q(0,y) = X^{1+i_0} y^{p+j_0} - \sum_{m=0}^{p-1} H_m(X) y^m.$$

The functional equation finally reads

$$K(x,y)Q(x,y) = (x^{1+i_0} - X^{1+i_0}) y^{p+j_0} - \sum_{m=0}^{p-1} (H_m(x) - H_m(X)) y^m.$$

Since the substitution of an algebraic series into a D-finite one gives another D-finite series, this equation shows that $Q(x,y)$ is D-finite.

∎

[5]In the square lattice case, the symmetry of the model in x and y makes this step unnecessary: once the holonomy of $Q(x,0)$ is proved, the holonomy of $Q(x,y)$ follows.

4. **Perspectives**

This paper was written in a rush, right after the material of Section 2 was found. This new proof of Kreweras' formula suggests numerous questions and research directions, which I would like to explore in the coming weeks (or months...). Here are some of these questions.

4.1. **Other starting points.** It was observed by Gessel in [18] that the method he used to prove Kreweras' result was hard to implement for a starting point different from the origin. The reason of this difficulty is that, unlike the method presented here, Gessel's approach *checks* the known expression of the generating function, but does not *construct* it. I am confident that the new approach of Section 2 can be used to solve such questions. If the starting point does not lie on the main diagonal, the $x - y$ symmetry is lost; the diagram of Section 2.1 also loses its symmetry, and gives *four* different equations between the *two* unknown functions $Q(x, 0)$ and $Q(0, y)$.

4.2. **Other algebraic walk models.** A close examination of the ingredients that make the proof of Section 2 work might help to construct other walk models which, for non-obvious reasons, would have an algebraic generating function. Note that for some degenerate sets of steps, like those of Figure 3, the quadrant condition is equivalent to a half-plane condition and thus yields an algebraic series.

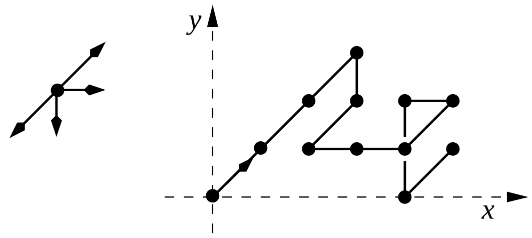


FIGURE 3. A degenerate set of steps.

I have started a systematic exploration of walks with few steps and only one up step: the non-trivial algebraic cases do not seem to be legion! However, I met in this exploration one model that seems to yield nice numbers (with a D-finite generating function) and for which the method of Section 2 "almost" works. I then realized that the same problem had been communicated to me, under a slightly different form, by Ira Gessel, a few months ago. I plan to explore this model further.

4.3. **Other equations.** Any combinatorial problem that seems to have an algebraic generating function and for which a linear functional equation with two "catalytic" variables (in the terminology of Zeilberger [42]) is available is now likely to be attacked via the method of Section 2. These conditions might seem very restrictive, but there is at least one such problem! The *vexillary* involutions, which were conjectured in 1995 to be counted by Motzkin numbers, satisfy the following equation:

$$\left(1 + \frac{t^2x^2y}{1-x} + \frac{t^2y}{1-y}\right) F(x, y; t) = \frac{t^2x^2y^2}{(1-ty)(1-txy)} + t\left(1 + \frac{ty}{1-y}\right) F(xy, 1; t) + \frac{t^2x^2y}{1-x} F(1, y; t).$$

The conjecture was recently proved via a difficult combinatorial construction [23]. I have been able to apply successfully the method of Section 2 to this equation [7].

4.4. **Random walks in the quarter plane.** Random walks in the quarter plane are naturally studied in probabilities. Given a Markov chain on the first quadrant, a central question is the determination of an/the invariant measure $(p_{i,j})_{i,j\geq 0}$. The invariance is equivalent to a linear equation satisfied by the series $P(x,y) = \sum p_{i,j}x^iy^j$, in which the variables x and y are "catalytic". A whole recent book is devoted to the solution of this equation in the case where the walk has small horizontal and vertical variations [14]. This book contains *one* example for which the series $P(x,y)$ is algebraic: no suprise, the steps of the corresponding walk are exactly Kreweras' steps... This result is actually due to Flatto and Hahn [17]. The equation satisfied by the series $P(x,y)$ does not work exactly like the equations for complete generating functions like $Q(x,y;t)$: roughly speaking, the third variable t is replaced by the additional constraint $P(x,y) = 1$. However, I hope that either a refinement of the enumeration problem that would contain the invariant distribution as a limit distribution, or a direct adaptation of the method of Section 2 to the context of $P(x,y)$, will give a new, simpler proof of Flatto and Hahn's result. Their proof is based on non-trivial complex analysis, and uses a parametrisation of the roots of the kernel by elliptic functions, which are *not* algebraic. It seems that a large detour is done to end up with an algebraic series. I hope my new approach will offer a significant shortcut, by staying in the field of algebraic series.

Acknowledgements. To my shame, I must recall that, in the lecture that I gave at FPSAC'01 in Phoenix, I mentioned (part of) Kreweras' result as a conjecture. I am very grateful to Ira Gessel who enlighted my ignorance by giving me the right references.

References

[1] D. André, Solution directe du problème résolu par M. Bertrand, *C. R. Acad. Sci. Paris* **105** (1887) 436–437.

[2] C. Banderier, M. Bousquet-Mélou, A. Denise, D. Gardy, D. Gouyou-Beauchamps and P. Flajolet, Generating functions for generating trees, to appear in *Discrete Math.*

[3] C. Banderier and P. Flajolet, Basic analytic combinatorics of directed lattice paths, to appear in *Theoret. Comput. Sci.*

[4] E. Barcucci, E. Pergola, R. Pinzani and S. Rinaldi, A bijection for some paths on the slit plane, *Adv. in Applied Math.* **26** no. 2 (2001) 89–96.

[5] M. Bousquet-Mélou, Multi-statistic enumeration of two-stack sortable permutations, *Electronic J. Combinatorics* **5** (1998) R21.

[6] M. Bousquet-Mélou, Walks on the slit plane: other approaches, *Advances in Applied Math.* **27** No. 2-3 (2001) 243–288.

[7] M. Bousquet-Mélou, The number of vexillary involutions, in preparation.

[8] M. Bousquet-Mélou, On (some) functional equations arising in enumerative combinatorics, in *Proceedings of FPSAC'01*, Arizona State University, pp. 83–89.

[9] M. Bousquet-Mélou and M. Petkovšek, Linear recurrences with constant coefficients: the multivariate case, *Discrete Math.* **225** (2000) 51–75.

[10] M. Bousquet-Mélou and M. Petkovšek, Walks confined in a quadrant are not always D-finite, submitted.

[11] M. Bousquet-Mélou and G. Schaeffer, Walks on the slit plane, preprint 2000, `arXiv:math.CO/0012230`. To appear in *Proba. Theory Related Fields.*

[12] W. G. Brown, On the existence of square roots in certain rings of power series, *Math. Annalen* **158** (1965) 82-89.

[13] P. Duchon, On the enumeration and generation of generalized Dyck words, *Discrete Math.* **225** (2000) 121–135.

[14] G. Fayolle, R. Iasnogorodski and V. Malyshev, *Random walks in the quarter plane*, Vol. 40 of *Applications of Mathematics: Stochastic Modelling and Applied Probabilities*, Springer-Verlag, Berlin Heidelberg, 1999.

[15] W. Feller, *An introduction to probability theory and its applications*, Vol.1, John Wiley and Sons, New York, 1950.

[16] P. Flajolet, Analytic models and ambiguity of context-free languages, *Theoret. Comput. Sci.* **49** (1987) 283–309.

[17] L. Flatto and S. Hahn, Two parallel queues created by arrivals with two demands I. *SIAM Journal of Applied Math.* **44** (1984) 1041–1053.

[18] I. M. Gessel, A probabilistic method for lattice path enumeration, *J. Stat. Planning Infer.* **14** (1986) 49–58.

[19] I. M. Gessel, A factorization for formal Laurent series and lattice path enumeration, *J. Combin. Theory. Ser. A* **28** (1980) 321–337.

[20] I. M. Gessel and D. Zeilberger, Random walk in a Weyl chamber, *Proc. Amer. Math. Soc.* **115** (1992) 27-31.

[21] I. P. Goulden and D. M. Jackson, *Combinatorial enumeration,* John Wiley and Sons, 1983.

[22] D. J. Grabiner, Random walk in an alcove of an affine Weyl group, and non-colliding random walks on an interval, in *Proceedings of FPSAC 2001*, Tempe, Arizona, pp. 217–225.

[23] O. Guibert, E. Pergola and R. Pinzani, Vexillary involutions are enumerated by Motzkin numbers, *Annals of Combinatorics* **5** (2001) 1–22.

[24] G. Kreweras, Sur une classe de problèmes liés au treillis des partitions d'entiers, *Cahiers du B.U.R.O.* **6** (1965) 5–105.

[25] J. Labelle, Langages de Dyck généralisés, *Ann. Sci. Math. Québec* **17** (1993) 53–64.

[26] J. Labelle and Y.-N. Yeh, Generalized Dyck paths, *Discrete Math.* **82** (1990) 1–6.

[27] G. F. Lawler, *Intersections of random walks*, Probabilities and its applications, Birkhäuser Boston, 1991.

[28] L. Lipshitz, The diagonal of a D-finite power series is D-finite, *J. Algebra* **113** (1988) 373–378.

[29] L. Lipshitz, D-finite power series, *J. Algebra* **122** (1989) 353–373.

[30] S. G. Mohanty, *Lattice path counting and applications*, Academic Press, 1979.

[31] T. V. Narayana, A partial order and its applications to probability theory, *Sankhya* **21** (1959) 91–98.

[32] H. Niederhausen, The ballot problem with three candidates, *Europ. J. of Combinatorics* **4** (1983) 161-167.

[33] H. Niederhausen, Lattice paths between diagonal boundaries, *Electronic J. Combinatorics* **5** (1998) R30.
[34] D. Poulalhon and G. Schaeffer, A bijection for loopless triangulations of a polygon with interior points, FPSAC 2002.
[35] G. Schaeffer, Random sampling of large planar maps and convex polyhedra, in *Proceedings of the 31th annual ACM Symposium on the Theory of Computing* (STOC'99), Atlanta, ACM press, 1999.
[36] F. Spitzer, *Principles of random walk*, The University Series in Higher Mathematics, Van Nostrand Company, Princeton, 1964.
[37] R. P. Stanley, Differentiably finite power series, *European J. Combin.* **1** (1980) 175–188.
[38] W. T. Tutte, A census of planar triangulations, *Canad. J. Math.* **14** (1962) 21–38.
[39] W. T. Tutte, A census of planar maps, *Canad. J. Math.* **15** (1963) 249–271.
[40] J. West, Sorting twice through a stack, *Theoret. Comput. Sci.* **117** (1993) 303–313.
[41] D. Zeilberger, A proof of Julian West's conjecture that the number of two-stack-sortable permutations of length n is $2(3n)!/((n+1)!(2n+1)!)$, *Discrete Math.* **102** (1992) 85–93.
[42] D. Zeilberger, The umbral transfer-matrix method: I. Foundations. *J. Comb. Theory, Ser. A*, **91** (2000) 451–463.

E-mail address: `bousquet@labri.fr`

CNRS, LaBRI, Université Bordeaux 1, 351 cours de la Libération, 33405 Talence Cedex, France

ON THE EQUIVALENCE PROBLEM FOR SUCCESSION RULES

SREČKO BRLEK*, ENRICA DUCHI, ELISA PERGOLA, AND SIMONE RINALDI

Abstract. The notion of *succession rule* (*system* for short) provides a powerful tool for the enumeration of many classes of combinatorial objects. Often, different systems exist for a given class of combinatorial objects, and a number of problems arise naturally. An important one is the equivalence problem between two different systems. In this paper, we show how to solve this problem in the case of systems having a particular form. More precisely, using a bijective proof, we show that the classical system defining the sequence of Catalan numbers is equivalent to a system obtained by linear combinations of labels of the first one.

Résumé Les *systèmes de réécriture* constituent des outils puissants pour l'énumération de nombreuses classes d'objets combinatoires. Souvent, il existe plusieurs systèmes pour une classe donnée et cela soulève le problème de l'équivalence des systèmes. Dans cet article, nous donnons une solution pour des systèmes ayant une forme particulière. Plus précisément, nous montrons que le système classique définissant les nombres de Catalan est équivalent à une infinité de systèmes obtenus par combinaisons linéaires des étiquettes de celui-ci.

1. Introduction

The notion of *succession rule* was introduced by Chung (et all.) in [6] as a compact notation for *generating trees*, and flourished later as a powerful tool for the enumeration of combinatorial objects (see for instance [1, 3, 8, 16]). More precisely, a *succession rule* Ω is a *system* $((b), \mathcal{R})$, consisting of an axiom and a set of *productions* or *rewriting rules* denoted

$$\Omega = \begin{cases} (b) & b \in \mathbb{N}^+, \\ (k) \rightsquigarrow (e_1(k))(e_2(k))\dots(e_k(k)), & e_i : \mathbb{N}^+ \to \mathbb{N}^+, k \in M \subseteq \mathbb{N}^+. \end{cases} \tag{1}$$

A system Ω is suitably represented by means of a *generating tree*, a rooted labelled where the root is labelled by the axiom (b), and a node labelled (k) produces k sons labelled by $e_1(k), \dots, e_k(k)$ respectively. Ω defines a non-decreasing sequence of positive integers $\{f_n\}_{n\geq 0}$, the number of nodes at level n (by convention, the root is at level 0), and the generating function of Ω is

$$f_\Omega(x) = \sum_{n\geq 0} f_n x^n \ .$$

The structure of the rewriting rules in a system is closely related to the sequence $\{f_n\}_n$, and this relationship has been studied in [1], for rational, algebraic and transcendental generating functions.

A well-known system is the one defining Schröder numbers [4], $(1, 2, 6, 22, 90, 394, \dots)$, (sequence M2898 in [15]):

$$\Omega_{\mathcal{S}} = \begin{cases} (2) \rightsquigarrow (3)(3) \\ (k) \rightsquigarrow (3)(4)\dots(k)(k+1)^2, & k \geq 3, \end{cases} \tag{2}$$

---

* On leave from LaCIM, Montréal, with the support of NSERC (Canada) and G.N.C.S., Istituto Nazionale dell'Alta Matematica, Italy.

where the power $(k+1)^2$ stands for the repetition $(k+1)(k+1)$. Often, the production of the axiom is omitted, when no confusion arises. In Fig.1 the first levels of the generating tree of (2) are shown. We refer to [3] for more details and examples. A system Ω is *finite* if

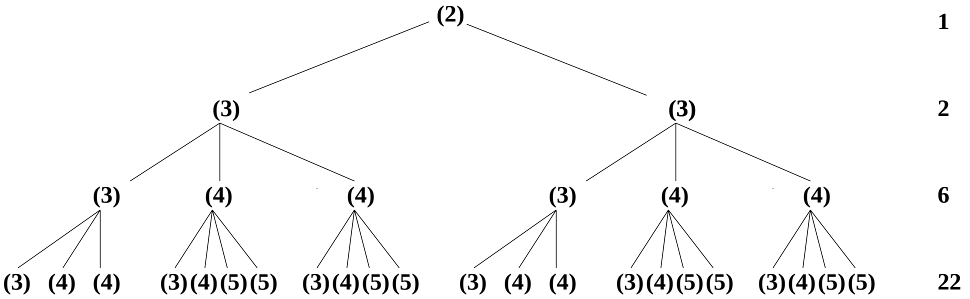


FIGURE 1. The first levels of the generating tree of (2), and its numerical sequence.

the number of labels in the productions is finite, that is, when $|M| < \infty$. In this particular case, the generating function is rational [1], and sometimes has an interpretation as a regular language or other combinatorial structure [8, 12].

A classical example of finite system is the one defining Fibonacci numbers, (M0692 in [15]):

(3) $$\Omega_{\mathcal{F}} = \begin{cases} (1); (1) \rightsquigarrow (2) \\ (2) \rightsquigarrow (1)(2). \end{cases}$$

A succession rule has a *factorial form*, if a finite modification of the set $\{1, 2, \ldots, k\}$ is reachable from k. More formally, a factorial succession rule has the form:

(4)
$$\Omega = \begin{cases} (b) \\ (k) \rightsquigarrow (r_0)(r_0+1)\ldots(k-c-1)(k+d_1)(k+d_2)\ldots(k+d_m), \qquad k \geq r_0 \geq 1, \end{cases}$$

with $c \geq 0$, $-c < d_1 \leq d_2 \leq \ldots \leq d_m > 0$, where the consistency principle of succession rules is satisfied imposing that $r_0 + c = m$; the rule in (2) is factorial.

Determining the generating function of a given system is not always an easy task [1]. Therefore, some recent papers focused on the development of some algebraic tools in order to study enumerative properties of succession rules, without computing the corresponding generating functions, by using a linear operator approach [8], or production matrices [7].

The study of these systems has been systematized by the italian school [3] in the so called ECO-method, from which we briefly recall some of the basics. Given a class $\mathcal{O}$ of combinatorial objects, we consider a fixed parameter $p : \mathcal{O} \rightarrow \mathbb{N}$, such that for all $n \in \mathbb{N}, \mathcal{O}_n = p^{-1}(n)$ is finite. If it is possible to define an operator

$$\vartheta : \mathcal{O}_n \longrightarrow 2^{\mathcal{O}_{n+1}},$$

performing "local expansions" on objects of size n (i.e. $\mathcal{O}_n = \{O \in \mathcal{O} : p(O) = n\}$), such that

(i) *for each* $O' \in \mathcal{O}_{n+1}$, *there exists* $O \in \mathcal{O}_n$ *such that* $O' \in \vartheta(O)$,
(ii) *for each* $O, O' \in \mathcal{O}_n$ *such that* $O \neq O'$, *then* $\vartheta(O) \cap \vartheta(O') = \emptyset$,

then the family of sets $\{\vartheta(O) : O \in \mathcal{O}_n\}$ is a partition of $\mathcal{O}_{n+1}$.
We refer to [3] for further details, proofs, definitions and examples. The parameter p being fixed, the recursive construction determined by ϑ is described by a *generating tree* [6], whose vertices are objects of $\mathcal{O}$; the objects having the same parameter value lie on the same level,

and the siblings of an object are the objects produced by ϑ: if $|\vartheta(P)| = k$ then the object P blossoms often according to a *system* (of the form (1)).

2. The equivalence problem

Two rules Ω_1 and Ω_2 are said to be *equivalent*, if they define the same number sequence,

$$\Omega_1 \cong \Omega_2 \iff f_{\Omega_1}(x) = f_{\Omega_2}(x) \ .$$

For instance, the reader can easily verify that the following rules are equivalent to (2), and define the Schröder numbers [4, 5]:

$$\Omega'_S = \begin{cases} (2) \\ (2k) \rightsquigarrow (2)(4)^2 \dots (2k)^2(2k+2), \end{cases}$$

$$\Omega''_S = \begin{cases} (2) \rightsquigarrow (3)(3) \\ (2k-1) \rightsquigarrow (3)^2(5)^2 \dots (2k-1)^2(2k+1), \end{cases}$$

$$\Omega'''_S = \begin{cases} (2) \\ (2^k) \rightsquigarrow (2)^{2^{k-1}}(4)^{2^{k-2}}(8)^{2^{k-3}} \dots (2^{k-1})^2(2^k)(2^{k+1}). \end{cases}$$

The *equivalence problem* consists in determining if two different systems are equivalent. In general, as mentioned recently by M. Robson [11], this problem is not decidable. However, there are classes of systems for which the answer is positive. The easy case of finite systems stems out from formal language theory. Indeed, A PD0L system is a triple [14]:

$$G = (\Sigma, h, w_0),$$

where $\Sigma = \{a_1, \dots, a_k\}$ is a k-letter alphabet, h is an endomorphism defined on the set Σ^+ of non-empty words, and $w_0 \in \Sigma^+$ is called the *axiom.* The length of a word $w \in \Sigma$ is denoted $|w|$. The *language* of G is defined by:

$$L(G) = \{h^i(w_0) : i \geq 0\}.$$

The function $f_G(n) = |h^n(w_0)|$, $n \geq 0$ is the *growth function* of G, and the sequence $|h^n(w_0)|$, $n \geq 0$ is its *growth sequence.* A *growth matrix* M associated to G is defined by,

$$M_G[i,j] = |h(a_j)|_{a_i},$$

where $|h(a_j)|_{a_i}$ is the number of occurrences of the letter a_i in $h(a_j)$. The growth sequence is then obtained by the generating function

$$f_G(x) = \frac{[10^{k-1}] \cdot \chi(M) \cdot (I - Mx)^{-1} \cdot [1^k]^t}{x^k \cdot \chi(M)} \tag{5}$$

where $\chi(M)$ is the characteristic polynomial of M, $[1^k]$ is the k-length vector with all entries 1, and $[10^{k-1}]$ has all entries 0 except the first which is 1. (see [13] for details).

Remark that any finite system Ω can be viewed as a particular PD0L system where the alphabet Σ is the set of labels of Ω, and h is defined by the productions of Ω, and $w_0 \in \Sigma$.

For instance, the rule (3) defines a PD0L system F, where $\Sigma = \{1, 2\}$, $w_0 = 1$, and

$$\begin{array}{rcl} h(1) & = & 2 \\ h(2) & = & 12 \end{array} \quad ; \quad M = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

The words in the language of F are $1, 2, 12, 212, 12212, 21212212, 1221221212212, \dots$, and its growth sequence is obtained from the matrix associated to G by computing the generating function (5)

$$f_G(x) = (1 + x + 2\,x^2 + 3\,x^3 + 5\,x^4 + 8\,x^5 + 13\,x^6 + 21\,x^7 + 34\,x^8 + 55\,x^9 + O\left(x^{10}\right)).$$

	1	2	3	4	5	6	7	...
1	1							...
2	1	1						...
3	2	2	1					...
4	5	5	3	1				...
5	14	14	9	4	1			...
6	42	42	28	14	5	1		...
7	132	132	90	48	20	6	1	...
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

TABLE 1. The Catalan triangle.

Now, two D0L systems are *growth equivalent* if they have the same generating function, which amounts to check if two polynomials are equal, and, consequently, the equivalence problem is decidable for the class of finite systems. However, the computation of the generating functions can be avoided, by checking the equality of the first few terms of the two sequences as stated below.

Theorem 1. *The equivalence problem is decidable for the class of finite systems.*

Proof. Let Ω_1 and Ω_2 be two finite succession rules having k_1 and k_2 labels respectively. In view of Theorem 3.3 [14] it is necessary and sufficient to check if the first $k_1 + k_2$ terms of the two sequences defined by Ω_1 and Ω_2 coincide. $\square$

For example, the finite rules:

$$\Omega_1 = \begin{cases} (2) \\ (2) \rightsquigarrow (2)(3) \\ (3) \rightsquigarrow (2)(3)(3), \end{cases} \qquad \Omega_2 = \begin{cases} (2) \\ (1) \rightsquigarrow (2) \\ (2) \rightsquigarrow (1)(4) \\ (4) \rightsquigarrow (1)(2)(4)(4), \end{cases} \tag{6}$$

both define odd index Fibonacci numbers, $1, 2, 5, 13, 34, 89, \ldots$ (sequence M1439 in [15]). Their equivalence can be verified by comparing the first 5 terms of the defined sequences.

In [1] the authors formalize and then apply a method, called the *kernel* method, in order to find a solution to the functional equation arising from a factorial system(4). The main result states that a factorial system has an algebraic generating function.

Theorem 2. *The equivalence problem is decidable for the rules having a factorial form.*

Proof. A classical result on the equality of algebraic generating functions in several commutative variables, shows that the equality is decidable (see [14] Theorem IV. 5.1) $\square$

3. An infinite set of rules for the Ballot numbers

For $k, n \in \mathbb{N}$, let $a_{n,k}$, be the set of Ballot numbers, defined by the recurrence,

$$\begin{aligned} &a_{1,1} = 1, \\ &a_{n+1,1} = \textstyle\sum_{j\geq 1} a_{n,j}, \\ &a_{n+1,k} = \textstyle\sum_{j\geq k-1} a_{n,j}, \qquad k \geq 2. \end{aligned}$$

They can conveniently be displayed in tabular form, as below, in a triangular array, sometimes known as the *Catalan triangle* shown in Table 1. For any positive integer h, a rule defining the sequence in the h-th column is known in literature:

$$\Omega^h = \begin{cases} (h) \\ (k) \rightsquigarrow (2)(3)\dots(k)(k+1). \end{cases} \tag{7}$$

Remark that for $h = 1$, we have the rules defining the Catalan numbers.
Let $h, \alpha \in \mathbb{N}^+$, and $\beta \in \mathbb{N}$. We first define the following rule:
(8)

$$\Omega^h_{\alpha,\beta} = \begin{cases} (h) \\ (1) \rightsquigarrow (2) \\ (2) \rightsquigarrow (2)(3) \\ \dots \qquad \dots \\ (\alpha+\beta-1) \rightsquigarrow (2)(3)\dots(\alpha+\beta) \\ (\alpha k+\beta) \rightsquigarrow (1)^k \dots (\alpha-1)^k(\alpha+1)\dots(\alpha+\beta)(2\alpha+\beta)\dots((k+1)\alpha+\beta), \ k \geq 1. \end{cases}$$

In the sequel we prove that, for $h \leq \alpha + \beta$, the system (8) is equivalent to the system (7), so the first can be viewed as a generalization of the second, where the labels have been linearly combined according to the positive coefficients α and β. Moreover, the first $\alpha + \beta$ levels of the two generating trees coincide. As a consequence, we obtain that (8) defines the numbers $\{a_{n,h} : n \geq 0\}$, for any α and β such that $h \leq \alpha + \beta$. In particular, for $h = 1$ we have an infinite set of succession rules defining Catalan numbers:
(9)

$$\Omega^1_{\alpha,\beta} = \begin{cases} (1) \\ (1) \rightsquigarrow (2) \\ (2) \rightsquigarrow (2)(3) \\ \dots \qquad \dots \\ (\alpha+\beta-1) \rightsquigarrow (2)(3)\dots(\alpha+\beta) \\ (\alpha k+\beta) \rightsquigarrow (1)^k \dots (\alpha-1)^k(\alpha+1)\dots(\alpha+\beta)(2\alpha+\beta)\dots((k+1)\alpha+\beta), \ k \geq 1. \end{cases}$$

Instead of using generating functions as in [1], we provide a bijective proof by the application of the ECO method.

3.1. **Dyck paths.** We consider lattice paths in the plane $\mathbb{Z} \times \mathbb{Z}$, starting from the origin $(0, 0)$, and using *rise steps* $\mathbf{x} = (1, 1)$ and *fall steps* $\overline{\mathbf{x}} = (1, -1)$. The set $\mathcal{D}$ of Dyck paths is the subset of $\Sigma^* = \{\mathbf{x}, \overline{\mathbf{x}}\}^*$ generated by the grammar

$$D := \epsilon + \mathbf{x}D\overline{\mathbf{x}}D, \tag{10}$$

and we refer to paths as words, in which the notions of prefix, suffix have the usual meaning.

A *Dyck path* of length $2n$ is a sequence remaining weakly above the x-axis. The *height* of a point $P = (P_x, P_y)$ is defined by $h(P) = P_y$. A Dyck path is said *elevated* or *primitive* if it can be written as $D = \mathbf{x}D'\overline{\mathbf{x}}$ with $D' \in \mathcal{D}$, and we denote the stripping operation by

$$D' = \mathrm{Top}(D).$$

Given two points P', P of a Dyck path D, the *factor* starting at P' and ending at P of the corresponding Dyck word is denoted $D[P'_x, P_x]$. By convention, $D[i, j] = \epsilon$ if $i \geq j$. The *insertion* of a word w in D at position i is defined by

$$\mathrm{insert}(D, w, i) = D[0, i] \cdot w \cdot D[i + 1, 2n] \ .$$

The last sequence of fall steps $\ell_d(D)$, or *last descent*, of D satisfies $\ell_d(D) = \overline{\mathbf{x}}^k$ for some $k \geq 1$, and $\mathbf{P}(D)$ is the set of its points. Finally, $|D|$ denotes the length of the word (number of steps). From the grammar (10), one can easily deduce the properties summarized in the next statement.

Proposition 1. *Every non empty Dyck $D = u\mathbf{x}\overline{\mathbf{x}}^k, k \geq 1$, satisfies the conditions*
(a) $D = D_1 D_2 \dots D_m$ *where* $\forall i, D_i$ *is primitive;*
(b) $\exists D'' = u\mathbf{x}^{k-1} \in \mathcal{D}$ *such that* $D = \mathrm{insert}(D'', \mathbf{x}\overline{\mathbf{x}}, |D''| - (k-1))$*;*
(c) $\forall$ *suffix* v *of* D, $\exists D'' \in \mathcal{D}, \exists u' \in \Sigma^*$, *such that* $u'v \in \mathcal{D}$ *and* $D = \mathrm{insert}(u'v, D'', |u'|) = u'D''v$*.*

We provide now a valuation on the set of Dyck paths,

$$\mathrm{Val} : \mathcal{D} \longrightarrow \{0, 1, 2\} \times \mathbb{N},$$

where the first term yields a partition

$$\mathcal{D} = \mathcal{D}^0 \cup \mathcal{D}^1 \cup \mathcal{D}^2, \quad \text{with } \mathcal{D}^i \cap \mathcal{D}^j = \emptyset, \ \forall i \neq j \ .$$

A valuation for Dyck paths. For $l, i, n \in \mathbb{N}$, a *Dyck path at level* i is the image of an ordinary Dyck path (at level 0) under the translation $(0,0) \mapsto (l,i)$, running from (l,i) to $(l+2n,i)$ and above the line $y = i$.

Let $\mathcal{D}(i)$ be the set of Dyck paths at level i, and $\mathbb{D} = \{\bigcup \mathcal{D}(i) : i \in \mathbb{N}\}$. By Proposition 1 (a), each path $D(i) \in \mathcal{D}(i)$ admits a unique decomposition in terms of *primitive* paths at level i,

$$D(i) = D(i,1)D(i,2)\dots D(i,m),$$

where $D(i,j)$ is the j-th *component*, and $\#(D(i)) = m$ is the number of components. In

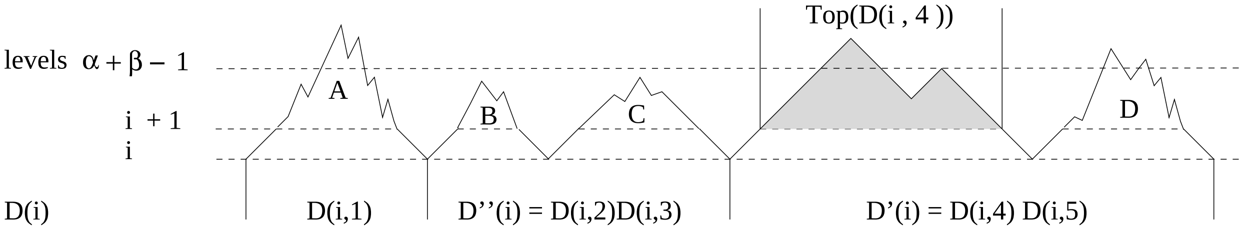


FIGURE 2. Decomposition of a Dyck path at level i and notations.

this decomposition, $D''(i)$ denotes the rightmost factor having height less than $\alpha+\beta-i-1$, while $D'(i)$ is the factor on the right of $D''(i)$.

The valuation is defined by $\mathrm{Val}(D) = \overline{\mathrm{Val}}(D, 0)$, where $\overline{\mathrm{Val}}(D, i)$ is defined by the following algorithm, where the variables are defined by the notations above.

Algorithm $\overline{\mathrm{Val}}(D, i)$**:** ;
 $D - D''$ is the path obtained by removing D'' from D; $\overline{\mathrm{Val}}(D, i)[1]$ refers to the first component of the valuation.
if $D = \epsilon$ **then:** return $(0, i)$
elseif $\#(D') = 0$ **then :** there are three cases
  if $\#(D'') > 1$ **then:** return $(0, i)$
  elseif: $\overline{\mathrm{Val}}(D - D'', i)[1] = 1$ **then** return $(2, i)$
  else: return $(0, i)$
else: let $D' = \prod_{j=1}^{m} D'(j)$; we have four cases:
  if: $\overline{\mathrm{Val}}(\prod_{j=1}^{m-1} D'(j), i)[1] = 1$ **then** return $(2, i)$
  elseif $i < \alpha - 2$ **then:** return $\overline{\mathrm{Val}}(\mathrm{Top}(D'_m), i+1))$
  elseif $|\mathbf{P}(D'_m)| < \alpha + \beta - i$ **then:** return $(0, i)$
  else : return $(1, i)$.
end $\overline{\mathrm{Val}}(D, i)$**.:**

3.2. **An ECO-system for $\mathcal{D}$.** We define now an ECO-system for the generation of Dyck paths, according to the rule $\Omega^1_{\alpha,\beta}$, that classifies the paths according to the valuation Val.

The ECO operator $\vartheta : \mathcal{D}_n \to 2^{\mathcal{D}_{n+1}}$, is defined inductively by setting $\epsilon \in \mathcal{D}^0$, and each path produced is classified in some $\mathcal{D}^i$, with an extra labelling of those in $\mathcal{D}^2$. So, in case of

$[D \in \mathcal{D}^0]$**:** /*See Figure 3 for an example with $\alpha = 3, \beta = 2$, and $|\mathbf{P}(D)| = 4$.*/

- for each point $P \in \mathbf{P}(D)$ do
  $D_{h(P)} \leftarrow \text{insert}(D, \mathbf{x\overline{x}}, P_x)$;
  if $h(P) < \alpha + \beta - 2$ then $D_{h(P)} \in \mathcal{D}^0$ else $D_{h(P)} \in \mathcal{D}^1$. /* classifying */

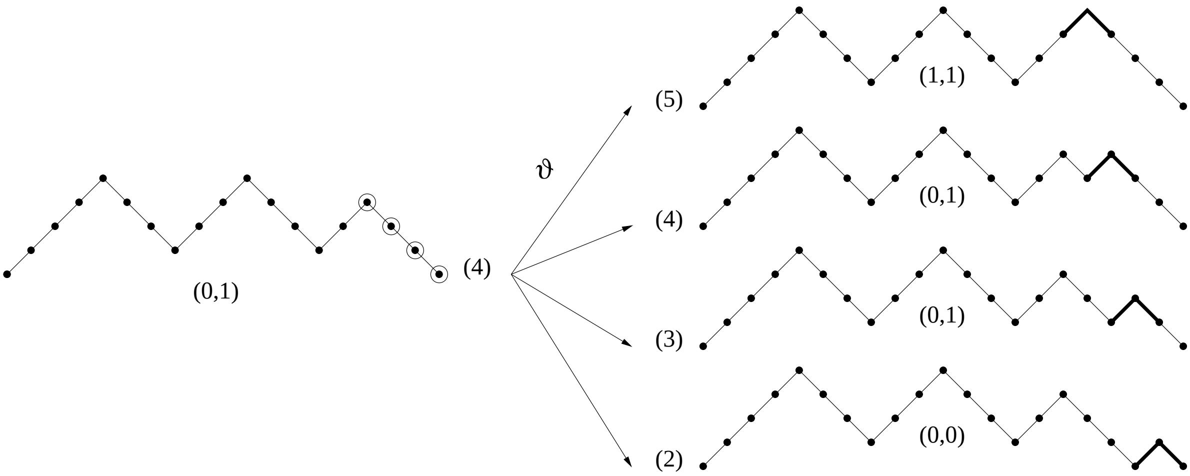


FIGURE 3. The operator ϑ applied to a path in $\mathcal{D}^0$.

Remark. For each path in the class $\mathcal{D}^0$ we have $D \in \mathcal{D}^0 \Longrightarrow h(\ell_d(D)) < \alpha + \beta - 1$.

$[D \in \mathcal{D}^1]$**:** /* See Figure 4 for an example with $\alpha = 3, \beta = 2$.*/
- for each point $P \in \mathbf{P}(D)$ such that $h(P) \geq \alpha + \beta - 1$ do
  $D_{h(P)} \leftarrow \text{insert}(D, \mathbf{x}\overline{\mathbf{x}}, P_x)$; $D_{h(P)} \in \mathcal{D}^1$;
  let P' be the leftmost point of D such that $P'P \in \mathcal{D}$;
  /* then $D = uD[P'_x, P_x]v$ with $uv \in \mathcal{D}$; this decomposition exists by Proposition 1(c) */
  for each point $Q \in \mathbf{P}(D)$ such that $0 \leq h(Q) \leq \alpha - 2$ do
    $D_{h(P),h(Q)} \leftarrow \text{insert}(uv, \mathbf{x}D[P'_x, P_x]\overline{\mathbf{x}}, Q_x)$; $D_{h(P),h(Q)} \in \mathcal{D}^2$;
    label$(D_{h(P),h(Q)}) \leftarrow h(Q)$; /* labelling*/
- for each point $P \in \mathbf{P}(D)$ such that $\alpha - 1 \leq h(P) \leq \alpha + \beta - 2$ do
  $D_{h(P)} \leftarrow \text{insert}(D, \mathbf{x}\overline{\mathbf{x}}, P_x)$;
  if $h(P) = \alpha + \beta - 2$ then $D_{h(P)} \in \mathcal{D}^1$ else $D_{h(P)} \in \mathcal{D}^0$.

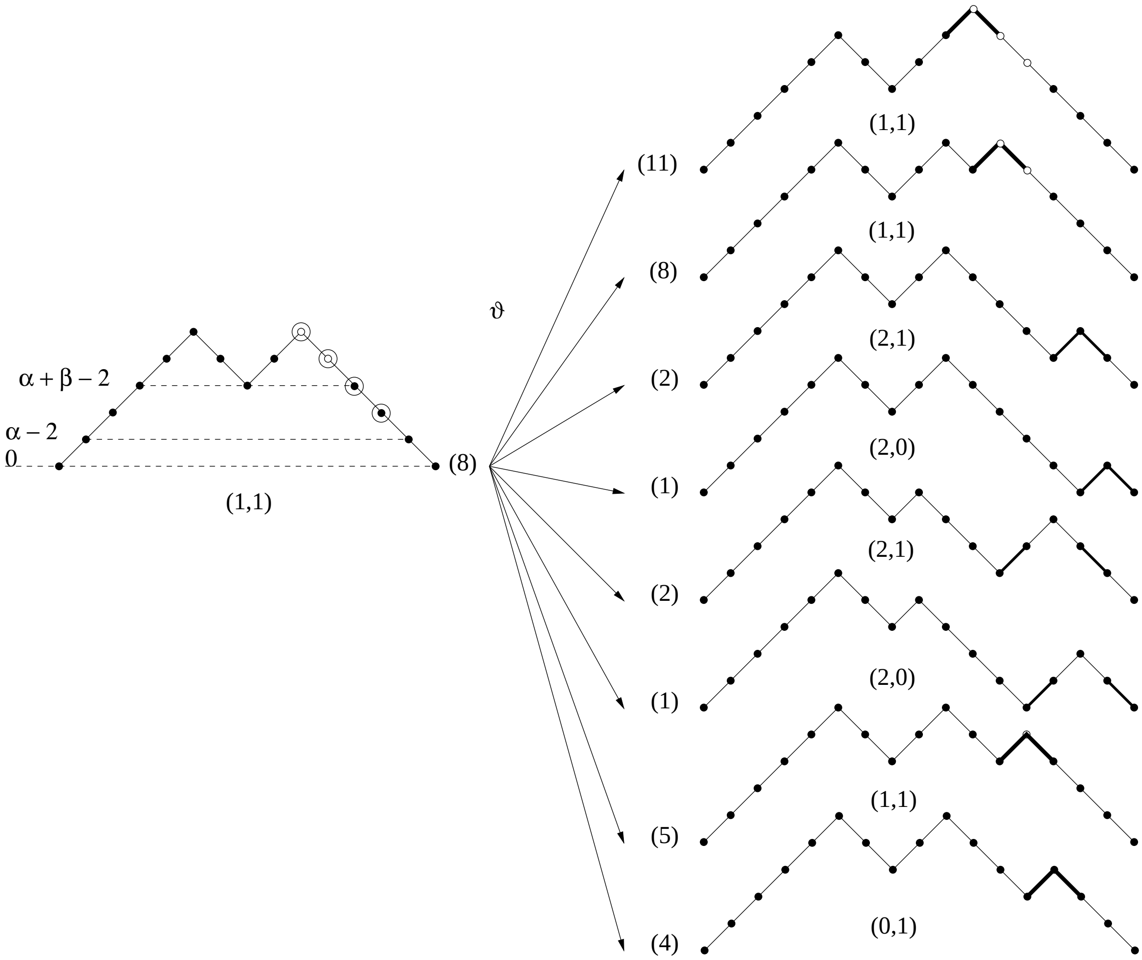


FIGURE 4. The operator ϑ applied to a path in $\mathcal{D}^1$.

Remark. For algorithmic efficiency, the statements of the form $D_{h(P)} \in \mathcal{D}^i$ should be viewed as an assignment of a label, which avoids a call of the valuation function. Moreover, observe that only the paths in $\mathcal{D}^2$ receive a label which is equal to the height of the insertion point.

$[D \in \mathcal{D}^2]$**:** See Figure 5 for an example with $\alpha = 3, \beta = 0$. These paths are labelled.
- for each point $P \in \mathbf{P}(D)$ such that $h(P) \leq \text{label}(D)$ do
  $D_{h(P)} \leftarrow \text{insert}(D, \mathbf{x}\overline{\mathbf{x}}, P_x)$;
  if $\text{label}(D) < \alpha + \beta - 2$ then $D_{h(P)} \in \mathcal{D}^0$ else $D_{h(P)} \in \mathcal{D}^1$.

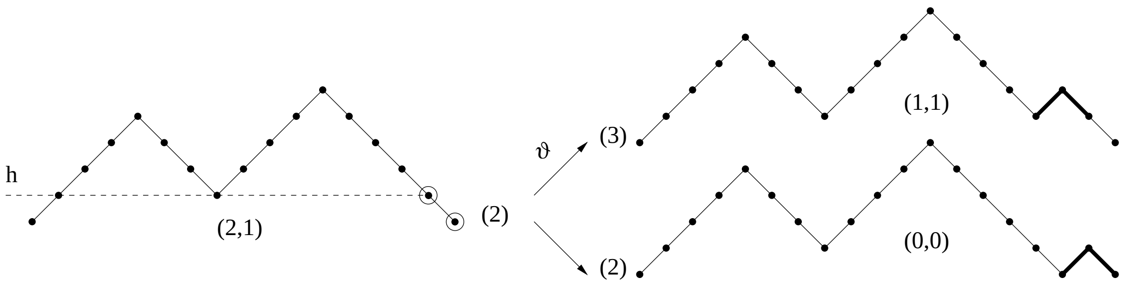


FIGURE 5. The operator ϑ applied to a path in $\mathcal{D}^2$.

It remains now to prove that the described construction generates all the Dyck paths (i) and that we have a partition (ii). This is achieved for both conditions by induction, that follows the inductive definition of ϑ.

(i) Let $D' \in \mathcal{D}_{n+1}$; then there exists $D \in \mathcal{D}_n$, such that $D' \in \vartheta(D)$:
- if $D' \in \mathcal{D}^0 \cup \mathcal{D}^1$, then $D' = u\mathbf{x}\overline{\mathbf{x}}^k$ and Proposition 1(b) identifies the last peak of D' to be removed, i.e.

$$D = u\overline{\mathbf{x}}^{k-1} \in \mathcal{D}_n ,$$

  such that $D' = \text{insert}(D, \mathbf{x}\overline{\mathbf{x}}, |D| - (k-1))$.
- if $D' \in \mathcal{D}^2$, from the valuation 3.1, we have

$$\text{Val}(D') = (2, |v|), \quad \text{and} \quad D' = uD''v ,$$

  where $D'' \neq \epsilon$, and $\#(D'') \geq 2$. Then, Proposition 1(a) provides the factorization

$$D'' = D''_1 \dots D''_{m-1} D''_m ,$$

  where $D''_m = \mathbf{x}\text{Top}(D''_m)\overline{\mathbf{x}}$. Let P_x be the position of the last peak of D''_{m-1}. Then,

$$D = \text{insert}(uD''_1 \dots D''_{m-1}v, \text{Top}(D''_m), P_x) .$$

(ii) Let D and $D' \in \mathcal{D}_n$, then $\vartheta(D) \cap \vartheta(D') = \emptyset$; when D and D' are such that ϑ performs the insertion of $\mathbf{x}\overline{\mathbf{x}}$ in their last descent, the result follows from the fact that, for each $P \in \mathbf{P}(D)$ and for each $P' \in \mathbf{P}(D')$, we have

$$\text{insert}(D, \mathbf{x}\overline{\mathbf{x}}, P_x) = \text{insert}(D', \mathbf{x}\overline{\mathbf{x}}, P'_x) \Longrightarrow D = D'.$$

When $\vartheta(D), \vartheta(D') \in \mathcal{D}^2$, we have two cases.
- $\text{Val}(\vartheta(D))[2] \neq \text{Val}(\vartheta(D'))[2] \Longrightarrow \vartheta(D) \neq \vartheta(D')$.
- $\text{Val}(\vartheta(D))[2] = \text{Val}(\vartheta(D'))[2]$; if $\vartheta(D) = \vartheta(D')$ then the construction (i) above shows that $D = D'$.

We are now in a position to state our main result.

Proposition 2. *Let $\alpha \in \mathbb{N}^+$ and $\beta \in \mathbb{N}$, then $\Omega^1_{\alpha,\beta} \cong \Omega^1$.*

Corollary 1. *Let $h, \alpha \in \mathbb{N}^+$, $\beta \in \mathbb{N}$ and $h \leq \alpha + \beta$. We have $\Omega^h_{\alpha,\beta} \cong \Omega^h$.*

Proof. It is a direct consequence of Proposition 2. Indeed, the rules $\Omega^h_{\alpha,\beta}$ and Ω^h both enumerate the class of Dyck paths beginning with h rise steps. □

Remark 1. *We then have an infinite set of systems defining Ballot numbers. In particular, the following define the Catalan numbers:*

$$\Omega_{2,0} = \begin{cases} (1) \\ (1) \rightsquigarrow (2) \\ (2k) \rightsquigarrow (1)^k(4)(6)\dots(2k)(2k+2); \end{cases} \tag{11}$$

$$\Omega_{3,1} = \begin{cases} (1) \\ (1) \rightsquigarrow (2) \\ (2) \rightsquigarrow (2)(3) \\ (3) \rightsquigarrow (2)(3)(4) \\ (3k+1) \rightsquigarrow (1)^k(2)^k(4)(7)\dots(3k+1)(3k+4). \end{cases} \tag{12}$$

3.3. **Computation of the valuation for Dyck paths.** For sake of completenes we provide a computation of the valuation described in section 3.1 applied to the path in Figure 6.

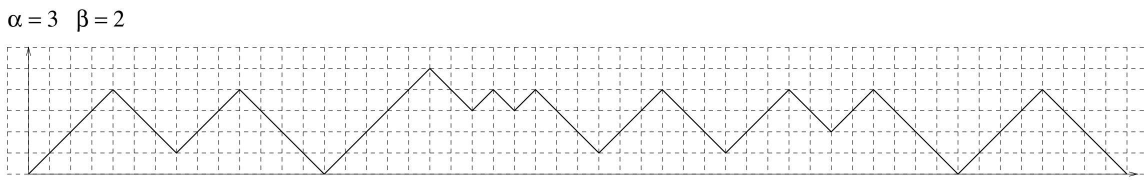


FIGURE 6. A Dyck D path with $Val(D) = (2, 0)$.

The path D factors as $D = D_1D_2D_3$, and it can be easily checked that

$\mathrm{Val}(D_1) = \overline{\mathrm{Val}}(D_1, 0) = (2, 1)$,

$\mathrm{Val}(D_1D_2) = \overline{\mathrm{Val}}(D_1D_2, 0) = (1, 1)$,

$\mathrm{Val}(D_1D_2D_3) = \overline{\mathrm{Val}}(D_1D_2D_3, 0) = (2, 0)$.

4. Concluding remarks and open questions.

The equivalence relation $\cong$ partitions a set $\mathcal{R} \subset \mathcal{S}$ of systems into equivalence classes, identified by the corresponding number sequence. For instance, if $\mathcal{R}$ is the class of *rational systems*, those having a rational generating function, we already know that finite systems are in it. On the other hand, there exists rational generating functions that are not the growth sequence of a D0L system ([13] Theorem III.4.11). Therefore, many problems arise naturally concerning

- finiteness of the equivalence classes: is $|[\Omega]_{\cong}| < \infty$ when Ω is finite? More generally, for a given rational generating function, is its class finite?
- the characterization of the rules in a given equivalence class;
- the extension of the decidability of equivalence for finite systems to a larger class, by using the same decision procedure;
- operations on rules (or trees) that provide equivalent systems.

By Theorem 1, the class of finite systems in included in the class of rational systems. On the other hand, Theorem III.4.11 of [13] characterizes the rational functions with integers coefficients that are the generating functions of D0L systems. Therefore, the following problem seems natural.

Conjecture 1: *Each rational system is equivalent to a finite one.*

Actually, a weaker statement is the following.

Conjecture 2: *A system counting a regular language is equivalent to a finite system.*

In some recent discussions with Cyril Banderier, of INRIA, we were led to speculate about *algebraic* systems.

Conjecture 3: *A system with algebraic generating function is equivalent to a factorial system.*

References

[1] C. Banderier, M. Bousquet-Mélou, A. Denise, P. Flajolet, D. Gardy and D. Gouyou-Beauchamps, On generating functions of generating trees, *Proceedings of 11th FPSAC, Barcelona* (1999), 40–52.

[2] E. Barcucci, A. Del Lungo, A. Frosini, S. Rinaldi, From rational functions to regular languages, *Proceedings of 12th FPSAC, Moscow* (2000), 633–644.

[3] E. Barcucci, A. Del Lungo, E. Pergola, R. Pinzani, ECO: a methodology for the Enumeration of Combinatorial Objects, *Journal of Difference Equations and Applications* **5** (1999), 435–490.

[4] E. Barcucci, A. Del Lungo, E. Pergola, R. Pinzani, Some combinatorial interpretations of q-analogs of Schröder numbers, *Annals of Combinatorics* **3** (1999), 173–192.

[5] E. Barcucci, E. Pergola, R. Pinzani, S. Rinaldi, Eco method and hill-free paths, *Seminaire Lotharingien de Combinatorie*, B46b (2001) 14 pp.

[6] F. R. K. Chung, R. L. Graham, V. E. Hoggatt, M. Kleimann, The number of Baxter permutations, *J. Combin. Theory Ser. A* **24** (1978), 382–394.

[7] E. Deutsch, Operations on production matrices, (Personal Communication).

[8] L. Ferrari, E. Pergola, R. Pinzani, S. Rinaldi, An algebraic characterization of the set of succession rules, *Theoret. Comp. Sci.* (to appear).

[9] O. Guibert, *Combinatoire des permutations à motifs exclus en liason avec mots, cartes planaries et tableaux de Young*, Thèse de l'Université de Bordeaux I, (1996).

[10] E. Pergola, R. Pinzani, S. Rinaldi, Approximating algebraic functions by means of rational ones, *Theoret. Comp. Sci.* **270** (2002), 643–657.

[11] M. Robson (Personal communication).

[12] S. Rinaldi, *Succession rules: the whole story*, Phd thesis, University of Florence, (2001).

[13] G. Rozenberg, A. Salomaa, *The mathematical theory of L systems*, Academic Press, London (1980).

[14] A. Salomaa and M. Soittola, *Automata-theoretic aspects of formal power series*, Springer-Verlag, New York (1978).

[15] N. J. A. Sloane and S. Plouffe, *The encyclopedia of integer sequences*, Academic Press, New York (1996).

[16] J. West, Generating trees and the Catalan and Schröder numbers, *Discrete Mathematics* **146** (1995), 247–262.

LaCIM, Université du Québec à Montréal, CP 8888, Succ. Centre-Ville, Montréal (QC), Canada H3C 3P8
E-mail address: `brlek@lacim.uqam.ca`

Dipartimento di Sistemi e Informatica, Via Lombroso 6/17, 50134 Firenze, Italy
E-mail address: {`duchi, elisa`}`@dsi.unifi.it`

Dipartimento di Matematica, Via del Capitano, 15, 53100 Siena, Italy
E-mail address: `rinaldi@unisi.it`

WORDS RESTRICTED BY PATTERNS WITH AT MOST 2 DISTINCT LETTERS

ALEXANDER BURSTEIN AND TOUFIK MANSOUR

Abstract. We find generating functions for the number of words avoiding certain patterns or sets of patterns with at most 2 distinct letters and determine which of them are equally avoided. We also find exact number of words avoiding certain patterns and provide bijective proofs for the resulting formulas.

Résumé. On obtient les séries génératrices pour les mots qui évitent certains motifs et certains ensembles de motifs, tous contenant au plus deux lettres différentes. On détermine ainsi des motifs équitablement évités. On donne enfin des formules d'énumération explicites dans certains cas particuliers, cas pour lesquels on donne aussi des preuves bijectives.

The main goal of this note to give analogies of enumerative results on certain classes of permutations characterized by *pattern-avoidance* in the symmetric group (see [SS]), and in the words on k letters (see [B]).

Pattern avoidance in the symmetric group. Let $\alpha \in S_n$ and $\tau \in S_k$ be two permutations. We say that α *contains* τ if there exists a subsequence $1 \le i_1 < i_2 < \cdots < i_k \le n$ such that $(\alpha_{i_1}, \dots, \alpha_{i_k})$ is order-isomorphic to τ; in such a context τ is usually called a *pattern*. We say that α *avoids* τ, or is τ*-avoiding*, if such a subsequence does not exist. The first paper devoted entirely to study of permutations avoiding certain patterns appeared in 1985 (see [SS]) and currently there exist more than 70 papers on this subject. These papers containing more than 5 analogies, for example: words (see [B] and references therein), generalized patterns (see [BS]), signed permutations (see [S]), and coloured permutations [M] and references therein). In the present paper we give another analogue for this problem.

Pattern avoidance in the words on k letters. Let $[k] = \{1, 2, \dots, k\}$ be a (totally ordered) alphabet on k letters. We call the elements of $[k]^n$ *words.* Consider two words, $\sigma \in [k]^n$ and $\tau \in [\ell]^m$. In other words, σ is an n-long k-ary word and τ is an m-long ℓ-ary word. Assume additionally that τ contains all letters 1 through ℓ. We say that σ contains an *occurrence* of τ, or simply that σ *contains* τ, if σ has a subsequence *order-isomorphic* to τ, i.e. if there exist $1 \le i_1 < \ldots < i_m \le n$ such that, for any relation $\phi \in \{<, =, >\}$ and indices $1 \le a, b \le m$, $\sigma(i_a)\phi\sigma(i_b)$ if and only if $\tau(a)\phi\tau(b)$. In this situation, the word τ is called a *pattern.* If σ contains no occurrences of τ, we say that σ *avoids* τ.

Up to now, most research on forbidden patterns dealt with cases where both σ and τ are permutations, i.e. have no repeated letters. Some papers (see Atkinson et al. [AH], Burstein [B], Regev [R] and references therein) also dealt with cases where only τ is a permutation. The natural analogue of avoiding permutations in $[k]^n$ is avoiding words.

In this paper, we consider some cases where forbidden patterns τ contain repeated letters. Just like [B], this paper is structured in the manner of Simion and Schmidt [SS], which was the first to systematically investigate forbidden patterns and sets of patterns.

1. Preliminaries

Let $[k]^n(\tau)$ denote the set of n-long k-ary words which avoid pattern τ. If T is a set of patterns, let $[k]^n(T)$ denote the set of n-long k-ary words which simultaneously avoid all patterns in T, that is $[k]^n(T) = \cap_{\tau\in T}[k]^n(\tau)$.

For a given set of patterns T, let $f_T(n,k)$ be the number of T-avoiding words in $[k]^n$, i.e. $f_T(n,k) = |[k]^n(T)|$. We denote the corresponding exponential generating function by $F_T(x;k)$; that is, $F_T(x;k) = \sum_{n\geq 0} f_T(n,k)x^n/n!$. Further, we denote the ordinary generating function of $F_T(x;k)$ by $F_T(x,y)$; that is, $F_T(x,y) = \sum_{k\geq 0} F_T(x;k)y^k$. The reason for our choices of generating functions is that $k^n \geq |[k]^n(T)| \geq n!\binom{k}{n}$ for any set of patterns with repeated letters (since permutations without repeated letters avoid all such patterns). We also let $G_T(n;y) = \sum_{k=0}^{\infty} f_T(n,k)y^k$, then $F_T(x,y)$ is the exponential generating function of $G_T(n;y)$.

We say that two sets of patterns T_1 and T_2 belong to the same *cardinality class*, or *Wilf class*, or are *Wilf-equivalent*, if for all values of k and n, we have $f_{T_1}(n,k) = f_{T_2}(n,k)$.

It is easy to see that, for each τ, two maps give us patterns Wilf-equivalent to τ. One map, $r : \tau(i) \mapsto \tau(m+1-i)$, where τ is read right-to-left, is called *reversal*; the other map, where τ is read upside down, $c : \tau(i) \mapsto \ell+1-\tau(i)$, is called *complementation*. For example, if $\ell = 3$, $m = 4$, then $r(1231) = 1321$, $c(1231) = 3213$, $r(c(1231)) = c(r(1231)) = 3123$. Clearly, $c\circ r = r\circ c$ and $r^2 = c^2 = (c\circ r)^2 = id$, so $\langle r, c\rangle$ is a group of symmetries of a rectangle. Therefore, we call $\{\tau,\ r(\tau),\ c(\tau),\ r(c(\tau))\}$ the *symmetry class* of τ.

Hence, to determine cardinality classes of patterns it is enough to consider only representatives of each symmetry class.

2. Two-letter patterns

There are two symmetry classes here with representatives 11 and 12. Avoiding 11 simply means having no repeated letters, so

$$f_{11}(n,k) = \binom{k}{n} n! = (k)_n, \qquad F_{11}(x;k) = (1+x)^k.$$

A word avoiding 12 is just a non-increasing string, so

$$f_{12}(n,k) = \binom{n+k-1}{n}, \qquad F_{12}(x;k) = \frac{1}{(1-x)^k}.$$

3. Single 3-letter patterns

The symmetry class representatives are 123, 132, 112, 121, 111. It is well-known [Kn] that

$$|S_n(123)| = |S_n(132)| = C_n = \frac{1}{n+1}\binom{2n}{n},$$

the nth Catalan number. It was also shown earlier by the first author [B] that

$$f_{123}(n,k) = f_{132}(n,k) = 2^{n-2(k-2)} \sum_{j=0}^{k-2} a_{k-2,j}\binom{n+2j}{n},$$

where

$$a_{k,j} = \sum_{m=j}^{k} C_m D_{k-m}, \quad D_t = \binom{2t}{t},$$

and

$$F_{123}(x,y) = F_{132}(x,y) = 1 + \frac{y}{1-x} + \frac{2y^2}{(1-2x)(1-y) + \sqrt{((1-2x)^2 - y)(1-y)}}.$$

Avoiding pattern 111 means having no more than 2 copies of each letter. There are $0 \le i \le k$ distinct letters in each word $\sigma \in [k]^n$ avoiding 111, $0 \le j \le i$ of which occur twice. Hence, $2j + (i-j) = n$, so $j = n - i$. Therefore,

$$f_{111}(n,k) = \sum_{i=0}^{k} \binom{k}{i}\binom{i}{n-i}\frac{n!}{2^{n-i}} = \sum_{i=0}^{k} \frac{n!}{2^{n-i}(n-i)!(2i-n)!}(k)_i = \sum_{i=0}^{k} B(i, n-i)(k)_i,$$

where $(k)_i$ is the falling factorial, and $B(r,s) = \dfrac{(r+s)!}{2^s(r-s)!s!}$ is the Bessel number of the first kind. In particular, we note that $f_{111}(n,k) = 0$ when $n > 2k$.

Theorem 1. $F_{111}(x;k) = \left(1 + x + \dfrac{x^2}{2}\right)^k$.

Proof. This can be derived from the exact formula above. Alternatively, let α be any word in $[k]^n(111)$. Since α avoids 111, the number of occurrences of the letter k in α is 0, 1 or 2. Hence, there are $f_{111}(n,k-1)$, $nf_{111}(n-1,k-1)$ and $\binom{n}{2}f_{111}(n-2,k-1)$ words α with 0, 1 and 2 copies of k, respectively. Hence

$$f_{111}(n,k) = f_{111}(n,k-1) + nf_{111}(n-1,k-1) + \binom{n}{2}f_{111}(n-2,k-1),$$

for all $n, k \ge 2$. Also, $f_{111}(n,1) = 1$ for $n = 0,1,2$, $f_{111}(n,1) = 0$ for all $n \ge 3$, $f_{111}(0,k) = 1$ and $f_{111}(1,k) = k$ for all k, hence the theorem holds. $\square$

Finally, we consider patterns 112 and 121. We start with pattern 121.

If a word $\sigma \in [k]^n$ avoids pattern 121, then it contains no letters other than 1 between any two 1's, which means that all 1's in σ, if any, are consecutive. Deletion of all 1's from σ leaves another word σ_1 which avoids 121 and contains no 1's, so all 2's in σ_1, if any, are consecutive. In general, deletion of all letters 1 through j leaves a (possibly empty) word σ_j on letters $j+1$ through k in which all letters $j+1$, if any, occur consecutively.

If a word $\sigma \in [k]^n$ avoids pattern 112, then only the leftmost 1 of σ may occur before a greater letter. The rest of the 1's must occur at the end of σ. In fact, just as in the previous case, deletion of all letters 1 through j leaves a (possibly empty) word σ_j on letters $j+1$ through k in which all occurrences of $j+1$, except possibly the leftmost one, are at the end of σ_j. We will call all occurrences of a letter j, except the leftmost j, *excess* j's.

The preceding analysis suggests a natural bijection $\rho : [k]^n(121) \to [k]^n(112)$. Given a word $\sigma \in [k]^n(121)$, we apply the following algorithm of k steps. Say it yields a word $\sigma^{(j)}$ after Step j, with $\sigma^{(0)} = \sigma$. Then Step j $(1 \le j \le k)$ is:

Step j. Cut the block of excess j's, then insert it immediately before the final block of all smaller excess letters of $\sigma^{(j-1)}$, or at the end of $\sigma^{(j-1)}$ if there are no smaller excess letters.

It is easy to see that, at the end of the algorithm, we get a word $\sigma^{(k)} \in [k]^n(112)$.

The inverse map, $\rho^{-1} : [k]^n(112) \to [k]^n(121)$ is given by a similar algorithm of k steps. Given a word $\sigma \in [k]^n(112)$ and keeping the same notation as above, Step j is as follows:

Step j. Cut the block of excess j's (which are at the end of $\sigma^{(j-1)}$), then insert it immediately after the leftmost j in $\sigma^{(j-1)}$.

Clearly, we get $\sigma^{(k)} \in [k]^n(121)$ at the end of the algorithm.

Thus, we have the following

Theorem 2. *Patterns 121 and 112 are Wilf-equivalent.*

We will now find $f_{112}(n,k)$ and provide a bijective proof of the resulting formula.

Consider all words $\sigma \in [k]^n(112)$ which contain a letter 1. Their number is

$$g_{112}(n,k) = f_{112}(n,k) - |\{\sigma \in [k]^n(112) : \sigma \text{ has no 1's}\}| = f_{112}(n,k) - f_{112}(n,k-1). \tag{1}$$

On the other hand, each such σ either ends on 1 or not.

If σ ends on 1, then deletion of this 1 may produce any word in $\bar{\sigma} \in [k]^{n-1}(112)$, since addition of the rightmost 1 to any word in $\bar{\sigma} \in [k]^{n-1}(112)$ does not produce extra occurrences of pattern 112.

If σ does not end on 1, then it has no excess 1's, so its only 1 is the leftmost 1 which does not occur at end of σ. Deletion of this 1 produces a word in $\bar{\sigma} \in \{2,\dots,k\}^{n-1}(112)$. Since insertion of a single 1 into each such $\bar{\sigma}$ does not produce extra occurrences of pattern 112, for each word $\bar{\sigma} \in \{2,\dots,k\}^{n-1}(112)$ we may insert a single 1 in $n-1$ positions (all except the rightmost one) to get a word $\sigma \in [k]^n(112)$ which contains a single 1 not at the end.

Thus, we have

$$\begin{aligned} g_{112}(n,k) &= f_{112}(n-1,k) + (n-1)|\{\sigma \in [k]^{n-1}(112) : \sigma \text{ has no 1's}\}| = \\ &= f_{112}(n-1,k) + (n-1)f_{112}(n-1,k-1). \end{aligned} \tag{2}$$

Combining (1) and (2), we get

$$f_{112}(n,k) - f_{112}(n,k-1) = f_{112}(n-1,k) + (n-1)f_{112}(n-1,k-1), \quad n \geq 1,\ k \geq 1. \tag{3}$$

The initial values are $f_{112}(n,0) = \delta_{n0}$ for all $n \geq 0$ and $f_{112}(0,k) = 1$, $f_{112}(1,k) = k$ for all $k \geq 0$.

Therefore, multiplying (6) by y^k and summing over k, we get

$$G_{112}(n;y) - \delta_{n0} - yG_{112}(n;y) = G_{112}(n-1;y) - \delta_{n-1,0} + (n-1)yG_{112}(n-1;y), \quad n \geq 1,$$

hence,

$$(1-y)G_{112}(n;y) = (1+(n-1)y)G_{112}(n-1;y), \quad n \geq 2.$$

Therefore,

$$G_{112}(n;y) = \frac{1+(n-1)y}{1-y}\, G_{112}(n-1;y), \quad n \geq 2. \tag{4}$$

Also, $G_{112}(0;y) = \dfrac{1}{1-y}$ and $G_{112}(1;y) = \dfrac{y}{(1-y)^2}$, so applying the previous equation repeatedly yields

$$G_{112}(n;y) = \frac{y(1+y)(1+2y)\cdots(1+(n-1)y)}{(1-y)^{n+1}}. \tag{5}$$

We have

$$\begin{aligned} \frac{1}{y} Numer(G_{112}(n;y)) &= (1+y)(1+2y)\cdots(1+(n-1)y) = y^n \prod_{j=0}^{n-1} \left(\frac{1}{y} + j\right) = \\ &= y^n \sum_{k=0}^{n} c(n,k) \left(\frac{1}{y}\right)^k = \sum_{k=0}^{n} c(n,k) y^{n-k} = \sum_{k=0}^{n} c(n,n-k) y^k, \end{aligned}$$

where $c(n,j)$ is the signless Stirling number of the first kind, and

$$yDenom(G_{112}(n;y)) = \frac{y}{(1-y)^{n+1}} = \sum_{k=0}^{\infty} \binom{n+k-1}{n} y^k,$$

so $f(n,k)$ is the convolution of the two coefficients:

$$f_{112}(n,k) = \left(c(n,n-k) * \binom{n+k-1}{n}\right) = \sum_{j=0}^{k} \binom{n+k-j-1}{n} c(n,n-j).$$

Thus, we have a new and improved version of Theorem 2.

Theorem 3. *Patterns 112 and 121 are Wilf-equivalent, and*

$$\begin{aligned} f_{121}(n,k) = f_{112}(n,k) &= \sum_{j=0}^{k} \binom{n+k-j-1}{n} c(n,n-j), \\ F_{121}(x,y) = F_{112}(x,y) &= \frac{1}{1-y} \cdot \left(\frac{1-y}{1-y-xy}\right)^{1/y}. \end{aligned} \tag{6}$$

We note that this is the first time that Stirling numbers appear in enumeration of words (or permutations) with forbidden patterns.

Proof. The first formula is proved above. The second formula can be obtained as the exponential generating function of $G_{112}(n;y)$ from the recursive equation (4). Alternatively, multiplying the recursive formula (3) by $x^n/n!$ and summing over n yields

$$\frac{d}{dx}F_{112}(x;k) = F_{112}(x;k) + (1+x)\frac{d}{dx}F_{112}(x;k-1).$$

Multiplying this by y^k and summing over $k \geq 1$, we obtain

$$\frac{d}{dx}F_{112}(x,y) = \frac{1}{1-y-yx}F_{112}(x,y).$$

Solving this equation together with the initial condition $F_{112}(0,y) = \dfrac{1}{1-y}$ yields the desired formula. $\square$

We will now prove the exact formula (6) bijectively. As it turns out, a little more natural bijective proof of the same formula obtains for $f_{221}(n,k)$, an equivalent result since $221 = c(112)$. This bijective proof is suggested by equation (3) and by the fact that $c(n,n-j)$ enumerates permutations of n letters with $n-j$ right-to-left minima (i.e. with j right-to-left nonminima), and $\binom{n+k-j-1}{n}$ enumerates nondecreasing strings of length n on letters in $\{0,1,\dots,k-j-1\}$.

Given a permutation $\pi \in S_n$ which has $n-j$ right-to-left minima, we will construct a word $\sigma \in [j+1]^n(221)$ with certain additional properties to be discussed later. The algorithm for this construction is as follows.

Algorithm 1. (1) Let $d = (d_1,\dots,d_n)$, where

$$d_r = \begin{cases} 0, & \text{if } r \text{ is a right-to-left minimum in } \pi, \\ 1, & \text{otherwise.} \end{cases}$$

(2) Let $s = (s_1,s_2,\dots,s_n)$, where $s_r = 1 + \sum_{i=1}^{r} d_r$, $r = 1,\dots,n$.
(3) Let $\sigma = \pi \circ s$ (i.e. $\sigma_r = s_{\pi(r)}$, $r = 1,\dots,n$). This is the desired word σ.

Example 1. Let $\pi = 621/93/574/8/10 \in S_{10}$. Then $n-j = 5$, so $j+1 = 6$, $d = 0100111010$, $s = 1222345566$, so the corresponding word $\sigma = 4216235256 \in [6]^{10}(221)$.

Note that each letter s_r in σ is in the same position as that of r in π, i.e. $\pi^{-1}(r)$.

Let us show that our algorithm does indeed produce a word $\sigma \in [j+1]^n(221)$.

Since π has $n-j$ right-to-left minima, only j of the d_r's are 1s, the rest are 0s. The sequence $\{s_r\}$ is clearly nondecreasing and its maximum, $s_n = 1 + 1 \cdot j = j+1$. Thus, $\sigma \in [j+1]^n$ and σ contains all letters from 1 to $j+1$.

Suppose now σ contains an occurrence of the pattern 221. This means π contains a subsequence bca or cba, $a < b < c$. On the other hand, $s_b = s_c$, so $0 = s_c - s_b = \sum_{r=b+1}^{c} d_r$, hence $d_c = 0$ and c must be a right-to-left minimum. But $a < c$ is to the right of c, so c is not a right-to-left minimum. Contradiction. Therefore, σ avoids pattern 221.

Thus, $\sigma \in [j+1]^n(221)$ and contains all letters 1 through $j+1$. Moreover, the leftmost (and *only* the leftmost) occurrence of each letter (except 1) is to the left of some smaller letter. This is because $s_b = s_{b-1}$ means $d_b = 0$, that is b is a right-to-left minimum, i.e. occurs to the right of all smaller letters. Hence, s_b is also to the right of all smaller letters, i.e. is a right-to-left minimum of σ. On the other hand, $s_b > s_{b-1}$ means $d_b = 1$, that is b is not a right-to-left minimum of π, so s_b is not a right-to-left minimum of σ.

It is easy to construct an inverse of Algorithm 1. Assume we are given a word σ as above. We will construct a permutation $\pi \in S_n$ which has $n-j$ right-to-left minima.

Algorithm 2.

(1) Reorder the elements of σ in nondecreasing order; call the resulting string s.
(2) Let $\pi \in S_n$ be the permutation such that $\sigma_r = s_{\pi(r)}$, $r = 1, \ldots, n$, given that $\sigma_a = \sigma_b$ (i.e. $s_{\pi(a)} = s_{\pi(b)}$) implies $\pi(a) < \pi(b) \Leftrightarrow a < b$). In other words, π is monotone increasing on positions of equal letters. Then π is the desired permutation.

Example 2. Let $\sigma = 4216235256 \in [6]^{10}(221)$ from our earlier example (so $j+1 = 6$). Then $s = 1222345566$, so looking at positions of 1s, 2s, etc., 6s, we get

$$\begin{aligned} \pi(1) &= 6 \\ \pi(\{2,5,8\}) &= \{2,3,4\} &&\Longrightarrow \quad \pi(2) = 2,\ \pi(5) = 3,\ \pi(8) = 4 \\ \pi(3) &= 1 \\ \pi(\{4,10\}) &= \{9,10\} &&\Longrightarrow \quad \pi(9) = 4,\ \pi(10) = 10 \\ \pi(6) &= 5 \\ \pi(\{7,9\}) &= \{7,8\} &&\Longrightarrow \quad \pi(7) = 7,\ \pi(9) = 8. \end{aligned}$$

Hence, $\pi = (6, 2, 1, 9, 3, 5, 7, 4, 8, 10)$ (in the one-line notation, not the cycle notation) and π has $n-j$ right-to-left minima: 10, 8, 4, 3, 1.

Note that the position of each s_r in σ is $\pi^{-1}(r)$, i.e. again the same as r has in π. Therefore, we conclude as above that π has $j+1-1 = j$ right-to-left nonminima, hence, $n-j$ right-to-left minima. Furthermore, the same property implies that Algorithm 2 is the inverse of Algorithm 1.

Note, however, that more than one word in $[k]^n(221)$ may map to a given permutation $\pi \in S_n$ with exactly $n-j$ right-to-left minima. We only need require that just the letters corresponding to the right-to-left nonminima of π be to the left of a smaller letter (i.e. not at the end) in σ. Values of 0 and 1 of d_r in Step 1 of Algorithm 1 are minimal increases required to recover back the permutation π with Algorithm 2. We must have $d_r \geq 1$ when we have to increase s_r, that is when s_r is not a right-to-left minimum of σ, i.e. when r is not a right-to-left minimum of π. Otherwise, we don't have to increase s_r, so $d_r \geq 0$.

Let $\sigma \in [k]^n(221)$, $\pi = Alg2(\sigma)$, $\tilde{\sigma} = Alg1(\pi) = Alg1(Alg2(\sigma)) \in [j+1]^n(221)$, and $\eta = \sigma - \tilde{\sigma}$ (vector subtraction). Note that $e_r = s_r(\sigma) - s_r(\tilde{\sigma}) \geq 0$ does not decrease (since $s_r(\sigma)$ cannot stay the same if $s_r(\tilde{\sigma})$ is increased by 1) and $0 \leq e_1 \leq \ldots \leq e_n \leq k-j-1$.

Since position of each e_r in η is the same as position of s_r in σ (i.e. $\eta_a = e_{\pi(a)}$, $e = e_1 e_2 \dots e_n$), the number of such sequences η is the number of nondecreasing sequences e of length n on letters in $\{0, \dots, k-j-1\}$, which is $\binom{n+k-j-1}{n}$.

Thus, $\sigma \in [k]^n(221)$ uniquely determines the pair (π, e), and vice versa. This proves the formula (6) of Theorem 3.

All of the above lets us state the following

Theorem 4. *There are 3 Wilf classes of multipermutations of length 3, with representatives 123, 112 and 111.*

4. Pairs of 3-letter patterns

There are 8 symmetric classes of pairs of 3-letters words, which are

$$\{111,112\},\{111,121\},\{112,121\},\{112,122\},\{112,211\},\{112,212\},\{112,221\},\{121,212\}.$$

Theorem 5. *The pairs* $\{111, 112\}$ *and* $\{111, 121\}$ *are Wilf equivalent, and*

$$F_{111,121}(x,y) = F_{111,112}(x,y) = \frac{e^{-x}}{1-y} \cdot \left(\frac{1-y}{1-y-xy}\right)^{1/y},$$

$$f_{111,112}(n,k) = \sum_{i=0}^{n} \sum_{j=0}^{k} (-1)^{n-i} \binom{n}{i} \binom{k+i-j-1}{i} c(i, i-j).$$

Proof. To prove equivalence, notice that the bijection $\rho : [k]^n(121) \to [k]^n(112)$ preserves the number of excess copies of each letter and that avoiding pattern 111 is the same as having at most 1 excess letter j for each $j = 1, \dots, k$. Thus, restriction of ρ to words with ≤ 1 excess letter of each kind yields a bijection $\rho \rightharpoonup_{111}$: $[k]^n(111, 121) \to [k]^n(111, 112)$.

Let $\alpha \in [k]^n(111, 112)$ contain i copies of letter 1. Since α avoids 111, we see that $i \in \{0, 1, 2\}$. Corresponding to these three cases, the number of such words α is $f_{111,112}(n, k-1)$, $nf_{111,112}(n-1, k-1)$ or $(n-1)f_{111,112}(n-2, k-1)$, respectively. Therefore,

$$f_{111,112}(n,k) = f_{111,112}(n,k-1) + nf_{111,112}(n-1,k-1) + (n-1)f_{111,112}(n-2,k-1),$$

for $n, k \geq 1$. Also, $f_{111,112}(n, 0) = \delta_{n0}$ and $f_{111,112}(0, k) = 1$, hence

$$F_{111,112}(x;k) = (1+x)F_{111,112}(x;k-1) + \int xF_{111,112}(x;k-1)dx,$$

where $f_{111,112}(0, k) = 1$. Multiply the above equation by y^k and sum over all $k \geq 1$ to get

$$F_{111,112}(x,y) = c(y)e^{-x} \cdot \left(\frac{1-y}{1-y-xy}\right)^{1/y},$$

which, together with $F_{111,112}(0, y) = \dfrac{1}{1-y}$, yields the generating function.

Notice that $F_{111,112}(x, y) = e^{-x}F_{112}(x, y)$, hence, $F_{111,112}(x; k) = e^{-x}F_{112}(x; k)$, so $f_{111,112}(n, k)$ is the exponential convolution of $(-1)^n$ and $f_{112}(n, k)$. This yields the second formula. □

Theorem 6. *Let* $H_{112,121}(x; k) = \sum_{n\geq 0} f_{112,121}(n, k)x^n$. *Then for any* $k \geq 1$,

$$H_k(x) = \frac{1}{1-x} H_{112,121}(x;k-1) + x^2 \frac{d}{dx} H_{112,121}(x;k-1),$$

and $H_{112,121}(x; 0) = 1$.

Proof. Let $\alpha \in [k]^n(112,121)$ such that contains j letters 1. Since α avoids 112 and 121, we have that for $j>1$, all j copies of letter 1 appear in α in positions $n-j+1$ through n. When $j=1$, the single 1 may appear in any position. Therefore,

$$f_{112,121}(n;k) = f_{112,121}(n;k-1) + nf_{112,121}(n-1,k-1) + \sum_{j=2}^{n} f_{112,121}(n-j;k-1),$$

which means that

$$\begin{aligned} f_{112,121}(n;k) = f_{112,121}(n-1;k) + f_{112,121}(n;k-1) \\ + (n-1)f_{112,121}(n-1,k-1) - (n-2)f_{112,121}(n-2,k-1). \end{aligned}$$

We also have $f_{112,121}(n;0)=1$, hence it is easy to see the theorem holds. □

Theorem 7. *Let $H_{112,211}(x;k) = \sum_{n\geq 0} f_{112,211}(n,k)x^n$. Then for any $k \geq 1$,*

$$H_{112,211}(x;k) = (1+x+x^2)H_{112,211}(x;k-1) + \frac{x^3}{1-x} + \frac{d}{dx}H_{112,211}(x;k-1),$$

and $H_{112,211}(x;0) = 1$.

Proof. Let $\alpha \in [k]^n(112,211)$ such that contains j letters 1. Since α avoids 112 and 211 we have that $j = 0,1,2,n$. When $j=2$, the two 1's must at the beginning and at the end. Hence, it is easy to see that for $j=0,1,2,n$ there are $f_{112,211}(n;k-1)$, $nf_{112,211}(n-1;k-1)$, $f_{112,211}(n-2;k-1)$ and 1 such α, respectively. Therefore,

$$f_{112,211}(n;k) = f_{112,211}(n;k-1) + nf_{112,211}(n-1,k-1) + f_{112,211}(n-2,k-1) + \delta_{n\geq 3}.$$

We also have $f_{112,121}(n;0)=1$, hence it is easy to see the theorem holds. □

Theorem 8. *Let $a_{n,k} = f_{112,212}(n,k)$, then*

$$a_{n,k} = a_{n,k-1} + \sum_{d=1}^{n}\sum_{r=0}^{k-1}\sum_{j=0}^{n-d} a_{j,r}a_{n-d-j,k-1-r}$$

and $a_{0,k}=1$, $a_{n,1}=1$.

Proof. Let $\alpha \in [k]^n(112,212)$ have exactly d letters 1. If $d=0$, there are $a_{n,k-1}$ such α. Let $d \geq 1$, and assume that $\alpha_{i_d} = 1$ where $d = 1,2,\ldots j$. Since α avoids 112, we have $i_2 = n+2-d$ (if $d=1$, we define $i_2 = n+1$), and since α avoids 212 we have that α_a, α_b are different for all $a < i_1 < b < i_2$. Therefore, α avoids $\{112,212\}$ if and only if $(\alpha_1,\ldots,\alpha_{i_1-1})$, and $(\alpha_{i_1+1},\ldots,\alpha_{i_2-1})$ are $\{112,212\}$-avoiding. The rest is easy to obtain. □

Theorem 9.

$$f_{112,221}(n,k) = \sum_{j=1}^{k} j\cdot j!\binom{k}{j}$$

for all $n \geq k+1$,

$$f_{112,221}(n,k) = n!\binom{k}{n} + \sum_{j=1}^{n-1} j\cdot j!\binom{k}{j}$$

for all $k \geq n \geq 2$, and $f_{112,221}(0,k)=1$, $f_{112,221}(1,k)=k$.

Proof. Let $\alpha \in [k]^n(112,221)$ and $j \leq n$ be such that $\alpha_1,\ldots,\alpha_j$ are all distinct and j is maximal. Clearly, $j \leq k$. Since α avoids $\{112,221\}$ and j is maximal, we get that the letters $\alpha_{j+1},\ldots,\alpha_n$, if any, must all be the same and equal to one of the letters $\alpha_1,\ldots,\alpha_j$. Hence, there are $j\cdot j!\binom{k}{j}$ such α if , for $j<n$ or $j=n>k$. For $j=n\leq k$, there are $n!\binom{k}{n}$ such α. Hence, summing over all possible $j=1,\ldots,k$, we obtain the theorem. □

Theorem 10.

$$f_{121,212}(n,k)=\sum_{j=0}^{k} j!\binom{k}{j}\binom{n-1}{j-1}$$

for $k \geq 0$, $n \geq 1$, and $f_{121,212}(0,k)=1$ for $k \geq 0$.

Proof. Let $\alpha \in [k]^n(121,212)$ contain exactly j distinct letters. Then all copies of each letter 1 through j must be consecutive, or α would contain an occurrence of either 121 or 212. Hence, α is a concatenation of j constant strings. Suppose the i-th string has length $n_i > 0$, then $n = \sum_{i=1}^{j} n_i$. Therefore, to obtain any $\alpha \in [k]^n(121,212)$, we can choose j letters out of k in $\binom{k}{j}$ ways, then choose any ordered partition of n into j parts in $\binom{n-1}{j-1}$ ways, then label each part n_i with a distinct number $l_i \in \{1,\dots,j\}$ in $j!$ ways, then substitute n_i copies of letter l_i for the part n_i $(i = 1,\dots,j)$. This yields the desired formula. □

Unfortunately, the case of the pair $(112,122)$ still remains unsolved.

5. Some triples of 3-letter patterns

Theorem 11.

$$F_{112,121,211}(x;k)=1+\frac{(e^x-1)((1+x)^k-1)}{x},$$

$$f_{112,121,211}(n,k)=\begin{cases}\displaystyle\sum_{j=1}^{n}\frac{1}{j!}\binom{n+1}{j}\binom{k}{n+1-j}, & n\geq 1,\\ 1, \quad n=0.\end{cases}$$

Proof. Let $\alpha \in [k]^n(112,121,211)$ contain j letters 1. For $j \geq 2$, there are no letters between the 1's, to the left of the first 1 or to the right of the last 1, hence $j = n$. For $j = 1$, $j = 0$ it is easy to see from definition that there are $nf_{112,121,211}(n-1,k-1)$ and $f_{112,121,211}(n,k-1)$ such α, respectively. Hence,

$$f_{112,121,211}(n,k)=f_{112,121,211}(n,k-1)+nf_{112,121,211}(n-1,k-1)+1,$$

for $n,k \geq 2$. Also, $a(n,1)=a(n,0)=1$, $a(0,k)=1$, and $a(1,k)=k$. If we let $b(n,k)=f_{112,121,211}(n,k)/n!$, then

$$b(n,k)=b(n,k-1)+b(n-1,k-1)+\frac{1}{n!}.$$

Let $b_k(x)=\sum_{n\geq 0} b(n,k)x^n$, then it is easy to see that $b_k(x)=(1+x)b_{k-1}(x)+e^x-1$. Since we also have $b_0(x)=e^x$, the theorem follows by induction. □

6. Some patterns of arbitrary length

6.1. **Pattern** $11\dots1$. Let us denote by $\langle a\rangle_l$ the word consisting of l copies of letter a.

Theorem 12. *For any $l,k \geq 0$,*

$$F_{\langle 1\rangle_l}(x;k)=\left(\sum_{j=0}^{l-1}\frac{x^j}{j!}\right)^k.$$

Proof. Let $\alpha \in [k]^n(\langle 1\rangle_l)$ contain j letters 1. Since α avoids $\langle 1\rangle_l$, we have $j \leq l-1$. If α contains exactly j letters of 1, then there are $\binom{n}{j} f_{\langle 1\rangle_l}(n-j,k-1)$ such α, therefore

$$f_{\langle 1\rangle_l}(n,k)=\sum_{j=0}^{l-1}\binom{n}{j} f_{\langle 1\rangle_l}(n-j,k-1).$$

We also have $f_{\langle 1\rangle_l}(n,k) = k^n$ for $n \le l-1$, hence it is easy to see the theorem holds. $\square$

In fact, [CS] shows that we have

$$f_{\langle 1\rangle_l}(n,k) = \sum_{i=1}^{n} M_2^{l-1}(n,i)(k)_i,$$

where $M_2^{l-1}(n,i)$ is the number of partitions of an n-set into i parts of size $\le l-1$.

6.2. **Pattern** $11\ldots121\ldots11$**.** Let us denote $v_{m,l} = 11\ldots121\ldots11$, where m (respectively, l) is the number of 1's on the left (respectively, right) side of 2 in $v_{m,l}$. In this section we prove the number of words in $[k]^n(v_{m,l})$ is the same as the number of words in $[k]^n(v_{m+l,0})$ for all $m, l \ge 0$.

Theorem 13. *Let $m, l \ge 0$, $k \ge 1$. Then for $n \ge 1$,*

$$f_{v_{m,l}}(n+1,k) - f_{v_{m,l}}(n,k) = \sum_{j=0}^{m+l-1} \binom{n}{j} f_{v_{m,l}}(n+1-j,k-1).$$

Proof. Let $\alpha \in [k]^n(v_{m,l})$ contain exactly j letters 1. Since the 1's cannot be part of an occurrence of $v_{m,l}$ in α when $j \le m+l-1$, these 1's can be in any j positions, so there are $\binom{n}{j} f_{v_{m,l}}(n,k-1)$ such α. If $j \ge m+l$, then the m-th through $(j-l+1)$-st (l-th from the right) 1's must be consecutive letters in α (with the convention that the 0-th 1 is the beginning of α and $(j+1)$-st 1 is the end of α). Hence, there are $\binom{n-j+m+l-1}{m+l-1} f_{v_{m,l}}(n-j,k-1)$ such α, and hence

$$f_{v_{m,l}}(n;k) = \sum_{j=0}^{m+l-1} \binom{n}{j} f_{v_{m,l}}(n-j,k-1) + \sum_{j=m+l}^{n} \binom{n-j+m+l-1}{m+l-1} f_{v_{m,l}}(n-j,k-1).$$

Hence for all $n \ge 1$,

$$f_{v_{m,l}}(n+1,k) - f_{v_{m,l}}(n,k) = \sum_{j=0}^{m+l-1} \binom{n}{j} f_{v_{m,l}}(n+1-j,k-1).$$

$\square$

An immediate corollary of Theorem 13 is the following.

Corollary 14. *Let $m, l \ge 0$, $k \ge 0$. Then for $n \ge 0$*

$$f_{v_{m,l}}(n,k) = f_{v_{m+l,0}}(n,k).$$

In other words, all patterns $v_{m,l}$ with the same $m+l$ are Wilf-equivalent.

Proof. We will give an alternative, bijective proof of this by generalizing our earlier bijection $\rho : [k]^n(121) \to [k]^n(112)$. Let $\alpha \in [k]^n(v_{m,l})$. Recall that α_j is a word obtained by deleting all letters 1 through j from α (with $\alpha_0 := \alpha$).

Suppose that α contains i letters $j+1$. Then all occurrences of $j+1$ from m-th through $(i-l+1)$-st, if any (i.e. if $j \ge m+l$), must be consecutive letters in α_j. We will denote as *excess* j's the $(m+1)$-st through $(i-l+1)$-st copies of j when $l > 0$, and m-th through i-th copies of j when $l = 0$.

Suppose that $m+l = m'+l'$. Then the bijection $\rho_{m,l;m',l'} : [k]^n(v_{m,l}) \to [k]^n(v_{m',l'})$ is an algorithm of k steps. Given a word $\alpha \in [k]^n(v_{m,l})$, say it yields a word $\alpha^{(j)}$ after Step j, with $\alpha^{(0)} := \alpha$. Then Step j $(1 \le j \le k)$ is as follows:

Step j.

(1) Cut the block of excess j's from ${\alpha^{(j-1)}}_{j-1}$ (which is immediately after the m-th occurrence of j), then insert it immediately after the m'-th occurrence of j if $l' > 0$, or at the end of ${\alpha^{(j-1)}}_{j-1}$ if $l' = 0$.
(2) Insert letters 1 through $j-1$ into the resulting string in the same positions they are in $\alpha^{(j-1)}$ and call the combined string $\alpha^{(j)}$.

Clearly,
$${\alpha^{(j)}}_j = {\alpha^{(j-1)}}_j = \ldots = {\alpha^{(0)}}_j \alpha_j$$
and at Step j, the j's are rearranged so that no j can be part of an occurrence of $v_{m',l'}$. Also, positions of letters 1 through $j-1$ are the same in $\alpha^{(j)}$ and $\alpha^{(j-1)}$, hence, no letter from 1 to j can be part of $v_{m',l'}$ in $\alpha^{(j)}$ by induction. Therefore, $\alpha^{(k)} \in [k]^n(v_{m',l'})$ as desired.

Clearly, this map is invertible, and $\rho_{m',l';m,l} = (\rho_{m,l;m',l'})^{-1}$. This ends the proof. □

Theorem 15. *Let $p \geq 1$ and $d_p(f(x)) = \int \ldots \int f(x)dx \ldots dx$ (and we define $d_0(f(x)) = f(x)$)). Then for any $k \geq 1$,*
$$F_{v_{p,0}}(x;k) - \int F_{v_{p,0}}(x;k)dx = \sum_{j=0}^{p-1}\left((-1)^j d_p(F_{v_{p,0}}(x;k-1)) \sum_{i=0}^{p-1-j} \frac{x^i}{i!}\right),$$
and $F_{v_{p,0}}(x;1) = e^x$, $F_{v_{p,0}}(0;k) = 1$.

Proof. By definition, we have $f_{v_{p,0}}(n,1) = 1$ for all $n \geq 0$ so $F_{v_{p,0}}(x;1) = e^x$. On the other hand, Theorem 13 yields immediately the rest of this theorem. □

Example 3. *For $p = 1$, Theorem 15 yields*
$$\sum_{n\geq 0} |[k]^n(12)| \frac{x^n}{n!} = e^x \sum_{j=0}^{k-1} \binom{k-1}{j} \frac{x^j}{j!},$$
which means that, for any $n \geq 0$
$$|[k]^n(12)| = \binom{n+k-1}{k-1}.$$
(cf. Section 2.)

Example 4. *For $p = 2$, Theorem 15 yields*
$$F_{112}(x;k) = e^x \cdot \int (1+x)e^{-x} F_{112}(x;k-1)dx,$$
and $F_{112}(x;0) = 1$.

Corollary 16. *For any $p \geq 0$*
$$F_{v_{p,0}}(x;2) = e^x \sum_{j=0}^{p} \frac{x^j}{j!}.$$

Proof. From Theorem 15, we immediately get that
$$F_{v_{p,0}}(x;2) - \int F_{v_{p,0}}(x;2)dx = e^x \sum_{j=0}^{p-1} (-1)^j \sum_{i=0}^{p-1-j} \frac{x^i}{i!},$$
which means that
$$e^x \frac{d}{dx}\left(e^{-x} F_{v_{p,0}}(x;2)\right) = e^x \sum_{j=0}^{p-1} \frac{x^j}{j!},$$
hence the corollary holds. □

References

[AH] M. Albert, R. Aldred, M.D. Atkinson, C. Handley, D. Holton, Permutations of a multiset avoiding permutations of length 3, preprint.

[BS] E. Babson and E. Steìngrimsson, Generalized permutation patterns and a classification of the Mahonian statistics, *Séminaire Lotharingien de Combinatoire*, **44** (2000), Article B44b.

[B] A. Burstein, Enumeration of words with forbidden patterns, Ph.D. thesis, University of Pennsylvania, 1998.

[CS] J.Y. Choi, J.D.H. Smith, Multi-restricted numbers and powers of permutation representations, preprint.

[Kn] D.E. Knuth, The Art of Computer Programming, 2nd edition, Addison Wesley, Reading, MA, 1973.

[M] T. Mansour, Pattern avoidance in coloured permutations, *Séminaire Lotharingien de Combinatoire*, **46** (2001), Article B46g.

[R] A. Regev. Asymptotics of the number of k-words with an ℓ-descent, *Electronic J. of Combinatorics* **5** (1998), #R15.

[S] R. Simion, Combinatrial statistics on type-B analogies of noncrossing partitions and restricted permutations, *Electronic J. of Combin.* **7** (2000) #R9.

[SS] R. Simion, F.W. Schmidt, Restricted Permutations, *Europ. J. of Combinatorics* **6** (1985), 383–406.

Department of Mathematics, Iowa State University, Ames, IA 50011-2064 USA
E-mail address: `burstein@math.iastate.edu`

LaBRI, Université Bordeaux, 351 cours de la Libération, 33405 Talence Cedex, France
E-mail address: `toufik@labri.fr`

EXPLICIT FORMULAE FOR SOME KAZHDAN-LUSZTIG POLYNOMIALS FOR S_n

FABRIZIO CASELLI

Abstract. We consider Kazhdan-Lusztig polynomials $P_{u,v}(q)$ indexed by permutations u, v having particular forms with regard to their monotonicity patterns. First we present a solution to the problem of computing $P_{u,v}(q)$ when $u, v \in S_n$ satisfy $u^{-1}(n) - v^{-1}(n) \leq 3$ by reducing it to a problem for Kazhdan-Lusztig polynomials for S_{n-1}. As a corollary we obtain an "almost" explicit expression for $P_{e,4\ldots n123}$. We also show explicit formulae for $P_{e,v}(q)$ when v is obtained by inserting n in a permutation of S_{n-1} that is allowed to rise only in the first and in the last position generalizing a theorem due to B. Shapiro, M. Shapiro and A. Vainshtein. As an application of this result we write down explicit formulae for $P_{e,\sigma(n-2)\sigma(n-1)\sigma(n)\, n-3\ldots 4\, \tau(1)\tau(2)\tau(3)}$ for $(\sigma, \tau) \in S_3 \times S_3 \setminus (e, e)$, where σ act on the set $\{n-2, n-1, n\}$ in the natural way, establishing, in particular, two conjectures due to F. Brenti and R. Simion. Our proofs are based on the fact that the polynomials under consideration (together with some others) satisfy some nice recurrence relations. Moreover they are purely combinatorial and make no use of the geometry of the corresponding Schubert varieties.

Résumé. Nous considérons polynômes de Kazhdan-Lusztig $P_{u,v}(q)$ indexés par permutations u, v avec des formes particulières par rapport à leurs andaments de monotonicité. D'abord nous proposons une solution au problème du calcul de $P_{u,v}(q)$ quand $u, v \in S_n$ vérifient $u^{-1}(n) - v^{-1}(n) \leq 3$ en le réduisant à un problème pour les polynômes de Kazhdan-Lusztig pour S_{n-1}. Comme première application de ce résultat nous obtiendrons une expression "presque" explicite pour $P_{e,4\ldots n123}$. Nous montrerons aussi des formules explicites pour $P_{e,v}(q)$ quand v est obtenu en insérant n dans une permutation de S_{n-1} qui a au plus 2 montées dans la première et la dernière position. Ce résultat généralise un théorème de B. Shapiro, M. Shapiro et A. Vainshtein. Comme application de ce résultat nous montrerons des formules explicites pour $P_{e,\sigma(n-2)\sigma(n-1)\sigma(n)\, n-3\ldots 4\, \tau(1)\tau(2)\tau(3)}$ avec $(\sigma, \tau) \in S_3 \times S_3 \setminus (e, e)$, oùσ agit sur l'ensemble $\{n-2, n-1, n\}$ de la façon naturelle, en prouvant deux conjectures de F. Brenti et R. Simion. Nos démonstrations sont fondeès sur le fait que les polynômes que nous considérons vérifient de bonnes relations de recurrence. De plus, elles sont purement combinatoires et n'utilisent pas la geometrie des variétés de Schubert correspondentes.

1. Introduction

In [KL1] Kazhdan and Lusztig defined, for every Coxeter system W, a family of polynomials, parametrized by pairs of elements of W, which have become known as the Kazhdan-Lusztig polynomials of W. These polynomials are intimately related to the Bruhat order of W and have proven to be of fundamental importance in representation theory and in the geometry of the Schubert varieties. We focus our attention to the case of the symmetric group, where these polynomials can be computed with some particular purely combinatorial rules (see, for example, [Br2, Corollary 4.6]. Despite the rather elementary recursion relations they satisfy, these polynomials are in general quite difficult to compute explicitly. In fact the only families of Kazhdan-Lusztig polynomials that are known correspond to situations where the geometry of the corresponding Schubert varieties is easier (see, for example, [LSc],[Bo], [P] and [SSV, Theorems 1 and 2]) , where the interval $[u, v]$ has some special shape (see, for example, [Br1, Corollaries 6.8 and 6.9] or when the shape of the indexing permutation lead in some natural way to the use of induction (see [BS, Corollary 3.2 and Theorem 3.3]). This work gives results in the direction of explicit formulae for the Kazhdan-Lusztig polynomials of the symmetric group when the indexing permutations are of particular forms.

The main results are the following. First we reduce the calculation of $P_{u,v}(q)$ when $u, v \in S_n$ satisfy $u^{-1}(n) - v^{-1}(n) \leq 3$ to an (easier) problem in S_{n-1}. Next we will show some applications of this result. In particular we obtain a formula for $P_{1\ldots n-3\,\sigma(n-2)\,\sigma(n-1)\,\sigma(n)\,,4\ldots n123}$, where σ is any permutation of the set $\{n-2, n-1, n\}$ that shows how these families of polynomials are intimately related among each other. Then we will focus our attention to permutations in S_n that are obtained from an element of S_{n-1} allowed to rise only in the first and in the last position by inserting n (or 1) anywhere in its complete notation. We write down some recurrence relations they satisfy and we obtain explicit formulae from these relations. Finally, as an application of this result, we find explicit formulae for $P_{e,\sigma(n-2)\sigma(n-1)\sigma(n)\,n-3\ldots 4\,\tau(1)\tau(2)\tau(3)}$ where $(\sigma, \tau) \in S_3 \times S_3 \setminus (e, e)$ act on the set $\{n-2, n-1, n, 1, 2, 3\}$ in the most natural way, establishing, in particular, two conjectures due to F. Brenti and R. Simion (see [BS, Conjecture 4.1 and 4.2]). The proofs rely on the special shape of the permutation under consideration that will allow us to deduce some easy recursions satisfied by these polynomials (together with some other Kazhdan-Lusztig polynomials) with no use of geometry.

In §2 we fix the notation and we provide the necessary preliminaries on the Bruhat order of S_n and some known results about the Kazhdan-Lusztig polynomials which are used in the proofs of our results.

2. Notation and preliminaries

In this section we collect some definitions and results that have been used in the proofs of this work.

We let $\mathbb{N} := \{0, 1, 2, 3, \ldots\}$ be the set of non-negative integers and for $a \in \mathbb{N}$ we let $[a] := \{1, 2, \ldots, a\}$ (where $[0] = \emptyset$) and for $a \in \mathbb{R}$, $\lfloor a \rfloor$ is the largest integer $\leq a$. Given $n, m \in \mathbb{N}$, $n \leq m$, we let $[n, m] := \{n, n+1, \ldots, m\}$. We write $S = \{a_1, \ldots, a_r\}_<$ to mean that $S = \{a_1, \ldots, a_r\}$ and $a_1 < \cdots < a_r$

For $i \in \mathbb{Z}$ we denote by

$$[i]_q := \sum_{j=0}^{i-1} q^j$$

so that $[n]_q = 0$ if $n \leq 0$. Given a polynomial $P(q)$ and $i \in \mathbb{N}$ we denote by $[q^i](P(q))$ the coefficient of q^i in $P(q)$.

Given a set T we let $S(T)$ be the set of all bijections of T. In particular, $S_n = S([n])$ is the symmetric group on n elements and we denote by e the identity of S_n. If $u \in S([n, n+k])$ for some $n, k \in \mathbb{N}$, then we write $u = u_1 u_2 \ldots u_{k+1}$ to mean that $u(n+i) = u_{i+1}$ for $i = 0, \ldots, k$, while we denote by s_i the transposition $(i, i+1)$. Given $\sigma, \tau \in S(T)$, we let $\sigma\tau := \sigma \circ \tau$, i.e. we compose permutations as functions, from right to left. Given $\sigma \in S_n$, the right descent set of σ is

$$D_R(\sigma) := \{i \in [n-1] : \sigma(i) > \sigma(i+1)\}$$

and the left descent set is

$$D_L(\sigma) := \left\{i \in [n-1] : \sigma^{-1}(i) > \sigma^{-1}(i+1)\right\}$$

and the length of σ is defined by the number of inversions:

$$\ell(\sigma) := \mathrm{inv}(\sigma) := \#\left\{(a, b) \in [n] \times [n] : a < b, \sigma(a) > \sigma(b)\right\}$$

For example, if $\sigma = 6\,3\,5\,2\,4\,1$ then $D_R(\sigma) = \{1,3,5\}$, $D_L(\sigma) = \{1,2,4,5\}$ and $\ell(\sigma) = 11$. Throughout this work we view S_n as a poset ordered by the strong Bruhat order. We are not going to define this order in the usual way (see [H, Section 5.9] for its definition), but we shall use the following characterization of it due to Ehresmann [E]. For $\sigma \in S_n$ and $j \in [n]$, let

$$\left\{\sigma^{j,1}, \dots, \sigma^{j,j}\right\}_< := \{\sigma(1), \dots, \sigma(j)\}$$

Theorem 2.1. *Let $\sigma, \tau \in S_n$. Then $\sigma \leq \tau$ if and only if $\sigma^{j,i} \leq \tau^{j,i}$ for all $1 \leq i \leq j \leq n-1$.*

We take the following fundamental result (see [H, Section 7.11] for a proof) as the definition of the Kazhdan-Lusztig polynomials:

Theorem 2.2. *There exists a unique family of polynomials $\{P_{u,v}(q),\, u,v \in S_n\} \subset \mathbb{Z}[q]$ such that:*
(1) $P_{u,v}(q) = 0$ if $u \not\leq v$
(2) $P_{u,v}(q) = 1$ if $u = v$
(3) If $u \leq v$ and $i \in D_R(v)$ then

$$P_{u,v}(q) = q^{1-c} P_{us_i, vs_i}(q) + q^c P_{u, vs_i}(q) - \sum_{\{z : i \in D(z)\}} q^{\frac{\ell(v)-\ell(z)}{2}} \mu(z, vs_i) P_{u,z}(q)$$

where, for $u, v \in S_n$,

$$\mu(u,w) := \begin{cases} \left[q^{\frac{1}{2}(\ell(w)-\ell(u)-1)}\right](P_{u,w}(q)) & \text{if } u < w \text{ and } \ell(w)-\ell(u) \text{ is odd} \\ 0 & \text{otherwise} \end{cases}$$

and $c = 1$ if $i \in D(u)$, and $c = 0$ otherwise.

The following result, whose proof can be found in [KL2], is a consequence of the role played by the Kazhdan-Lusztig polynomials in the geometry of Schubert varieties and no combinatorial proof of it is known so far :

Theorem 2.3. *Let $u, v \in S_n$. Then $P_{u,v}(q) \in \mathbb{N}[q]$.*

Two important consequences of Theorem 2.2 are the following:

Proposition 2.4. *Let $u, v \in S_n$ such that $u < v$. Then*

$$\deg(P_{u,v}(q)) \leq \frac{1}{2}(\ell(v) - \ell(u) - 1)$$

and, if $i \in D_R(v)$

$$P_{u,v}(q) = P_{us_i, v}(q)$$

It should be remarked that Theorem 2.2 and the second part of Proposition 2.4 can be reformulated in a similar way using left descents instead of right descents. An immediate consequence of Proposition 2.4 is the following

Corollary 2.5. *Let $z, w \in S_n$, $z \leq w$, be such that $\mu(z,w) \neq 0$ and $\ell(w) - \ell(z) > 1$. Then $D_R(z) \supseteq D_R(w)$ and $D_L(z) \supseteq D_L(w)$.*

Two other properties that we often use are the following (see [H, Theorem 7.9 and Corollary 7.14] for proofs:

Proposition 2.6. *Let $u, v \in S_n$. Then*

$$\begin{aligned} P_{u,v}(q) &= P_{u^{-1}, v^{-1}}(q) \\ &= P_{w_0 u w_0, w_0 v w_0}(q) \end{aligned}$$

where $w_0 = n \dots 2\,1$ is the longest element of S_n.

Let $w \in S_n$. We denote by $\overline{w}$ the permutation of S_{n-1} obtained from w by suppressing the value n from its notation. Then the following result can be obtained as a corollary of Theorem 2.2 and will be of fundamental importance in the rest of our work.

Proposition 2.7. *Let $u, v \in S_n$ such that n occurs in the same position in both u and v. Then*

$$P_{u,v}(q) = P_{\overline{u},\overline{v}}(q)$$

It should be mentioned that Proposition 2.7 can be stated in a "dual" version when 1 occur in the same position in both u and v.

The next result gives an explicit formula for $P_{u,v}$ when v is allowed to rise only in the first and in the last place (see [SSV]) and will be generalized in §4:

Theorem 2.8. *Let $u, v \in S_n$, $u \leq v$, be such that $[2, n-2] \subseteq D_R(v)$. Then*

$$P_{u,v}(q) = \begin{cases} 1 & \textit{if } v(1) < v(n) \textit{ or } v(n) \leq u(1) \textit{ or } v(1) \geq u(n) \\ 1 + q^{v(1)-v(n)} & \textit{otherwise} \end{cases}$$

We conclude this section with an easy characterization of the permutations that gives rise to Kazhdan-Lusztig polynomials equal to 1 (see [LSa] for a proof). Let $\tau \in S_m$ and $\sigma \in S_n$ with $n \geq m$. We say that σ avoids τ if there is no subsequence $1 \leq i_1 < \cdots < i_m \leq n$ such that

$$\sigma\left(i_{\tau(1)}\right) < \cdots < \sigma\left(i_{\tau(m)}\right)$$

Theorem 2.9. *Let $v \in S_n$. Then*

$$P_{u,v}(q) = 1 \ \forall u \leq v \Longleftrightarrow v \textit{ avoids both } 3412 \textit{ and } 4231$$

3. A reduction theorem

Definition. Let $u, v \in S_n$. Then we set

$$d(u, v) := u^{-1}(n) - v^{-1}(n)$$

Note that by Theorem 2.1, if $u \leq v$ we have $d(u, v) \geq 0$.

We are going to reduce the calculation of $P_{u,v}(q)$ to a problem for Kazhdan-Lusztig polynomials for S_{n-1} when $d(u, v) \leq 3$. We have already seen that if $d(u, v) = 0$ then $P_{u,v}(q) = P_{\overline{u},\overline{v}}(q)$, so we may focus our attention to the case $d(u, v) > 0$.

The next results, for $d(u, v) = 1$ or 2, represent a reformulation and a generalization of a theorem due to F. Brenti and R. Simion (see [BS, Theorem 3.1]).

Theorem 3.1. *Let $u, v \in S_n$ such that $u \leq v$ and $i = v^{-1}(n)$. Then*
(1) If $d(u, v) = 1$

$$P_{u,v}(q) = P_{\overline{u},\overline{v}}(q)$$

(2) If $d(u, v) = 2$

$$P_{u,v}(q) = \begin{cases} q^{1-c} P_{us_i,vs_i}(q) + q^c P_{u,vs_i}(q) & \textit{if } i+1 \notin D_R(v) \\ P_{\overline{u},\overline{v}}(q) & \textit{if } i+1 \in D_R(v) \end{cases}$$

where $c = 1$ if $i \in D(u)$ and $c = 0$ otherwise.

Note that the first part of Theorem 3.1 follows easily from Proposition 2.4 and Proposition 2.7.

Suppose now that $d(u, v) = 3$ and again set $i = v^{-1}(n)$. To fix the ideas we write

$$u := \ldots u_i u_{i+1} u_{i+2} n \ldots$$

and

$$v := \ldots n v_{i+1} v_{i+2} v_{i+3} \ldots$$

If $v_{i+2} > v_{i+3}$ then, by Proposition 2.4, we may swap u_{i+2} and n in u and hence we go back to the case $d(u,v) = 2$. So, with no lack of generality, we may suppose $v_{i+2} < v_{i+3}$, i.e. $i+2 \notin D_R(v)$. We would like to use Theorem 2.2 taking i as a right descent for v. The next result will allow us to simplify the sum in that formula in this case.

Proposition 3.2. *Let $u, v \in S_n$ such that $u \leq v$, $d(u,v) = 3$, $i = v^{-1}(n)$ and $i+2 \notin D_R(v)$. Then the application $F : z \mapsto \overline{z}$ establishes a bijection between*

$$\left\{ \begin{array}{c} z \in S_n \text{ such that } z \geq u \\ i \in D_R(z),\ \mu(z, vs_i) \neq 0 \\ \text{and } \ell(vs_i) - \ell(z) > 1 \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} z \in S_{n-1} \text{such that } z \geq \overline{u} \\ i, i+1 \in D_R(z) \\ \text{and } \mu(z, \overline{v}) \neq 0 \end{array} \right\}$$

Moreover, for z in the set of the left-hand side, we have $\mu(z, vs_i) = \mu(\overline{z}, \overline{v})$, $\ell(v) - \ell(z) = \ell(\overline{v}) - \ell(\overline{z}) + 3$ and $P_{u,z}(q) = P_{\overline{u},\overline{z}}(q)$.

We are now ready to state the main result of this section.

Theorem 3.3. *Let $u, v \in S_n$ such that $u \leq v$, $d(u,v) = 3$, $i = v^{-1}(n)$ and $i+2 \notin D_R(v)$. Then*

$$\begin{aligned} P_{u,v}(q) &= q^{1-c} P_{us_i, vs_i}(q) + q^c P_{u, vs_i}(q) - \sum_{\{z \in S_{n-1} : i, i+1 \in D_R(z)\}} q^{\frac{\ell(\overline{v}) - \ell(z) + 3}{2}} \mu(z, \overline{v}) P_{\overline{u}, z}(q) \\ &- \varepsilon_0 q P_{\overline{u}, \overline{v}}(q) - \varepsilon_1 q P_{\overline{u}, \overline{v} s_{i+1}} \end{aligned}$$

where

$$\varepsilon_j := \begin{cases} 0 & v_{i+1} < v_{i+j+2} \\ 1 & \text{otherwise} \end{cases}$$

for $j = 0, 1$ and, as usual, $c = 1$ if $i \in D_R(u)$ and $c = 0$ otherwise.

Note that the polynomials in the first two summands in the right-hand side of the previous formula are indexed by permutations x, y verifying $d(x,y) = 2$.

It should be mentioned that both Theorems 3.1 and 3.3 can also be stated in a "dual" version when $u, v \in S_n$ satisfy $\widetilde{d}(u,v) := v^{-1}(1) - u^{-1}(1) \leq 3$.

The next example shows us that, unfortunately, there could be many summands different from 0 in the previous sum.

Example 3.4. Let $n \geq 5$ and $v := 3 \ldots (n-2)\, n\, (n-1)\, 1\, 2$ and hence $\overline{v} = 3 \ldots (n-2)\, (n-1)\, 1\, 2$ and $u = e$. Then it is easy to check that for every $i \in [3, n-2]$, $(1,i)\overline{v}$ gives rise to a non-zero summand in Theorem 3.3.

Before showing the first application of Theorem 3.3, let us state the following:

Problem 3.5. Let $i, n \in \mathbb{N}$, $1 \leq i \leq n$. Find an explicit formula for

$$P_{e, i \ldots n\, 1 \ldots i-1}(q)$$

Problem 3.5 is trivial for $i = 1$ and it is an easy consequence of Theorem 2.9 for $i = 2$. In 1992 M. Haiman conjectured a formula to solve this problem for $i = 3$ which has been proved by F. Brenti and R. Simion in 2000 (see [BS], Corollary 3.2):

Theorem 3.6. *Let*

$$F_n(q) := \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n-i}{i} q^i$$

be the q-analogue of the Fibonacci number. Then $\forall n \geq 3$

$$P_{e, 3 \ldots n\, 1\, 2}(q) = F_{n-2}(q)$$

If we use Theorems 2.2 and 3.1 then we can deduce

$$P_{s_{n-1},3\ldots n\,1\,2}(q) = P_{e,3\ldots n-1\,1\,2}$$
$$P_{e,3\ldots n\,1\,2}(q) = P_{e,3\ldots n-1\,1\,2}(q) + qP_{s_{n-2},3\ldots n-1\,1\,2}(q)$$

This can be restated in the following way:

Proposition 3.7. *Let $n \geq 3$. Then*

$$\begin{pmatrix} P_{e,3\ldots n\,1\,2}(q) \\ P_{s_{n-1},3\ldots n\,1\,2}(q) \end{pmatrix} = \begin{pmatrix} 1 & q \\ 1 & 0 \end{pmatrix}^{n-3} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

The next result is a direct consequence of Theorem 3.3 and is the starting point to the solution of Problem 3.5 for $i = 4$.

Lemma 3.8. *Let $n \in \mathbb{N}$, $n \geq 4$, $v_n \in S_n$, $v_n := 4\ldots n\,1\,2\,3$ and $u \leq v$. Then*

$$P_{u,v_n} = q^{1-c}P_{us_{n-3},v_ns_{n-3}}(q) + q^cP_{u,v_ns_{n-3}}(q)$$

where $c = 1$ if $n-3 \in D_R(u)$ and $c = 0$ otherwise.

Now let S_3 act on the set $\{n-2, n-1, n\}$ by identifying $1, 2$ and 3 with $n-2, n-1$ and n respectively. Then, $\forall \sigma \in S_3$ we set

$$\Phi_\sigma(q) := P_{1\ldots(n-3)\sigma(n-2)\sigma(n-1)\sigma(n),v_n}$$

In 2001 Billey and Warrington (see [BW, Theorem 4]) gave a generating function description of $\Phi_{123}(q)$.

The following theorem shows an explicit formula for all $\Phi_\sigma(q)$, $\sigma \in S_3$ at the same time and in particular it solves Problem 3.5 for $i = 4$.

Theorem 3.9. *Let $n \in \mathbb{N}$ $n \geq 4$.*

$$\begin{pmatrix} \Phi_{123}(q) \\ \Phi_{132}(q) \\ \Phi_{213}(q) \\ \Phi_{231}(q) \\ \Phi_{312}(q) \\ \Phi_{321}(q) \end{pmatrix} = \begin{pmatrix} 1 & q & q & q^2 & 0 & 0 \\ 1 & 0 & q & 0 & 0 & 0 \\ 1 & q & 0 & 0 & q & q^2 \\ 0 & 1 & 0 & 0 & q & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}^{n-4} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

4. Explicit formulae

The main target of this section is to find an explicit formula for all polynomials $P_{u,v}$, $u, v \in S_n$, when $D_R(\overline{v}) \supseteq [2, n-2]$, $v^{-1}(n) \neq 1$ and u have some particular shape depending on that of v. This result will turn out to be a generalization of Theorem 2.8 and will allow us to prove two conjectures due to F. Brenti and R. Simion.

With this purpose we fix $x, y, n \in \mathbb{N}$ such that $x, y \in [2, n-1]$ and $x \neq y$. We denote by σ_0 the unique element v of S_{n-1} such that $v(1) = x$, $v(n-1) = y$ and $[2, n-2] \subseteq D_R(v)$.

For any $i \in [2, n]$ we denote by v_i the unique permutation of S_n satisfying the following two conditions:

$$\overline{v_i} = \sigma_0$$
$$v_i^{-1}(n) = i$$

We also set, $\forall a, i \in \mathbb{N}$, $2 \leq a \leq i \leq n$,

$$u_{i,a} := s_as_{a+1}\cdots s_{i-1}$$

so that, in particular, $u_{i,i} = e$ $\forall i$.

For example, for $n = 6$, $x = 4$ and $y = 2$ we have $v_4 = 4\,5\,3\,6\,1\,2$, $v_5 = 4\,5\,3\,1\,6\,2$ and $u_{4,2} = 1\,3\,4\,2\,5\,6$.

We denote by $R_{i,a}(q) := P_{u_{i,a},v_i}(q)$. We also set, for the notation convenience, $R_{n+1,a} = 0$.
The fundamental result about this family of Kazhdan-Lusztig polynomials is the following:

Theorem 4.1. *Let $2 \leq a \leq i \leq n-1$. Then the polynomials $R_{i,a}(q)$ verify the following relations:*
If $x > y$

$$\begin{aligned} R_{i,a}(q) &= qR_{i+1,a}(q) + R_{i+1,i+1}(q) - qR_{i+2,i+2}(q) \\ &\quad -\delta_{i,n-y-1}q - \delta_{i,n-x}q^{x-y+1} \end{aligned}$$

and if $x < y$

$$\begin{aligned} R_{i,a}(q) &= qR_{i+1,a}(q) + R_{i+1,i+1}(q) \\ &\quad -qR_{i+2,i+2}(q) - \delta_{i,n-y}q \end{aligned}$$

where $\delta_{i,j}$ is the usual Kronecker symbol.

Theorem 4.1 can be used to find explicit formulae for the polynomials $R_{i,a}(q)$ using a double induction on n and $n-i$.

Corollary 4.2. *Suppose $x > y$. Then if we set:*

$$H_i(q) = q\,[n-y-i]_q + q^{n-y-1}\,[x-i]_q + q^{x-y+1}\,[n-x+1-i]_q + q^{n-y}\,[y-1-i]_q$$

we have:

$$R_{i,a}(q) = \begin{cases} (1+q^{x-y})\,[n-1-i]_q - H_i(q) & \text{if } a \in [2, y-1] \\ [n-i]_q + q^{x-y}\,[n-1-i]_q - H_i(q) & \text{if } a \in [y, x] \\ (1+q^{x-y})\,[n-i]_q - H_i(q) & \text{if } a \in [x+1, n-1] \end{cases}$$

Corollary 4.3. *Suppose $x < y$. Then if we set*

$$K_i(q) := q\,[n-y+1-i]_q + q^{n-y}\,[y-1-i]_q$$

we have

$$R_{i,a}(q) = \begin{cases} [n-1-i]_q - K_i(q) & \textit{if } a \in [2, y-1] \\ [n-i]_q - K_i(q) & \textit{if } a \in [y, n-1] \end{cases}$$

Note that in this case the polynomials $R_{i,a}$ don't depend on x.

Remark 4.4. We could have defined $R_{i,1}(q)$, in a similar way, but we have chosen not to do it because they satisfy a slight different recursion (and the general discussion would have been more complicated) and hence we have preferred to suppose $a \geq 2$.

Note that by an easy argument on induction on n based on Theorems 2.4 and 2.9, in order to prove Theorem 2.8 it's enough to show it when $u = e$ and $v = \sigma_0$. Hence our result provides also a combinatorial proof of this theorem which was originally proved by B. Shapiro, M. Shapiro and A. Vainshtein in a geometric way.
We are now going to show further explicit formulae for some other families of Kazhdan-Lusztig polynomials whose proofs still use Theorems 3.1 and 3.3 as well as other recursive relations similar to that of Theorem 4.1.
With this purpose we let $n \in \mathbb{N}$, $n \geq 6$ and S_3 act at the same time on $\{1,2,3\}$ in the usual way and on $\{n-2, n-1, n\}$ as it did in the §3,
i.e. in the natural way identifying $n-2, n-1$ and n with $1, 2$ and 3 respectively.

Definition. $\forall(\sigma,\tau) \in S_3 \times S_3$ we denote by $D_{\sigma,\tau}(q)$ the following Kazhdan-Lusztig polynomial:

$$D_{\sigma,\tau}(q) := P_{e,\sigma(n-2)\,\sigma(n-1)\,\sigma(n)\,n-3\ldots 4\,\tau(1)\,\tau(2)\,\tau(3)}(q)$$

Theorem 4.5. $\forall n \geq 6$ *the following formulae hold:*

(1) $D_{123,321}(q) = D_{321,123}(q) = 1$
(2) $D_{132,321}(q) = D_{321,132}(q) = D_{321,213}(q) = D_{213,321}(q) = 1$
(3) $D_{231,321}(q) = D_{321,312}(q) = 1$
(4) $D_{321,321}(q) = 1$
(5) $D_{312,321}(q) = D_{321,231}(q) = 1 + q$
(6) $D_{231,312}(q) = 1 + q^{n-3}$
(7) $D_{213,312}(q) = D_{231,132}(q) = D_{231,132}(q) = D_{132,312}(q) = 1 + q^{n-4}$
(8) $D_{132,213}(q) = D_{213,132}(q) = 1 + q^{n-5}$
(9) $D_{132,132}(q) = D_{213,213}(q) = 1 + q^{n-5}(1 + q)$
(10) $D_{123,312}(q) = D_{231,123}(q) = 1 + 2q^{n-4}$
(11) $D_{123,132}(q) = D_{132,123}(q) = D_{213,123}(q) = D_{123,213}(q) = 1 + q^{n-5}(2 + q)$
(12) $D_{123,231}(q) = D_{312,123}(q) = (1 + 2q^{n-5})(1 + q)$
(13) $D_{231,231}(q) = D_{312,312}(q) = 1 + q + q^{n-4}$
(14) $D_{132,231}(q) = D_{312,132}(q) = D_{312,213}(q) = D_{213,231}(q) = (1 + q)(1 + q^{n-5})$
(15) $D_{312,231}(q) = (1 + q)^2(1 + q^{n-5})$

All the equalities among the $D_{\sigma,\tau}(q)$'s in each row of Theorem 4.5 are due to Theorem 2.6 while equations 1,2,3 and 4 follows directly from Theorem 2.9. Equations 6,7,8,9,10 and 11 are particular cases of the explicit formulae arising from Theorem 4.1, while all the others need a further proof. Equations 10 and 11 were conjectured by F. Brenti and R. Simion (see [BS, Conjectures 4.1 and 4.2]) while equation 6 had been conjectured by M. Haiman and already proved by B. Shapiro, M. Shapiro and A. Vainshtein as a particular case of Theorem 2.8.

The only missing case from Theorem 4.5 is $D_{123,123}$. This has turned out to be much more difficult than the others and will be treated apart in a joint work of the author and M. Marietti.

References

[Bo] B.D. Boe, "Kazhdan-Lusztig polynomials for hermitian symmetric spaces", *Trans. Amer. Math. Soc.* **309** (1988), 279-294.

[Br1] F. Brenti, "A combinatorial formula for Kazhdan-Lusztig polynomials", *Invent. Math.* **118** (1994), 371-394.

[Br2] F. Brenti, "Combinatorial expansions of Kazhdan-Lusztig polynomials", *J. London Math. Soc.* **55** (1997), 448-472.

[BS] F. Brenti and R. Simion, "Explicit formulae for some Kazhdan-Lusztig polynomials", *J. Alg. Comb.* **11** (2000), 187-196.

[BW] S. Billey and G. Warrington, "Kazhdan-Lusztig polynomials for 321-hexagon avoiding permutations", *J. Alg. Comb.* **13** (2001), 111-136.

[E] C. Ehresmann, "Sur la topologie de certains espaces homogènes", *Ann. Math.* **35** (1934), 396-443.

[H] J.E. Humphreys, "Reflection groups and Coxeter groups", Cambridge Univ. Press, Cambridge, 1990.

[LSa] V. Lakshmibai and B. Sandhya, "Criterion for smoothness of Schubert varieties in $SL(n)/B$", *Proc. Indian Acad. Sci. Math. Sci.* **100** (1990), 45-52.

[LSc] A. Lascoux and M.P. Schützenberger, " Polynômes de Kazhdan-Lusztig pour les grassmanniennes", Young tableaux and Schur functions in algebra and geometry, *Astérisque* **87-88** (1981), 249-266.

[KL1] D. Kazhdan and G. Lusztig, "Representations of Coxeter groups and Hecke algebras", *Inv. Math.* **53** (1979), 165-184.

[KL2] D. Kazhdan and G. Lusztig, "Schubert varieties and Poincaré duality, Geometry of the Laplace operator", *Proc. Symp. Pure Math.* **34**, Amer.Math. Soc., Providence, RI, 1980, 185-203.

[P] P. Polo, "Construction of arbitrary Kazhdan-Lusz tig polynomials in symmetric groups", *Representation Theory* **3** (1999), 90-104.

[SSV] B. Shapiro, M. Shapiro and A. Vainshtein, "Kazhdan-Lusztig polynomials and incomplete varieties of flags", *Discr. Math.* **180** (1998), 345-355.

Dipartimento di Matematica "G. Castelnuovo", Università di Roma "La Sapienza", P.le A. Moro 5, 00185, Roma, Italy

E-mail address: caselli@mat.uniroma1.it

SKEW OSCILLATING SEMI-STANDARD TABLEAUX (EXTENDED ABSTRACT)

SEUL HEE CHOI

ABSTRACT. We introduce an analogue of the Robinson-Schensted correspondence for skew oscillating semi-standard tableaux which generalize the correspondence for skew oscillating tableaux. We give the geometric construction for skew oscillating semi-standard tableaux and examine its combinatorial properties.

RÉSUMÉ. Nous introduisons une construction analogue de la correspondance de Robinson-Schensted pour les tableaux semi-standards oscillants gauches qui généralise la correspondance pour les tableaux oscillants gauches. Nous donnons la construction géométrique pour les tableaux semi-standards oscillants gauches et exeminons ses propriétés combinatoires.

1. Introduction

The Robinson-Schensted correspondence between permutations and pairs of standards tableaux of the same shape is introduced by Robinson ([6]) and it is given by Schensted ([10]) a little different form. After generalization by Knuth ([5]) to generalized permutations and pairs of semi-standards tableaux, various analogues of the Robinson-Schensted correspondence have been produced on different kinds of tableaux ([8],[13],[3],[9]).

More recently, Dulucq and Sagan ([4]) have given the Robinson-Schensted correspondence for oscillating tableaux and skew oscillating tableaux.

In this article, we extended the properties and constructions of analogue of Robinson-Schensted correspondence in [4] to skew oscillating semi-standard tableaux. In sections 2, we give basic definitions of generalised biwords and skew oscillating semi-standard tableaux. An algorithm of Robinson-Schensted for skew oscillating semi-standard tableaux is given in section 3, which is an extension of the algorithm of Robinson-Schensted correspondence for skew oscillating standard tableaux given in [4]. Then we give a geometric construction of a generalized biword due to Viennot, Chauve and Dulucq ([1],[8],[13]).

2. Definition and notations

Let $\lambda = (\lambda_1, ..., \lambda_k)$, $\lambda_1 \leq ... \leq \lambda_k$, be a partition of n such that $\sum_{i=1}^{k} \lambda_i = n$. The partition λ can be displayed a Ferrers diagram with the part λ_i in the row i. If $\mu \subseteq \lambda$ then the corresponding skew shape λ/μ is the set $\{c | c \in \lambda, c \notin \mu\}$. If $|\lambda/\mu| = n$ then we write $\lambda/\mu \vdash n$ and say that λ/μ is a skew partition of n. A skew semi-standard tableau S of shape λ/μ is a labeling of the cells of λ/μ with positive integers so that the rows are strictly increasing and the columns are weakly increasing. $\emptyset_\alpha$ denotes the empty tableaux of the shape α (or a skew tableau of the shape α/α). $T(\lambda/\mu)$ denotes the set of skew semi-standard tableaux of shape λ/μ.

$S(i,j)$ denotes the label of the cell in the i^{th} row and j^{th} column of a skew semi-standard tableau S so that $k \in S$ means $k = S(i,j)$ for some i, j. $\overline{T}(\lambda/\mu)$ denotes the set of tableaux of shape λ/μ with rows strictly decreasing and columns weakly decreasing. For example, when $\lambda = (5,4,3,1)$ $\mu = (2,2)$, the two following tableaux belong to $T(\lambda/\mu)$ and $\overline{T}(\lambda/\mu)$ respectively.

$$\begin{array}{ccccc} 1 & & & & \\ 1 & 3 & 6 & & \\ \square & \square & 2 & 5 & \\ \square & \square & 2 & 3 & 5 \end{array} \qquad \begin{array}{ccccc} 1 & & & & \\ 5 & 4 & 3 & & \\ \square & \square & 6 & 3 & \\ \square & \square & 7 & 3 & 2 \end{array}$$

Four kinds of insertions and deletions in a skew semi-stanaard tableau ([1],[4]) are defined below. Let S be a skew semi-standard tableau of shape λ/μ.

1. The *external insertion* inserts an integer x in S by using the Knuth-Robinson-Schensted algorithm([2],[5]). We denote the new tableau obtained after this insertion by $ExtI(S, x)$. The inverse process is called *external deletion*, denoted by $ExtD(S, x)$, which ends with the expulsion of an integer out of S.

2. The *internal insertion* occurs only in a cell (u, v) of S such that $(u, v) \notin \mu$ and it belongs to one of three cases: (i) $(u-1, v) \in \mu$ and $(u, v-1) \in \mu$, (ii) $v = 1$ and $(u-1, v) \in \mu$, (iii) $u = 1$ and $(u, v-1) \in \mu$.

The *internal insertion* of the cell (u, v) inserts the integer x contained in $S(u, v)$ from the row $u+1$ using the external insertion algorithm. We denote the new tableau by $IntI(S, (u, v))$. The external deletion is called *internal deletion* if it ends in filling a cell of μ. $IntD(S, (u, v))$ denotes the internal deletion.

3. The *empty insertion* adds an empty cell (u, v) in S such that $(u, v) \notin \lambda$, satisfying (i)$(u-1, v) \in \mu$ and $(u, v-1) \in \mu$, (ii)$u = 1$, $(u, v-1) \in \mu$ or (iii)$v = 1$, $(u-1, v) \in \mu$. $EmpI(S, (u, v))$ denotes the new tableau obtained after this insertion and the inverse process is called *empty deletion*, denoted by $EmpD(S, (u, v))$.

4. A cell can simply be *attached* or *erased* using neither the insertion algorithms nor the deletion algorithms.

Example 2.1.

$$\text{P} = \begin{array}{cccc} \square & & & \\ \square & 2 & & \\ \square & 2 & 4 & \\ \square & \square & 3 & 7 \end{array}$$

$$\text{ExtI(P,6)} = \begin{array}{cccc} \square & & & \\ \square & 2 & & \\ \square & 2 & 4 & 7 \\ \square & \square & 3 & 6 \end{array} \qquad \text{ExtD(P,3)} = \begin{array}{cccc} \square & & & \\ \square & 2 & & \\ \square & 2 & & \\ \square & \square & 4 & 7 \end{array}$$

$$\text{IntI(P,(2,2))} = \begin{array}{cccc} \square & 2 & & \\ \square & 2 & & \\ \square & \square & 4 & \\ \square & \square & 3 & 7 \end{array} \qquad \text{IntD(P,(3,2))} = \begin{array}{cccc} \square & & & \\ \square & & & \\ \square & 2 & 4 & \\ \square & 2 & 3 & 7 \end{array}$$

$$\text{EmpI(P,(5,1))} = \begin{array}{cccc} \square & & & \\ \square & & & \\ \square & 2 & & \\ \square & 2 & 4 & \\ \square & \square & 3 & 7 \end{array} \qquad \text{EmpD(P,(4,1))} = \begin{array}{cccc} \square & 2 & & \\ \square & 2 & 4 & \\ \square & \square & 3 & 7 \end{array}$$

A skew oscillating semi-standard tableau of length n is a sequence of semi-standard tableaux $P = (P_0, P_1, ..., P_n)$ where P_k is obtained from P_{k-1} by an insertion or a deletion of a cell.

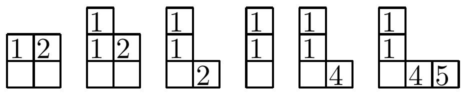
$\Theta_n(\alpha/\gamma \to \beta/\mu)$ denotes the set of skew oscillating tableaux $P = (P_0, P_1, ..., P_n)$ of length n satisfying the following conditions:

(1) the shape of P_0 is α/γ, and the shape of P_n is β/μ,

(2) P_k is obtained from P_{k-1} by attaching a cell with a label (this is not by the insertion algorithms) or a deletion of a cell by external deletion, internal deletion or empty deletion.

(3) if $x_i, x_j, ..., x_m$ are inserted respectively in $P_i, P_j, ..., P_m$, $i < j < ... < m$, then $x_i \leq x_j \leq ... \leq x_m$.

For example, if $\alpha = (2, 2)$, $\gamma = (2)$, $\beta = (3, 1, 1)$ and $\mu = (1)$ then the following tableau belongs to $\Theta_5(\alpha/\gamma \to \beta/\mu)$. 1 is inserted in P_1, 4 in P_4 and 5 in P_5.



For a $P \in \Theta_n(\alpha/\gamma \to \beta/\mu)$, we define a set of nondecreasing sequences of positive integers $I(P) = \cup_{j \in N} I_j$, where $I_j = \{j_0, j_1, j_2, ..., j_n\}$, $j_0 = 0 \leq j_1 \leq ... \leq j_n$ and $j_k = x$ if $P_k = P_{k-1} + (u, v)$ with $P_k(u, v) = x$ for $1 \leq k \leq n$. An I_j of the example above is $\{0, 1, j_2, j_3, 4, 5\}$, where j_2, j_3 are positive integers.

A skew oscillating semi-standard tableau of $\Theta(\emptyset\mu \to \lambda/\mu)$ having only insertion steps, is a skew semi-standard tableau of shape λ/μ, the label of a cell being given by its creation.

A generalized biword π of size $2n$ is a sequence of vertical pairs of positive integers $\pi = \left(\begin{array}{cccc} u_1 & u_2 & ... & u_n \\ v_1 & v_2 & ... & v_n \end{array} \right)$ where $u_1 \geq u_2 \geq ... \geq u_k$, $u_i \geq v_i$ for $i = 1, ..., k$, and $v_i \geq v_2$ if $u_i = u_j$. $\hat{\pi}$ denotes the

top row of π and $\check{\pi}$ its bottom row.

GB denotes the set of generalized biwords. The size of π is $2\times$ the number of pairs of $\binom{u_i}{v_i}$, or $|\pi| = 2n$. GB_{2n} denotes the set of generalized biwords of size $2n$.

3. Generalized biwords and skew oscillating semi-standard tableaux

We give a description of an algorithm to examine the relation between skew oscillating semi-standard tableaux and the triples $(S, U, \pi) \in \cup_{\mu \subseteq \alpha \cap \beta}[T(\beta/\mu) \times \overline{T}(\alpha/\mu)] \times GB$.

Algorithm **OSCIL**

(a)The input is $(S, U, \pi) \in \cup_{\mu \subseteq \alpha \cap \beta}[T(\beta/\mu) \times \overline{T}(\alpha/\mu)] \times GB$,

(b) The output is (P, I) where $P \in \Theta_n(\emptyset_\alpha \to \beta/\mu)$, and $I = \{i_0 = 0, i_1, i_2, ..., i_n\} \in I(P)$, i.e., I satisfies the following conditions :

(1) $i_1, ..., i_n$ being a nondecreasing sequence of positive integers,

(2) if we obtain P_k from P_{k-1} by attaching a cell (u, v) with label a, that is, $P_k = P_{k-1} + (u, v)$, with $P_k(u, v) = a$, then $i_k = a$.

We construct a sequence of nonnegative integers $I = \{i_0, i_1, i_2, ..., i_n\}$ as follows : let $i_0 = 0$ and $i_1, ..., i_n$ be the rearranged elements of S, U, $\hat{\pi}$ and $\check{\pi}$ in nondecreasing order. We have $n = |S| + |U| + |\pi|$.

Let $P_n = S$.
For k from n to 1 :
(a) if there is a cell $P_k(u, v) = i_k$, then erase this cell to obtain P_{k-1},
(b) else if the pair (i_k, x) belongs to π, then $P_{k-1} = ExtI(P_k, x)$,
(c)else if $U(u, v) = i_k$ and $P_k(u, v)$ exists (with label x), then $P_{k-1} = IntI(P_k, (u, v))$,
(d) else $P_{k-1} = EmpI(P_k, (u, v))$.
The tableaux P_k have respective shapes λ_k/μ_k. $P = (P_0, ..., P_n)$ and $I = \{i_0, i_1, ..., i_n\}$ satisfy that $i_k = a$ when $P_k = P_{k-1} + (u, v)$, with $P_k(u, v) = a$, so $I \in I(P)$.

Algorithm $\mathbf{OSCIL^{-1}}$.
(a) The input is (P, I) where $P \in \Theta_n(\emptyset_\alpha \to \beta/\mu)$ with $\mu \subseteq \alpha \cap \beta$ and $I \in I(P)$.
(b) The output is a triple $(S, U, \pi) \in \cup_{\mu \subseteq \alpha \cap \beta}[T(\beta/\mu) \times \overline{T}(\alpha/\mu)] \times GB$.

Let $\pi = \emptyset$, and $U_0 = P_0$.
For k from 1 to n:
(a)if $P_k = P_{k-1} + (u, v)$, then $U_i = U_{i-1}$,
(b) else $(P_k = P_{k-1} - (u, v))$, we have three cases :
(b_1) if the deletion is external (x ejected out of P_{k-1}), then add the pair (i_k, x) to π, $U_k = U_{k-1}$,

(b_2) else if it is internal (the cell $P_{k-1}(u, v)$ with label x is erased), then label the cell $U_{i-1}(u, v)$ with i_k to obtain U_k,

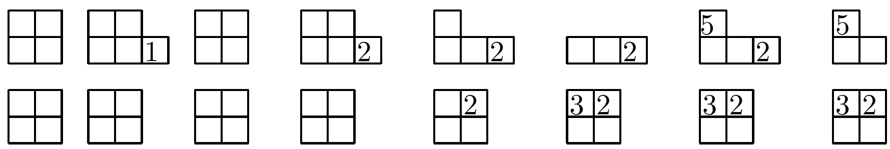
(b_3) else label the cell $U_{k-1}(u, v)$ with i_k to obtain U_k.
Finally, we obtain $S = P_n$, $U = U_n$ and $\pi \in GB$

Example 3.1.

k =	0	1	2	3	4	5	6	7
i_k =	0	1	1	2	2	3	5	6

$P_k =$

$U_k =$



π_k =	$\emptyset$	$\emptyset$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 6 & 1 \\ 2 & 1 \end{pmatrix}$

Theorem 1. *Let α, β be fixed partitions. There is a bijection Φ from triples (S, U, π) of*

$\cup_{\mu\subseteq\alpha\cap\beta}[T(\beta/\mu)\times\overline{T}(\alpha/\mu)]\times GB$ *to* (P,I) *with a skew oscillating semi-standard tableau* P *of* $\Theta_n(\emptyset_\alpha\to\beta/\mu)$, $n=|S|+|U|+|\pi|$ *and* $I=\{i_0,i_1,i_2,...,i_n\}\in I(P)$

Proof: For a triple $(S,U,\pi)\in\cup_{\mu\subseteq\alpha\cap\beta}[T(\beta/\mu)\overline{T}(\alpha/\mu)\times GB]$, we obtain directly a nondecreasing sequence $I=\{i_0,i_1,i_2,...,i_n\}$ with $i_0=0$ and $\{i_1,i_2,...,i_n\}$ rearranging the elements of S, U, $\hat{\pi}$ and $\check{\pi}$. A skew oscillating semi-standard tableaux $P\in\Theta_n$ results by applying the algorithm $OSCIL$.

To give the inverse operation, we construct a nondecreasing sequence $I=\{i_0,i_1,i_2,...,$ $i_n\}$ from $P\in\Theta_n(\emptyset_\alpha\to\beta/\mu)$ as follows : (1) $i_0=0$ (2) if $P_k=P_{k-1}+(u,v)$, with $P_k(u,v)=x$ then $i_k=x$, else i_k is a positive integer satisfying $i_{k-1}\le i_k\le i_{k+1}$, so $I\in I(P)$. Next, we construct a sequence $(S_0,U_0,\pi_0)=(P_0,P_0,\emptyset)$, $(S_1,U_1,\pi_1),...,(S_n,U_n,\pi_n)=(S,U,\pi)$ from P and $I(P)$ by applying the algorithm $OSCIL^{-1}$. The algorithm $OSCIL^{-1}$ corresponds exactly to the inverse construction of the cell produced by the algorithm $OSCIL$. So (S,U,π) is in bijection with (P,I). Example 3.2 shows the application of the algorithm $OSCIL$ and $OSCIL^{-1}$. $\diamondsuit$

Remark 3.2 If the skew oscillating semi-standard tableaux $P\in\Theta_n(\emptyset_\alpha\to\beta/\alpha)$ has only insertion steps, the bijection Φ is $\Phi^{-1}(P,I)=(P_n,\emptyset_\alpha,\emptyset)$

$\overline{\Theta}_n(\beta/\mu\to\alpha/\gamma)$ denotes the set of skew oscillating tableaux of length n, $Q=(Q_0,Q_1,...,$ $Q_n)$, satisfying the following conditions:

(1) the shape of Q_0 is β/μ, and the shape of Q_n is α/γ,

(2) Q_k is obtained from Q_{k-1} by erasing of a labelled cell (not by deletion algorithms) or an insertion of a cell by external insertion, internal insertion or empty insertion.

(3) if $x_i,x_j,...,x_m$ are deleted respectively from $Q_i,Q_j,...Q_m$, $i<j<...<m$, then $x_i\ge x_j\ge...\ge x_m$.

We know that $P=(P_0,P_1,...,P_n)\in\Theta_n(\alpha/\gamma\to\beta/\mu)$ if and only if $\overline{P}=(P_n,P_{n-1},...,P_0)\in\overline{\Theta}_n(\beta/\mu\to\alpha/\gamma)$.

We define a set of nonincreasing sequences of positive integers $\overline{I}(Q)=\cup_{j\in N}\overline{I}_j$,
$\overline{I}_j=\{\overline{j}_1,\overline{j}_2,...,\overline{j}_n\}$ for $Q\in\overline{\Theta}_n(\beta/\mu\to\alpha/\gamma)$ as follows:
(1) $\overline{j}_1\ge\overline{j}_2\ge...\ge\overline{j}_n$
(2) if $Q_{k+1}=Q_k-(u,v)$ with $Q_k(u,v)=x$, then $\overline{j}_k=x$.

Theorem 2. *Let* π *be a generalized biword of size* $2n$ *and* α *be an empty partition (of shape* α*). There is a bijection* Φ_{RS} *from pairs* $(\emptyset_\alpha,\pi)$ *to* $\{(P,I_1),(Q,\overline{I}_2))\}$ *of* $\cup_\beta[\{\Theta_n(\emptyset_\alpha\to\beta/\alpha)\times I(P)\},\times\{\overline{\Theta}_n(\emptyset_\alpha\to\beta/\alpha)\times\overline{I}(Q)\}]$.

Proof: According to the previous theorem, we obtain $(P_0=\emptyset_\alpha,...,P_n...,P_{2n}=\emptyset_\alpha)$ with $I=\{i_0,i_1,...i_n,..,i_{2n}\}$, and the result $(P,I_1)=(P_0,P_1,...,P_n(\text{of shape}\beta/\alpha),\{i_0,i_1,...,i_n\})$ with $I_1\in I(P)$, and $(Q,\overline{I}_2)=(P_{2n},P_{2n-1},...,P_n(\text{of shape}\beta/\alpha),\{i_{2n},i_{2(n-1)},...,i_n\})$. Therefore $Q\in\overline{\Theta}_n(\emptyset_\alpha\to\beta/\alpha)$ and $\overline{I}_2\in\overline{I}(Q)$ $\diamondsuit$

Taking an empty initial and final semi-standard tableaux in the theorem 1 and 2, we have an analog of Robinson-Schensted correspondence for oscillating semi-standard tableaux, as stated in the following results.

Corollaire 1. *Let β be fixed partitions and n a fixed integer. There is a bijection $\Phi_\emptyset$ from pairs (S,π) of $T(\beta)\times GB$ such that $n=|S|+|\pi|$ to pairs (P,I) with P of $\Theta_n(\emptyset\to\beta)$ and $I\in I(P)$.*

Corollaire 2. *Let n be a fixed integer. There is a bijection $\Phi_{RS_\emptyset}$ from generalized biwords π of GB_{2n} to pairs $\{(P,I_1),(Q,\overline{I}_2)\}$ of $\cup_\beta[(\Theta_n(\emptyset\to\beta)\times I(P))\times(\overline{\Theta}_n(\emptyset\to\beta)]\times\overline{I}(Q)]$.*

4. Geometric description of a generalized biword

In this section, we represent a geometric description of a generalized biword in the the first quadrant of the Cartesian plane by applying the geometric construction of Viennot in [1] and [13] for standard tableaux. We obtain an oscillating semi-standard tableau from the geometric description of
a generalized biword.

First, we present a method to standardize a generalized biword. Let $\mathbb{N}$ be a set of positive integers. For a given generalized biword $\pi=\begin{pmatrix} u_1 & u_2 & ... & u_n \\ v_1 & v_2 & ... & v_n\end{pmatrix}$, we propose a new alphabet $\mathbb{N}\cup\{j^{(h)}:j,h\in\mathbb{N}\}$ such that

$$...<j<j^{(1)}<j^{(2)}<...<j+1<j+1^{(1)}<j+1^{(2)}<...$$

If $u_j=u_{j+1}=...=u_{j+m}=v_{i_1}=v_{i_2}=...=v_{i_k}$, $i_1<i_2<...<i_k$, in π then we translate $v_{i_k}\to u_j$, $v_{i_{k-1}}\to u_j^{(1)}$, ..., $u_j\to u_j^{(m+k-1)}$.

For example,

$$\pi=\begin{pmatrix} 5 & 5 & 4 & 3 & 3 & 3 \\ 4 & 2 & 1 & 2 & 2 & 1\end{pmatrix}\leftrightarrow\tau=\begin{pmatrix} 5^{(1)} & 5 & 4^{(1)} & 3^{(2)} & 3^{(1)} & 3 \\ 4 & 2^{(2)} & 1^{(1)} & 2^{(1)} & 2 & 1\end{pmatrix}$$

The translation from π to τ is bijective and is called **standardization of** π. It is denoted by $\tau=\varphi(\pi)$. We know that τ has the properties of the generalized biword on the new alphabet $N\cup\{j^{(h)}:j,h\in N\}$.

Then, we represent $\tau=\varphi(\pi)$ in the place of π in the part $\{0,1,2,...,n\}\times\{0,1,...,n\}$ of the Cartesian plane as follows:

- Define a map $\Psi:$ abscissas x $(x=0,1,2,...,n)\to\{$ the greatest element of $\hat{\pi}+1\}\cup\hat{\tau}$ with

$$\Psi(x)=\begin{cases}\text{the greatest element of } \hat{\pi}+1 & \text{if } x=0\\ x^{th} \text{ greatest element of } \hat{\tau} & \text{else}\end{cases}$$

- Define a map $\Gamma:$ abscissas y $(y=0,1,2,...,n)\to\{0\}\cup\check{\tau}$ with

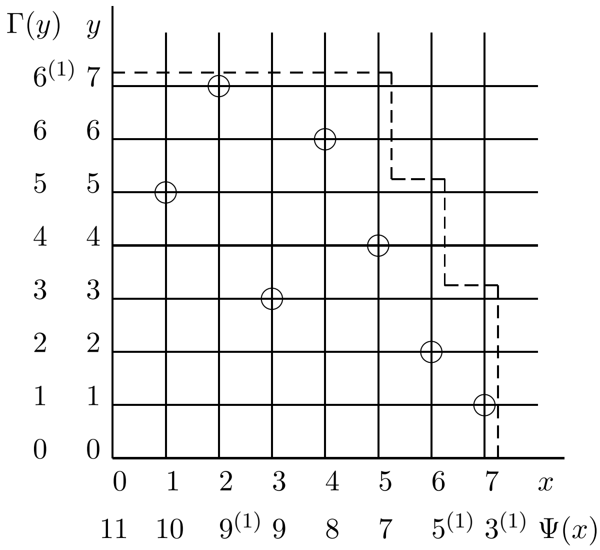
$$\Gamma(x)=\begin{cases}0 & \text{if } y=0\\ y^{th} \text{ lowest element of } \check{\tau} & \text{else}\end{cases}$$

- We define valid domain as the set of points (x,y) such that $\Psi(x)\geq\Gamma(y)$.
- For each pair (u_k,v_k) of τ, we set up the point $(\Psi^{-1}(u_k),\Gamma^{-1}(v_k))$ which is in the valid domain.

Example 4.1. For a generalized biword π of GB_{14} we obtain $\tau = \varphi(\pi)$ by the standardization of π :

$$\pi = \begin{pmatrix} 10 & 9 & 9 & 8 & 7 & 5 & 3 \\ 5 & 6 & 3 & 6 & 4 & 2 & 1 \end{pmatrix} \to \tau = \begin{pmatrix} 10 & 9^{(1)} & 9 & 8 & 7 & 5^{(1)} & 3^{(1)} \\ 5 & 6^{(1)} & 3 & 6 & 4 & 2 & 1 \end{pmatrix}.$$

Here we give the representation of τ. The dashed line describes the limit of valid domain, which is slightly extended on the figure for readability.



Description of $\tau = \varphi(\pi)$

Figure 4.1

From the Figure 4.1, we construct an oscillating semi-standard tableaux $T = \{T_0 = \emptyset, T_1, T_2, ..., T_{2n} = \emptyset\}$ as follows:

(a) let $A = \{\Psi(x_i)\}_{i=1..n}$ and $B = \{\Gamma(y_i)\}_{i=0..n}$. $I = \{i_0, ..., i_{2n}\}$ describes A and B lined up in nondecreasing order.

(b) For k from $2n$ to 1:
if $i_k \in A$ and $(\Psi^{-1}(i_k), y)$ is SW-corner of a Shadow line, then $T_{k-1} = ExtI(T_k, \Gamma(y))$.
else if $i_k \in B$ and $(x, \Gamma^{-1}(i_k))$ is SW-corner of a Shadow line, then $T_{k-1} = T_k - (u, v)$ with $T_k(u, v) = i_k$.

Figure 4.2 is the skew oscillating semi-standard tableaux corresponding to τ in Example 4.1. If $\Phi(\emptyset, \emptyset, \pi) = (P, J)$ in theorem 1, then we know that (P, J) is exactly equal to (T, I) if the exponents of contents of T and I are removed, resulting in the following theorem.

Theorem 3. *Let π be a generalized biword and $\tau = \varphi(\tau)$ and $D(\tau)$ be a description of τ with shadow lines in a Cartesian plane. Then there is a bijection from (T,I) of $\Theta_{2n}(\emptyset \to \emptyset) \times I(T)$*

to $D(\tau)$ *of* $\{D(\tau)|\tau=\varphi(\pi),\pi\in GB_{2n}\}$*. If* $\Phi(\emptyset,\emptyset,\pi)=(P,J)$*, then* (P,J) *is equal to* (T,I) *in removing the exponents of contents of* T *and* I*.*

Definition 1. *The shadow* $S(\tau)$ *of a generalized biword* τ *is the set of points* (x,y) *such that there is a point* (x',y') *of the representation of* τ *with* $x'\le x$*,* $y'\le y$*.*

Shadow lines of τ are defined recursively. The first shadow line L_1 of τ is the boundary of $S(\tau)$. To construct the shadow line L_{i+1} of τ remove the points of the representation of τ lying on L_i and construct the shadrow line of the remaining points. This procedure ends when there is no remaining point on the plane. The SW-coners of a shadow line are the points of the representation of τ located on this line ([8],[13]). The NE-coners of a shadow line are the points (x,y) of the shadow line such that $(x+1,y)$ and $(x,y+1)$ are not a part of this shadow line.

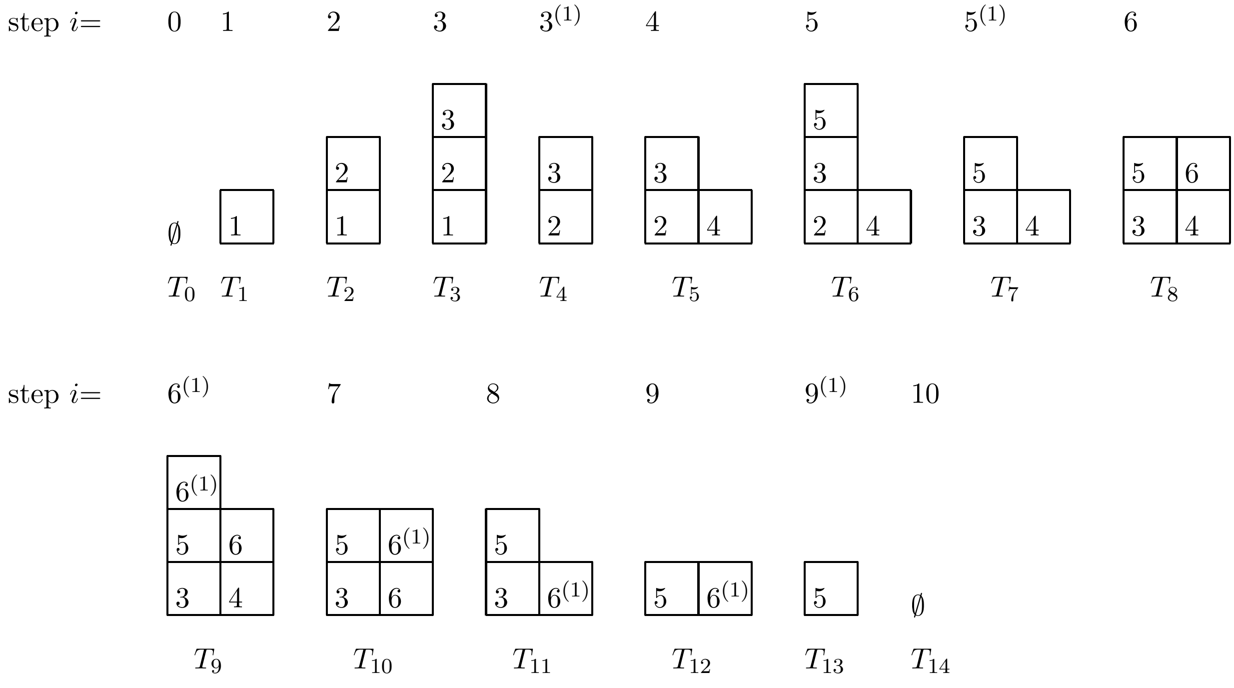


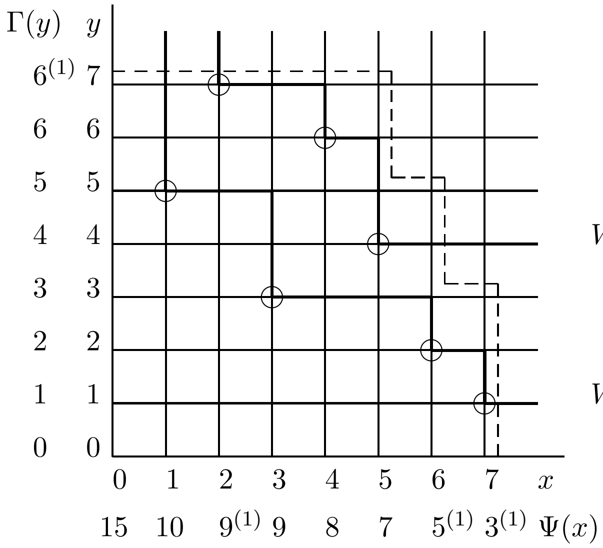
Figure 4.2

Definition 2. *The* k^{th} *skeleton of a generalized biword defined recursively by*

1. $\tau^{(1)}=\tau$

2. $\tau^{(k+1)}=\begin{pmatrix}\Psi(a_1) & \Psi(a_2) & \dots & \Psi(a_m)\\ \Gamma(b_1) & \Gamma(b_2) & \dots & \Gamma(b_n)\end{pmatrix}$ *where* $(a_1,b_1),\dots,(a_m,b_m)$ *are the NE-corners of* $\tau^{(k)}$*. The shadow diagram of* τ *is the set of shadow lines of all the skeletons* $\tau^{(k)}$ *of* τ*. The shadow lines of* $\tau^{(k)}$ *are denoted by* $W_j^{(k)}$*.*

Example 4.2. Let π be the generalized biword of size $2n$ defined in example 4.1. Here we have the description of shadow lines $W_j^{(1)}$, $j=1,2$, of $\tau=\varphi(\pi)$.



Description of generalized biword $\tau = \varphi(\pi)$ with shadow lines

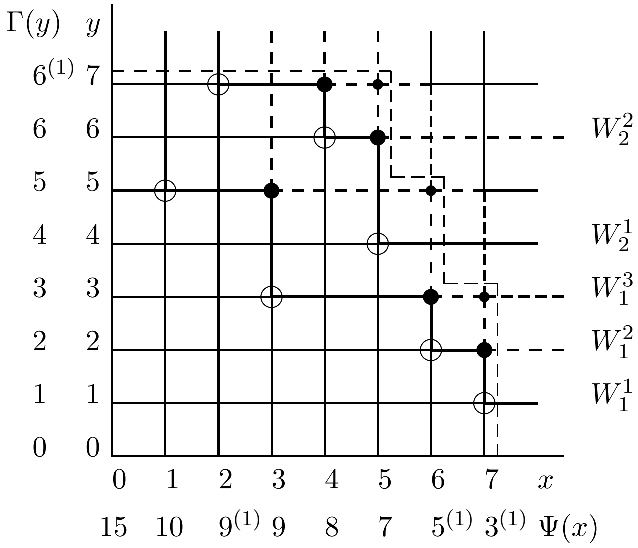
Figure 4.3

We can see that the shadow line W_1^1 in Figure 4.3 describes the behaviour of the first cell of the first row during the construction of $T_{14}, T_{13}, ..., T_0$. The shadow line W_1^1 has four SW-corners at (1,5), (3,3), (6,2) and (7,1). For the SW-corner (1,5), with $\Psi(1) = 10$ and $\Gamma(5) = 5$, followed by (3,3) with $\Psi(3) = 9$ and $\Gamma(3) = 3$. During the construction of the tableaux T_{14} to T_0, the first cell of first row is created during step 10 with label 5, this label is replaced during step 9 by the label 3. The label 3 is replaced during step $5^{(1)}$ by the label 2 and during step $3^{(1)}$ by the label 1, because $\Psi(6) = 5^{(1)}$ and $\Gamma(2) = 2$, $\Psi(7) = 3^{(1)}$ and $\Gamma(1) = 1$. The cell is deleted during step 1.

In the same way the shadow line W_j^i describes the behaviour of the j^{th} cell of the i^{th} row. So the theorem 4.1 in [1] is satisfied for a generalized biword and an oscillating semi-standard tableau as follows.

Theorem 4. *Let π be a generalized biword of size $2n$ and τ the standardization of π, i.e. $\tau = \varphi(\pi)$. If $\Phi(\emptyset, \emptyset, \tau) = (T, I)$ then the shadow line $W_j^{(i)}$ of τ describes the behavior of the j^{th} cell of the i^th row of the tableaux $T_{2n}, ..., T_0$ in the following way :*

1. *a SW-corner (x, y) indicates that during step $\Psi(x)$, the label $\Gamma(y)$ fills in this cell,*
2. *when the line leaves the valid domain at (x, y), this cell is deleted during step $\Gamma(y)$,*
3. *otherwise, the cell remains unchanged.*



Description of a biword π

Figure 4.4

From Figure 4.2 the generalized biword π in Example 4.1 is in bijection with the following pairs $((P, I_1), (Q, I_2))$, where $(P, I_1) \in (\Theta_n(\emptyset \to \beta), I(P))$ and $(Q, I_2) \in (\overline{\Theta}_n(\emptyset \to \beta), \overline{I}(Q))$:

$I_1 =$ 0 1 2 3 3 4 5 5

P = $\emptyset$ $\begin{array}{|c|}\hline 1\\\hline\end{array}$ $\begin{array}{|c|}\hline 2\\\hline 1\\\hline\end{array}$ $\begin{array}{|c|}\hline 3\\\hline 2\\\hline 1\\\hline\end{array}$ $\begin{array}{|c|}\hline 3\\\hline 2\\\hline\end{array}$ $\begin{array}{|c|c}\cline{1-1} 3 & \\\hline 2 & \multicolumn{1}{c|}{4}\\\hline\end{array}$ $\begin{array}{|c|c}\cline{1-1} 5 & \\\cline{1-1} 3 & \\\hline 2 & \multicolumn{1}{c|}{4}\\\hline\end{array}$ $\begin{array}{|c|c}\cline{1-1} 5 & \\\hline 3 & \multicolumn{1}{c|}{4}\\\hline\end{array}$

$I_2 =$ 10 9 9 8 7 6 6 5

Q = $\emptyset$ $\begin{array}{|c|}\hline 5\\\hline\end{array}$ $\begin{array}{|c|c|}\hline 5 & 6\\\hline\end{array}$ $\begin{array}{|c|c}\cline{1-1} 5 & \\\hline 3 & \multicolumn{1}{c|}{6}\\\hline\end{array}$ $\begin{array}{|c|c|}\hline 5 & 6\\\hline 3 & 6\\\hline\end{array}$ $\begin{array}{|c|c}\cline{1-1} 6 & \\\hline 5 & \multicolumn{1}{c|}{6}\\\hline 3 & \multicolumn{1}{c|}{4}\\\hline\end{array}$ $\begin{array}{|c|c|}\hline 5 & 6\\\hline 3 & 4\\\hline\end{array}$ $\begin{array}{|c|c}\cline{1-1} 5 & \\\hline 3 & \multicolumn{1}{c|}{4}\\\hline\end{array}$

5. References

[1] C. CHAUVE and S. DULUCQ, *Properties of the Robinson-Schensted correspondence for oscillating and skew oscillating tableaux*, Technical Report RR-1211-99, LaBRI, Université Bordeaux I, 1999.

[2] S. H. CHOI and D. GOUYOU-BEAUCHAMPS, *Enumeration of generalized Young tableaux with bounded height*, Theoretical Computer Science, 117 (1993) 137-151.

[3] M. DELEST, S. Dulucq and L. FAVREAU, *An analogue to the Robinson-Schensted correspondence for oscillating tableaux* In Séminaire Lotharingien de Combinatoire

(Alghero,1988), Univ. Louis Pasteur, Strasbourg, 1988.

[4] S. Dulucq and B. E. *Sagan. La correspondance de Robinson-Schensted pour les tableaux oscillants gauches* Discrete Math., 139(1-3):129-142, 1995. Formal power series and algebraic combinatorics(Montreal, PQ, 1992).

[5] D. E. KNUTH, *Permutations, matrices and generalized Young tableaux*, Pacific J. Math., 34 (1970) 709-727.

[6] G. De B. ROBINSON, *On representation of the symmetric group*, Amer. J. Math. 60(1968), 745-760.

[7] B. E. SAGAN, *Ssifted tableaux, Schur Q-functions, and a conjecture of R. Stanley*, J. Combin. Theory Ser. A 45, (1987) 62-103.

[8] B. E. SAGAN, *The symmetric group: representations, combinatorial algorithms, and symmetric functions* Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1991.

[9] B. E. SAGAN and R.P. STANLEY,*Robin-Schensted algorithms for Skew Tableaux*, J. Combin. Theory Ser. A 55, (1990) 161-193.

[10] C. SCHENSTED, *Longest increasing and decreasing subsequences*, Canad. J. Math., 13 (1961) 179-191.

[11] M.-P. Schützenberger, *La correspondance de Robinson*, In Combinatoire et representation du group symetrique, pages 59-113, Lecture Notes in Math., Vol. 579, berlin,1977.

[12] S. Sundaram. *The cauchy identity for Sp(2n)*, J. Combin. Theory Ser. A, 53(2): 209-238, 1990.

[13] G. Viennot, *Une forme géometrique de la correspondance de Robinson-Schensted*, In Combinatoire et representation du group symetrique, pages 29-58, Lecture Notes in Math., Vol. 579, Springer berlin,1977.

Department of Mathematics, Jeonju University, Chon-ju 560-759
E-mail address: chois@www.jeonju.ac.kr

EFFECTIVE D-FINITE SYMMETRIC FUNCTIONS (EXTENDED ABSTRACT)

FRÉDÉRIC CHYZAK, MARNI MISHNA, AND BRUNO SALVY

Abstract. Many combinatorial generating functions can be extracted from symmetric functions. Gessel has outlined a large class of symmetric functions for which the resulting generating functions are D-finite. We extend Gessel's work by providing algorithms that compute differential equations these generating functions satisfy. Examples of applications to k-regular graphs and Young tableaux with repeated entries are given.

Résumé. De nombreuses fonctions génératrices combinatoires s'expriment en termes de fonctions symétriques. Gessel a décrit une grande classe de fonctions symétriques pour lesquelles les fonctions génératrices extraites sont D-finies. Nous étendons ce travail de Gessel en donnant des algorithmes qui calculent des équations différentielles satisfaites par ces fonctions génératrices. Nous donnons des exemples d'application aux graphes k-réguliers et aux tableaux de Young avec entrées répétées.

Introduction

A power series in one variable is called differentiably finite, or simply D-finite, when it is solution of a linear differential equation with polynomial coefficients. This differential equation turns out to be a convenient data structure for expressing information related to the series and many algorithms operate directly on this differential equation. In particular, univariate D-finite power series are closed under sum, product, Hadamard product, Borel transform,... and algorithms computing the corresponding differential equations are known (see for instance [13]). Moreover, the coefficient sequence of a univariate D-finite power series satisfies a linear recurrence, which makes it possible to compute many terms of the sequence efficiently. These closure properties are implemented in computer algebra systems [10, 12]. Also, the mere knowledge that a series is D-finite gives information concerning its asymptotic behaviour. Thus, whether it be for algorithmic or theoretical reasons, it is often important to know whether a given series is D-finite or not, and it is useful to compute the corresponding differential equation when possible.

D-finiteness extends to power series in several variables: a power series is called D-finite when the vector space spanned by the series and its derivatives is finite-dimensional. Again, this class enjoys many closure properties and algorithms are available for computing the systems of linear differential equations generating the corresponding operator ideals [1, 2]. Algorithmically, the key tool is provided by Gröbner bases in rings of linear differential operators and an implementation is available in Chyzak's `Mgfun` package[1]. An additional, very important closure operation on multivariate D-finite power series is definite integration. It can be computed by an algorithm called *creative telescoping*, due to Zeilberger [15]. Again, this method takes as input (linear) differential operators and outputs differential operators (in fewer variables) satisfied by the definite integral. It turns out that the algorithmic realisation of creative telescoping has several common features with the algorithms we introduce here.

[1]This package is part of the algolib library available at http://algo.inria.fr/packages.

Beyond the multivariate case, Gessel considered the case of infinitely many variables and laid the foundations of a theory of D-finiteness for symmetric functions [3]. He defines a notion of D-finite symmetric series and obtains several closure properties. The motivation for studying D-finite symmetric series is that new closure properties occur and can be exploited to derive the D-finiteness of usual multivariate or univariate power series. Thus, the main application of [3] is a proof of the D-finiteness for several combinatorial counting functions. This is achieved by describing the counting functions as combinations of coefficients of D-finite symmetric series, which can then be computed by way of a scalar product of symmetric functions. Under certain conditions, the scalar product is D-finite, where D-finiteness is that of (usual) multivariate power series. Most of Gessel's proofs are not constructive. In this article, we give algorithms that compute the resulting systems of differential equations. Besides Gessel's work, these algorithms are inspired by methods used by Goulden, Jackson and Reilly in [5]. Finally, Gröbner bases are used to help make these methods into algorithms. An outcome is a simplification of the original methods.

This article is organized as follows. After recalling the necessary part of Gessel's work in Section 1, we present the algorithm for computing the differential equations satisfied by the scalar product in Section 2. The example of k-regular graphs is detailed in Section 3. We treat a variant of Young tableaux where each element is repeated k times in Section 4. (These are in bijection with a generalisation of involutions [7].) Then special cases where the algorithm can be further tuned are described in Section 5. This extended abstract does not contain the (technical) proofs of termination of the algorithms.

1. Symmetric D-finite Functions

In this section we recall the facts we need about symmetric functions, D-finite functions and symmetric D-finite functions.

1.1. **Symmetric functions.** We refer the reader to [9] for all definitions and notation related to symmetric functions.

Denote by $\lambda = (\lambda_1, \ldots, \lambda_k)$ a partition of the integer n (this means that $n = \lambda_1 + \cdots + \lambda_k$ and $\lambda_1 \geq \cdots \geq \lambda_k > 0$). Partitions serve as indices for the four principal symmetric function families that we use: homogeneous (h_λ), power (p_λ), monomial (m_λ), elementary (e_λ), and Schur (s_λ). These are functions in the infinite set of variables, $x_1, x_2, \ldots$ over a field K of characteristic 0. When the set of indices is restricted to the partitions of n, any of these families forms a basis for the vector space of symmetric polynomials of degree n in $x_1, \ldots, x_n$. Moreover, the set of p_i's, indexed by $i \in \mathbb{N}$, forms a basis of the ring of symmetric functions $\Lambda = K[p_1, p_2, \ldots]$.

Generating series of symmetric functions live in the larger ring of symmetric series $K[t][\![p_1, p_2, \ldots]\!]$. There, we have the generating series of homogeneous and elementary functions:

$$H(t) = \sum_n h_n t^n = \exp\left(\sum_i p_i \frac{t^i}{i}\right),$$

$$E(t) = \sum_n e_n t^n = \exp\left(\sum_i (-1)^i p_i \frac{t^i}{i}\right).$$

1.2. **Scalar product and coefficient extraction.** The ring of symmetric series is endowed with a scalar product defined as a bilinear symmetric form such that the bases (h_λ) and (m_λ) are dual to each other:

$$\langle m_\lambda, h_\mu \rangle = \delta_{\lambda\mu}, \tag{1}$$

where $\delta_{\lambda\mu} = 1$ if $\lambda = \mu$ and 0 otherwise.

An alternative notation for partitions is $\lambda = 1^{n_1} \cdots k^{n_k}$, which means that i occurs n_i times in λ, for $i = 1, 2, \dots, k$. Then the normalization constant

$$z_\lambda := 1^{n_1} n_1! \cdots k^{n_k} n_k!$$

plays the role of the square of a norm of p_λ in the following important formula:

$$\langle p_\lambda, p_\mu \rangle = \delta_{\lambda\mu} z_\lambda. \tag{2}$$

The scalar product is thus a basic tool for coefficient extraction. Indeed, if we write $F(x_1, x_2, \dots)$ in the form $\sum_\lambda f_\lambda m_\lambda$, then the coefficient of $x_1^{\lambda_1} \cdots x_k^{\lambda_k}$ in F is $f_\lambda = \langle F, h_\lambda \rangle$, by (1). Moreover, when $\lambda = 1^n$, the identity $m_\lambda = p_\lambda$ yields a simple way to compute this coefficient when F is written in the p_λ basis:

Theorem 1 (Gessel, Goulden & Jackson)**.** *Let θ be the homomorphism from the ring of symmetric functions to the ring of formal power series in t defined by $\theta(p_1) = t$, $\theta(p_n) = 0$ for $n > 1$. Then if F is a symmetric function,*

$$\theta(F) = \sum_{n=0}^{\infty} a_n \frac{t^n}{n!},$$

where a_n is the coefficient of $x_1 \cdots x_n$ in F.

Gessel also provides an analog for this theorem when $\lambda = 1^n 2^m$.

1.3. **Plethysm.** Plethysm is a way to compose symmetric functions. It can be defined by its action on the p_i's: $p_n \circ (a p_m) = a^n p_{nm}$, for any $a \in K$ and then extended to all of Λ by $f \circ (gh) = (f \circ g)(f \circ h)$, $f \circ (g + h) = (f \circ g) + (f \circ h)$ and $p_n \circ g = g \circ p_n$.

1.4. **D-finiteness of multivariate series.** Recall that a series $F \in K[\![x_1, \dots, x_n]\!]$ is *D-finite* in $x_1, \dots, x_n$ when the set of all partial derivatives $\partial^{i_1 + i_1 + \cdots + i_n} F / \partial x_1^{i_1} \cdots \partial x_n^{i_n}$ spans a finite-dimensional vector space over the field $K(x_1, \dots, x_n)$.

The properties we need here are summarized in the following theorem.

Theorem 2. (1) *The set of D-finite power series forms a K-subalgebra of $K[\![x_1, \dots, x_n]\!]$ for the usual product of series;*
(2) *If F is D-finite in $x_1, \dots, x_n$ then for any subset of variables $x_{i_1}, \dots, x_{i_k}$ the specialized function $F|_{x_{i_1} = \cdots = x_{i_k} = 0}$ is D-finite in the remaining variables;*
(3) *If $P(x)$ is a polynomial in $x_1, \dots, x_n$, then $\exp(P(x))$ is D-finite in $x_1, \dots, x_n$;*
(4) *If F and G are D-finite in the variables $x_1, \dots, x_{m+n}$, then the Hadamard product $F \times G$ with respect to the variables $x_1, \dots, x_n$ is D-finite in $x_1, \dots, x_{m+n}$.*

(Recall that the Hadamard product of two series is $\sum_{\alpha \in \mathbb{N}^k} a_\alpha \mathbf{u}^\alpha \times \sum_{\beta \in \mathbb{N}^k} b_\beta \mathbf{u}^\beta = \sum_{\alpha \in \mathbb{N}^k} a_\alpha b_\alpha \mathbf{u}^\alpha$, where $\mathbf{u}^\alpha = u_1^{\alpha_1} \cdots u_k^{\alpha_k}$.)

These properties are classical. The first three are elementary, the last one relies on more delicate questions of dimension and is due to Lipshitz [8].

1.5. **D-finite symmetric functions.** The definition of D-finiteness of series in an infinite number of variables is achieved by generalizing the property Theorem 2.2: $F \in K[\![x_1, x_2, \dots]\!]$ is called *D-finite* in the x_i if the specialization of all but a finite choice S of variables to 0 is D-finite for any choice of S.

In this case, all the properties in Theorem 2 hold in the infinite multivariate case.

The definition is then *specialized* to symmetric series by considering the ring of symmetric series $K[\![p_1, p_2, \dots]\!]$. Thus a symmetric series is called *D-finite* when it is D-finite in the p_i's.

Theorem 2.4 has the following very important consequence:

Theorem 3 (Gessel)**.** *Let f and g in $\big(K[[t_1,\dots,t_k]]\big)[[p_1,p_2,\dots]]$ be D-finite in the p_i's and t_j's, and suppose that g involves only finitely many of the p_i's. Then $\langle f,g\rangle$ is D-finite in the t_j's as long as it is well defined as a power series.*

1.6. **Effective D-finite symmetric closures.** Our work consists in making this theorem effective by giving an algorithm (in Section 2) producing linear differential equations annihilating $\langle f,g\rangle$ from an input consisting in generators of ideals of differential operators annihilating g and the specialization of f in the finite number of p_i's required by g.

Providing algorithms that manipulate linear differential equations amounts to making effective the closure properties of univariate D-finite series; similarly, algorithms operating on systems of linear differential operators make effective the closure properties of multivariate D-finite series. Our title is thus motivated by the fact that our algorithm makes it possible to compute all the information that can be predicted from D-finiteness.

In our examples, we make use of symmetric series that are built by plethysm. Closure properties are given by Gessel, but in our example we only need a simple consequence of Theorem 2.3, namely that if g is a polynomial in the p_i's, then $H\circ g$ and $E\circ g$ are D-finite.

2. Algorithm

We now give a new algorithm to compute scalar products of D-finite symmetric which satisfy the hypotheses of Theorem 3. When the number of t_j's is 1, the output is a single differential equation for which the available computer algebra algorithms might find a closed-form solution. In most cases however, no such solution exists and we are content with a differential equation out of which useful information can be extracted.

The basic tool we use here are noncommutative Gröbner bases in Weyl algebras. An introduction to this topic can be found in [11]. We work in an extension $\mathbb{A}$ of the Weyl algebra $K\langle p_1,\dots,p_n,\mathbf{t},\partial_1,\dots,\partial_n,\mathbf{d_t}\rangle$ in which the coefficients of the differential operators are still polynomials in the p_i's but now rational in $\mathbf{t}$. Here, $\mathbf{t}=t_1,\dots,t_k$, and $\mathbf{d_t}=d_{t_1},\dots,d_{t_k}$. Suppose F and G belong to $K[\mathbf{t}][[p_1,\dots,p_n]]$ and are D-finite symmetric series as in Theorem 3. In particular, they both satisfy systems of linear differential equations with coefficients in $K(\mathbf{t})[p_1,\dots,p_n]$. We can write these equations as elements of $\mathbb{A}$ acting on F and G. The sets I_F (resp. I_G) of all operators of $\mathbb{A}$ annihilating F (resp. G) is then a *left ideal* of $\mathbb{A}$. Given as input Gröbner bases of I_F and I_G, our algorithm outputs nontrivial elements in the annihilating left ideal of $\langle F,G\rangle$ in $K\langle\mathbf{t},\mathbf{d_t}\rangle$.

We first outline the algorithm for the special case when $F\in K[[p_1,\dots,p_n]]$, i.e., does not involve $\mathbf{t}$. Then, if $\phi\in I_F$,

$$0=\langle 0,G\rangle=\langle\phi(F),G\rangle=\langle F,\phi^*(G)\rangle\,,$$

where ϕ^* is the adjoint of ϕ with respect to the scalar product. (This can be computed from $p_i^*=i\partial_i,\partial_i^*=p_i/i$ respectively with $(\partial_i p_i)^*=\partial_i p_i$ [9]). Thus our aim is to determine $\beta\in R=I_F^*+I_G$ which is a polynomial in only the variables t and ∂_t, that is $\beta\in R\cap K\langle\mathbf{t},\mathbf{d_t}\rangle$.

Note however that while I_G is a left ideal, I_F^* is a *right* ideal and the sum of their elements does not form an ideal. This problem is very similar to the problem of creative telescoping: given an ideal I_F and a variable p, the aim in the first step of this method is to determine an element of $\partial\mathbb{A}+I_F$ that does not involve p. There also, $\partial\mathbb{A}$ is a right ideal. The algorithm we present thus has a nonfortuitous resemblance with that of [14].

The structure of R that we can use however, is that of a vector space over $K(\mathbf{t})$. (We could also use a structure of module over $K\langle\mathbf{t},\mathbf{d_t}\rangle$, but this will not generalize to the case when F depends on $\mathbf{t}$.) The idea is then to use linear algebra in this vector space to eliminate the ∂_i and p_i in R. Roughly speaking, we incrementally generate lines in a matrix corresponding to elements of R, and perform Gaussian elimination to get rid of the monomials involving ∂_i's or p_i's.

We generate elements of R iteratively by considering monomials α in increasing order for a monomial ordering such as $T = \text{degrevlex}(\mathbf{d_t}, p_1, \partial_1, \ldots, p_n, \partial_n)$ (total degree refined by reverse lexicographic order). Then for each α, we get two new elements of R using I_F and I_G. Next, these add two "lines" in a matrix (and for sufficiently large α only one "column") where we perform Gaussian elimination to cancel columns corresponding to monomials involving the p_i's or ∂_i's.

We now state the algorithm more formally. Then we give an example in the next section. After this example, we describe the modifications necessary to handle the general case and show how special cases can be handled more efficiently.

Algorithm 1 (Scalar Product)**.** Input: $F \in K[\![p_1, \ldots, p_n]\!]$ and $G \in K[\mathbf{t}][\![p_1, \ldots, p_n]\!]$.
Output: A differential equation satisfied by $\langle F, G\rangle$.

(1) Determine $\mathcal{G}_F$ and $\mathcal{G}_G$, Gröbner bases for I_F and I_G in $\mathbb{A}$ with respect to some term order T;
(2) Set $B := \{\}$;
(3) Iterate through each monomial $\alpha \in K[p_1, \ldots, p_n, \partial_1, \ldots, \partial_n, \mathbf{d}_t]$ incrementally with respect to the order T;
   (a) Determine $\alpha_F := \alpha - \alpha'$ where α' is the adjoint of α^* reduced with respect to $\mathcal{G}_F$. Insert this into the basis B;
   (b) Determine α_G, the reduction of α with respect to $\mathcal{G}_G$, and insert into the basis B;
   (c) If B contains an element β that has only t_j's and d_{t_j}'s, break and return β.

The operator $*$ is the adjunction operator described earlier. The reduction with respect to either Gröbner basis $\mathcal{G}_G$ or $\mathcal{G}_F$ is a multivariate analogue of the remainder in Euclidean division. It is such that for any α, α- (the reduction of α with respect to $\mathcal{G}$) belongs to the ideal generated by $\mathcal{G}$.

The insertion into the basis B performs the Gaussian reduction of α with respect to the "lines" already in B and returns the new value of B. In practice, B can be handled (not inefficiently) by a computation of Gröbner basis over a module with respect to a term order that eliminates the p_i's and ∂_i's. The insertion corresponds to reducing with respect to this basis and updating it.

3. Example: k-regular Graphs

This example is taken from [3] and [5]. After introducing its combinatorial motivation, we describe in detail how our algorithm deals with it.

A generating function for all simple graphs labelled with integers from $\mathbb{N} \setminus \{0\}$, $\mathcal{G}$ is:

$$G(\mathbf{x}) = \sum_{G \in \mathcal{G}} \prod_{(i,j) \in E(G)} x_i x_j = \prod_{i<j} (1 + x_i x_j),$$

as each edge $(i, j) \in E(G)$ is either in the graph or not. Similarly, we can make a generating function for graphs with multiple edges

$$G'(\mathbf{x}) = \prod_{i<j} \frac{1}{(1 - x_i x_j)}.$$

Clearly both of these are symmetric functions, and in fact, $G(\mathbf{x}) = H \circ (e_2(\mathbf{x}))$ and $G'(\mathbf{x}) = E \circ (e_2(\mathbf{x}))$. These can be rewritten in terms of the p_i's:

$$G = \exp\left(\sum_i p_i + p_{2i}/2\right) \qquad \text{and} \qquad G' = \exp\left(\sum_i (-1^i)p_i + p_{2i}/2\right).$$

In any given term, the degree of x_i gives the valency of node i. So, for example, the coefficient $g_n = [x_1 \cdots x_n]G(\mathbf{x})$ gives the number of 1-regular graphs, or perfect matchings on the complete graph on n vertices, and in general the coefficient $g_n^{(k)} = [x_1^k \cdots x_n^k]G(\mathbf{x})$ gives the number of k-regular graphs on n vertices. Since coefficient extraction amounts to a scalar product, the generating function of k-regular graphs is given by

$$(3) \quad G_k(t) = \sum g_n^{(k)} t^n/n! = \left\langle G, \sum_n h_{k^n} t^n/n! \right\rangle = \left\langle G, \sum_n (h_k t)^n/n! \right\rangle = \langle G, \exp(h_k t)\rangle .$$

Now, as $h_k = \sum_{\lambda \vdash k} p_\lambda/z_\lambda$ (where the sum is over all partitions λ of k), the exponential generating function of these numbers $H^{(k)}(t) = \sum_t h_{k^n} t^n/n! = \exp\left(t \sum_{\lambda\vdash n} p_\lambda/z_\lambda\right)$ is an exponential in a finite number of p_i's. By Theorem 2.3, this is D-finite. Further, as a result of scalar product property (2), we can rewrite equation (3) as

$$(4) \qquad G_k(t) = \left\langle \exp\left(\sum_{i \text{ even}, i\le k} (-1)^{i/2}\frac{p_i^2}{2i} + \frac{p_i}{i} + \sum_{i \text{ odd}, i \le k} \frac{p_i^2}{2i}\right), \exp\left(t \sum_{\lambda \vdash k} \frac{p_\lambda}{z_\lambda}\right)\right\rangle$$

and now by Theorem 3 this generating function $G_k(t)$ is D-finite.

3.1. **Computation for $k = 2$.** In this section we calculate $G_2(t)$, beginning with equation (4):

$$G_2(t) = \left\langle \exp((p_1^2 - p_2)/2 - p_2^2/4), \exp(t(p_1^2 + p_2)/2)\right\rangle .$$

Assign $f = \exp((p_1^2 - p_2)/2 - p_2^2/4)$ and $g = \exp(t(p_1^2 + p_2)/2)$. The input of the algorithm consists in the following Gröbner bases, with respect to the degrevlex$(t, dt, p_1, \partial_1, p_2, \partial_2)$ term ordering, which express the first order differential equations satisfied by f and g:

$$\mathcal{G}_f = \{2\partial_2 + p_2 + 1, \partial_1 - p_1\} \qquad \text{and} \qquad \mathcal{G}_g = \{2\partial_2 - t, p_1^2 + p_2 - 2d_t, \partial_1 - tp_1\}.$$

Note that the elements of $\mathcal{G}_f$ are self-adjoint.

After a few first steps which we omit here, we obtain

$$B = \{p_2 + t + 1, p_1, 2\partial_2 - t, \partial_1, p_1^2 - 2d_t - t - 1\}.$$

We illustrate a typical insertion step by considering the monomial $\alpha = p_1\partial_1$. First we compute $\alpha_f = p_1\partial_1 - p_1^2 + 1$ and $\alpha_g = p_1\partial_1 + tp_2 - 2td_t$. Next, α_f is inserted. Its leading monomial $p_1\partial_1$ adds a new "column" and the rest of the line is reduced by rewriting p_1^2 using the last element of B. This step leads to

$$B := B \cup \{p_1\partial_1 - 2d_t - t\}.$$

Then the algorithm inserts α_g. Its leading monomial $p_1\partial_1$ is already present in B which leads to a first reduction into $tp_2 - 2(1-t)d_t - t$. Then the new leading term is tp_2 which can be reduced by the first element of B and thus we get

$$B := B \cup \{2(1-t)\partial_t + t^2\}.$$

This new element involves t and d_t only and thus we have found the classical differential equation

$$2(1-t)G_2'(t) - t^2 G_2(t) = 0.$$

	2
ϕ_0	$-t^2$
ϕ_1	$-2t+2$
ϕ_2	0
	3
ϕ_0	$t^3(2t^2+t^4-2)^2$
ϕ_1	$-3(t^{10}+6t^8+3t^6-6t^4-26t^2+8)$
ϕ_2	$-9t^3(2t^2+t^4-2)$
	4
ϕ_0	$-t^4(t^5+2t^4+2t^2+8t-4)^2$
ϕ_1	$-4(t^{13}+4t^{12}-16t^{10}-10t^9-36t^8-220t^7-348t^6$ $-48t^5+200t^4-336t^3-240t^2+416t-96)$
ϕ_2	$16t^2(t-1)^2(t^5+2t^4+2t^2+8t-4)(t+2)^3$

Table 1. Differential equation $\phi_2 G_k'' + \phi_1 G_k' + \phi_0 G_k = 0$ satisfied by $G_k(t)$, $k = 2, 3, 4$.

Table 1 summarizes the results by the same algorithm for $k = 2, 3, 4$. These match with the results in [5].

4. Hammond Series

In the example above, it turned out that apart from the monomials of degree 1, only the monomials p_1^2 and $p_1\partial_1$ were necessary to reach the solution. However, depending on the term order, the algorithm might well consider many monomials before it adds the ones that eliminate the p_i's and d_i's. The problem becomes far more serious as the number of monomials increases. It turns out that in the frequent case when the scalar product is of the type $\langle F, H^{(k)}(t)\rangle$ it is possible to modify the approach and eliminate the p_i and the ∂_i in a more efficient manner using the *Hammond series* (or H-series) introduced by Goulden, Jackson and Reilly in [5]. In particular, their H-series theorem is useful.

For $F \in K[\![p_1, p_2, \dots]\!]$, the Hammond series of F, is defined as

$$\mathcal{H}(F) = \left\langle F, \sum_{\lambda} h_\lambda z_\lambda^{-1} \mathbf{t}^\lambda \right\rangle,$$

where the sum is over all partitions λ and $\lambda = 1^{m_1} \cdots k^{m_k}$ implies $\mathbf{t}^\lambda = t_1^{m_1} \cdots t_k^{m_k}$.

Observe that the generating function for k-regular graphs is $G_k(t) = \mathcal{H}(G)(0, \dots, 0, t, 0, \dots)$ where the t occurs in position k. This is true for any generating function which takes the form $\langle F, H^{(k)}(t)\rangle$, for some F.

The H-series theorem states that $\mathcal{H}(\partial_{p_n} F)$ and $\mathcal{H}(p_n F)$ can be expressed in terms of the $\partial_{t_i}\mathcal{H}(F)$'s. In terms of Gröbner bases, this corresponds to introducing the additional variables $t_1, \dots, t_k$ instead of $t = t_k$ alone and work with the generating series $\mathcal{H}_k(t_1, \dots, t_k)$ of the $h_\lambda z_\lambda^{-1}$ over partitions whose largest part is k, instead of the univariate $\mathcal{H}_k = H^{(k)}(t)$. The H-series theorem therefore implies that for an appropriate term order, there is a Gröbner basis of the set $I_{\mathcal{H}_k}$ of all operators of $\mathbb{A}$ annihilating H_k, with elements of the form

$$p_i - P_i(\mathbf{t}, \mathbf{d_t}), \quad \partial_i - Q_i(\mathbf{t}, \mathbf{d_t}), \qquad i = 1, \dots, k. \tag{5}$$

The algorithm is modified as follows.

Algorithm 2 (Hammond Series)**.** Input: An integer k, and $F \in K[\![p_1, \dots, p_n]\!]$.
Output: A differential equation satisfied by $\mathcal{H}(F)(0, \dots, 0, t, 0, \dots)$ where t is in the kth position.

(1) Compute $\mathcal{G}_F$ a Gröbner basis for I_F the left ideal annihilating F in $\mathbb{A}$;
(2) Compute $\mathcal{G}_{\mathcal{H}_k}$ a Gröbner basis of the form (5);
(3) For each $\alpha \in \mathcal{G}_F$, compute $r_\alpha \in K[\mathbf{t}, \mathbf{d}_t]$ as the reduction of α^* by $\mathcal{G}_{\mathcal{H}_k}$. Let R_1 be the set of r_α's;
(4) For i from 1 to $k-1$ eliminate ∂_i from R_i and set $t_i = 0$ in the resulting polynomials; call R_{i+1} the new set;
(5) Return R_k.

After step (3), all the p_i's and ∂_i's have been eliminated and thus we have a set of generators of a D-finite ideal annihilating $\langle F, \mathcal{H}_k \rangle$. Then, in order to obtain differential equations satisfied by the specialization at $t_1 = \cdots = t_{k-1} = 0$, step (4) proceeds in order by eliminating differentiation with respect to t_i and then setting $t_i = 0$ in the remaining operators.

Note that the Gröbner basis of step (2) can be precomputed for the required k's (but most of the time is actually spent in step (4)).

In order to compute the elimination in step (4), one should not compute a Gröbner basis for an elimination order, since this would in particular perform the unnecessary computation of a Gröbner basis of the eliminated ideal. Instead, one can modify the main loop in the Gröbner basis computation so that it stops as soon as sufficient elimination has been performed or revert to skew elimination by the non-commutative version of the extended Euclidean algorithm as described in [2]. This is the method we have adopted in the example session given in Appendix B.

This calculation is comparatively rapid since the size of the basis is greatly reduced. Further, it reduces as it progresses, on account of setting variables to 0. We can compute the case of 4-regular graphs in a second, in place of a couple of minutes using the general algorithm, although the 5-regular expression requires significantly more time computationally.

5. Example: k-regular Tableaux and Generalized Involutions

Another family of combinatorial objects whose generating function can be resolved with this method is a certain class of Young tableaux.

Standard Young tableaux are in direct correspondence with many different combinatorial objects. For example, Stanley [13] has studied the link between standard tableaux and paths in Young's lattice, the lattice of partitions ordered by inclusion of diagrams. This link was generalized by Gessel [4] to tableaux with repeated entries. Gessel remarks that such paths have arisen in the work of Sundaram and the combinatorics of representations of symplectic groups.

Here we consider Young tableaux in which each entry appears k times. The tableaux are column strictly increasing and row weakly increasing. A Young tableaux with these properties is called k-regular. These correspond to paths in Young's lattice with steps of length k. The set of k-regular tableaux of size kn are also in bijection with symmetric $n \times n$ matrices with non-negative entries and each row sum equal to k.

Gessel notes that for fixed k, the generating series of the number of k-regular tableaux is D-finite [3]. Our method makes this effective.

The weight of a tableau is $\mu = (\mu_1, \dots, \mu_k)$ where μ_1 is the number of 1s, μ_2 is the number of 2s, etc. Thus, a k-regular tableau of size kn has weight k^n. Two observations

	1
ϕ_0	$-(t-1)$
ϕ_1	1
ϕ_2	0
	2
ϕ_0	$t^2(t-2)$
ϕ_1	$-2(t-1)^2$
ϕ_2	0
	3
ϕ_0	$(t^{11}+t^{10}-6t^9-4t^8+11t^7-15t^6+8t^5-2t^3+12t^2-24t-24)$
ϕ_1	$-3t(t^{10}-2t^8+2t^6-6t^5+8t^4+2t^3+8t^2+16t-8)$
ϕ_2	$9t^3(-t^2-2+t+t^4)$
	4
	(See Appendix A)

Table 2. The differential equation $\phi_2 Y_k^{(2)}(t) + \phi_1 Y_k(t)' + \phi_0 Y_k(t) = 0$ satisfied by $Y_k(t)$, $k = 1, \ldots, 4$.

from [9] are essential. First, $[x_1^{\mu_1} \cdots x_k^{\mu_k}]s_\lambda$ is the number of (column strictly increasing, row weakly increasing) tableaux with weight μ. Secondly,

$$\sum_\lambda s_\lambda = H \circ (e_1 + e_2) = \exp\left(\sum_i p_i^2/2i + \sum_{i \text{ odd}} p_i\right),$$

which is D-finite.

Define now $y_n^{(k)}$ to be the number of k-regular tableaux of size kn, and let Y_k be the generating series of these numbers:

$$Y_k(t) = \sum_n y_n^{(k)} t^k.$$

The previous two observations imply

$$Y_k(t) = \left\langle \exp\left(\sum_{i=1}^k p_i^2/2i + \sum_{i \text{ odd}}^k p_i\right), H^{(k)}(t)\right\rangle,$$

where, as before, $H^{(k)}(t) = \sum_n h_{k^n} t^n$. This problem is well suited to our methods since again we treat an exponential of a polynomial in the p_i's.

Calculating the equations for $k = 1, 2, 3, 4$ is rapid with either Algorithm 1 or Algorithm 2. The resulting differential equations are listed in Table 2. For $k = 1, 2$ these results accord with known results [6, 13]. The first few values of $y_n^{(k)}$ are summarized in the following table.

k	$y_0^{(k)}, y_1^{(k)}, y_2^{(k)}, \ldots$
1	1, 1, 2, 4, 10, 26, 76, 232, 764, 2620, 9496, 35696, 140152, 568504
2	1, 1, 3, 11, 56, 348, 2578, 22054, 213798, 2313638, 27627434
3	1, 1, 4, 23, 214, 2698, 44288, 902962, 22262244
4	1, 1, 5, 42, 641, 14751, 478711, 20758650, 1158207312

Table 3. $y_n^{(k)}$, The number of k-regular tableaux of size kn

6. General Case

So far, we have concentrated on the special case when only one of the D-finite symmetric functions whose scalar product is sought involves the variables $\mathbf{t}$. While this is the more useful case in many applications, it is possible to modify our algorithm in order to accommodate t_j's in both functions and thus make effective the full power of Theorem 3.

The new difficulty is that for each t_i, ∂_{t_i} is no longer self-adjoint. Instead, the usual product rule applies:

$$\partial_{t_i} \langle F, G\rangle = \langle \partial_{t_i} F, G\rangle + \langle F, \partial_{t_i} G\rangle ,$$

and one rewrites $\langle \partial_{t_i} F, G\rangle$ as $-\langle F, \partial_{t_i} G\rangle + \partial_{t_i} \langle F, G\rangle$. The idea is then to manipulate operators in *two* sets of ∂'s, the usual one and a new one that we denote ∂_{g_i}.

A monomial $\mathbf{t}^\alpha \partial_{\mathbf{g}}^\beta \mathbf{d}_{\mathbf{t}}^\gamma$ thus acts on $\langle F, G\rangle$ by

$$\mathbf{t}^\alpha \partial_{\mathbf{g}}^\beta \mathbf{d}_{\mathbf{t}}^\gamma \langle F, G\rangle = \mathbf{t}^\alpha \mathbf{d}_{\mathbf{t}}^\gamma \left\langle F, \mathbf{d}_{\mathbf{t}}^\beta G\right\rangle .$$

The action of polynomials is defined from this by linearity.

The algorithm consists as before in an iteration over monomials α in increasing order (with no ∂_{g_i} involved). For each such monomial, its adjoint is computed by $d_{t_i}^* = d_{t_i} - \partial_{g_i}$ while as before $p_i^* = i\partial_i$ and $\partial_i^* = p_i/i$. This amounts to expressing $0 = \langle \alpha F, G\rangle$ as a linear combination of $\mathbf{d}_{\mathbf{t}}^{\mathbf{i}} \langle F, \beta G\rangle$ with coefficients in $K[\mathbf{t}]$. The rest of the algorithm proceeds as before by performing Gaussian elimination over $K(\mathbf{t})$. This is summarized in the following algorithm

Algorithm 3 (General Scalar Product)**.** Input: $F, G \in K[\mathbf{t}][\![p_1, \ldots, p_n]\!]$.
Output: A differential equation satisfied by $\langle F, G\rangle$.

(1) Compute $\mathcal{G}_F$ and $\mathcal{G}_G$, Gröbner bases for I_F and I_G in $\mathbb{A}$ with respect to the order T;
(2) Replace dt_i's by ∂g_i's in $\mathcal{G}_G$;
(3) Use the rule $d_{t_i}^* = d_{t_i} - \partial_{g_i}$ in all adjoint computations;
(4) Apply Algorithm Scalar Product where the elimination in B has to eliminate the ∂g_i's besides the p_i's and ∂_i's.

Appendix A. 4-regular Young Tableaux

The differential equation satisfied by $Y_4(t)$ is

$$-64t^4(t-2)^2(t+1)^4\alpha(t)Y_4^{(3)}(t) + 16t^2(t-2)(t+1)^2\beta(t)Y_4^{(2)}(t) - 4\gamma(t)Y_4'(t) + \delta(t)Y_4(t) = 0$$

where $\alpha(t), \beta(t), \gamma(t), \delta(t)$ are irreducible polynomials given by

$$\begin{aligned}
\alpha(t) &= t^{14} - t^{13} - 5t^{12} - 7t^{11} + 6t^{10} + 35t^9 + 39t^7 - 50t^6 - 162t^5 - 92t^4 \\
&\quad + 228t^3 + 424t^2 + 248t + 48 \\
\beta(t) &= t^{29} - 3t^{28} - 16t^{27} + 24t^{26} + 147t^{25} + 14t^{24} - 770t^{23} - 666t^{22} + 1416t^{21} \\
&\quad + 3567t^{20} - 916t^{19} - 16598t^{18} + 17766t^{17} + 40678t^{16} - 102556t^{15} - 53272t^{14} \\
&\quad + 390656t^{13} + 364080t^{12} - 707936t^{11} - 1406336t^{10} - 552544t^9 + 1397664t^8 + 2020864t^7 \\
&\quad + 176256t^6 - 916864t^5 + 304896t^4 + 1283328t^3 + 877056t^2 + 253440t + 27648 \\
\gamma(t) &= t^{28} - t^{27} - 14t^{26} - 20t^{25} + 111t^{24} + 278t^{23} - 196t^{22} - 1216t^{21} - 1384t^{20} + 2765t^{19} \\
&\quad + 3170t^{18} - 3400t^{17} + 12140t^{16} + 15588t^{15} - 70280t^{14} - 108946t^{13} + 121796t^{12} \\
&\quad + 349056t^{11} + 116992t^{10} - 481704t^9 - 706320t^8 + 3040t^7 + 581184t^6 + 158688t^5 \\
&\quad - 297408t^4 - 173952t^3 + 22272t^2 + 35712t + 6912 \\
\delta(t) &= 2t^{21} - 3t^{20} - 17t^{19} - 2t^{18} + 74t^{17} + 105t^{16} - 108t^{15} - 172t^{14} - 252t^{13} + 432t^{12} \\
&\quad - 667t^{11} + 1500t^{10} + 7336t^9 - 3772t^8 - 23056t^7 - 20584t^6 + 15504t^5 + 38160t^4 \\
&\quad + 17904t^3 - 4512t^2 - 5568t - 1152.
\end{aligned}$$

Appendix B. Sample Maple Session for 3-regular Graph Computation

The following Maple session indicates the high-level routines required to program Algorithm 2. It requires the library `algolib`, which is available at `http://algo.inria.fr/packages/`.

```
with(Ore_algebra): with(Mgfun): with (Groebner): # load the packages
# Determine the DE satisfied by the generating function for 3 regular graphs
k:=3:
Fp:= exp(1/2*p1^2-1/4*p2^2-1/2*p2):
Gp:=exp(1/6*t3*p1^3+1/2*t2*p1^2+t1*p1+1/2*t3*p2*p1+1/2*t2*p2+1/3*t3*p3):
# define the variables
vars:= seq(p.i, i=1..k): dvars:= seq(d.i, i=1..k):
tvars:= seq(t.i, i=1..k): dtvars:= seq(dt.i, i=1..k):

# define the algebra
A:= diff_algebra(seq([dvars[i], vars[i]], i=1..k),
seq([dtvars[i], tvars[i]], i=1..k), polynom={vars}):
At:= diff_algebra(seq([dtvars[i], tvars[i]], i=1..k)):

# define the term orders
T[g]:=termorder(A, lexdeg([dvars, vars],[dtvars])):
T[f]:=termorder(A,tdeg(vars, dvars, dtvars)):

#define the systems
sys[g]:=dfinite_expr_to_sys(exp_g, F(seq(p.i::diff, i=1..k),
        seq(t.i::diff, i=1..k))):
newsys[g]:=subs([seq(diff(F(vars,tvars),vars[i])=dvars[i],i=1..k),
                 seq(diff(F(vars, tvars), tvars[i])=dtvars[i], i=1..k),
                 F(vars,tvars)=1],sys[g]):
```

```
#find the Groebner basis for G
GB[g]:=gbasis(newsys[g],T[g]);

# do the same for F
sys[f]:=dfinite_expr_to_sys(exp_f, F(seq(p.i::diff, i=1..k))):
newsys[f]:=subs([seq(diff(F(vars),vars[i])=dvars[i],i=1..k),
F(vars)=1],sys[f]);
GB[f]:=gbasis(newsys[f],T[f]);

# define the adjoint and reduction procedures
star:= x->subs([seq(d.i=1/i*p.i, i=1..k),seq(p.i=d.i*i, i=1..k)],x):
rdc[f] := x->star(star(x)-map(normalf, star(x), GB[f], T[f]));
rdc[g] := x->normalf(x, GB[g], T[g]);

# reduce the Groebner basis of F
for pol in GB[f] do m[pol]:=rdc[g](pol) od:

# small optimization: we will always try to reduce with respect to a
# linear term when possible
lpol:=[seq(m[i],i=subsop(1=NULL,GB[f])),m[GB[f][1]]:

for  indelim from k-1 by -1 to 1 do
   # eliminate dt.indelim
   for j from 2 to nops(lpol) do
       newpol[j]:=skew_elim(lpol[j],lpol[1],dt.indelim,At) od;
   # set t.indelim = 0
   lpol:=map(primpart,subs(t.indelim=0,[seq(newpol[j],j=2..nops(lpol))]),
           [dtvars])
od:

# the only term left is the correct one
ode:=op(lpol):
# convert to recurrence
diffeqtorec({applyopr(ode, F(t.k), At), F(0)=1}, F(t.k), a(n)):
# calculate some terms
rectoproc(,a(n),list)(20):[seq([i]*(i-1)!,i=1..nops())];

  [1, 0, 0, 0, 1, 0, 70, 0, 19355, 0, 11180820, 0, 11555272575, 0,
        19506631814670, 0, 50262958713792825, 0,
        187747837889699887800, 0, 976273961160363172131825]
```

References

[1] Frédéric Chyzak. *Fonctions holonomes en calcul formel.* Phd thesis, École polytechnique, 1998. INRIA, TU-0531. 227 pages.

[2] Frédéric Chyzak and Bruno Salvy. Non-commutative elimination in Ore algebras proves multivariate holonomic identities. *Journal of Symbolic Computation*, 26(2):187–227, August 1998.

[3] Ira M. Gessel. Symmetric functions and P-recursiveness. *Journal of Combinatorial Theory,* Series A, 53:257–285, 1990.

[4] Ira M. Gessel. Counting paths in Young's lattice. *Journal of Statistical Planning and Inference*, 34(1):125–134, 1993.

[5] I. P. Goulden, D. M. Jackson, and J. W. Reilly. The Hammond series of a symmetric function and its application to P-recursiveness. *SIAM Journal on Algebraic and Discrete Methods*, 4(2):179–193, 1983.

[6] Hansraj Gupta. Enumeration of symmetric matrices. *Duke Mathematical Journal*, 35:653–659, 1968.

[7] Donald E. Knuth. Permutations, matrices, and generalized Young tableaux. *Pacific Journal of Mathematics*, 34:709–727, 1970.

[8] L. Lipshitz. The diagonal of a D-finite power series is D-finite. *Journal of Algebra*, 113:373–378, 1988.

[9] Ian Grant Macdonald. *Symmetric functions and Hall polynomials*. Oxford Mathematical Monographs. Oxford University Press, 2nd edition, 1995.

[10] Christian Mallinger. *Algorithmic Manipulations and Transformations of Univariate Holonomic Functions and Sequences*. Master thesis, RISC, Johannes Kepler Universität Linz, Austria, August 1996. Available at the URL:
`http://www.risc.uni-linz.ac.at/research/combinat/risc/publications/`.

[11] Mutsumi Saito, Bernd Sturmfels, and Nobuki Takayama. *Gröbner deformations of hypergeometric differential equations*. Springer-Verlag, Berlin, 2000.

[12] Bruno Salvy and Paul Zimmermann. Gfun: a Maple package for the manipulation of generating and holonomic functions in one variable. *ACM Transactions on Mathematical Software*, 20(2):163–177, 1994.

[13] Richard P. Stanley. *Enumerative combinatorics*, volume 2. Cambridge University Press, 1999.

[14] Nobuki Takayama. An algorithm of constructing the integral of a module — an infinite dimensional analog of Gröbner basis. In *Symbolic and algebraic computation*, pages 206–211. ACM, 1990. Proceedings of ISSAC'90, Kyoto.

[15] Doron Zeilberger. The method of creative telescoping. *Journal of Symbolic Computation*, 11:195–204, 1991.

Projet Algorithmes, INRIA Rocquencourt, France
E-mail address: `{Frederic.Chyzak,Bruno.Salvy}@inria.fr`

LaCIM, Université du Québec à Montréal, Canada
E-mail address: `mishna@math.uqam.ca`

PARTITIONS AND COMPOSITIONS DEFINED BY (IN)EQUALITIES (EXTENDED ABSTRACT)

SYLVIE CORTEEL AND CARLA D. SAVAGE

Abstract. We consider sequences of integers $(\lambda_1, \dots, \lambda_k)$ defined by a system of linear inequalities with integer coefficients. We show that when the constraints are strong enough to guarantee that all λ_i are nonnegative, the weight generating function for the integer solutions has a finite product form $\prod_i (1 - q^{b_i})^{-1}$, where the b_i are positive integers that can be computed from the coefficients of the inequalities. The results are proved bijectively and are used to give several examples of interesting identities for integer partitions and compositions. The method can be adapted to accomodate equalities along with inequalities and can be used to obtain multivariate forms of the generating function. Our initial results were conjectured thanks to the Omega package [6].

We generalize the method to handle special cases with rational coefficients (including lecture hall partitions) and obtain new identities involving partitions and compositions defined by the ratio of consecutive parts. In particular, we obtain a surprising result about "anti-lecture hall" compositions.

Résumé. Nous considérons des suites d'entiers $(\lambda_1, \dots, \lambda_k)$ définies par un système d'inégalités linéaires à coefficients entiers. Nous montrons que si les solutions du système sont toujours des suites d'entiers positifs ou nuls alors la série génératrice des solutions selon leur poids est le produit $\prod_{i=0}^{k-1}(1-q^{b_i})^{-1}$. Les b_i sont des entiers positifs que l'on calcule à partir du système. Les résultats sont démontrés bijectivement et donnent des identités intéressantes pour les partitions et les compositions. La méthode peut être adaptée aux systèmes d'(in)égalités et permet aussi d'obtenir des séries génératrices multivariées. Les résultats initiaux ont été conjecturés grâce au package Omega [6].

Nous généralisons la méthode pour manipuler des cas avec des coefficients rationnels (comme les lecture hall partitions) et obtenons des nouvelles identités pour les partitions et compositions définies par le rapport entre parts successives. En particulier nous obtenons un résultat surprenant pour les "anti-lecture hall" compositions.

1. Introduction

For a sequence $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$ of integers, define the *weight* of λ to be $\lambda_1 + \cdots + \lambda_k$ and call each λ_i a *part* of λ. If a sequence λ of weight n has all parts nonnegative, we call it a *composition* of n into k nonnegative parts and if, in addition, λ is a nonincreasing sequence, we call it a *partition* of n into at most k parts. In the remainder of the paper we will consider that $\lambda_i = 0$ if $i < 0$ or $i > k$.

In this paper we want to study partitions and compositions into k nonnegative parts defined by equalities and inequalities. This work was motivated by results of the form :

- Given a positive integer r, the partitions $\lambda = (\lambda_1, \dots, \lambda_k)$ of n which satisfy $\lambda_i \geq r\lambda_{i+1}$ for $1 \leq i \leq k$ have weight generating function $\prod_{i=0}^{k-1}\left(1 - q^{1+r+\cdots+r^i}\right)^{-1}$ [14].
- Given a positive integer r the weight generating function of the partitions λ with at most k parts and $\lambda_i \geq \sum_{j=1}^{r}(-1)^{j+1}\binom{r}{j}\lambda_{j+i}$, $1 \leq i < k$ is : $\prod_{i=0}^{k-1}\left(1 - q^{\binom{i+r}{r}}\right)^{-1}$. See [2, 13, 18].

- The weight generating function of the partitions λ with at most k parts and $\lambda_i/(k-i+1) \geq \lambda_{i+1}/(k-i)$, $1 \leq i < k$ is : $\prod_{i=0}^{k-1}\left(1-q^{2i+1}\right)^{-1}$. See the *Lecture Hall Theorem* in [10].

More generally, we consider integer sequences λ of length k satisfying $\lambda_i \geq \sum_{j=1}^{k-i} a_{ij}\lambda_{i+j}$ where the a_{ij} guarantee that all $\lambda_i \geq 0$. We show in Section 2 that when the a_{ij} are all integers, the weight generating function for these compositions is $\prod_{i=0}^{k-1}(1-q^{b_i})^{-1}$, where the $b = (b_0, \ldots, b_{k-1})$ is a sequence of positive integers that may be readily derived from the a_{ij}. Several generalizations and a linear algebra proof are included.

In Section 3, we consider rational coefficients a_{ij}. We show how to use the results of Section 2 to give an explicit form for the generating function for *any* set of compositions defined by the ratio of consecutive parts:

$$\lambda_1 \geq \frac{n_1}{d_1}\lambda_2; \quad \lambda_2 \geq \frac{n_2}{d_2}\lambda_3; \quad \lambda_3 \geq \frac{n_3}{d_3}\lambda_4; \quad \ldots; \quad \lambda_{k-1} \geq \frac{n_{k-1}}{d_{k-1}}\lambda_k; \quad \lambda_k \geq 0,$$

This result has been implemented in Maple and our experiments have led to several interesting results. We focus here on some related to the Lecture Hall Theorem.

In [10], Bousquet-Mélou and Eriksson considered the set L_k of partitions λ into at most k parts satisfying $\lambda_i/(k-i+1) \geq \lambda_{i+1}/(k-i)$, for $1 \leq i < k$, and proved the following *Lecture Hall Theorem*:

$$\sum_{\lambda \in L_k} q^{|\lambda|} = \prod_{i=0}^{k-1} \frac{1}{1-q^{2i+1}}. \tag{1}$$

This result was generalized in [11] to an (m,l)-*Lecture Hall Theorem* $(m, l \geq 2)$ for partitions into at most k parts satisfying $\lambda_i/a_{k-i+1} \geq \lambda_{i+1}/a_{k-i}$, for $1 \leq i < k$, where $\{a_i\}$ is the (m,l)-sequence defined by:

$$a_{2i} = la_{2i-1} - a_{2i-2}; \qquad a_{2i-1} = ma_{2i-2} - a_{2i-3}, \quad i \geq 2,$$

with the initial conditions $a_1 = 1$ and $a_2 = l$, $m, l \geq 2$. A different approach in [12] led to the *Refined Lecture Hall Theorem* (setting $u = v = 1$ gives (1)) :

$$\sum_{\lambda \in L_k} q^{|\lambda|} u^{|\lceil\lambda\rceil|} v^{o(\lceil\lambda\rceil)} = \prod_{i=1}^{k} \frac{1+uvq^i}{1-u^2q^{k+i}},$$

with $\lceil\lambda\rceil = (\lceil\lambda_1/(k-i+1)\rceil, \lceil\lambda_2/(k-i)\rceil, \ldots, \lceil\lambda_k/k\rceil)$ and $o(\lambda)$ is the number of odd parts in λ.

In Section 4, we show a slight generalization of (1) in which the constraints on λ_1 can be modified. In a footnote in [11], Bousquet-Mélou and Erikson note that their proof of the (m,l)-Lecture Hall Theorem can be simplified. In Section 5, we shall describe the resulting short and elegant proof of the (m,l)-Lecture Hall Theorem and show that it can be generalized to compositions when $m = 1$ and $l > 3$ or $m > 3$ and $l = 1$.

So, for one example, compositions satisfying,

$$\frac{\lambda_1}{12} \geq \frac{\lambda_2}{5} \geq \frac{\lambda_3}{8} \geq \frac{\lambda_4}{3} \geq \frac{\lambda_5}{4} \geq \frac{\lambda_6}{1}$$

have generating function

$$[(1-q)(1-q^4)(1-q^5)(1-q^7)(1-q^{11})(1-q^{17})]^{-1}.$$

Finally, in Section 6, we prove another new result, the *Anti-Lecture Hall Theorem* for the set A_k of compositions into at most k parts satisfying $\lambda_i/i \geq \lambda_{i+1}/(i+1)$, for $1 \leq i < k$:

$$\sum_{\lambda \in A_k} q^{|\lambda|} = \prod_{i=1}^{k} \frac{1+q^i}{1-q^{i+1}}. \tag{2}$$

In fact, we prove the following refinement of (2):

$$\sum_{\lambda \in A_k} q^{|\lambda|} u^{|\lfloor \lambda \rfloor|} v^{o(\lfloor \lambda \rfloor)} = \prod_{i=1}^{k} \frac{1+uvq^i}{1-u^2q^{i+1}}.$$

where $\lfloor \lambda \rfloor = (\lfloor \lambda_1/1 \rfloor, \lfloor \lambda_2/2 \rfloor, \ldots, \lfloor \lambda_k/k \rfloor)$. The bijective proof we give follows the idea of Yee's beautiful proof [17] of the Refined Lecture Hall Theorem.

Because of space constraints we give only sketches of the proofs in this extended abstract.

Our initial results were conjectured thanks to experiments with the Omega package [6, 4], a Mathematica implementation of the Omega operator defined by MacMahon [15] as:

$$\underset{\geq}{\Omega} \sum_{s_1=-\infty}^{\infty} \cdots \sum_{s_r=-\infty}^{\infty} A_{s_1,\ldots,s_r} x_1^{s_1} \ldots x_r^{s_r} = \sum_{s_1=0}^{\infty} \cdots \sum_{s_r=0}^{\infty} A_{s_1,\ldots,s_r}.$$

This operator was then not used for 85 years except by Stanley in 1973 [16]. A few years ago Andrews revived this operator [1, 2] and used it in [1] to give a second proof of the Lecture Hall Theorem. In conjunction with Paule and Riese, he implemented the operator in the Omega package and together they have continued to identify the power of the Omega operator for such combinatorial problems as magic squares [9], hypergeometric multisums [3], constrained compositions [8], plane partitions diamonds [5], and k-gons partitions [7].

2. Integer coefficients

Let $A[1...k-1, 1..k-1]$ be an upper diagonal matrix of integers and let P_A be the set of sequences $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ satisfying

$$\begin{array}{rcll} \lambda_1 & = & \displaystyle\sum_{j=1}^{k-1} A[1,j]\lambda_{j+1} & \\ \lambda_i & \geq & \displaystyle\sum_{j=i}^{k-1} A[i,j]\lambda_{j+1} & \text{for } 2 \leq i \leq k-1 \\ \lambda_k & \geq & 0. & \end{array} \tag{3}$$

Define a sequence of matrices $A^{(t)}, 1 \leq t \leq k-1$, as follows: $A^{(1)} = A$ and $A^{(t)}$ is a matrix of dimension $(k-t) \times (k-t)$, whose entries are defined by the recurrence :

$$\begin{array}{rcl} A^{(t)}[1,j] & = & (A^{(t-1)}[1,1]+1)A^{(t-1)}[2,j+1] + A^{(t-1)}[1,j+1] \\ A^{(t)}[i,j] & = & A^{(t-1)}[i+1,j+1] \ \text{ if } 2 \leq i \leq k-t \ . \end{array}$$

Then each matrix $A^{(t)}$ is an upper diagonal matrix of integers and we can show the following.

Lemma 1. *If every element of P_A is a composition, then the same is true of $P_{A^{(t)}}$ for $1 \leq t \leq k-1$.*

The main result of this section is the following.

Theorem 1. *Let $P_A(k)$ be the set of sequences $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ satisfying (3). Let $P_A(n,k)$ be the set of sequences $P_A(k)$ of weight n. If every $\lambda \in P_A$ is a composition, then the weight generating function for $P_A(k)$ is*

$$\sum_{n=0}^{\infty} |P_A(n,k)| q^n = \prod_{i=1}^{k-1} \frac{1}{1-q^{b_i}}, \tag{4}$$

where $b_i = A^{(i)}[1,1] + 1 > 0$.

Sketch of proof. The mapping

$$\Theta(\lambda) = (b_1 s_1, b_2 s_2, \ldots, b_{k-1} s_{k-1})$$

where

$$s_i = \lambda_{i+1} - \sum_{j=i+1}^{k-1} A[i,j]\lambda_{j+1}$$

can be shown to be a bijection from $P_A(n,k)$ to the set of sequences $r_1, r_2, \ldots, r_{k-1}$ of weight n in which r_i is a nonnegative multiple of b_i. In the case that the b_i are distinct positive integers, this can be viewed alternatively as a bijection with partitions of n into parts in $\{b_1, \ldots, b_{k-1}\}$. □

Example 1. The partitions λ of n with at most k parts and with $\lambda_1 = \sum_{i=1}^{k-1} \lambda_{i+1}$ comprise $P_A(n,k)$ where, for $1 \le i \le j \le k-1$, $A[i,j] = 1$ if $i = 1$ or $i = j$ and $A[i,j] = 0$ otherwise. Then $b_i = 2i$ for $1 \le i \le k-1$ so by Theorem 1, the generating function is $\prod_{i=1}^{k-1}(1-q^{2i})^{-1}$.

Example 2. There is a one-to-one correspondence between sequences $\lambda_1, \ldots, \lambda_k$ of weight n satisfying

$$\begin{aligned} \sum_{j=0}^{k-1} (-1)^j \lambda_{1+j} &= 0 \\ \sum_{j=0}^{k-i} (-1)^j \lambda_{i+j} &\ge 0 \qquad \text{for } 2 \le i \le k \end{aligned}$$

and the set of compositions of n into $k-1$ even parts. The λ satisfying the constraints are compositions and they comprise the set $P_A(n,k)$ where A is the $(k-1) \times (k-1)$ upper diagonal matrix defined by $A[i,j] = (-1)^{i+j}$ which has $b_1 = b_2 = \cdots = b_{k-1} = 2$.

Corollary 1. *Under the constraints of Theorem 1, if we now allow $\lambda_1 \ge \sum_{j=1}^{k-1} A[1,j]\lambda_{j+1}$ the generating function becomes*

$$\prod_{i=0}^{k-1} \frac{1}{1-q^{b_i}}$$

where $b_0 = 1$.

Linear algebra approach. Our method of proof makes the bijection of Theorem 1 explicit. However, as suggested by one of the referees, a linear algebra proof can also be illuminating. Following the argument of the referee, we embed the constraint matrix, A, into a $k \times k$ matrix, B, where $B[i,j] = 0$ if $i \ge j$ and $B[i,j] = A[i, j-1]$ otherwise. Hence B is strictly upper diagonal and thus, nilpotent : $B^k = 0$. Considering λ as a column vector, the matrix inequality, $\lambda \ge B\lambda$, describes the system (3), with the first constraint changed to inequality. Then

$$\lambda(I-B) = [s_0, s_1, \ldots s_{k-1}]^T$$

for the s_i defined in the proof of Theorem 1. Iterating this identity yields

$$\begin{aligned}\lambda &= (1+B+B^2+\cdots+B^{k-1})[s_0,s_1,\ldots s_{k-1}]^T\\ &= (I-B)^{-1}[s_0,s_1,\ldots s_{k-1}]^T,\end{aligned}$$

since $B^k=0$. So, $I-B$ is invertible, and, in particular, it is an upper diagonal matrix of *integers* whenever B, and therefore, A, is. Hence the recurrences of $A^{(t)}$ are equivalent to taking the inverse of a matrix. Let $C=(I-B)^{-1}$. Then, furthermore, all the elements of P_A are compositions iff all the entries in C are nonnegative. Let $h=(1,1,\ldots,1)$ be the row vector of length k containing only ones. The weight of the composition λ is $|\lambda|=h\lambda=hC[s_0,s_1,\ldots s_{k-1}]^T$. Define $p=hC$. Then the weight generating function of P_A is

$$\sum_{\lambda\in P_A} q^{\lambda_1+\cdots+\lambda_k} = \sum_{s_0,\ldots,s_{k-1}\ge 0} q^{p_1s_0+\cdots+p_ks_{k-1}} = \prod_{j=1}^{k}\frac{1}{1-q^{p_j}}, \tag{5}$$

where p_j (corresponding to b_{j-1}) is $p_j=C[1,j]+C[2,j]+\cdots+C[j,j]$. □

Example 3. The weight generating function of the partitions λ with at most k parts and with $\lambda_1\ge 2\lambda_2+\sum_{i=2}^{k-1}\lambda_{i+1}$ and $\lambda_2\ge\sum_{i=2}^{k-1}\lambda_{i+1}$ is :

$$\frac{1}{(1-q)(1-q^3)}\prod_{i=1}^{k-2}\frac{1}{1-q^{5i}}.$$

This follows from Corollary 1: since λ is a partition, for $i>2$, $A[i,i]=1$ and $A[i,j]=0$ for $2<j<i$. Also, $A[1,1]=2$ and $A[i,j]=1$ for $i=1,2$ and $j\ge 2$, so $b_0=1$, $b_1=3$, and $b_i=5(i-1)$ for $2\le i\le k-1$.

Example 4. The generating function of the partitions λ with at most k parts and $\lambda_i\ge i\sum_{j=i+1}^{k}\lambda_j$, $1\le i\le k$, is : $\prod_{i=1}^{k}(1-q^{i!})^{-1}$.

We can generalize Theorem 1 to allow the constraints of the matrix A to be satisfied with equality for any specified set of λ_i. Given a set $S\subseteq\{0,1,\ldots,k-1\}$, let $P_A(k;S)$ be the set of sequences $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_k)$ satisfying, for $1\le i\le k$:

$$\lambda_i=\sum_{j=i}^{k-1}A[i,j]\lambda_{j+1}\quad \text{if } i-1\in S$$

$$\lambda_i\ge\sum_{j=i}^{k-1}A[i,j]\lambda_{j+1}\quad \text{if } i-1\notin S.$$

Let $P_A(n,k;S)$ be the set of sequences in $P_A(k;S)$ of weight n.

Corollary 2. *If all elements of P_A are compositions, the weight generating function for $P_A(k;S)$ is*

$$\sum_{n=0}^{\infty}|P_A(n,k;S)|q^n = \prod_{i=0,i\notin S}^{k-1}\frac{1}{1-q^{b_i}},$$

where $b_0=1$ and for $i\ge 1$, $b_i=A^{(i)}[1,1]+1$. □

As noted by the referee, the inverse problem can be solved completely in the domain of linear algebra if $b_1=1$. Given a sequence $(b_1,\ldots,b_k)$ of positive integers, such that $b_1=1$, you can always construct an upper triangular matrix C with ones on the diagonal and nonnegative entries such that the sum of the entries in the jth column is b_j. Then the matrix $B=1-C^{-1}$ contains in its northeast corner a $(k-1)\times(k-1)$ constraint matrix A such that P_A has generating function $\prod_{i=1}^{k}(1-q^{b_i})^{-1}$.

3. Rational Coefficients

In this section we would like to generalize our results to allow some of the elements of the constraint matrix to be rational. In particular, we will find an explicit form of the generating function for the set of integer sequences $\lambda_1, \ldots, \lambda_k$ satisfying the constraints:
(6)

$$\lambda_1 \geq c_1 \lceil \frac{n_1}{d_1}\lambda_2 \rceil + \sum_{i=2}^{k} c_i\lambda_i; \quad \lambda_2 \geq \frac{n_2}{d_2}\lambda_3; \quad \lambda_3 \geq \frac{n_3}{d_3}\lambda_4; \quad \ldots; \quad \lambda_{k-1} \geq \frac{n_{k-1}}{d_{k-1}}\lambda_k; \quad \lambda_k \geq 0,$$

where for $1 \leq i \leq k-1$, n_i and d_i are positive integers and the c_i are any integers which make the first constraint strong enough to guarantee that $\lambda_1 \geq 0$.

For $i = 1, \ldots, k$, let $a_i = \prod_{j=1}^{i-1} d_j \prod_{t=i}^{k-1} n_t$. Then $a_i/a_{i+1} = n_i/d_i$, so the system (6) above is equivalent to:

$$\lambda_1 \geq c_1\lceil \frac{a_1}{a_2}\lambda_2\rceil + \sum_{i=2}^{k} c_i\lambda_i \quad \text{and} \tag{7}$$

$$\lambda_2/a_2 \geq \lambda_3/a_3 \geq \cdots \geq \lambda_{k-1}/a_{k-1} \geq \lambda_k/a_k.$$

Theorem 2. *Given a sequence of positive integers $a_1, \ldots, a_k$ and a sequence of integers $c_1, \ldots, c_k$, consider the set of sequences $\lambda_1, \ldots, \lambda_k$ satisfying (7). As long as the c_i are integers which guarantee that $\lambda_1 \geq 0$, the weight generating function is*

$$\frac{\sum_{z_2=0}^{a_2-1}\sum_{z_3=0}^{a_3-1}\cdots\sum_{z_k=0}^{a_k-1} q^{c_1\lceil \frac{a_1 z_2}{a_2}\rceil + \sum_{i=2}^{k}(c_i+1)z_i} \prod_{i=1}^{k-2} q^{b_i\lceil \frac{z_{i+2}}{a_{i+2}} - \frac{z_{i+1}}{a_{i+1}}\rceil}}{\prod_{i=0}^{k-1}(1-q^{b_i})}, \tag{8}$$

where $b_0 = 1$, $b_1 = c_1a_1$ and $b_i = c_1a_1 + (c_2+1)a_2 + \cdots + (c_{i+1}+1)a_{i+1}$, for $2 \leq i \leq k-1$.

Sketch of proof. For $2 \leq i \leq k$, let $\lambda_i = a_ix_i + z_i$, where $x_i \geq 0$ and $0 \leq z_i < a_i$. Rearrange the sequence $\lambda_1, \lambda_2, \ldots, \lambda_k$ by decreasing part λ_i by $(a_i-1)x_i + z_i$ and increasing λ_1 by the same amount for $2 \leq i \leq k$ to get a new sequence $x_1, x_2, \ldots, x_k$, of the same weight satisfying

$$\begin{aligned} x_1 &\geq (c_1a_1 + (c_2+1)a_2 - 1)x_2 + \sum_{i=3}^{k}((c_i+1)a_i - 1)x_i + s_0 \\ x_i &\geq x_{i+1} + s_{i-1} \quad \text{for } 2 \leq i \leq k-1 \\ x_k &\geq 0 \end{aligned}$$

where

$$s_0 = c_1\lceil a_1z_2/a_2\rceil + \sum_{i=2}^{k}(c_i+1)z_i,$$

and

$$s_i = \lceil z_{i+2}/a_{i+2} - z_{i+1}/a_{i+1}\rceil$$

for $1 \leq i \leq k-2$.. By the method of Theorem 1, we can show that the generating function for fixed $s_0, s_1, \ldots s_{k-2}$ is

$$\frac{\prod_{i=0}^{k-2} q^{b_is_i}}{\prod_{i=0}^{k-1}(1-q^{b_i})},$$

where $b_0 = 1$, $b_1 = c_1a_1$, and $b_i = c_1a_1 + (c_2+1)a_2 + (c_3+1)a_3 + \cdots + (c_{i+1}+1)a_{i+1}$, for $2 \leq i \leq k-1$. Summing over all possible sequences $s_0, s_1, \ldots s_{k-2}$ as the z_i vary independently from 0 to $a_i - 1$ gives the result. $\square$

Example 5. By Theorem 2, sequences $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ satisfying $\lambda_1 \geq 2\lambda_2 - \lambda_3 + 2\lambda_4$ and $\lambda_2/3 \geq \lambda_3/2 \geq \lambda_4/1 \geq 0$ have generating function

$$\frac{1+q^3+q^6+q^9+q^{12}+q^6}{(1-q)(1-q^9)(1-q^9)(1-q^{12})} = \frac{(1+q^3+q^6)(1+q^6)}{(1-q)(1-q^9)(1-q^9)(1-q^{12})} = \frac{1}{(1-q)(1-q^3)(1-q^6)(1-q^9)}.$$

In some cases, the numerator of the generating function of Theorem 2 can be easily shown to factor, giving results such as the following.

Corollary 3. *Suppose the sequence $a_1, a_2, \ldots, a_k$ has the property that for $1 \leq i \leq k-1$ if $a_i > 1$, then $a_{i+1} = 1$. Then there is a one-to-one correspondence between the compositions $\lambda_1, \ldots, \lambda_k$ of n satisfying*

$$\lambda_1/a_1 \geq \lambda_2/a_2 \geq \lambda_3/a_3 \geq \cdots \geq \lambda_{k-1}/a_{k-1} \geq \lambda_k/a_k$$

and the partitions of n into parts in

$$\{1, b_1, b_1+1, \ldots, b_{k-1}\},$$

where $b_0 = 1$, $b_1 = a_1 + a_2$ and $b_{i+1} = b_i + a_{i+2}$ for $1 \leq i \leq k-2$, such that at most one part can appear from each of the sets

$$S_i = \{p | b_i + 1 \leq p \leq b_{i+1} - 1\}.$$

Proof. The generating function (8) becomes

$$\frac{\prod_{i=0:b_{i+1}-b_i>1}^{k-2}(1+q^{b_i+1}+q^{b_i+2}+\cdots+q^{b_{i+1}-1})}{\prod_{i=0}^{k-1}(1-q^{b_i})}.$$

□

Example 6. By Corollary 4, compositions of n satisfying

$$\lambda_1 \geq \lambda_2 \geq \lambda_3/2 \geq \lambda_4 \geq \lambda_5/2 \geq \cdots \geq \lambda_{2k} \geq \lambda_{2k+1}/2$$

are in one-to-one correspondence with the set of partitions of n into parts of size at most $3k+1$ in which parts divisible by 3 can appear at most once.

In some cases, the numerator of the generating function of Theorem 2 does factor, but it is not as easily shown. We consider some examples of this type in Section 5 on lecture hall compositions and in Section 6 on anti-lecture hall compositions.

4. Two Variable Generating Functions

In their study of lecture hall partitions, Bousquet-Mélou and Eriksson found it very useful to consider the 2-variable (odd/even weighted) generating function of the set of partitions satisfying the lecture hall constraints. We show here how our method can be adapted to get multivariable generating functions for compositions satisfying linear constraints, using the two-variable case as an example.

Let $A[1..k-1, 1..k-1]$ be an upper diagonal matrix of integers such that P_A is a set of compositions. For $\lambda \in P_A$, let $|\lambda|_o = \lambda_1 + \lambda_3 + \lambda_5 + \cdots$ and and $|\lambda|_e = \lambda_2 + \lambda_4 + \lambda_6 + \cdots$. Define two sequences of matrices $O^{(t)}$, and $E^{(t)}$, $1 \leq t \leq k-1$ so that $O^{(t)} + E^{(t)} = A^{(t)}$ as follows. $O^{(t)}$ and $E^{(t)}$ are $(k-t) \times (k-t)$ matrices satisfying

$O^{(1)}[i,j] = A^{(1)}[i,j]$ if i is odd, otherwise, $O^{(1)}[i,j] = 0$;

$E^{(1)}[i,j] = A^{(1)}[i,j]$ if i is even, otherwise, $E^{(1)}[i,j] = 0$;

For $t > 1$,

$$\begin{aligned} O^{(t)}[1,j] &= (O^{(t-1)}[1,1] + 1 - (t \bmod 2))A^{(t-1)}[2,j+1] + O^{(t-1)}[1,j+1], \\ O^{(t)}[i,j] &= O^{(t-1)}[i+1,j+1] \text{ if } i \geq 2 \ ; \\ E^{(t)}[1,j] &= (E^{(t-1)}[1,1] + (t \bmod 2))A^{(t-1)}[2,j+1] + E^{(t-1)}[1,j+1], \\ E^{(t)}[i,j] &= E^{(t-1)}[i+1,j+1] \text{ if } i \geq 2 \ ; \end{aligned}$$

We can prove the following 2-variable version of Theorem 1.

Theorem 3. *The odd/even weighted generating function for the compositions in P_A is:*

$$\sum_{\lambda \in P_A} x^{|\lambda|_o} y^{|\lambda|_e} = \prod_{i=0}^{k-1} \frac{1}{1 - x^{o_i} y^{e_i}},$$

where sequences $o_0, o_1, \ldots, o_{k-1}$ and $e_0, e_1, \ldots, e_{k-1}$ are defined by $o_0 = 1$; $e_0 = 0$ and for $t > 0$,

$$o_t = O^{(t)}[1,1] + 1 - (t \bmod 2) \qquad e_t = E^{(t)}[1,1] + (t \bmod 2).$$

Note from referee. We can refine Theorem 1 via the enumeration in (5) by associating with λ the monomial $\prod q_i^{\lambda_i}$, is possible :

$$\sum_{\lambda \in P_A} \prod_i q_i^{\lambda_i} = \prod_{j=1}^{k} \frac{1}{1 - \prod_{i \leq j} q_i^{C[i,j]}}.$$

From Theorem 3, imitating the bijective proof of Theorem 2, we can get the odd/even generating function $G_k(x,y)$ for the compositions satisfying the constraints:

$$\frac{\lambda_1}{a_1} \geq \frac{\lambda_2}{a_2} \geq \frac{\lambda_3}{a_3} \geq \cdots \geq \frac{\lambda_{k-1}}{a_{k-1}} \geq \frac{\lambda_k}{a_k} \geq 0, \tag{9}$$

Theorem 4.

$$G_k(x,y) = \frac{\sum_{z_2=0}^{a_2-1} \cdots \sum_{z_k=0}^{a_k-1} x^{\lceil \frac{a_1 z_2}{a_2} \rceil + z_3 + z_5 + \cdots} y^{z_2 + z_4 + \cdots} \prod_{i=1}^{k-2} (x^{o_i} y^{e_i})^{\lceil \frac{z_{i+2}}{a_{i+2}} - \frac{z_{i+1}}{a_{i+1}} \rceil}}{\prod_{i=0}^{k-1} (1 - x^{o_i} y^{e_i})}, \tag{10}$$

where $o_0 = 1$, $e_0 = 0$, and for $1 \leq i \leq k-1$ $o_i = a_1 + a_3 + a_5 + \cdots + a_{2\lfloor (i-1)/2 \rfloor + 1}$ and $e_i = a_2 + a_4 + a_6 + \cdots + a_{2\lceil i/2 \rceil}$.

Proof. Omitted. □

We can now add more conditions on the first part when $a_k \geq a_{k-1} \geq \cdots \geq a_1$:

Theorem 5. *If $G_k(x,y) = H_k(x,y)/(1-x)$ is the generating function given in (10), then whenever $a_1 \geq a_2 \geq \cdots \geq a_k$, for any $l \geq 1$ and $j \geq 2 - l$, $H(q^l, q^j)$ is the generating function for the partitions satisfying*

$$\begin{aligned} \lambda_1 &= l \lceil a_1 \lambda_2 / a_2 \rceil + (j-1)(\lambda_2 + \lambda_4 + \lambda_6 + \cdots) + (l-1)(\lambda_3 + \lambda_5 + \lambda_7 + \cdots) \\ \frac{\lambda_2}{a_2} &\geq \frac{\lambda_3}{a_3} \geq \cdots \geq \frac{\lambda_{k-1}}{a_{k-1}} \geq \frac{\lambda_k}{a_k} \geq 0. \end{aligned} \tag{11}$$

Proof. If $j = l = 1$, the system (11) is the same as (9) and the generating function is $H(x,y) = G(x,y)/(1-x)$. Suppose λ satisfies (9), with $|\lambda|_o = l$ and $|\lambda|_e = m$. To transform λ into a composition satisfying (11), we increase the first part by $(l-1)|\lambda|_o + (j-1)|\lambda|_e$ to get λ'. The conditions on j and l and the a_i guarantee that this increase is positive. Then $|\lambda'| = |\lambda| + (l-1)|\lambda|_o + (j-1)|\lambda|_e = |\lambda|_o + |\lambda|_e + (l-1)|\lambda|_o + (j-1)|\lambda|_e = l|\lambda|_o + j|\lambda|_e$. □

We can use Theorem 5 to generalize the Lecture Hall Partition theorem of Bousquet-Mélou and Eriksson.

Corollary 4. *For $l \geq 1$ and $j \geq 2-l$, the generating function for the sequences $\lambda_1, \ldots, \lambda_k$ satisfying $\lambda_1 \geq l\lceil k\lambda_2/(k-1)\rceil + (j-1)(\lambda_2+\lambda_4+\lambda_6+\cdots) + (l-1)(\lambda_3+\lambda_5+\lambda_7+\cdots)$ and $\frac{\lambda_2}{k-1} \geq \frac{\lambda_3}{k-2} \geq \cdots \geq \frac{\lambda_{k-1}}{2} \geq \frac{\lambda_k}{1}$ is*

$$\frac{1}{(1-q)}\prod_{i=0}^{k-1}\frac{1}{(1-q^{il+ij+l})}$$

Proof. Direct, as $G_k(x,y) = \prod_{i=0}^{k-1}\frac{1}{1-x^{i+1}y^i}$ [10]. □

Example 7. There is a one-to-one correspondence between the compositions of n satisfying $\lambda_1 \geq 3\lceil k\lambda_2/(k-1)\rceil + 2\sum_{j=1}^{k-1}(-1)^j\lambda_{1+j}$ and $\frac{\lambda_2}{k-1} \geq \frac{\lambda_3}{k-2} \geq \cdots \geq \frac{\lambda_{k-1}}{2} \geq \frac{\lambda_k}{1} \geq 0$ and the partitions of n into odd parts less than or equal to $2k+1$.

5. Generalized Lecture Hall Compositions

Given integers m, l, define the (m,l)-sequence $a^{(m,l)} = (a_1, a_2, \ldots)$ by the recurrence

$$a_{2i} = la_{2i-1} - a_{2i-2}, \quad a_{2i-1} = ma_{2i-2} - a_{2i-3}, \quad i \geq 2,$$

with the initial conditions $a_1 = 1$ and $a_2 = l$. Let $L_k^{(m,l)}$ be the set of compositions into k nonnegative parts satisfying, for $1 \leq i < k$,

$$\lambda_i \geq \frac{a_{k-i+1}}{a_{k-i}}\lambda_{i+1}.$$

Note that $a^{(2,2)} = (1,2,3,4,\ldots)$ and $L_k^{(2,2)}$ is the set of lecture hall partitions of [10]. When $m, l > 1$, then $a^{(m,l)}$ is nondecreasing and $L_k^{(m,l)}$ is a set of partitions which we call generalized lecture hall partitions. However, if $m = 1$ and $l > 3$ or $l = 1$ and $m > 3$, the sequence $a^{(m,l)}$ is an infinite, but non-monotone, sequence of positive integers and $L_k^{(m,l)}$ is a set of compositions, but not partitions.

Let $G_k^{(m,l)}(x,y)$ be the odd/even weighted generating function for $L_k^{(m,l)}$, that is

$$G_k^{(m,l)}(x,y) = \sum_{\lambda \in L_k^{(m,l)}} = x^{|\lambda_o|}y^{|\lambda_e|}.$$

Theorem 6 below was proved in [11] for $m, l > 1$. We give a short proof and extend it to generalized lecture hall compositions ($m = 1, l > 3$ or $l = 1, m > 3$). We note that it implies a straightforward bijective proof of the theorem.

Theorem 6. *For integers m, l satisfying $m, l > 1$ or $m = 1$ and $l > 3$, or $l = 1$ and $m > 3$, $G_1^{(m,l)}(x,y) = 1/(1-x)$ and*

$$G_{2k+1}^{(m,l)}(x,y) = G_{2k}^{(m,l)}(x^m y, x^{-1})/(1-x), \quad G_{2k}^{(m,l)}(x,y) = G_{2k-1}^{(m,l)}(x^l y, x^{-1})/(1-x), \quad n > 0.$$

Note that this gives for $k > 0$

$$G_{2k}^{(m,l)}(x,y) = \prod_{i=1}^{2k}\frac{1}{1-x^{a_i}y^{b_{i-1}}} \quad \text{and} \quad G_{2k-1}^{(m,l)}(x,y) = \prod_{i=1}^{2n-1}\frac{1}{1-x^{b_i}y^{a_{i-1}}}$$

where $a_0 = 0$, $(a_1, a_2, \ldots) = a^{(m,l)}$, and $(b_1, b_2, \ldots) = a^{(l,m)}$.

Proof of theorem. We follow the method of Bousquet-Mélou and Eriksson, bypassing the detour to reduced lecture hall partitions, as suggested in a note on page 10 of [11]. We fix m, l and drop the superscripts (m,l) to simplify notation. Let $L_{k+1,=}$ denote the partitions

$\lambda \in L_k$ with $\lambda_1 = \lceil a_k\lambda_2/a_{k-1}\rceil$. We do this by establishing a bijection ϕ_k between partitions $\lambda \in L_k$ and partitions $\gamma \in L_{k+1,=}$ with $|\gamma_e| = m|\lambda_e| - |\lambda_o|$ if k is even and $|\gamma_e| = l|\lambda_e| - |\lambda_o|$ if k is odd, and with $|\gamma_o| = |\lambda_e|$. The bijection and its reverse are presented below :

$\phi_k \;:\; \lambda \to \gamma$ $\gamma_e \leftarrow \lambda_o$ $\gamma_1 \leftarrow \lceil \frac{a_{k+1}}{a_k}\lambda_1\rceil$ For i from 1 to $\lfloor (k+1)/2\rfloor$ $\gamma_{2i+1} \leftarrow \lceil \frac{a_{k-2i+1}}{a_{k-2i}}\lambda_{2i+1}\rceil +$ $\lfloor \frac{a_{k+1-2i}}{a_{k+2-2i}}\lambda_{2i-1}\rfloor - \lambda_{2i}$.	$\phi_k^{(-1)} \;:\; \gamma \to \lambda$ $\lambda_o \leftarrow \gamma_e$ For i from 1 to $\lfloor k/2\rfloor$ $\lambda_{2i} \leftarrow \lceil \frac{a_{k-2i+1}}{a_{k-2i}}\gamma_{2i+2}\rceil +$ $\lceil \frac{a_{k+1-2i}}{a_{k+2-2i}}\gamma_{2i}\rceil - \gamma_{2i+1}$.

As in [11] we can show that for the (m,l) sequence a and for any $j \geq 0$

$$\lfloor \frac{a_{i-1}}{a_i} j\rfloor + \lceil \frac{a_{i+1}}{a_i} j\rceil = \begin{cases} mj & \text{if } i \text{ even} \\ lj & \text{if } i \text{ odd} \end{cases}$$

$|\gamma_o| = m|\lambda_o| - |\lambda_e|$ if k even and $|\gamma_o| = l|\lambda_o| - |\lambda_e|$ if k odd and $|\gamma_e| = |\lambda_o|$.

Also,

$$\gamma_1 = \lceil a_{k+1}\gamma_2/a_k\rceil.$$

It remains to show that for $k \geq i \geq 2$, $\gamma_i \geq \frac{a_{k+2-i}}{a_{k+1-i}}\gamma_{i-1}$. Note that consecutive parts $\lambda_{2i-1}, \lambda_{2i}, \lambda_{2i+1}$ in λ, map by ϕ_k to the consecutive parts

$$\gamma_{2i} = \lambda_{2i-1}, \qquad \gamma_{2i+1} = \lceil \frac{a_{k+1-2i}}{a_{k-2i}}\lambda_{2i+1}\rceil + \lfloor \frac{a_{k-2i+1}}{a_{k+2-2i}}\lambda_{2i-1}\rfloor - \lambda_{2i}, \qquad , \gamma_{2i+2} = \lambda_{2i+1}$$

in γ. As $\lambda_{2i} \geq \frac{a_{k+1-2i}}{a_{k-2i}}\lambda_{2i+1}$, we get that

$$\lceil a_{k+1-2i}\lambda_{2i+1}/a_{k-2i}\rceil - \lambda_{2i} \leq 0$$

and

$$\gamma_{2i+1} \leq a_{k-2i+1}\lambda_{2i-1}/a_{k+2-2i} = a_{k-2i+1}\gamma_{2i}/a_{k+2-2i}.$$

As $\lambda_{2i-1} \geq a_{k+2-2i}\lambda_{2i}/a_{k+1-2i}$, we get that

$$\lfloor a_{k-2i+1}\lambda_{2i-1}/a_{k+2-2i}\lambda_{2i-1}\rfloor - \lambda_{2i} \geq 0,$$

and

$$\gamma_{2i+1} \geq a_{k+1-2i}\lambda_{2i+1}/a_{k-2i} = a_{k+1-2i}\gamma_{2i+2}/a_{k-2i}.$$

We have our conditions. □

The bijection Now we show that this gives a straighforward bijection ψ_k between Lecture Hall partitions in L_k and partitions into parts in $c = \{c_1, c_2, \ldots, c_k\}$ with $c_i = a_i + b_{i-1}$ if k is even and $c_i = a_{i-1} + b_i$ otherwise. Let λ be a partition in L_k and μ its image by ψ_k. We denote by $\mu(i)$ the multiplicity of the part c_i in μ. The bijection and its reverse are presented below :

$\psi_k \; : \; \lambda \to \mu$
  $\mu \leftarrow$ empty partition
  For i from k downto 1 do
    $j \leftarrow \lambda_1 - \lceil a_i/a_{i-1}\lambda_2 \rceil$
    $\lambda_1 \leftarrow \lambda_1 - j$
    $\mu(k-i+1) \leftarrow j$
    If $i > 1$ then
      $\lambda \leftarrow \phi_{i-1}^{-1}(\lambda)$.

$\psi_k^{-1} \; : \; \mu \to \lambda$
  $\lambda \leftarrow$ empty partition
  For i from 1 to k do
    $\lambda_1 \leftarrow \lambda_1 + \mu(k-i+1)$
    If $i < k$ then
      $\lambda \leftarrow \phi_i(\lambda)$.

Example 8. For $m = l = 2$, $a = (1, 2, 3, 4, \ldots)$ and for $\lambda = (6, 4, 2, 1)$ we get $\mu = (5, 5, 3)$.
Example 9. For $m = 1$ and $l = 4$, $a = (1, 4, 3, 8, \ldots)$ and for $\lambda = (12, 4, 5, 1)$ we get $\mu = (5, 4, 4, 4, 4, 1)$.

6. Anti-Lecture Hall Compositions

In this section we study the sequences $(\lambda_1, \ldots, \lambda_k)$ defined by

$$\frac{\lambda_1}{1} \geq \frac{\lambda_2}{2} \geq \ldots \frac{\lambda_k}{k} \geq 0.$$

We call these sequences *anti-lecture hall* because if they represent the heights of the rows in an amphitheater (as it was done for the Lecture Hall partitions) only the students of the first row are guaranteed to see the professor! We want to show that these compositions have a surprising behavior, in particular, their weight generating function is

$$\prod_{i=1}^{k} \frac{1+q^i}{1-q^{i+1}}. \tag{12}$$

Let A_k be the set of anti-lecture hall compositions into k nonnegative parts. Given $\lambda \in A_k$, we can write λ as $((x_1, \ldots, x_k), (z_1, z_2, \ldots, z_k))$ where $\lambda_i = ix_i + z_i$ with $0 \leq z_i \leq i-1$, $1 \leq i \leq k$. Note that $\lambda \in A_k$ if and only if $x_1 \geq x_2 \geq \ldots x_k \geq 0$ and if $x_i = x_{i+1}$ then $z_i \geq z_{i+1}$. Moreover $|\lambda| = \sum_{i=1}^{k} z_i + ix_i$.

The main result of this section is the following theorem (setting $u = v = 1$ gives (12)).

Theorem 7.

$$\sum_{\lambda \in A_k} q^{|\lambda|} u^{|x|} v^{o(x)} = \prod_{i=1}^{k} \frac{1+uvq^i}{1-u^2 q^{i+1}}.$$

where $x = (x_1, \ldots, x_k)$ and $o(x)$ is the number of odd parts of the partition x.

Proof. Let D_k be the set of partitions into distinct parts less than or equal to k. Let E_k be the subset of A_k where all the x_i are even. To show the theorem we must give two bijections :

- A bijection between A_k and $D_k \times E_k$ such that if (α, β) is the image of λ then $|\alpha| + |\beta| = |\lambda|$, $l(\alpha) + |\lfloor \beta \rfloor| = |x|$, and $l(\alpha) = o(x)$, where $\lfloor \beta \rfloor = (\lfloor \beta_1/1 \rfloor, \ldots, \lfloor \beta_k/k \rfloor)$ and $l(\alpha)$ denotes the number of positive parts of α.

- A bijection between the set E_k and the set F_k of partitions into parts in the set $\{2, 3, \ldots, k+1\}$ such that if μ is the image of λ then $|\mu| = |\lambda|$ and $l(\mu) = \sum_{i=1}^{k} x_i/2$.

The first bijection will show that : $H(u, v, q) = \prod_{i=1}^{k}(1 + uvq^i)E_k(u, q)$ where $E_k(u, q) = \sum_{\lambda \in E_k} q^{|\lambda|} u^{|x|}$. The second bijection will show that $E_k(u, q) = \prod_{i=2}^{k+1}(1 - u^2 q^i)^{-1}$.

We construct the first bijection :

$A_k \to D_k \times E_k \qquad \lambda \to (\alpha, \beta)$
$\alpha \leftarrow$ empty partition
$\beta \leftarrow \lambda$
While one of the x_i is odd
$d \leftarrow \max\{i \mid x_i \text{ odd}\}$
$i \leftarrow \min\{j \geq d \mid j = k \text{ or } x_d - 1 > x_{j+1} \text{ or } z_d \geq z_{j+1}\}$
$x_d \leftarrow x_d - 1; \quad temp \leftarrow z_d,$
$z_j \leftarrow z_{j+1} - 1, \quad d \leq j \leq i - 1$
$z_i \leftarrow temp; \quad \alpha \leftarrow \alpha \cup i$

Consider the f^{th} iteration of the loop. Let d_f and i_f be the indices chosen during that iteration. A careful look at the algorithm shows that $d_f > d_{f+1}$ and that $i_f > i_{f+1}$. As $\alpha_f = i_f$ we get that α is a partition into distinct parts. Moreover it is clear that β is in E_k as each iteration decreases by 1 only the odd x_i. Finally we must note that $|\alpha| + |\beta| = |\lambda|$ and $l(\alpha) = o(x)$. The reverse bijection is easy to construct.

We now give the second bijection between the set E_k and the set F_k.

$E_k \to F_k \qquad \lambda \to \mu$
$m(k+1) = z_k + \frac{kx_k}{2}$
$m(i) = z_{i-1} - z_i + \frac{(i-1)(x_{i-1} - x_i)}{2}, \ 2 \leq i \leq k$
where $m(i)$ is the multiplicity of the part i in μ.

It is easy to see that we can reconstruct λ from μ. Note that $m(i)$ is always nonnegative as $z_{i-1} < i - 1$ and if $z_{i-1} < z_i$ then $x_{i-1} > x_i$, that is, $(x_{i-1} - x_i)(i-1)/2 \geq i - 1$. Now we must show $|\mu| = |\lambda|$ and that the number of parts of μ is equal to the sum of the x_i divided by 2.

$|\mu| = z_k(k+1) + \sum_{i=2}^{k} i(z_{i-1} - z_i) + k(k+1)x_k/2 + \sum_{i=2}^{k}(i-1)i(x_{i-1} - x_i)/2 = \sum_{i=1}^{k}(z_i + ix_i)$

$l(\mu) = \sum_{i=2}^{k+1} m(i) = z_k + \sum_{i=2}^{k}(z_{i-1} - z_i) + kx_k/2 + \sum_{i=2}^{k} = (i-1)(x_{i-1} - x_i)/2 = \sum_{i=1}^{k} x_i/2.$ □

Example 10. Starting with $\lambda = ((7, 6, 4, 3, 3, 2), (0, 1, 1, 2, 2, 3))$ we apply the first bijection and get $\beta = ((6, 6, 4, 2, 2, 2), (0, 0, 1, 2, 2, 2))$ and $\alpha = (6, 4, 2)$. Then we apply the second bijection and get $\mu = (7, 7, 7, 7, 7, 7, 7, 7, 4, 4, 3))$. We can check that $|\lambda| = |\mu| + |\alpha| = 79$, $|x| = 2l(\mu) + l(\alpha) = 25$ and $o(x) = l(\alpha) = 3$.

Acknowledgments. Some of this work was done while the authors were visiting UQAM and we are grateful to LaCIM-UQAM for support. Thanks also to the referees for all of their constructive comments and especially to the referee who supplied the linear algebra approach in Section 2.

References

[1] George E. Andrews. MacMahon's partition analysis. I. The lecture hall partition theorem. In *Mathematical essays in honor of Gian-Carlo Rota (Cambridge, MA, 1996)*, pages 1–22. Birkhäuser Boston, Boston, MA, 1998.

[2] George E. Andrews. MacMahon's Partition Analysis II: Fundamental theorems. *Annals of Combinatorics*, 4(3-4):327–338, 2000.

[3] George E. Andrews and Peter Paule. MacMahon's partition analysis. IV. Hypergeometric multisums. *Sém. Lothar. Combin.*, 42, 1999. The Andrews Festschrift (Maratea, 1998).

[4] George E. Andrews, Peter Paule, and Axel Riese. MacMahon's Partition Analysis VI: A new reduction theorem. *Submitted*, 2000.

[5] George E. Andrews, Peter Paule, and Axel Riese. MacMahon's Partition Analysis VIII: Plane partitions diamonds. *Submitted*, 2000.

[6] George E. Andrews, Peter Paule, and Axel Riese. MacMahon's Partition Analysis III: The Omega package. *European J, Comb.*, 22(7):887–904, 2001.
[7] George E. Andrews, Peter Paule, and Axel Riese. MacMahon's Partition Analysis IX: k-gon partitions. *Bull. Austral. Math. Soc.*, 64(2):321–329, 2001.
[8] George E. Andrews, Peter Paule, and Axel Riese. MacMahon's Partition Analysis VII: Constrained compositions. *Submitted*, 2001.
[9] George E. Andrews, Peter Paule, Axel Riese, and Volker Strehl. MacMahon's partition analysis V: Bijections, recursions, and magic squares. In *Algebraic combinatorics and applications (Gößweinstein, 1999)*, pages 1–39. Springer, Berlin, 2001.
[10] Mireille Bousquet-Mélou and Kimmo Eriksson. Lecture hall partitions. *Ramanujan J.*, 1(1):101–111, 1997.
[11] Mireille Bousquet-Mélou and Kimmo Eriksson. Lecture hall partitions. II. *Ramanujan J.*, 1(2):165–185, 1997.
[12] Mireille Bousquet-Mélou and Kimmo Eriksson. A refinement of the lecture hall theorem. *J. Combin. Theory Ser. A*, 86(1):63–84, 1999.
[13] Rod Canfield, Sylvie Corteel, and Pawel Hitczenko. Random partitions with non negative r^{th} differences. *Adv. Applied Maths*, 27:298–317, 2001.
[14] D. R. Hickerson. A partition identity of the Euler type. *Amer. Math. Monthly*, 81:627–629, 1974.
[15] Percy A. MacMahon. *Combinatory analysis*. Chelsea Publishing Co., New York, 1960. Two volumes (bound as one).
[16] Richard P. Stanley. Linear homogeneous Diophantine equations and magic labelings of graphs. *Duke Math. J.*, 40:607–632, 1973.
[17] Ae Ja Yee. On the refined lecture hall theorem. *Discrete Math.*, 2001. To appear.
[18] Doron Zeilberger. Sylvie Corteel's one line proof of a partition theorem generated by Andrews-Paule-Riese's computer. *Shalosh B. Ekhad's and Doron Zeilberger's Very Own Journal*, 1998.

CNRS PRiSM, UVSQ, France
E-mail address: syl@prism.uvsq.fr

Department of Computer Science, North Carolina State University, Raleigh, NC 27695-8206, USA
E-mail address: savage@csc.ncsu.edu

THE ART OF NUMBER GUESSING: WHERE COMBINATORICS MEETS PHYSICS

JAN DE GIER

In my talk I will discuss some surprising and interesting applications of enumeration problems to physics. With the help of known results for enumerations, exact formulae for certain physically interesting quantities can be guessed from small sets of data. An example is given by the following.

Consider the algebra with generators $\{e_i\}_{i=1}^{2n-1}$ subject to the following relations

$$e_i^2 = e_i \quad e_i e_{i\pm 1} e_i = e_i \quad [e_i, e_j] = 0 \text{ for } |i-j| > 1. \tag{1}$$

This algebra is called the Temperley-Lieb (TL) algebra. I will only be concerned with the left ideal generated by the action of TL on $I_0 = \prod_{i=1}^{n} e_{2i-1}$.

From a physical point of view, an important object is formed by the following operator which is called the Hamiltonian,

$$H^{(n)} = \sum_{j=1}^{2n-1} (1 - e_j). \tag{2}$$

In a particular representation of the TL algebra, the Hamiltonian H models the physics of a chain of interacting atoms with a two-valued internal degree of freedom, otherwise known as the XXZ spin chain. In another representation, it is closely related to percolation and also to a one-dimensional stochastic system. The Hamiltonian describes the equations of motion for a physical system. It is important to consider its spectrum and in particular its lowest eigenvalue $\lambda^{(n)}$ and its corresponding eigenstate $\psi^{(n)}$. For the Hamiltonian (2), $\lambda^{(n)} = 0$ and $\psi^{(n)}$ therefore obeys

$$H^{(n)} \psi^{(n)} = 0. \tag{3}$$

Let us consider two examples. For $n = 2$ the left ideal contains only two words, I_0 and $e_2 I_0$. The action of H_2 on these two words is easily calculated and one finds that on the basis of these two words, $\psi^{(2)} = (2, 1)$. For $n = 3$ there are five words in the left ideal: I_0, $e_2 I_0$, $e_4 I_0$, $e_2 e_4 I_0$ and $e_3 e_2 e_4 I_0$. The eigenvector with eigenvalue zero is given by $\psi^{(3)} = (11, 5, 5, 4, 1)$. One can go on like this and, normalising the smallest element to 1, find that the largest element $\psi_{\max}^{(n)} = 2, 11, 170 = 2 \cdot 5 \cdot 17, 7429 = 17 \cdot 19 \cdot 23, \ldots$ and that the sum of its elements $Z^{(n)}$ follows the sequence $3, 26 = 2 \cdot 13, 646 = 2 \cdot 17 \cdot 19, 45885 = 3 \cdot 5 \cdot 7 \cdot 19 \cdot 23, \ldots$. These calculations have been done up to $n = 9$ and it is found that all elements of the sequences factorize into small primes, and moreover that they are related to the well known enumeration problem of alternating-sign matrices. It is conjectured that the general formulae for $\psi_{\max}^{(n)}$ and $Z^{(n)}$ are given by

$$\psi_{\max}^{(n)} = \prod_{j=1}^{n-1} (3j+1) \frac{(2j)!(6j)!}{(4j)!(4j+1)!} \qquad Z^{(n)} = \prod_{j=0}^{n-1} (3j+2) \frac{(2j+1)!(6j+3)!}{(4j+2)!(4j+3)!} \tag{4}$$

Department of Mathematics and Statistics, The University of Melbourne, Parkville VIC 3010, Australia

VALIDATION TEMPORELLE D'APPLICATIONS TEMPS-RÉEL STRICTES AU MOYEN D'AUTOMATES FINIS ET DE SÉRIES GÉNÉRATRICES

JEAN-PHILIPPE DUBERNARD AND DOMINIQUE GENIET

Résumé. Nous avons mis en place une technique de validation temporelle d'application temps-réel qui s'exprime en termes de langages rationnels. Nous enrichissons ici cette technique en associant aux langages rationnels des fonctions génératrices, pour donner au modèle une expressivité suffisante pour valider temporellement des systèmes temps réel généraux, c'est à dire comportant des composantes périodiques et de type *alarme*, et destinés à fonctionner sur des cibles multi-processeurs. La contribution principale de cet article est de montrer que l'on peut valider un système temps-réel avec des tâches périodiques et apériodiques à partir de sa seule composante périodique.

Abstract. We set a technique, based on regular languages, to validate hard real-time software. Here, we upgrade this technique by associating the regular languages with generating functions, to give the model enough expressivity to validate general hard real-time systems, i.e. systems composed of both periodic and sporadic (or *alarm*) tasks, and designed for multi-processor target architectures. The main contribution of this article is to show that we can validate a real-time system, which contains periodic and aperiodic tasks, only from its periodic component.

1. Introduction

Un système informatique est temps-réel lorsque, devant assurer le pilotage d'un procédé physique, il se doit de fonctionner à une vitesse compatible avec l'évolution du procédé dans son environnement physique.

1.1. Les systèmes temps-réel.

Par essence même, un tel système est donc d'une part **réactif** (car devant prendre en compte, à la volée, les informations entrantes qui lui indiquent l'évolution du procédé dans son environnement physique), et d'autre part **concurrent** (car l'ensemble des opérations relatives au pilotage –observations de l'environnement et actes de conduite– doivent être réalisées simultanément). Il est donc naturellement composé d'un ensemble de **tâches** élémentaires, chacune d'elles implémentant une réaction (ou une partie de réaction) que le système devra fournir à une sollicitation extérieure (qui s'exprime sous la forme d'un **événement**).

Certaines informations entrant dans le système suivent des flots réguliers : ce sont, généralement, des données transmises par des capteurs périodiques. Les tâches de lecture et de traitement de ces informations doivent donc être synchronisées avec les capteurs en question. Ce sont des tâches **périodiques**. Si le système comporte n tâches périodiques, nous les appelons ici $(\tau_i)_{i\in[1,n]}$. Dans le modèle classique de tâches temps-réel [1], la spécification temporelle de τ_i est constituée par la donnée de quatre caractéristiques :

- La **date de première activation** (notée $\mathbf{r_i}$) est l'instant de création de τ_i (l'origine des temps est $\min\limits_{i\in[1,n]}(r_i)$)
- La **période** (notée $\mathbf{T_i}$) est le délai séparant deux activations successives de τ_i

– Le **délai critique** (noté $\mathbf{D_i}$) est le délai séparant la date d'activation d'une occurrence de τ_i de la date à laquelle l'exécution de cette occurrence doit être terminée (cette date est nommée **échéance** de l'occurrence de τ_i)
– La **durée d'exécution** (notée $\mathbf{C_i}$) est la durée pendant laquelle chaque occurrence de τ_i devra disposer d'un processeur pour pouvoir terminer son exécution

Notons que *date de première activation*, *période* et *délai critique* sont imposés par le procédé à piloter. À ce titre, ce sont des paramètres hérités. À l'opposé, la *durée d'exécution* est un paramètre synthétisé à partir des caractéristiques du processeur cible et du corps de τ_i.

Les signaux d'alarmes, ou encore les interventions d'un éventuel superviseur, constituent des flots d'informations entrantes non-périodiques. Les tâches de prise en compte de ces informations, activées uniquement lorsque c'est nécessaire, sont appelées tâches **sporadiques** ou **apériodiques**. Une tâche sporadique est caractérisée par la connaissance d'un délai minimum séparant deux de ses occurrences successives. Dans le cas où un tel délai ne peut être garanti, la tâche est dite apériodique. Si le système comporte p tâches sporadiques [1], nous les appelons ici $(\alpha_i)_{i\in[1,p]}$.

1.2. **Techniques de validation.** Avant sa mise en exploitation, tout système doit être validé, de façon à garantir que son comportement correspond aux besoins affichés. Ceci est particulièrement vrai pour les applications de contrôle-commande. Pour les applications temps-réel, la validation comporte deux aspects. D'une part la classique validation fonctionnelle, mais également la validation temporelle [2] : il s'agit de garantir que, quelle que soit l'évolution du procédé, le système temps-réel est apte à traiter dans les temps (c'est à dire *en respectant ses contraintes temporelles*) l'ensemble des signaux susceptibles de se présenter. Le pilotage d'un système temps-réel est classiquement abordé suivant deux approches différentes (voir une synthèse fournie dans [2]) : l'ordonnancement **en ligne** et l'ordonnancement **hors-ligne**.

1.2.1. *Politiques en ligne.* Le pilotage en ligne consiste à implanter dans la machine cible un ordonnanceur qui s'appuie sur une politique spécifique (basée sur des priorités, où sur les paramètres temporels des tâches, par exemple) : chaque événement (signal ou commutation) voit l'ordonnanceur élire **à la volée** la prochaine tâche en fonction de sa politique. Cette approche a motivé un grand nombre de travaux ([3], par exemple, pour les systèmes périodiques, [4] propose des politiques adaptées à la gestion des ressources critiques, on trouvera dans [5] une synthèse sur l'intégration des sporadiques dans cette approche). Si l'approche en ligne est souple, elle est toutefois limitée par sa non-optimalité [3] dès que les tâches sont interdépendantes (c'est à dire à peu près tout le temps, dans les cas réels !).

1.2.2. *Politiques hors-ligne.* En toute généralité, le problème de décision de l'ordonnançabilité d'un système temps-réel est NP-complet [6]. Il est donc illusoire de rechercher une politique en ligne optimale dans le cas général. Cet état de fait a motivé la mise en place de stratégies basées sur l'énumération exhaustive des ordonnancements possibles d'une configuration donnée : les stratégies hors ligne. Ici, l'énumération permet, dès lors que la configuration est ordonnançable, d'exhiber une séquence valide qui, implantée dans un séquenceur, pilotera l'application.

[1]Nous nous focalisons ici sur les systèmes constitués de tâches périodiques et sporadiques.

[2]Un *bon* résultat rendu hors délai est faux.

[3]Un algorithme d'ordonnancement est **optimal** si il n'échoue que sur les configurations de tâches qui ne peuvent être ordonnancées (voir [2] pour une définition formelle de l'optimalité).

Pour les systèmes périodiques [4], la production d'une telle séquence suppose la connaissance de la durée minimale de simulation. [7] en fournit une borne, et [2] généralise ce résultat.

Les méthodes hors-ligne s'appuient sur des modèles de tâches à base de systèmes à états (automates et réseaux de Petri). Certains sont à temps explicite [8] [9] [10] (on parle de modèles **temporisés**), d'autres à temps implicite [11]. Les modèles à temps explicite sont exploités via des techniques de simulation, ceux à temps implicite via des techniques d'analyse. Dans les deux cas, un premier axe d'étude consiste à produire des techniques de modélisation exhaustives (i.e. qui prennent en compte l'ensemble des configurations réelles possibles). Par ailleurs, les complexités de décision étant exponentielles, un axe majeur des travaux est la mise en œuvre de techniques d'amélioration des temps de calcul (essentiellement par la détection précoce de branches correspondant à des cas d'échec). Les systèmes périodiques interdépendants pour lesquels les tâches ont une durée d'exécution fixée sont généralement pris en compte. Ce cas de figure est tout à fait irréaliste, puisque dès que le corps d'exécution d'une tâche intègre un test du type *If... Then... Else...*, on sort du modèle. Dans [12], nous montrons que notre approche permet d'étendre l'analyse aux systèmes de tâches comportant des instructions de choix.

1.2.3. *Validation.* Valider une application temps-réel (i.e. un système de tâches) consiste à prouver que, quelle que soit l'évolution du procédé qu'elle pilote, elle ne se trouvera jamais en situation de violation de ses contraintes temporelles. Dans le cas où l'objectif est un pilotage en ligne, la validation s'obtient à l'aide de techniques de simulation. Dans le cas hors-ligne, on utilise des techniques de model-checking pour produire une séquence d'ordonnancement en accord avec les contraintes spécifiées.

1.3. **Cadre du travail.** Nous avons introduit dans [13] une technique de modélisation de tâches temps-réel à base d'automates finis, et nous avons montré qu'elle prend en compte les configurations de tâches périodiques interdépendantes à durée d'exécution fixe. Nous étendons la validité de la méthode au cas des tâches à durées variables dans [12], et une première approche pour les sporadiques est proposée dans [14].

Le temps y étant implicite, notre modèle présente, par rapport aux autres approches orientées modèles (automates temporisés [15], réseaux de Petri [2]) un certain nombre d'avantages. D'une part, il permet de disposer de processus de décision basés sur les techniques d'analyse des automates finis, dont la puissance, l'efficacité et la modularité sont établies depuis longtemps. Par exemple, le résultat central de [2], qui établit la cyclicité des ordonnancements en environnement mono processeurs de systèmes de tâches interdépendants, se retrouve, dans notre approche, comme un corollaire de la rationnalité de l'ensemble des comportements valides d'une application temps-réel. La rationnalité de cet ensemble persistant en multi-processeur, nous pouvons affirmer que la cyclicité des ordonnancements persiste en multi-processeurs. D'autre part, l'aspect comportemental de notre approche permet l'expression de propriétés assez fines, plus difficilement exprimables par d'autres approches, soit parce qu'elles nécessitent des manipulations équationnelles élaborées (approches *temporisées*), soit parce qu'elles imposent des définitions structurelles assez lourdes (approches *réseaux de Petri*, par exemple).

Ici, après une synthèse de la modélisation et une réflexion sur les méthodes de calcul les plus adaptées, nous finalisons l'enrichissement de notre modèle proposé dans [14], dans l'optique de valider, à partir de la seule composante périodique de notre modèle, des systèmes constitués de tâches périodiques et sporadiques. Cet enrichissement s'appuie sur des séries

[4] i.e. uniquement constitués de tâches périodiques.

formelles non commutatives, auxquelles nous associons une sémantique temporelle. Nous établissons ensuite la modularité de ce modèle étendu.

Ce modèle possède une expressivité suffisante pour l'étude de systèmes temps-réel quelconques **et** réels. Par rapport aux modèles temporisés [10][15], il présente l'intérêt d'impliciter le temps, et donc d'utiliser les techniques de model checking les plus performantes qui soient : celles du modèle des automtes finis déterministes.

1.4. **Plan de l'article.** En section 2, nous présentons le modèle temporel utilisé pour la validation de systèmes de tâches. En section 3, nous montrons comment prendre en compte les phénomènes d'interdépendance, et la synchronisation. En section 4, nous montrons comment, à l'aide de séries commutatives et non-commutatives, nous validons des systèmes constitués de tâches périodiques et sporadiques à partir de la seule modélisation de la composante périodique du système.

2. Modélisation de tâches à contraintes strictes

Nous nous intéressons ici à la validation temporelle d'applications. Notre modèle doit donc être à même de répondre à des questions centrées sur l'aspect opérationnel (respect des échéances, donc), et non sur l'aspect fonctionnel (ce que fait le programme).

2.1. **Tâches périodiques.** Dans un premier temps, nous considérons les tâches à durée d'exécution fixes. Intéressons nous ici à la tâche τ_i [5], dont les paramètres temporels sont r_i, C_i, D_i et T_i. À partir de sa date d'activation (un entier de $r_i + T_i\mathbb{N}$), chaque occurrence de τ_i doit impérativement se voir allouer un processeur pendant exactement C_i unités de temps sur sa période. En représentant par le symbole a_i l'attribution d'un processeur à τ_i pendant une unité de temps, et par le symbole $\bullet$ la suspension de τ_i pendant une unité de temps, tout mot [6] de $a_i^{C_i} ⧢ \bullet^{D_i-C_i}$ correspond, sur tout intervalle temporel du type $[r_i + k.T_i, r_i + k.T_i + D_i[$, à une configuration d'allocation d'un processeur à τ_i compatible avec ses contraintes temporelles. Cet ensemble est rationnel. Prendre en compte le fait que, sur l'intervalle $[r_i + k.T_i + D_i, r_i + (k+1).T_i[$, τ_i est toujours inactive, revient à suffixer tout mot de $a_i^{C_i} ⧢ \bullet^{D_i-C_i}$ par $\bullet^{T_i-D_i}$. La concaténation de ces deux langages est le langage $a_i^{C_i} ⧢ \bullet^{D_i-C_i} \bullet^{T_i-D_i}$, également rationnel. Sur le plan opérationnel, τ_i est la suite $(\tau_{ij})_{j\in\mathbb{N}}$ de ses occurrences. Donc, allouer globalement à τ_i un processeur de manière compatible avec ses contraintes temporelles revient à allouer à chacune des occurrences τ_{ij} un processeur de manière compatible avec ses contraintes temporelles. Chaque mot ω de $a_i^{C_i} ⧢ \bullet^{D_i-C_i} \bullet^{T_i-D_i}$ est de longueur T_i : il décrit une configuration d'allocation **valide** [7] du processeur à une occurrence de τ_i. Donc, si $(\omega_j)_{j\in[1,n]} \in \left(\left(a_i^{C_i} ⧢ \bullet^{D_i-C_i}\right) \bullet^{T_i-D_i}\right)^n$, le mot $\omega_1.\omega_1 \ldots \omega_n$ modélise une configuration d'allocation du processeur valide pour n occurrences de τ_i. D'une façon générale, tout mot ω de $\left(\left(a_i^{C_i} ⧢ \bullet^{D_i-C_i}\right) \bullet^{T_i-D_i}\right)^*$ modélise une configuration d'allocation de processeur valide pour τ_i sur n'importe quel intervalle temporel du type $[r_i + k.T_i, r_i + k.T_i + |\omega|[$, et donc, notamment, sur l'intervalle $[r_i, r_i + |\omega|[$. Étant donné que τ_i est inactive sur l'intervalle $[0, r_i[$, tout mot $\bullet^{r_i}\omega$, où $\omega \in \left(\left(a_i^{C_i} ⧢ \bullet^{D_i-C_i}\right) \bullet^{T_i-D_i}\right)^*$,

[5] Nous ne considérons que le cas des tâches non réentrantes, c'est à dire des tâches ne pouvant avoir deux occurences fonctionnant simultanément.

[6] L'opération de **mélange** de mots (notée ⧢) se définit de la façon suivante

$$\begin{cases} \forall a \in \Sigma, a ⧢ \epsilon = \epsilon ⧢ a = \{a\} \\ \forall (a, b, \omega_1, \omega_2) \in \Sigma^2 \times (\Sigma^*)^2, a.\omega_1 ⧢ b.\omega_2 = a.(\omega_1 ⧢ b.\omega_2) \cup b.(a.\omega_1 ⧢ \omega_2) \end{cases}$$

[7] c'est à dire compatible avec les contraintes temporelles de la tâche.

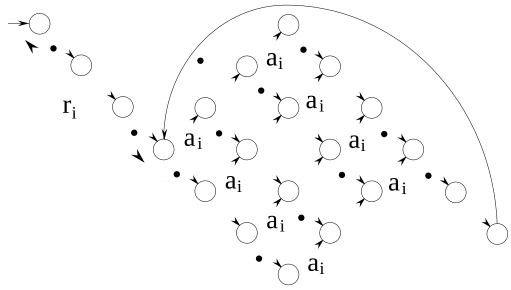


FIG. 1. Comportements de la tâche

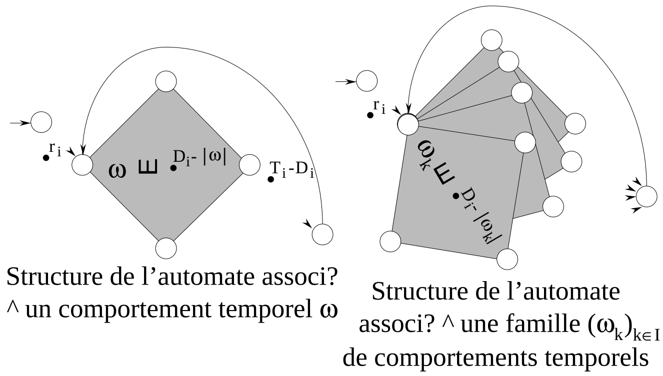


FIG. 2. Allure de l'automate associé â une tâche à durée variable

modélise une configuration d'allocation de processeur valide pour τ_i sur l'intervalle $[0, r_i + |\omega|[$. Le mot ω pouvant être de longueur arbitrairement grande (le langage est une étoile de langage rationnel), il modélise les configurations d'allocation valides pour n'importe quelle durée de vie de la tâche. Le langage $\bullet^{r_i}\left(\left(a_i^{C_i}\text{Ш}\bullet^{D_i-C_i}\right)\bullet^{T_i-D_i}\right)^*$ est rationnel, et reconnu par l'automate présenté en Figure 1. Le problème de l'ordonnancement temps-réel consiste, à un instant t donné de la vie de la tâche, à pouvoir prédire ses possibilités d'évolution dans un environnement donné. Bien entendu, le passé de la tâche est connu. Pour ce qui nous intéresse ici, ce passé est simplement l'historique des allocations CPU de τ_i un mot ω de $\{\bullet, a_i\}^*$. Par construction, ω est de la forme $\omega_1.\mu$, où $\omega_1 \in \bullet^{r_i}\left(\left(a_i^{C_i}\text{Ш}\bullet^{D_i-C_i}\right)\bullet^{T_i-D_i}\right)^*$ et

$$\exists \nu \in \{\bullet, a_i\}^* \text{ tel que } \mu\nu \in \left(a_i^{C_i}\text{Ш}\bullet^{D_i-C_i}\right)\bullet^{T_i-D_i}$$

Donc, ω est un préfixe d'un mot de $\bullet^{r_i}\left(\left(a_i^{C_i}\text{Ш}\bullet^{D_i-C_i}\right)\bullet^{T_i-D_i}\right)^*$. Par ailleurs, par construction également, quel que soit l'instant f appartenant au futur]t,$+\infty$[, il existe un mot ω_2 dans $\left(\left(a_i^{C_i}\text{Ш}\bullet^{D_i-C_i}\right)\bullet^{T_i-D_i}\right)^*$, et tel que $\omega\mu\nu\omega_2$ soit de longueur supérieure à f. À tout instant t, le passé ω de τ_i est donc un mot du centre [8] de $\bullet^{r_i}\left(\left(a_i^{C_i}\text{Ш}\bullet^{D_i-C_i}\right)\bullet^{T_i-D_i}\right)^*$. Réciproquement, par définition, tout mot du centre est le passé d'une configuration valide d'allocation de processeur à la tâche.

Considérons maitenant le cas où la tâche τ_i comporte des instructions de choix. Soient $(\omega_k)_{k\in I}$ les mots correspondant à des comportements possibles de τ_i (au sens de la trace du programme associé). Nous notons L_{C_i} l'ensemble de ces mots. Cet ensemble est fini, car la

[8] Le centre de L est l'ensemble des préfixes de L indéfiniment prolongeables dans L.

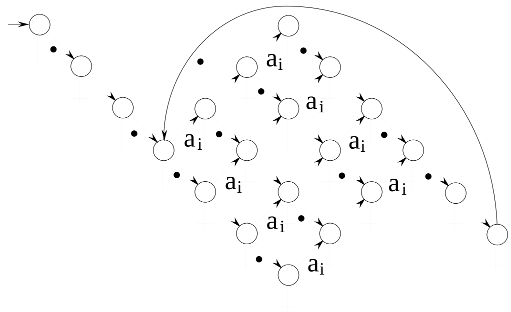


FIG. 3. Comportements temporels valides de la tâche

durée d'exécution de τ_i est majorée par C_i. I est donc fini également. Les comportements des différentes occurrences de τ_i correspondent toutefois à des ω_k non nécessairement égaux. Un comportement temporel valide de τ_i est donc, globalement, élement de $Centre\left(\bullet^{r_i}\left(\bigcup_{k\in I}\left(\omega_k ш \bullet^{D_i-|\omega_k|}\right)\bullet^{T_i-D_i}\right)^*\right)$. I étant un ensemble fini, ce langage est rationnel : il est accepté par l'automate dont l'allure est présenté en Figure 2.Droite.

Définition 1. *On appelle* **comportement temporel valide** *de la tâche τ_i tout mot ω_k appartenant au langage $Centre\left(\bullet^{r_i}\left(\bigcup_{k\in I}\left(\omega_k ш \bullet^{D_i-|\omega_k|}\right)\bullet^{T_i-D_i}\right)^*\right)$.*

Un comportement temporel valide de τ_i modélise donc une configuration d'allocation de processeur à τ_i pour laquelle on est en mesure de garantir que τ_i dispose d'au moins une possibilité de continuer son exécution en respectant ses contraintes temporelles. Dans la suite, nous notons $L(\tau_i)$ ou, plus succinctement, L_i, l'ensemble des comportements temporels valides de τ_i. Ce langage est rationnel, et reconnu par l'automate présenté en Figure 3.

Remarque

Dans [12], nous montrons que ce raisonnement reste valide pour les systèmes constitués de tâches périodiques et sporadiques.

3. Modélisation de systèmes de tâches

Une application temps-réel est constituée d'un ensemble de tâches, indépendantes ou non. Nous nous intéressons ici aux applications temps-réel à contraintes strictes, c'est à dire uniquement constituées de tâches soumises à des contraintes temporelles strictes. Une telle application est donc définie par la donnée de deux familles de tâches $(\tau_i)_{i\in[1,n]}$ et $(\alpha_i)_{i\in[1,p]}$: les α_i sont des sporadiques, et les τ_i des périodiques.

3.1. Fonctionnement simultané de l'ensemble des tâches.

Un système temps-réel étant réactif, l'ensemble des tâches qui le composent fonctionnent simultanément. Pour représenter cette simultanéité, nous utilisons les **produits de Hadamard** de langages : si $(\Sigma_i)_{i\in[1,n]}$ sont des alphabets, le produit de Hadamard est le morphisme de concaténation

$$(\Sigma_i)_{i\in[1,n]} \longrightarrow \prod_{i=1}^{n}\Sigma_i : (a_i)_{i\in[1,n]} \longmapsto \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

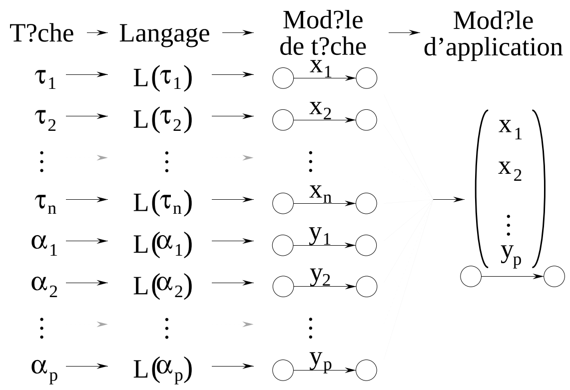


FIG. 4. Des produits de Hadamard pour modéliser les systèmes de tâches

La sémantique associée au produit de Hadamard d'une famille de lettres (chaque lettre représentant l'état d'activité d'une tâche) est la simultanéité [9]. Dans la suite, nous notons $\overset{i=n}{\underset{i=1}{\Omega}} L_i$ le langage produit de Hadamard des langages $(L_i)_{i\in[1,n]}$. Notons que le produit de Hadamard d'une famille de centres de langages rationnels reste un centre de langages rationnels [17]. La notion de comportement temporel valide, définie plus haut pour les tâches, s'étend naturellement aux systèmes de tâches.

Définition 2. *On appelle comportement temporel valide d'une application temps-réel à contraintes strictes constituée de deux familles de tâches $(\tau_i)_{i\in[1,n]}$ et $(\alpha_i)_{i\in[1,p]}$ indépendantes tout mot de* $\left(\overset{i=n}{\underset{i=1}{\Omega}} L(\tau_i)\right) \Omega \left(\overset{i=p}{\underset{i=1}{\Omega}} L(\alpha_i)\right)$

Dans la suite de ce travail, nous notons $L((\tau_i)_{i\in[1,n]}, (\alpha_i)_{i\in[1,p]})$ cet ensemble.

3.2. **Interdépendance des tâches.** Lorsque les tâches communiquent ou partagent des ressources critiques, l'ensemble des comportements temporels valides de l'application est un sous-ensemble de $L\left((\tau_i)_{i\in[1,n]}, (\alpha_i)_{i\in[1,p]}\right)$: certaines périodes d'activité de certaines tâches se trouvent retardées, du fait de l'exclusion mutuelle sur l'accès aux ressources, ou du fait de l'attente de messages. Certains comportements temporels valides dans le cas de l'indépendance des tâches ne sont plus valides.

Pour collecter l'ensemble des comportements valides, nous utilisons la technique proposée par Arnold et Nivat [18]. L'idée consiste à

(1) Définir une tâche virtuelle ρ_R pour chaque élément critique R (communication par message ou par rendez-vous, partage de ressource), dont l'objet est de tracer les états de l'élément en question (trace des états de la ressource, par exemple)

(2) Calculer le produit de Hadamard $H = L\left((\tau_i)_{i\in[1,n]}, (\alpha_i)_{i\in[1,p]}, (\rho_R)_{\text{R élément critique}}\right)$

(3) Construire l'ensemble S des vecteurs de lettres correspondant aux configurations instantannées valides (une seule attribution d'une ressource critique donnée à un instant donné, par exemple) : ce sont les seuls vecteurs dont l'apparition dans tout mot de H est autorisée

(4) Calculer $H \cap S^*$, qui collecte l'ensemble des comportements du système de tâches compatibles avec les contraintes de synchronisation

[9]Dans le cadre de la modélisation de systèmes concurrents, cette approche a été introduite par [16].

3.3. **Application au calcul d'ordonnançabilité.** L'intersection de langages n'est pas interne dans la classe des centres de langages rationnels. Dans notre étude, cette propriété traduit le fait que le partage de ressources (par exemple) entre n tâches soumises à des contraintes strictes peut amener certaines d'entre elles en situation de faute temporelle. Ce résultat est bien connu, et a motivé dans le passé un grand nombre de travaux [4]. Notre but est ici de décider de l'ordonnançabilité du système de tâches dans ce contexte d'interdépendance. Nous avons montré dans [13][19][14] que cette approche permet une résolution extrêmement efficace de ce problème.

La stratégie de décision repose sur des propriétés caractéristiques du langage associé au système de tâches. Dans le cas où les tâches sont à durées fixes [13], l'ordonnançabilité est acquise dès lors que le langage est non vide. Dans le cas où certaines (ou toutes) tâches sont à durées d'exécution variables [12], la décision d'ordonnançabilité est une caractéristique du langage :

Théorème 3. [12] *Un ensemble de tâches $(\tau_i)_{i\in[1,n]}$ est globalement ordonnançable si et seulement si* [10]

$$\forall\, (x_i)_{i\in[1,n]} \in \prod_{i=1}^{n} (X(\tau_i)), \exists \omega \in Centre\left(L\left((\tau_i)_{i\in[1,n]}\right)\right) \text{ tel que}$$

$$\forall i \in [1,n], \pi_{\neg\bullet}(\pi_i(\omega)) = x_i$$

où $(X(\tau_i))$ désigne l'alphabet asocié à la tâche τ_i.

4. Utilisation de séries génératrices pour améliorer le calcul de l'ordonnançabilité d'un système intégrant des sporadiques

Cette méthode de calcul d'ordonnançabilité est utilisable pour des systèmes quelconques de tâches : présence simultanée de périodiques et de sporadiques, interdépendance de tâches, ordonnancements toujours préemptif, toujours non-préemptif, ou bien globalement préemptif avec des sections localement non-préemptives (sections critiques d'accès aux ressources critiques, par exemple), etc.

Toutefois, l'analyse d'études de cas réels fait apparaître des configurations qui, bien que prises en compte par notre méthodologie, méritent une étude spécifique. C'est, par exemple, le cas des systèmes de tâches dans lesquels les sporadiques sont indépendantes des autres tâches [11]. Modulo un enrichissement du modèle, nous montrons ci-dessous que, dans ce cas, il est possible de décider de l'ordonnançabilité du système complet à partir de la seule connaissance de $Centre\left(L\left((\tau_i)_{i\in[1,n]}\right)\right)$.

Cet enrichissement est basé sur l'utilisation de certaines fonctions génératrices afin de prédire s'il est possible de traiter l'activation d'une tâche sporadique à un instant donné. Dans un premier temps, nous décrivons comment calculer ces fonctions génératrices. Puis, nous mettons en évidence une de leur propriété qui va nous permettre d'étendre la modularité du modèle originel à celui enrichi par les séries génératrices.

[10] Pour tout $i \in [1,n]$, on note π_i le morphisme $(x_j)_{i\in[1,n]} \to x_i$. Pour tout alphabet Σ et tout $A \subset \Sigma$, on note π_A le morphisme $\left\{\begin{array}{l} x \in A \Leftrightarrow \pi_A(x) = x \\ x \notin A \Leftrightarrow \pi_A(x) = \varepsilon \end{array}\right.$ et par $\pi_{\neg A}$ le morphisme $\left\{\begin{array}{l} x \in A \Leftrightarrow \pi_{\neg A}(x) = \varepsilon \\ x \notin A \Leftrightarrow \pi_{\neg A}(x) = x \end{array}\right.$

[11] Donc indépendantes entre elles, et indépendantes des périodiques.

4.1. **Des séries génératrices pour prévoir l'avenir.** Considérons un système de tâches [12] partitionné en n tâches périodiques $(\tau_i)_{i\in[1,n]}$ et p tâches sporadiques $(\alpha_i)_{i\in[1,p]}$. Pour éviter toute confusion, notons X_i^p les caractéristiques temporelles concernant les périodiques et X_i^s celles concernant les sporadiques. Chaque τ_i et chaque α_j sont à durée d'exécution fixe. Chaque α_j est indépendante de l'ensemble des autres tâches du système. Nous supposons ici que le système réduit à sa composante périodique est ordonnançable, c'est à dire que $Centre\left(L\left((\tau_i)_{i\in[1,n]}\right)\right) \neq \emptyset$, et qu'il vérifie le thórème 3. Le système global (c'est à dire composé des périodiques **et** des sporadiques) est ordonnançable si, pour tout $j \in [1,p]$ et tout $t \in \left[\min\limits_{i\in[1,n]}(r_i^p), +\infty\right[\cap \left\{ \begin{array}{c}\text{Dates d'activation}\\ \text{possibles de } \alpha_j\end{array}\right\}$, l'activation de α_j à l'instant t n'amème pas le système en situation de faute temporelle. Désignons par $A\left((\tau_i)_{i\in[1,n]}\right)$ l'automate déterministe minimal d'acceptation du langage $Centre\left(L\left((\tau_i)_{i\in[1,n]}\right)\right)$.

(1) Pour chaque état i de $A\left((\tau_i)_{i\in[1,n]}\right)$, nous devons pouvoir déterminer à l'avance quels sont les chemins issus de i, et permettant l'ordonnancement de chacune des α_j susceptibles d'être activées lorsque la composante périodique du système se trouve dans l'état i. Il nous faut donc associer à chaque transition t issue de i l'information $\left(D_j^s, K_{t,D_j^s}\right)_{j\in[1,p]}$, où K_{t,D_j^s} est le nombre maximal d'instants d'oisiveté du processeur dans les D_j^s prochaines unités de temps, si la transition t est choisie (voir Figure 5).

(2) Dans le cas où l'ordonnancement s'effectue sur multi-processeurs, cette seule information n'est cependant pas suffisante : k tâches peuvent être amenées à fonctionner simultanénement, il peut donc, à chaque unité de temps, y avoir jusqu'à k processeurs oisifs de front. Comme les tâches sont non parallélisables, on doit différencier *instants simultanés d'oisiveté* et *instants consécutifs d'oisiveté*. Dans le premier cas, deux instants d'oisiveté ne peuvent être alloués à la même tâche, alors qu'ils le peuvent dans le second cas. L'information à associer à t est donc $\left(D_j^s, \left(K_{t,j,D_j^s}\right)_{j\in[0,k]}\right)$, où K_{t,j,D_j^s} est le nombre d'unités de temps de l'intervalle [13] $\left[0, D_j^s\right]$ pendant lesquelles exactement j processeurs sont simultanément libres. On remarque que, ramené au cas mono-processeur, cette information correspond exactement à la durée d'oisiveté recherchée.

(3) Mais la seule donnée de cette information rend le modèle non modulaire, dans le sens où lors de la construction du produit de Hadamard d'une famille $(A_i)_{i\in[1,n]}$ d'automates finis, le calcul de l'information associée à une transition T du produit $\overset{i=n}{\underset{i=1}{\Omega}} A_i$ n'est pas réalisable à partir de la seule donnée des informations associées aux transitions $(t_i)_{i\in[1,n]}$ des automates A_i. Considérons, par exemple, les transitions t_1 et t_2, respectivement associées aux informations suivantes :

t_1: On dispose, dans les trois unités de temps à venir, d'un processeur oisif pendant trois unités de temps –on a deux processeurs en tout– et de deux processeurs simultanément libres pendant deux unités de temps. La configuration

[12] On rappelle que les tâches constituant les systèmes étudiés ne sont pas parallélisables.

[13] L'origine des temps est ici, évidemment, l'instant auquel le système périodique se trouve dans l'état i.

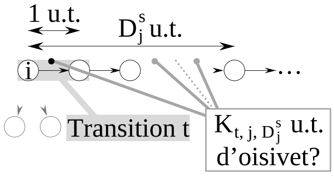


FIG. 5. Informations nécessaires à la décision d'ordonnançabilité d'une sporadique

d'oisiveté dans les trois unités de temps à venir est donc élément de l'ensemble

$$\left\{ \begin{pmatrix}\bullet\\ \bullet\end{pmatrix}\begin{pmatrix}\bullet\\ \bullet\end{pmatrix}\begin{pmatrix} \\ \bullet\end{pmatrix},\ \begin{pmatrix}\bullet\\ \bullet\end{pmatrix}\begin{pmatrix} \\ \bullet\end{pmatrix}\begin{pmatrix}\bullet\\ \bullet\end{pmatrix},\ \begin{pmatrix} \\ \bullet\end{pmatrix}\begin{pmatrix}\bullet\\ \bullet\end{pmatrix}\begin{pmatrix}\bullet\\ \bullet\end{pmatrix} \right\}$$

t_2**:** On sait que, dans les trois unités de temps à venir, le processeur –il n'y en a qu'un– est oisif pendant une seule unité de temps. La configuration d'oisiveté dans les trois unités de temps à venir est donc élément de l'ensemble

$$\{(\bullet)(\ )(\ ),(\ )(\bullet)(\ ),(\ )(\ )(\bullet)\}$$

Appelons T la transition de l'automate produit correspondant à la simultanéité de t_1 et t_2. En fonction des configurations respectives d'oisiveté de t_1 et de t_2, la configuration d'oisiveté de T est l'un des éléments de l'ensemble

$$\left\{ \begin{array}{ccc}
\begin{pmatrix}\bullet\\ \bullet\\ \bullet\end{pmatrix}\begin{pmatrix} \\ \bullet\\ \bullet\end{pmatrix}\begin{pmatrix} \\ \\ \bullet\end{pmatrix}, & \begin{pmatrix}\bullet\\ \bullet\\ \bullet\end{pmatrix}\begin{pmatrix} \\ \\ \bullet\end{pmatrix}\begin{pmatrix} \\ \bullet\\ \bullet\end{pmatrix}, & \begin{pmatrix} \\ \bullet\\ \bullet\end{pmatrix}\begin{pmatrix} \\ \bullet\\ \bullet\end{pmatrix}\begin{pmatrix} \\ \bullet\\ \bullet\end{pmatrix}, \\
\begin{pmatrix} \\ \bullet\\ \bullet\end{pmatrix}\begin{pmatrix}\bullet\\ \bullet\\ \bullet\end{pmatrix}\begin{pmatrix} \\ \\ \bullet\end{pmatrix}, & \begin{pmatrix} \\ \\ \bullet\end{pmatrix}\begin{pmatrix}\bullet\\ \bullet\\ \bullet\end{pmatrix}\begin{pmatrix} \\ \bullet\\ \bullet\end{pmatrix}, & \begin{pmatrix} \\ \bullet\\ \bullet\end{pmatrix}\begin{pmatrix} \\ \\ \bullet\end{pmatrix}\begin{pmatrix}\bullet\\ \bullet\\ \bullet\end{pmatrix}, \\
 & \begin{pmatrix} \\ \\ \bullet\end{pmatrix}\begin{pmatrix} \\ \bullet\\ \bullet\end{pmatrix}\begin{pmatrix}\bullet\\ \bullet\\ \bullet\end{pmatrix} &
\end{array} \right\}$$

Or, les différents éléments de cet ensemble correspondent à des configurations d'oisiveté qui n'offrent pas les mêmes capacités d'ordonnançabilité pour d'éventuelles tâches sporadiques. L'information associée à t_1 et t_2 n'est donc pas suffisante pour calculer l'information à associer à T : nous enrichissons donc cette information en associant à chaque transition l'information $\left(D,(Oisiv_i)_{i\in[1,D]}\right)$, où $D = \underset{\alpha\in\left\{(\alpha_i)_{i\in[1,p]}\right\}}{Max}(D_\alpha)$, et $Oisiv_i \in \mathbb{N}$ est le nombre de processeurs oisifs lors de la $i^{\text{ème}}$ unité de temps à venir. Cette information répond à ce problème : si t_1 est associée à l'information $\left(D,(Oisiv_{1i})_{i\in[1,D]}\right)$, et t_2 à l'information $\left(D,(Oisiv_{2i})_{i\in[1,D]}\right)$, alors, T est associée à l'information $\left(D,(Oisiv_{1i}+Oisiv_{2i})_{i\in[1,D]}\right)$ (notons qu'il s'agit d'un morphisme, que nous appelons // dans la suite).

Nous associons donc à chaque état ⓢ de l'automate fini une fonction $F_s(y) \in \mathbb{N} \ll (p_i)_{i\in[0,\pi]} \gg [y]$, dont la sémantique est la suivante :

- π est le nombre de processeurs disponibles
- L'indéterminée y trace la longueur des mots, c'est à dire le temps : la présence de y^a dans un monôme indique que l'on s'intéresse aux mots de longueur a. Comme, dans notre modèle, *longueur=durée*, cela signifie que l'on s'intéresse aux comportements de durée a.

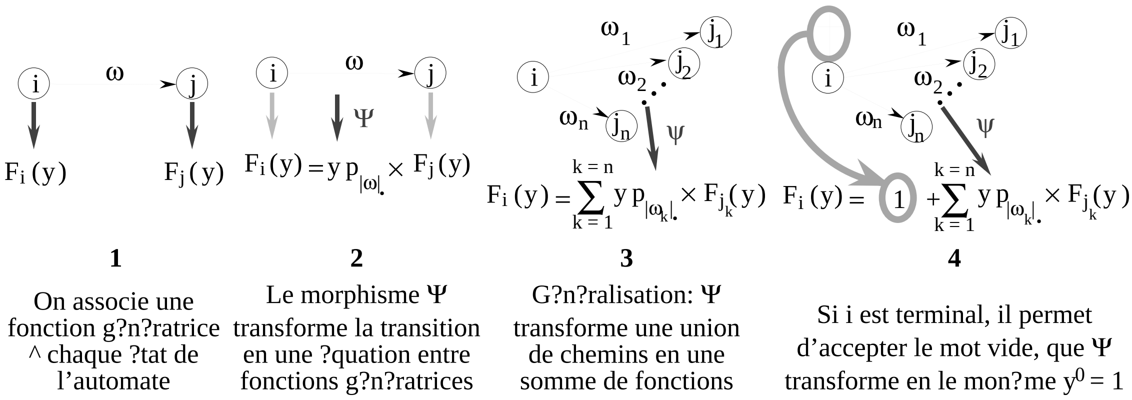


FIG. 6. Transformation de l'automate fini en un système linéaire

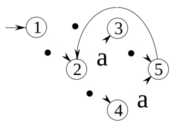


FIG. 7. Un exemple

– La suite des indéterminées p_i modélise le vecteur $(Oisiv_i)_{i\in[1,D]}$: la présence de p_i en $j^{\text{ème}}$ position dans la suite indique qu'à la $j^{\text{ème}}$ unité de temps à venir, on disposera d'exactement i processeurs oisifs. Par exemple, le facteur $p_1p_0p_2p_1$ modélise la configuration

Instant	1	2	3	4
Nombre de processeurs oisifs	1	0	2	1

L'automate est transformé en un système linéaire $M \times F = T$, où M est une matrice à coefficients polynomiaux, F un vecteur d'inconnues, et T un vecteur de polynômes de degrés nuls (voir Figure 6). Le vecteur de fractions rationnelles $(F_i(y))_{i\in[1,N]}$ (N est le nombre d'états de l'automate fini) est obtenu par la résolution du système linéaire.

Considérons par exemple l'automate représenté sur la Figure 7. On obtient le système :

$$\left\{\begin{array}{rcl} F_1(y) & = & p_1yF_2(y)+1 \\ F_2(y) & = & p_0yF_3(y)+p_1yF_4(y)+1 \\ F_3(y) & = & p_1yF_5(y)+1 \\ F_4(y) & = & p_0yF_5(y)+1 \\ F_5(y) & = & p_1yF_2(y)+1 \end{array}\right.$$

La résolution de ce système fournit :

$$\begin{cases} F_1(y) &= 1 + \dfrac{p_1 y + (p_1 p_0 + p_1^2) y^2}{1 - y^3 (p_1 p_0 p_1 + p_1^2 p_0)} \\ \\ F_2(y) &= 1 + y(p_0 + p_1) + y^2 (p_0 p_1 + p_1 p_1) \dfrac{p_1 y + (p_1 p_0 + p_1^2) y^2}{1 - y^3 (p_1 p_0 p_1 + p_1^2 p_0)} \\ \\ F_3(y) &= 1 + y p_1 + y p_1 \dfrac{p_1 y + (p_1 p_0 + p_1^2) y^2}{1 - y^3 (p_1 p_0 p_1 + p_1^2 p_0)} \\ \\ F_4(y) &= 1 + y p_0 + y p_0 \dfrac{p_1 y + (p_1 p_0 + p_1^2) y^2}{1 - y^3 (p_1 p_0 p_1 + p_1^2 p_0)} \\ \\ F_5(y) &= 1 + \dfrac{p_1 y + (p_1 p_0 + p_1^2) y^2}{1 - y^3 (p_1 p_0 p_1 + p_1^2 p_0)} \end{cases}$$

Une fois les fonctions génératrices déterminées, lorsque l'on veut savoir si une tâche sporadique de caractéristiques (C_s, D_s, P_s), se déclenchant à l'instant t, peut être traitée par le système, il suffit alors d'observer la fonction génératrice F associée à l'état de l'automate dans lequel se trouve le système à l'instant t :

- on développe F à l'ordre $D_s + 1$;
- on recherche, dans le coefficient de y^{D_s}, s'il existe un monôme indiquant au moins C_s unités de temps pendant lesquelles un processeur est oisif ;
- s'il en existe un, la tâche sporadique peut être prise en compte, sinon non.

4.2. **Caractérisation des fractions rationnelles.** En raison de leur mode de calcul, la caractérisation des fonctions associées aux transitions repose intégralement sur une caractérisation des F_i. Ce sont donc ces fonctions que nous allons étudier. Un premier résultat nous fournit une caractérisation de ces fonctions sur laquelle nous nous appuierons ensuite pour établir le résultat central de notre travail, dans l'optique de valider des systèmes temps-réel.

Théorème 4. *La fonction génératrice F_i, associée à l'état ⓘ, et obtenue par transformation de l'automate fini d'acceptation de $L\left((\tau_i)_{i\in[1,n]}\right)$ suivant le morphisme Ψ, est de la forme*

$$N_1(y) + \frac{N_2(y)}{1 - y^P D}.$$

N_1, N_2 et D sont des polynômes à coefficients dans $\mathbb{N} \ll (p_j)_{j\in[0,\pi]} \gg$ et $P = \underset{i\in[1,n]}{ppcm}(T_i^p)$.

Preuve

L'automate fini vérifie les deux propriétés suivantes :

- Les chemins joignant l'état ⓓ à l'état ⓐ partagent la même longueur : [2] établit que, en configuration mono-processeur, et lorsque les tâches sont à durées fixes, cette longueur est inférieure à $\underset{i\in[1,n]}{Max}(r_i) + \underset{i\in[1,n]}{ppcm}(T_i)$.
- Les chemins joignant l'état ⓐ à ⓔ partagent également la même longueur. Par construction de l'automate (que les tâches soient ou non à durées variables), cette longueur est $\underset{i\in[1,n]}{ppcm}(T_i)$.

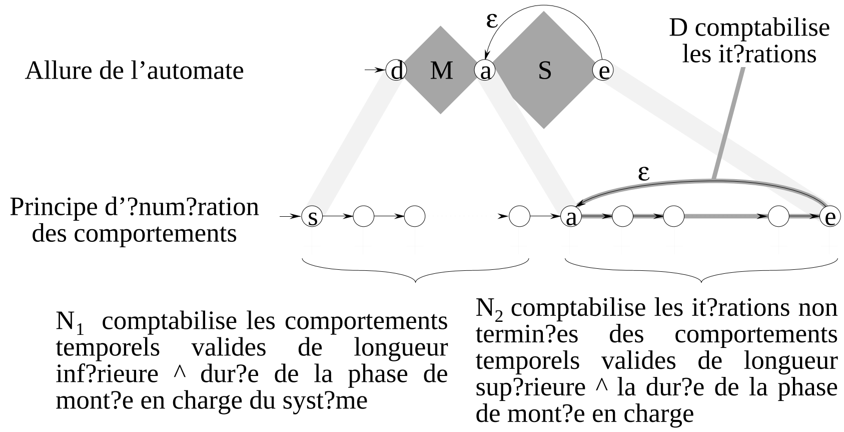


FIG. 8. Allure de l'automate

Soit ⓢ un état de l'automate. La structure de l'automate (voir Figure 8) entraîne que ◯ s se trouve soit sur (au moins) un chemin reliant ⓓ à ⓐ, soit sur un chemin reliant ⓐ à ⓔ. Donc, tout chemin issu de ⓢ est étiqueté par un mot ω, de la forme $\mu\nu\rho$, où

– μ est la trace d'un chemin reliant ⓢ à ⓔ
– ν est la trace d'un chemin reliant ⓐ à ⓐ. À ce titre, il est donc élément d'un langage étoilé
– ρ est la trace d'un chemin, de longueur inférieure à $\underset{i\in[1,n]}{ppcm}(T_i)$, joignant ⓐ à l'un (quelconque) des états atteignables depuis ⓐ

Notons P l'entier $\underset{i\in[1,n]}{ppcm}(T_i)$, l la longueur $|\mu|$, et $m_s(i)$ le polynôme associé aux chemins de longueur i, issus de ⓢ. Si $(\Sigma, Q, d, Q, \delta)$ est l'automate d'acceptation de $Centre\Big((\tau_i)_{i\in[1,n]}\Big)$, appelons L_s le langage accepté par l'automate $(\Sigma, Q, s, Q, \delta)$. On a $m_s(i) = \sum_{\omega\in L_s\cap\Sigma^i} m_{s,(\beta_h(\omega))_{h\in[1,k]}} \prod_{h=1}^{k} x_h^{\beta_h(\omega)}$, où $m_{s,(\beta_h(\omega))_{h\in[1,k]}}$ est le nombre de mots ω de L_s, de longueur i, correspondant à des ordonnancements pour lesquels on dispose, pour tout h, de $\beta_h(\omega)$ unités de temps pendant lesquelles h processeurs sont simultanément oisifs. L'état ⓐ est atteint depuis ⓢ en $l-1$ étapes. Calculons alors $F_s(y)$.

$$\begin{aligned} F_s(y) &= \sum_{i\geq 0} m_s(i)y^i \\ &= \sum_{i=0}^{l-1} m_s(i)y^i + \sum_{i\geq l} m_s(i)y^i. \end{aligned}$$

Or, tout chemin de longueur supérieure à l issu de ⓢ se décompose en

– Un chemin de ⓢ à ⓐ, de longueur l
– Un chemin issu de ⓐ

Les premiers chemins sont énumérés par $m_s(l)y^l$ et les seconds par $F_a(y)$. Donc

$$F_s = \sum_{i=0}^{l-1} m_s(i)y^i + m_s(l)y^l F_a(y)$$

Calculons alors $F_a(y)$.

Tous les chemins issus de ⓐ sont contenus dans une boucle de longueur P contenant ⓐ.

Donc, tout chemin issu de ⓐ de longueur supérieure à P débute par un chemin de longueur P de ⓐ à ⓐ. En utilisant ces remarques pour exprimer $F_a(y)$, on obtient :

$$\begin{aligned}
F_a(y) &= \sum_{i\geq 0} m_a(i)y^i \\
&= \sum_{i=0}^{P-1} m_a(i)y^i + \sum_{i\geq P} m_a(i)y^i \\
&= \sum_{i=0}^{P-1} m_a(i)y^i + \sum_{i\geq P} \left(m_a(P)y^P m_a(i-P)y^{i-P}\right) \\
&= \sum_{i=0}^{P-1} m_a(i)y^i + m_a(P)y^P \left(\sum_{i\geq 0} m_a(i)y^i\right) \\
&= \sum_{i=0}^{P-1} m_a(i)y^i + m_a(P)y^P F_a(y)
\end{aligned}$$

Donc

$$F_a(y) = \frac{\sum_{i=0}^{P-1} m_a(i)y^i}{1 - m_a(p)y^P}$$

et

$$F_s(y) = \sum_{i=0}^{l-1} m_s(i)y^i + \frac{m_s(l)y^l \sum_{i=0}^{p-1} m_a(i)y^i}{1 - m_a(P)y^P}$$

En posant $N_1(y) = \sum_{i=0}^{l-1} m_s(i)y^i$, $N_2(y) = m_s(l)y^l \sum_{i=0}^{P-1} m_a(i)y^P$, et $D = m_a(P)$, on obtient :

$$F_s(y) = N_1(y) + \frac{N_2(y)}{1 - Dy^P}$$

□

Remarques

- La construction des différents polynômes entraîne les inégalités suivantes :

$$\begin{gathered} deg(N_1, y) < l, \\ deg(N_2, y) < P + l - 1. \end{gathered}$$

- La sémantique naturellement associée aux polynômes s'exprime, en termes de chemins, de la façon suivante :
  - $N_1(y)$ correspond à la partie acyclique des chemins issus de ⓢ
  - $N_2(y)$ correspond à la fonction génératrice des chemins issus de ⓢ, passant par ⓐ , mais ne décrivant pas totalement un cycle de L_a
  - D correspond à la fonction génératrice de chemins de longueur P issus de ⓐ

4.3. **Modularité du modèle temporel enrichi.** Notre objectif est ici d'établir (voir Figure 9) que le calcul de la fonction génératrice associée à un état donné $\left\{(s_i)_{i\in[1,n]}\right\}$ du produit de Hadamard ne repose que sur la donnée des fonctions associées aux états $(s_i)_{i\in[1,n]}$ des automates *composantes*.

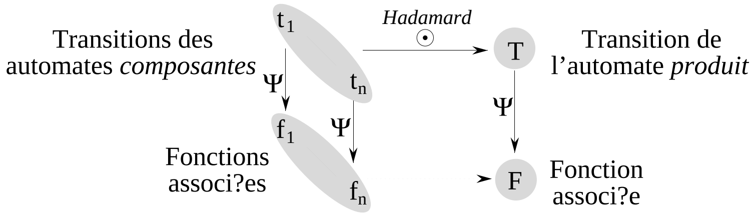


FIG. 9. Compatibilité des fonctions génératrices avec le produit de Hadamard

À partir de la famille $(A_i)_{i\in[1,n]}$ des automates associés aux tâches $(\tau_i)_{i\in[1,n]}$ (chaque A_i est un automate d'acceptation de $L(\tau_i)$), on calcule l'automate d'acceptation $A\left((\tau_i)_{i\in[1,n]}\right)$ de l'ensemble des comportements temporels valides du système $(\tau_i)_{i\in[1,n]}$ sur la base d'un produit de Hadamard des $(A_i)_{i\in[1,n]}$: chaque transition t de $A\left((\tau_i)_{i\in[1,n]}\right)$ est obtenue à partir du seul n-uplet $(t_i)_{i\in[1,n]}$ de transitions de $(A_i)_{i\in[1,n]}$ dont elle représente l'exécution simultanée. Pour le modèle enrichi, nous montrons que cette propriété est conservée : la fonction génératrice associée à t est calculable à partir de la seule donnée du n-uplet des fonctions génératrices associées aux $(t_i)_{i\in[1,n]}$.

Notre technique de calcul s'appuie sur la théorie des automates à multiplicités : l'idée consiste à associer à la fonction génératrice de chaque t_i un automate à multiplicité B_i puis, à partir de manipulations algébriques sur les B_i, à produire l'automate B associé à t. À partir de B, on est alors capable d'exprimer la fonction résultat [14].

4.3.1. *Automates à multiplicités.* Un K-sous-ensemble S de E est un morphisme de E dans K (où K est un semi-anneau) [17]. On appelle automate à multiplicité dans K tout 5-uplet Б=(A, Q, I, T, E) où A est un alphabet, Q un ensemble fini (les états de Б), I (les états initiaux) et T (les états terminaux) des K-sous-ensembles de Q et E un K-sous-ensemble de $Q \times A \times Q$ (les transitions et leur multiplicité).

4.3.2. *Représentation linéaire d'automates à multiplicités.* L'école de Schützenberger [17] a introduit la notion de représentation linéaire (λ, μ, γ), $\lambda \in K^{1\times n}$, $\mu \in K^{n\times n}$ et $\gamma \in K^{n\times 1}$ [20]. L'équivalence se fait de la façon suivante . Pour toute lettre $a \in A$, on construit une matrice $Q \times Q$, notée $\mu(a)$, telle que $\mu_{i,j}(a) = E(i, a, j)$ pout tout $(i, j) \in Q^2$. On définit ensuite $\lambda = (\lambda_1, \dots \lambda_n)$ et $\gamma = (\gamma_1, \dots, \gamma_n)$ de telle sorte que :

$$\lambda_i = \begin{cases} \text{0 si l'état } i \text{ de l'automate n'est pas initial} \\ \text{multiplicité associée à l'entrée dans l'état } i \text{ sinon} \end{cases}$$

$$\gamma_i = \begin{cases} \text{0 si l'état } i \text{ de l'automate n'est pas terminal} \\ \text{multiplicité associée à la sortie de l'état } i \text{ sinon} \end{cases}$$

Si l'on considère I comme une matrice ligne $K^{1\times Q}$ et T comme une matrice colonne $K^{Q\times 1}$, le comportement de Б est donné par

$$\mathrm{C}(\mathbf{B}) = \sum_{w\in A^*} IE(w)Tw,$$

qui est la série formelle reconnue par Б. Si l'on définit un étiquetage $Q = \{q_1, \dots, q_n\}$, on obtient, via cette correspondance, un triplet (λ, μ, γ) ($\lambda \in K^{1\times n}$, $\mu : A \to K^{n\times n}$, $\gamma \in K^{n\times 1}$) qui est appelé représentation linéaire de $\mathrm{C}(\mathbf{B})$.

[14] Les auteurs remercient Gérard Duchamp pour ses conseils sur ce sujet.

Soit $K \ll A \gg$ l'ensemble des séries formelles non commutatives sur un alphabet fini A et à coefficients dans K. Une série $S = \sum_{w \in A^*} (S,w)w$ est reconnaissable si et seulement si il existe une représentation linéaire (λ, μ, γ), de telle sorte que, pour tout $w \in A^*$, on a $(S,w) = \lambda\mu(w)\gamma$. Dans la suite, on notera cette propriété $S : (\lambda, \mu, \gamma)$ [21].
Ce travail est réalisé dans $K = \mathbb{N} \ll (p_j)_{j\in[0,\pi]} \gg\ll y \gg$. Si l'on ne considère que les indices des lettres p (par exemple $(1,0,2)$ au lieu de $p_1p_0p_2$), on se ramène au semi-anneau $K = (\mathbb{N} \cup \{\infty\})^{\mathbb{N}}$ muni des lois $(max, +)$, où $+$ correspond au morphisme // défini dans la section 4.1 et où max donne la structure de semi-anneau mais n'est pas utilisé dans notre étude.

4.3.3. *Automate à multiplicités associé à une fonction génératrice.* Considérons la fonction génératrice $F(y) = N(y)\,(yQ(y))^* = \frac{N(y)}{1-yQ(y)}$ associée à l'un des états (disons l'état ⓘ) de l'automate. N et Q sont des éléments de $\mathbb{N} \ll (p_i)_{i\in[0,\pi]} \gg [y]$. Notons $N(y) = \sum_{i=0}^{m} n_i y^i$ et $Q(y) = \sum_{i=0}^{P} q_i y^i$. Étant donné que N est obtenu à partir d'une réduction au même dénominateur (fusion des polynômes N_1 et N_2 –voir Théorème 4–), on a $P < m$.

Le résultat suivant fournit la représentation linéaire pour les fonctions génératrices.

Théorème 5. *Soient* $F(y) = \dfrac{N(y)}{1-yQ(y)} \in \mathbb{N} \ll (p_i)_{i\in[0,\pi]} \gg [y]$, *où*
$N(y) = \sum_{i=0}^{P} n_i y^i$, $Q(y) = \sum_{i=0}^{m} q_i y^i$, *et* $r = max(deg(N,y), deg(Q,y)) + 1$
Alors la représentation linéaire de $F(y)$ *est le triplet* (λ, μ, γ) *défini par :*

$$\lambda = \left(n_0, \ldots, n_P, 0_{1\times(r-P-1)}\right),\ \gamma = \begin{pmatrix} 1 \\ 0_{(r-1)\times 1} \end{pmatrix},\ \mu \in \mathbb{N} \ll (p_i)_{i\in[0,\pi]} \gg^{r\times r}$$

et

$$\mu(y) = \left(\begin{array}{ccccc|c} q_0 & \cdots & \cdots & \cdots & q_m & \\ 1 & 0 & \cdots & \cdots & 0 & \\ 0 & \ddots & \ddots & & \vdots & 0_{(r-m-1)\times(m+1)} \\ \vdots & \ddots & \ddots & \ddots & \vdots & \\ 0 & \cdots & 0 & 1 & 0 & \\ \hline & & 0_{(m+1)\times(m+1)} & & & 0_{(r-m-1)\times(r-m-1)} \end{array}\right)$$

Preuve

Dans un premier temps, construisons l'automate reconnaissant $yQ(y)$, c'est à dire $y\sum_{k=0}^{m} q_k y^k$.
Chacun des monômes de cette expression est de la forme $q_k y^{k+1}$. Considéré comme un mot, ce monôme étiquette le chemin (unique) dont l'origine est l'état initial $\left(\frac{y^{m-k+1}}{D}\right)$ de l'automate présenté en Figure 10, et l'arrivée le seul état terminal de cet automate. Ici, la multiplicité q_k, qui étiquette la flèche d'entrée dans l'automate, intervient au même titre qu'une lettre. Chaque état initial permet l'acceptation d'un monôme. Le langage accepté par cet automate est donc $\bigcup_{k=0}^{m} \left\{q_k y^{k+1}\right\}$. Le polynôme canoniquement associé à ce langage

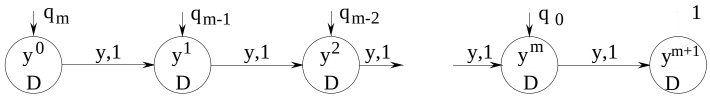


FIG. 10. Automate reconnaissant $yQ(y)$

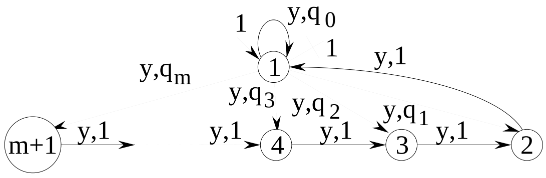


FIG. 11. Automate reconnaissant $(yQ(y))^*$

est donc $yQ(y)$. Sa representation linéaire est

$$\lambda^1 = (0, q_0, \dots, q_m),\ \gamma^1 = \begin{pmatrix} 1 \\ 0_{(m+1)\times 1} \end{pmatrix}$$

$$\text{et } \mu^1(y) = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ 1 & \ddots & & & \vdots \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

Pour obtenir la représentation linéaire de $(yQ(y))^* = \dfrac{1}{1 - yQ(y)}$, on utilise la forme linéaire de l'automate associé à un langage, présentée en section 4.3.2. Dans [22], il est établi que, si $S : (\lambda, \mu, \gamma)$ (de dimension m) est la représentation linéaire d'un automate, la représentation linéaire de S^* est [15] :

$$\left((0_{1\times m}, 1)\,, \left(\begin{array}{c|c} \mu(y) + \gamma\lambda\mu(y) & 0_{m\times 1} \\ \hline \lambda\mu(y) & 0 \end{array} \right), \begin{pmatrix} \gamma \\ 1 \end{pmatrix} \right)$$

En appliquant cette forme au calcul de la forme linéaire de $(yQ(y))^*$, on obtient :

$$\lambda^2 = (0_{1\times(m+2)}, 1),\ \gamma^2 = \begin{pmatrix} 1 \\ 0_{(m+1)\times 1} \\ 1 \end{pmatrix},\ \mu^2 \in \mathbb{N} \ll (p_j)_{j\in[0,\pi]} \gg^{(m+3)\times(m+3)}$$

$$\text{et } \mu^2(y) = \left(\begin{array}{cccccc|c} q_0 & \cdots & \cdots & \cdots & q_m & 0 & 0 \\ 1 & 0 & \cdots & \cdots & 0 & \vdots & \vdots \\ 0 & \ddots & \ddots & & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & 0 & 0 \\ \hline q_0 & \cdots & \cdots & \cdots & q_m & 0 & 0 \end{array} \right)$$

Cette forme correspond à un automate qui, après minimisation, se présente sous la forme proposée en Figure 11. En termes de langages, $(yQ(y))^* = \frac{1}{1-yQ(y)}$.

Modifions cet automate de façon que l'état initial soit (i), de multiplicité [16] m. Chaque

[15]Ce résultat n'est valide que lorsque $\lambda\gamma = 0$, ce qui est le cas dans notre étude.

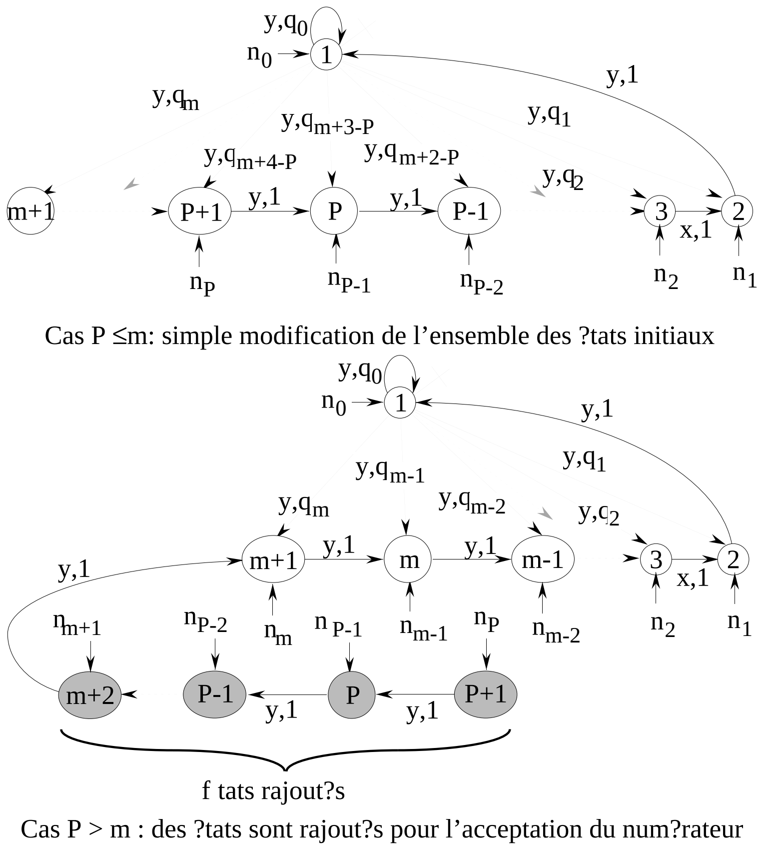


FIG. 12. Automate reconnaissant F(y)

monôme m reconnu par l'automate d'acceptation de $y(Q(y))^*$ est maintenant reconnu multiplié par my^i (coefficient étiquetant le seul chemin de ⓘ à ①, l'état terminal). L'automate modifié accepte donc $\frac{my^i}{1-yQ(y)}$.

Nous effectuons cette opération pour tout $i \in [0, P]$: cela modifie l'automate reconnaissant $(yQ(y))^*$ de telle sorte que l'état *i+1* soit initial et de multiplicité n_i. Notons que l'entrée dans l'état *1* est alors valuée par n_0 et non plus par 1. L'automate obtenu reconnait alors $\dfrac{1}{1-yQ(y)}\displaystyle\sum_{i=0}^{P} n_i y^i$, c'est à dire $\frac{N(y)}{1-yQ(y)}$. Remarquons que, dans certains cas, cette opération impose de rajouter des états manquants à la structure de l'automate initial. Ces nouveaux états sont alors liés par des transitions $(j,(y,1),j-1))$. Le nombre d'états de cet automate est finalement $r = Max\,(Deg(N,y), Deg(Q,y)) + 1$. L'automate correspondant est présenté en Figure 12. Sa représentation linéaire (λ, μ, γ) est :

$$\lambda = \left(n_0, \ldots, n_P, 0_{1\times(r-P-1)}\right),\ \gamma = \begin{pmatrix} 1 \\ 0_{(r-1)\times 1} \end{pmatrix},\ \mu \in \mathbb{N} \ll (p_j)_{j\in[0,\pi]} \gg^{r\times r}$$

[16] L'état ① n'est donc plus initial.

et

$$\mu(y)=\left(\begin{array}{ccccc|c} q_0 & \cdots & \cdots & \cdots & q_m & \\ 1 & 0 & \cdots & \cdots & 0 & \\ 0 & \ddots & \ddots & & \vdots & 0_{(r-m-1)\times(m+1)} \\ \vdots & \ddots & \ddots & \ddots & \vdots & \\ 0 & \cdots & 0 & 1 & 0 & \\ \hline \multicolumn{5}{c|}{0_{(m+1)\times(m+1)}} & 0_{(r-m-1)\times(r-m-1)} \end{array}\right)$$

□

Une fois déterminée la forme linéaire des fonctions génératrices que nous étudions, nous pouvons réaliser le produit de Hadamard de ces fonctions à l'aide de la propriété suivante :

Proposition 6. [22] *Soit $R : (\lambda^r, \mu^r, \gamma^r)$ (resp. $S : (\lambda^s, \mu^s, \gamma^s)$). Une représentation du produit de Hadamard est :*

$$R \odot S : (\lambda^r \otimes \lambda^s, \mu^r \otimes \mu^s, \gamma^r \otimes \gamma^s)$$

où $\otimes$ désigne le produit tensoriel.

Nous rappelons que, dans ce résultat, le produit tensoriel est l'opération de composition élémentaire des fonctions. Ici, il s'agit du morphisme // défini dans la section 4.1

Grâce à la proposition 6[22], lorsque nous réalisons un produit de Hadamard de langages (ou des automates associés), il n'est plus obligatoire de recalculer totalement le système des fonctions génératrices asociées, mais uniquement d'effectuer les produits de Hadamard de ces fonctions nécessaires. De ce fait, la modularité du modèle à automates se trouve étendue au modèle enrichi.

Le résultat des produits de Hadamard des fonctions génératrices étant exprimé sous forme de représentation linéaire, il nous faut maintenant déterminer l'expression algébrique de la série génératrice associée.

Soit $S(y)$ une série génératrice de représentation linéaire (λ, μ, γ).
Par définition, $S(y) = \sum_{i\geq 0}(S, y^i)y^i$. Comme $(S, y^i) = \lambda\mu(y^i)y^i\gamma$, on obtient :

$$S(y) = \sum_{i\geq 0} \lambda\mu(y^i)y^i\gamma = \lambda\left(\sum_{i\geq 0}\mu(y^i)y^i\right)\gamma = \lambda\left(\sum_{i\geq 0}(\mu(y)y)^i\right)\gamma = \lambda\,(\mu(y)y)^*\,\gamma.$$

Puisque $y\mu(y)$ est une matrice carrée, on calcule $(y\mu(y))^*$ en appliquant récursivement le résultat suivant [20] : soit $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ une matrice $n \times n$ structurée en blocs carrés a, b, c et d ; alors,

$$M^* = \left(\begin{array}{c|c} (a+bd^*c)^* & a^*b\,(ca^*b+d)^* \\ \hline d^*c\,(a+bd^*c)^* & (ca^*b+d)^* \end{array}\right)$$

5. Conclusion

L'utilisation des langages rationnels pour l'analyse d'ordonnançabilité s'est donc révélée très fructueuse. D'une part, le modèle est parfaitement adapté aux tâches à durées variables, pour lesquels les résultats établis dans [13] restent valides, et nous obtenons un critère de décision d'ordonnançabilité pour les systèmes de tâches à durée variables, interdépendants, et en environnement multi-processeurs. À notre connaissance, à ce jour, un tel critère n'a été obtenu par aucune autre approche. Par ailleurs, cette approche étend le résultat de la cyclicité des ordonnancements d'un système temps-réel (établi par [2] pour les systèmes périodiques en environnement mono-processeur) d'une part aux systèmes de tâches à durées variables, et d'autre part aux environnement multi-processeur, sous l'hypothèse de migration totale.

Pour les systèmes intégrant des tâches sporadiques, notre méthodologie fournit des résultats similaires. Ces tâches pouvant se modéliser à l'aide de langages rationnels, les résultats de décision d'ordonnançabilité et de cyclicité obtenus pour les systèmes périodiques persistent pour les systèmes intégrant des sporadiques. Par ailleurs, lorsque les tâches sporadiques sont indépendantes des tâches périodiques, nous utilisons un modèle enrichi (mis en place dans [14]) qui permet, à partir de la seule composante périodique du système, de décider des capacités du système à supporter la surcharge occasionnée par l'activation de tâches sporadiques. Nous fournissons ici une méthode de calcul des fonctions génératrices associées aux transitions compatible avec le produit de Hadamard de langages rationnels.

Actuellement, nous étudions les capacités d'analyse de notre modèle pour les systèmes distribués. Par ailleurs, les études comportementales du procédé de calcul semblent montrer qu'une explosion combinatoire reflète souvent une sous-charge du système analysé. Nous recherchons des raffinements du modèle, dans le but de limiter cette explosion, tout en conservant la puissance des outils d'analyse de langages rationnels. Dans le même ordre d'idée, nous étudions des modes de calcul des fonctions génératrices permettant de limiter la lourdeur du calcul. Le fait que les dénominateurs, par exemple, soient communs à l'ensemble des fonctions est un premier résultat ; la modularité facilitera grandement les calculs de produits. À moyen terme, nous désirons utiliser ces fonctions d'une part dans un but d'analyse de résistance à la surcharge, ou de résistance aux pannes, et d'autre part dans une optique d'analyse statistique des fautes temporelles contrôlées : l'idée sous-jacente est la mise en place d'une **mesure** de validité temporelle d'un système temps-réel, qui permette d'évaluer continûment sa validité, plutôt que de manière binaire, comme c'est le cas actuellement. À plus long terme, notre objectif est de mettre en place une méthodologie d'aide à la spécification temporelle d'applications temps-réel qui intègre l'aspect *gestion des risques* : les risques correspondront à des faiblesses connues et tolérées, qui pourront faire partie de la spécification temporelle.

Références

[1] C.L. Liu and J.W. Layland. Scheduling algorithms for multiprogramming in a hard real-time environment. *Journal of the ACM*, 20(1) :46–61, 1973.

[2] E. Grolleau. *Ordonnancement Temps-Réel Hors-Ligne Optimal à l'Aide de Réseaux de Petri en Environnement Monoprocesseur et Multiprocesseur.* PhD thesis, Univ. Poitiers, 1999.

[3] A. Choquet-Geniet. Ordonnancement des applications temps-réel. In *Actes de l'école d'éte École Temps Réel'99*, pages 53–68. ENSMA, 1999.

[4] T.P. Baker. Stack-based scheduling of real-time processes. *the Journal of Real-Time Systems*, 3 :67–99, 1991.

[5] J. Delacroix. Ordonnancement de tâches apériodiques. In *Actes de l'école d'éte E.T.R'99*, pages 69–82, 1999.

[6] S.K. Baruah, L.E. Rosier, and R.R. Howell. Algorithms and complexity concernig the preemptive scheduling of periodic, real-time tasks on one processor. *Real-Time Systems*, pages 301–324, 1990.

[7] J.Y.T. Leung and M.L. Merill. A note on preemptive scheduling of periodic real-time tasks. *Information Processing Letters*, 11(3) :115–118, 1980.

[8] R. Alur and D. Dill. A theory of timed automata. *Theoretical Computer Science*, 126 :183–235, 1994.

[9] G. Florin and S. Natkin. Les réseaux de petri stochastiques. *Techniques et Sciences Informatiques*, 4(1) :143–160, February 1985.

[10] C. Ramchandani. Analysis of asynchronous concurrent systems by petri nets. Technical report, Massachuchet Institute of Technology, project MAC, 1974.

[11] A. Choquet-Geniet, D. Geniet, and F. Cottet. Exhaustive computation of the scheduled task execution sequences of a real-time application. In *Proc. of Symposium on Formal Techniques in Real Time and Fault Tolerent Systems'96*, number 1135 in L.N.C.S., pages 246–262. Springer-Verlag, October 1996.

[12] D. Geniet and G. Largeteau. Validation temporelle de systèmes de tâches temps-réel strictes à durées variables à l'aide de langages. In *Proc. of Mod'elisation des Systèmes Réactifs' 2001*, pages 243–258. Hermes, October 2001.

[13] D. Geniet. Validation d'applications temps-réel à contraintes strictes à l'aide de langages rationnels. In *RTS'2000*, pages 91–106. Teknea, 2000.

[14] D. Geniet and J.P. Dubernard. Ordonnancement de tâches sporadiques à contraintes strictes à l'aide de séries génératrices. In *Proc. of Real Time Systems'01*, pages 149–166. Teknea, 2001.

[15] L. Aceto, P. Bouyer, A. Burgue, and K. G. Larsen. The power of reachability testing for timed automata. In *Proc. of* 18^{th} *Conf. Found. of Software Technology and Theor. Comp. Sci.*, LNCS 1530, pages 245–256. Springer-Verlag, December 1998.

[16] A. Arnold and M. Nivat. Comportements de processus. Technical Report 82-12, Univ. Paris 7, 1982.

[17] S. Eilenberg. *Automata Languages and machines*, volume A. Academic Press, 1976.

[18] A. Arnold. *Finite transition systems.* Prentice Hall, 1994.

[19] G. Largeteau, D. Geniet, and J.P. Dubernard. Validation of distributed periodic real-time systems using can protocol with finite automata. In *Proc. of* 5^{th} *S.C.I.* I.I.I.S., 2001.

[20] J. Berstel and C. Reutenauer. Rational series ans their languages. In *EATCS Monographs on Theoretical Computer Science*. Springer-Verlag, 1988.

[21] M. Fliess. Sur divers produits de séries formelles. *Bull. Sc. Math.*, 104 :181–191, 1974.

[22] G. Duchamp, M. Flouret, É. Laugerotte, and J.G. Luque. Direct and multiple laws for automata with multiplicities. *Theoretical Computer Science*, 267, 2001.

LABORATOIRE D'INFORMATIQUE FONDAMENTALE ET APPLIQUÉE DE ROUEN, UNIVERSITÉ DE ROUEN, PLACE ÉMILE BLONDEL, F-76821 MONT-SAINT-AIGNAN CÉDEX

E-mail address: `Jean-Philippe.Dubernard@univ-rouen.fr`

LABORATOIRE D'INFORMATIQUE SCIENTIFIQUE ET INDUSTRIELLE, UNIVERSITÉ DE POITIERS & E.N.S.M.A., TÉLÉPORT 2, 1 AV. C. ADER, BP 40109, F-86961 FUTUROSCOPE CHASSENEUIL CÉDEX

E-mail address: `dgeniet@ensma.fr`

INEQUALITIES FOR POLYTOPES AND ZONOTOPES

RICHARD EHRENBORG

Abstract. We prove that the coefficients of the **cd**-index of a convex polytope are increasing when replacing $\mathbf{c}^2$ with $\mathbf{d}$. That is, for P an n-dimensional convex polytope and u and v two **cd**-monomials such that the sum of their degrees is $n-2$, we have $[u\mathbf{c}^2 v]\Psi(P) \leq [u\mathbf{d}v]\Psi(P)$. This yields $(5 \cdot n - \sqrt{5})/25 \cdot \phi^n$ linear inequalities on the flag f-vector of an n-dimensional polytope, where ϕ is the golden ratio. We prove similar but stronger results for the **cd**-index of zonotopes.

Résumé. Nous prouvons que les coefficients de l'index **cd** d'un polytope convexe sont croissant lorsque l'on remplace $\mathbf{c}^2$ par $\mathbf{d}$. En d'autres termes, pour un polytope convexe de dimension n et deux monômes u et v tels que la somme de leurs degrés est $n-2$, on a $[u\mathbf{c}^2 v]\Psi(P) \leq [u\mathbf{d}v]\Psi(P)$. On produit $(5 \cdot n - \sqrt{5})/25 \cdot \phi^n$ inégalités linéaires pour le vecteur-f de drapeaux d'un polytope de dimension n, où ϕ est le nombre d'or. Nous prouvons des résultats similaires mais plus forts pour les index **cd** de zonotopes.

1. Introduction

The flag f-vector of a convex polytope contains all the enumerative incidence information between the faces. Thus to classify the set of all possible flag f-vectors is one of great open problems in discrete geometry. To date only partial results to this problem have been obtained. The case when the polytopes are simplicial (and dually, simple) the problem reduces to classifying the f-vectors of simplicial polytopes. This major step was solved by the combined effort of Billera and Lee [14] and Stanley [33]. Returning to the general case, the classification of flag f-vectors of three-dimensional polytopes was done by Steinitz [38] about one hundred years ago. Going up one dimension to four-dimensional polytopes, the case is still open. The article by Bayer [4] contains what is known for four dimensions.

The first step toward classifying flag f-vectors was taken by Bayer and Billera [6]. They observed that there are linear redundancies in the entries of the flag f-vector of a polytope. The relations holding between the entries of the flag f-vector are known as the generalized Dehn-Somerville relations. These relations imply that flag f-vectors of polytopes all lie in a subspace of dimension F_n, where F_n is the nth Fibonacci number ($F_0 = F_1 = 1$, $F_n = F_{n-1} + F_{n-2}$).

The next natural step is to look for linear inequalities that the flag vectors of polytopes satisfy. One such example is the toric g-vector. The entries of the toric g-vector are linear combinations of the entries of the flag f-vector. Stanley [35] proved that the toric g-vector of rational polytope is non-negative. Lately, there has been a lot of work of extending this result to non-rational polytopes; see [1, 2, 3, 19, 40]. More inequalities were obtained by Kalai by convoluting these inequalities together [27]. However, this is far from being all of the linear inequalities that the flag f-vector satisfies; see the work of Stenson [39].

A different direction of work has involved the **cd**-index. The **cd**-index $\Psi(P)$ of a polytope P is a non-commutative polynomial in two variables $\mathbf{c}$ and $\mathbf{d}$ whose coefficients are linear combinations of the entries of the flag f-vector. Thus a linear inequality among the coefficients of the **cd**-index implies a linear inequality among the entries of the flag f-vector. Moreover, the **cd**-index encodes the flag f-vector without linear redundancies. Another way to express this fact is that the set of **cd**-monomials offers a basis to the subspace of Fibonacci dimension. The existence of the **cd**-index was conjectured by Fine and proved by Bayer and Klapper [9].

Stanley [36] proved that the **cd**-index of a polytope has non-negative coefficients. This was the first important result which showed that the **cd**-index will play an important role in obtaining inequalities. The next step was taken by Billera and Ehrenborg who proved that the **cd**-index over all n-dimensional polytopes is minimized coefficientwise on the n-dimensional simplex Δ_n [10]. This gives a sharpening of the inequalities obtained by Stanley.

In this paper we will continue this vein of work. For an n-dimensional polytope P the **cd**-index is homogeneous of degree n, where the variable $\mathbf{c}$ has degree 1 and $\mathbf{d}$ has degree 2. We prove that the **cd**-index of a polytope P satisfies the family of inequalities

$$[u\mathbf{c}^2 v]\Psi(P) \;\leq\; [u\mathbf{d}v]\Psi(P), \tag{1.1}$$

where u and v are two **cd**-monomials such that $\deg(u) + \deg(v) = n - 2$. That is, when replacing a $\mathbf{c}^2$ with $\mathbf{d}$, the coefficient in $\Psi(P)$ increases. In total this yields $\sum_{i+j=n-2} F_i \cdot F_j \sim (5 \cdot n - \sqrt{5})/25 \cdot \phi^n$ linear inequalities, where ϕ is the golden ratio. Note however, that the $n - 1$ inequalities when the left-hand side is the coefficient of $\mathbf{c}^n$ are surpassed by the fact that the **cd**-index is minimized on the simplex.

There are quadratic inequalities known on the entries of the flag f-vector. Two large classes of quadratic inequalities are given by Braden and MacPherson [18] and Billera and Ehrenborg [10]. However, quadratic inequalities are not as fundamental as linear inequalities. That is, the set of flag f-vectors of convex polytopes seems to have as a first good approximation the cone determined by linear inequalities. Very little is known about this issue and it deserves a deeper study.

A second question of interest is to study flag f-vectors of zonotopes. Yet again, the first step is to consider linear relations. Since zonotopes are a subclass of polytopes, they satisfy the generalized Dehn-Somerville relations. Billera, Ehrenborg and Readdy proved that zonotopes satisfy no more linear relations [12]. The next step is to consider linear inequalities. Billera, Ehrenborg and Readdy proved that among all n-dimensional zonotopes (and more generally, the dual of the lattice of regions of oriented matroids), the n-dimensional cube minimizes the **cd**-index coefficientwise [11].

We prove two improvements of the inequality (1.1) for zonotopes. For an n-dimensional zonotope Z we have

$$[u\mathbf{d}v]\Psi(Z) - [u\mathbf{c}^2 v]\Psi(Z) \;\geq\; [u\mathbf{d}v]\Psi(\square_n) - [u\mathbf{c}^2 v]\Psi(\square_n), \tag{1.2}$$

for any two **cd**-monomials u and v such that $\deg(u)+\deg(v) = n-2$ and where $\square_n$ denotes the n-dimensional cube. That is, the increase in going from the coefficient of $u\mathbf{c}^2 v$ to $u\mathbf{d}v$ is greater than or equal to the corresponding increase in the cube. The second improvement is the following class of inequalities:

$$[\mathbf{c}^k\mathbf{d}v]\Psi(Z) - 2\cdot[\mathbf{c}^{k+2}v]\Psi(Z) \;\geq\; [\mathbf{c}^k\mathbf{d}v]\Psi(\square_n) - 2\cdot[\mathbf{c}^{k+2}v]\Psi(\square_n) \;\geq\; 0,$$

where k is a non-negative integer and v a **cd**-monomials such that $k+\deg(v) = n-2$. This improves the inequality (1.2) when u is a power of $\mathbf{c}$, that is, $u = \mathbf{c}^k$, by inserting a factor of 2 in front of the coefficient of $\mathbf{c}^{k+2}v$.

2. Preliminaries

A *partially ordered set (poset)* P is ranked if there is a rank function $\rho : P \longrightarrow \mathbb{Z}$ such that when x is covered by y then $\rho(y) = \rho(x) + 1$. Furthermore, we call P graded of rank n if it is ranked and has a minimal element $\hat{0}$ and a maximal element $\hat{1}$ such that $\rho(\hat{0}) = 0$

and $\rho(\hat{1}) = n$. Define the interval $[x, y]$ to be the subposet $\{z \in P : x \leq z \leq y\}$. Observe that the interval $[x, y]$ is also a graded poset of rank $\rho(x, y) = \rho(y) - \rho(x)$.

Let P be a graded poset of rank $n + 1$. For $S = \{s_1 < s_2 < \cdots < s_k\}$ a subset of $\{1, \ldots, n\}$, define f_S to be the number chains $\hat{0} = x_0 < x_1 < \cdots < x_{k+1} = \hat{1}$ where the rank of the element x_i is s_i for $1 \leq i \leq k$. These 2^n values constitute the *flag f-vector* of the poset P. Define the *flag h-vector* of P by the two equivalent relations $h_S = \sum_{T \subseteq S} (-1)^{|S-T|} f_T$ and $f_S = \sum_{T \subseteq S} h_T$. There has been a lot of recent work in understanding the flag f-vectors of graded posets and Eulerian posets; see [5, 8, 13].

For S a subset of $\{1, \ldots, n\}$ define the monomial $u_S = u_1 u_2 \cdots u_n$, where $u_i = \mathbf{a}$ if $i \notin S$ and $u_i = \mathbf{b}$ if $i \in S$. Define the **ab**-*index* of a graded poset P of rank $n + 1$ to be the sum

$$\Psi(P) = \sum_S h_S \cdot u_S.$$

A poset P is Eulerian if every interval $[x, y]$, where $x \neq y$, has the same number of elements of odd rank as the number of elements of even rank. This condition states that every interval $[x, y]$ satisfies the Euler-Poincaré relation. The condition of being Eulerian is also equivalent to that the Möbius function $\mu(x, y)$ is given by $(-1)^{\rho(x,y)}$. Two examples of Eulerian posets are the strong Bruhat order and face lattices of convex polytopes.

The following result was conjectured by Fine and proved by Bayer and Klapper [9].

Theorem 2.1. *Let P be an Eulerian poset. Then the* **ab**-*index of P, $\Psi(P)$, can be written in terms of* $\mathbf{c} = \mathbf{a} + \mathbf{b}$ *and* $\mathbf{d} = \mathbf{a} \cdot \mathbf{b} + \mathbf{b} \cdot \mathbf{a}$.

When $\Psi(P)$ is expressed in terms of $\mathbf{c}$ and $\mathbf{d}$ it is called the **cd**-*index* of poset P. There exist several proofs of this result in the literature; see [9, 15, 21, 25, 36]. The **cd**-index has been extraordinarily useful for flag vector computations; see [7, 11, 24]. Moreover, this basis is now emerging as a key tool for obtaining linear inequalities for the entries of the flag f-vector; see [10, 22, 36].

3. Polytopes

Let P be an n-dimensional convex polytope. The face lattice of P is an Eulerian poset of rank $n+1$. We denote the **cd**-index of the face lattice of the polytope P by $\Psi(P)$. A *regular cellular ball/sphere* Γ is a finite regular CW complex such that its underlying space $|\Gamma|$ is a ball/sphere. More generally, the face poset of a regular cellular sphere Γ is an Eulerian poset and we denote its **cd**-index by $\Psi(\Gamma)$.

One of the useful techniques for proving results about polytopes is shellings. In their seminal paper Bruggesser and Mani [20] proved that all polytopes are shellable. They did this by providing a special class of shellings called *line shellings*.

Recall that a *shelling* of an n-dimensional polytope P is an ordering of its facets $F_1, \ldots, F_m$ such that for $2 \leq r \leq m$, the complex $\Lambda = (F_1 \cup \cdots \cup F_{r-1}) \cap F_r$ is pure of dimension $n - 2$ and there exists a shelling $G_1, \ldots, G_k$ of the facet F_r such that $\Lambda = G_1 \cup \cdots \cup G_s$ for some s. For a line shelling of the polytope the shelling of F_r is also a line shelling. Another important observation is that for a line shelling the set $\partial(F_1 \cup \cdots \cup F_r)$ is combinatorially equivalent to an $(n-1)$-polytope. The set $\partial(F_1 \cup \cdots \cup F_r)$ is not itself an $(n-1)$-polytope. However, it can be projected onto a hyperplane such that the image is the boundary of a polytope with the same combinatorial type. We refer the reader to the article by Bruggesser and Mani for more details on line shellings.

When Stanley studied the **cd**-index in [36], he introduced a different notion of shelling, called spherical shelling or S-shelling for short. However, the line shellings of Bruggesser and Mani are also S-shellings. Hence it is enough for us to only consider line shellings.

To a regular cellular ball Γ of dimension n there are two regular cellular spheres associated with it. Namely, the boundary and the semi-suspension. The boundary $\partial\Gamma$ is an $(n-1)$-dimensional sphere. The *semi-suspension* Γ' is the n-dimensional regular cellular sphere obtained by attaching a new cell σ to Γ such that $\partial\sigma = \partial\Gamma$.

The following lemma is essential to our argument. It was proved by Billera and Ehrenborg (see [10, Lemma 4.2]) based on results of Stanley [36].

Lemma 3.1. *Let P be a polytope with a line shelling $F_1, \ldots, F_m$ and let $1 \leq r \leq m-1$. Let Λ be given by $(F_1 \cup \cdots \cup F_{r-1}) \cap F_r$. Then we have*

$$\Psi((F_1 \cup \cdots \cup F_r)') - \Psi((F_1 \cup \cdots \cup F_{r-1})') \;=\; (\Psi(F_r) - \Psi(\Lambda')) \cdot \mathbf{c} + \Psi(\partial\Lambda) \cdot \mathbf{d}.$$

On **cd**-polynomials there is a natural ordering by letting $z \leq w$ if and only if $w - z$ only has non-negative coefficients. We now define two stronger order relations $\preceq$ and $\preceq'$.

Definition 3.2. *Let z and w be two* **cd***-polynomials.*

(1) *Define the relation $z \preceq w$ if for all* **cd***-monomials u and v we have*

$$[u\mathbf{d}v]\, z - [u\mathbf{c}^2v]\, z \;\leq\; [u\mathbf{d}v]\, w - [u\mathbf{c}^2v]\, w.$$

(2) *Define the relation $z \preceq' w$ if for all* **cd***-monomials u and v, where v is not a power of* **c***, we have*

$$[u\mathbf{d}v]\, z - [u\mathbf{c}^2v]\, z \;\leq\; [u\mathbf{d}v]\, w - [u\mathbf{c}^2v]\, w.$$

A few observations are in order. When z and w are homogeneous of degree n, we may restrict u and v to be **cd**-monomials such that the sum of the degrees of u and v is $n-2$. Both order relations are transitive. Also, $\preceq$ is the stronger relation, that is, $z \preceq w$ implies $z \preceq' w$. Observe that both of these order relations are preserved under addition, that is, $z_1 \preceq w_1$ and $z_2 \preceq w_2$ implies $z_1 + z_2 \preceq w_1 + w_2$ and the similar addition rule for $\preceq'$ holds.

We call a **cd**-polynomial w *rising* if is satisfies $0 \preceq w$, that is, for all **cd**-monomials u and v we have that

$$[u\mathbf{c}^2v]\, w \;\leq\; [u\mathbf{d}v]\, w.$$

The rising homogeneous **cd**-polynomials of degree n thus form a cone in the linear space of homogeneous **cd**-polynomials of degree n.

Lemma 3.3. *If $z \succeq' 0$ and $w \succeq 0$ then $z \cdot \mathbf{c} + w \cdot \mathbf{d} \succeq' 0$.*

Proof. Observe that $z \succeq' 0 \Longrightarrow z \cdot \mathbf{c} \succeq' 0$ and $w \succeq 0 \Longrightarrow w \cdot \mathbf{d} \succeq 0 \Longrightarrow w \cdot \mathbf{d} \succeq' 0$. The lemma follows by adding these two conclusions. $\square$

We are now able to present the main theorem.

Theorem 3.4. *Let P be an n-dimensional polytope and let $F_1, \ldots, F_m$ be a line shelling of the polytope P. Then*

(a) *The* **cd***-index $\Psi(P)$ is rising.*

(b) *The following string of inequalities holds:*

$$0 \preceq' \Psi(F_1') \preceq' \Psi((F_1 \cup F_2)') \preceq' \cdots \preceq' \Psi((F_1 \cup \cdots \cup F_{m-1})') = \Psi(P).$$

Proof. The proof is by induction on the dimension n. We will show the implications $(a) \Longrightarrow (b) \Longrightarrow (a)$. However, in the step $(a) \Longrightarrow (b)$ we will use the result in (a) for lower dimensional polytopes. The induction basis is $n = 0$, and it is enough to observe that the **cd**-index of a point is rising.

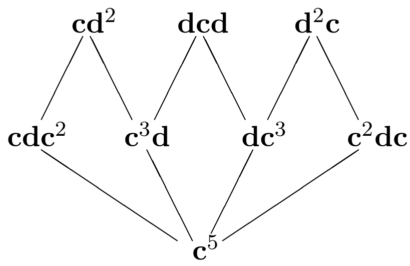


FIGURE 1. The eight **cd**-monomials of degree 5 in the partial order.

We next prove $(a) \Longrightarrow (b)$ by induction. By Lemma 3.1 we have that

$$\Psi((F_1 \cup \cdots \cup F_r)') - \Psi((F_1 \cup \cdots \cup F_{r-1})') \quad = \quad (\Psi(F_r) - \Psi(\Lambda')) \cdot \mathbf{c} + \Psi(\partial\Lambda) \cdot \mathbf{d},$$

where $\Lambda = (F_1 \cup \cdots \cup F_{r-1}) \cap F_r$. By induction we know that $\Psi(F_r) - \Psi(\Lambda') \succeq' 0$. Now consider the set $\partial\Lambda$. We know that Λ is the union of facets of F_r that form the beginning of a line shelling. Thus $\partial\Lambda$ is combinatorially equivalent to an $(n-2)$-dimensional polytope and hence by induction $\Psi(\partial\Lambda)$ is rising. Now by Lemma 3.3 the result follows.

We prove $(b) \Longrightarrow (a)$ by three cases. The first case when v is not a power of $\mathbf{c}$ follows directly by transitivity of all the order relations in (b). For the second case when v is a power of $\mathbf{c}$ and u is not, the result follows by applying the first case to the dual polytope. Finally, the third case is when both u and v are powers of $\mathbf{c}$. However, the result is immediate from the fact that the simplex has the smallest **cd**-index coefficientwise among all polytopes and that the simplex has positive coefficients; see [10, 24]. $\square$

We remark that Stanley's proof of the non-negativity of the **cd**-index of polytopes follows the same outline as the proof of Theorem 3.4. Just replace every occurrence of the two orders $\preceq$ and $\preceq'$ with the order $\leq$ and the proof of the non-negativity appears.

Corollary 3.5. *Let P be a polytope. Then the* **cd***-monomial with the largest coefficient in $\Psi(P)$ has no consecutive* **c***'s.*

It is natural to consider the partial order on the set of **cd**-monomials of degree n, where the cover relation is to replace an occurrence of $\mathbf{c}^2$ with $\mathbf{d}$. See Figure 1 for the poset in the case $n = 5$. This poset is simplicial, that is, it is the face lattice of a simplicial complex. This simplicial complex has been studied earlier; see [16, Corollary 2]. Every **cd**-monomial corresponds to a face having dimension equal to the number of **d**'s in the monomial minus one. The **cd**-monomials with no consecutive **c**'s correspond to facets. It is easy to observe that the dimensions of facets range between $\lfloor (n+1)/3 \rfloor - 1$ and $\lfloor n/2 \rfloor - 1$. Thus the simplicial complex is pure only when $n \leq 3$ or when $n = 5$. Compare this with the slightly misleading remark before Corollary 2 in [16].

4. ZONOTOPES

In this section we will improve the main inequality for zonotopes. Let $\square_n$ denote the n-dimensional cube.

Theorem 4.1. *Let Z be an n-dimensional zonotope, or more generally, let Z be the dual of the lattice of regions of an oriented matroid. Then the* **cd***-index $\Psi(Z)$ satisfies the inequality*

$\Psi(Z) \succeq \Psi(\square_n)$. *That is,*

$$[u\mathbf{d}v]\Psi(Z) - [u\mathbf{c}^2v]\Psi(Z) \;\geq\; [u\mathbf{d}v]\Psi(\square_n) - [u\mathbf{c}^2v]\Psi(\square_n),$$

for any two **cd***-monomials* u *and* v *such that* $\deg(u) + \deg(v) = n - 2$.

We will only prove this theorem for zonotopes. The proof for oriented matroids carries through exactly the same with the geometric language replaced with oriented matroid language.

Let ω be the linear map from $\mathbb{Z}\langle\mathbf{a},\mathbf{b}\rangle$ to $\mathbb{Z}\langle\mathbf{c},\mathbf{d}\rangle$ defined on an **ab**-monomial by replacing each occurrence of **ab** with $2\mathbf{d}$ and then replacing the remaining variables by **c**. Now we have fundamental theorem of computing the **cd**-index of a zonotope [11].

Theorem 4.2. *Let* Z *be a zonotope and* $\mathcal{H}$ *its associated central hyperplane arrangement. Let* L *be the intersection lattice of the hyperplane arrangement* $\mathcal{H}$ *and* $\Psi(L)$ *the* **ab***-index of the lattice* L*. Then the* **cd***-index of the zonotope and the sum of the* **cd***-indices of all the vertex figures of the zonotope are given by*

$$\begin{aligned} \Psi(Z) &= \omega(\mathbf{a} \cdot \Psi(L)), \\ \sum_v \Psi(Z/v) &= 2 \cdot \omega(\Psi(L)), \end{aligned}$$

where v *ranges over all vertices of the zonotope* Z*.*

The first identity is [11, Theorem 3.1]. The second identity follows from the first and using the linear map h defined in Section 8 in [11].

It remains to compute the **ab**-index of the intersection lattice L. We do this using R-labelings. For more details, see [11, Section 7] and [17, 32, 34]. Linearly order the hyperplanes in the arrangement $\mathcal{H}$, that is, $\mathcal{H} = \{H_1, \ldots, H_m\}$. Mark each edge $x \prec y$ in the Hasse diagram of the lattice L with the smallest (in the given linear order) hyperplane H such that intersecting x with H gives y. That is,

$$\lambda(x, y) = \min\{i \;:\; x \cap H_i = y\}.$$

For a maximal chain $c = \{\hat{0} = x_0 \prec x_1 \prec \cdots \prec x_n = \hat{1}\}$ define its *descent set* $D(c)$ by

$$D(c) = \{i \;:\; \lambda(x_{i-1}, x_i) > \lambda(x_i, x_{i+1})\}.$$

Then we have the following result; see Section 7 in [11].

Theorem 4.3. *The* **ab***-index of intersection lattice* L *is given by*

$$\Psi(L) = \sum_c u_{D(c)},$$

where the sum ranges over all maximal chains c *in the lattice* L*.*

Lemma 4.4. *Let* w *and* z *be rising non-negative* **cd***-polynomials. Then the* **cd***-polynomial* $w \cdot \mathbf{d} \cdot z$ *is also rising.*

Proposition 4.5. *Let* Z *be an* n*-dimensional zonotope and let* Z' *be the zonotope obtained by taking the Minkowski sum of* Z *with a line segment in the affine span of* Z*. Then the difference* $\Psi(Z') - \Psi(Z)$ *is rising.*

Proof. Let $\mathcal{H}$ and $\mathcal{H}'$ be the associated hyperplanes arrangements and let H be the new hyperplane. Let $\mathcal{H}'$ inherit the linear order of $\mathcal{H}$ with the new hyperplane H attached at the end of the linear order. Similarly, let L and L' be the corresponding intersection lattices. Observe that every maximal chain in L is also a maximal chain in L'. Also observe that there is no maximal chain in L' whose last label is H. Hence the difference in the **ab**-indices between the two intersection lattices is

$$\begin{aligned}\Psi(L') - \Psi(L) &= \sum_{c} u_{D(c)} \\ &= \sum_{\hat{0}<x\prec y} \Psi([\hat{0},x]) \cdot \mathbf{ab} \cdot \Psi([y,\hat{1}]) + \sum_{\hat{0}=x\prec y} \mathbf{b} \cdot \Psi([y,\hat{1}]),\end{aligned}$$

where the first sum is over all maximal chains c containing the label H and the next two sums is over edges $x \prec y$ in the Hasse diagram of L' having the label H. Applying the map $w \longmapsto \omega(\mathbf{a} \cdot w)$ we obtain

$$\Psi(Z') - \Psi(Z) = \sum_{\hat{0}<x\prec y} \omega(\mathbf{a} \cdot \Psi([\hat{0},x])) \cdot 2\mathbf{d} \cdot \omega(\Psi([y,\hat{1}])) + \sum_{\hat{0}\prec y} 2\mathbf{d} \cdot \omega(\Psi([y,\hat{1}])). \tag{4.1}$$

Observe that the term $\omega(\mathbf{a} \cdot \Psi([\hat{0},x]))$ is the **cd**-index of a zonotope and hence is rising by Theorem 3.4. Similarly, the term $\omega(\Psi([y,\hat{1}]))$ is one half of the sum of **cd**-indices of the vertex figures of a zonotope and hence is also rising. Now by Lemma 4.4 and the property that being rising is preserved under addition, the result follows. □

Proof of Theorem 4.1. Observe that any n-dimensional zonotope is obtained from the n-dimensional cube $\square_n$ by Minkowski adding line segments. Thus the result follows from Proposition 4.5. □

The second improvement of the inequalities is when comparing the coefficients of $\mathbf{c}^k\mathbf{d}v$ and $\mathbf{c}^{k+2}v$, that is, when u is a power of $\mathbf{c}$. For ease in notation, we introduce a third order relation.

Definition 4.6. *Define the order relation $z \preceq'' w$ on the* **cd***-polynomials z and w by*

$$[\mathbf{c}^k\mathbf{d}v]z - 2 \cdot [\mathbf{c}^{k+2}v]z \;\leq\; [\mathbf{c}^k\mathbf{d}v]w - 2 \cdot [\mathbf{c}^{k+2}v]w.$$

We call w weakly 2-rising *if $0 \preceq'' w$.*

Theorem 4.7. *Let Z be an n-dimensional zonotope, or more generally, let Z be the dual of the lattice of regions of an oriented matroid. Then the* **cd***-index $\Psi(Z)$ satisfies the inequalities*

$$\Psi(Z) \succeq'' \Psi(\square_n) \succeq'' 0.$$

That is, for all non-negative integers k and **cd***-monomials v such that $k + \deg(v) = n - 2$, we have*

$$[\mathbf{c}^k\mathbf{d}v]\Psi(Z) - 2 \cdot [\mathbf{c}^{k+2}v]\Psi(Z) \;\geq\; [\mathbf{c}^k\mathbf{d}v]\Psi(\square_n) - 2 \cdot [\mathbf{c}^{k+2}v]\Psi(\square_n) \;\geq\; 0.$$

Observe that Theorem 4.7 gives $\sum_{j=0}^{n-2} F_j = F_n - 1$ inequalities.

The proof of Theorem 4.7 consists of the following lemma and two propositions.

Lemma 4.8. *Let w be a weakly 2-rising* **cd***-polynomial and z a* **cd***-polynomial with non-negative coefficients. Then the* **cd***-polynomial $w \cdot \mathbf{d} \cdot z$ is also weakly 2-rising.*

Proposition 4.9. *The* **cd***-index of the n-dimensional cube $\square_n$ is weakly 2-rising.*

Proof. Proof by induction on n. The induction basis is $n = 0$ which is directly true. The induction step is based on the Purtill recursion for the **cd**-index of the n-dimensional cube; see [23, 30] or [24, Proposition 4.2]:

$$\Psi(\square_{n+1}) \quad = \quad \Psi(\square_n) \cdot \mathbf{c} + \sum_{i=0}^{n-1} 2^{n-i} \cdot \binom{n}{i} \cdot \Psi(\square_i) \cdot \mathbf{d} \cdot \Psi(\Delta_{n-i-1}).$$

Observe that the sum is weakly 2-rising by Lemma 4.8. However, the term $\Psi(\square_n) \cdot \mathbf{c}$ is not weakly 2-rising. It does satisfy the inequality $2 \cdot [\mathbf{c}^{k+2}v]\ \Psi(\square_n) \cdot \mathbf{c} \leq [\mathbf{c}^k\mathbf{d}v]\ \Psi(\square_n) \cdot \mathbf{c}$ for $0 \leq k \leq n-2$ but not for $k = n-1$. Hence $\Psi(\square_{n+1})$ satisfies the weakly 2-rising inequalities for $k \leq n-2$. To complete the proof it is enough to verify the $k = n-1$ case for $\Psi(\square_{n+1})$. This is straightforward since this amounts to stating that the cube $\square_{n+1}$ has at least four facets, which is true for cubes in dimension two and higher. $\square$

Proposition 4.10. *Let Z be an n-dimensional zonotope and let Z' be the zonotope obtained by taking the Minkowski sum of Z with a line segment in the affine span of Z. Assume that all zonotopes of dimension $n-1$ and less have their* **cd***-indices to be weakly 2-rising. Then the order relation $\Psi(Z) \preceq'' \Psi(Z')$ holds.*

Proof. The proof follows the same outline as the proof of Proposition 4.5. Observe that each term in equation (4.1) is weakly 2-rising by Lemma 4.8. Since the property of being weakly 2-rising is preserved under addition, the result follows. $\square$

We now prove Theorem 4.7.

Proof of Theorem 4.7. The proof is by induction. The induction basis is $n = 0$ which is straightforward. For the induction step assume that every zonotope of dimension k less than n satisfies the inequality $\Psi(\square_k) \preceq'' \Psi(Z)$. Especially, we know that the **cd**-index of a lower dimensional zonotope is weakly 2-rising. Thus by Proposition 4.10 we know that $\Psi(Z) \preceq'' \Psi(Z')$ holds for n-dimensional zonotopes. Now the theorem follows from Propositions 4.9. $\square$

5. Concluding remarks

Stanley conjectured that the **cd**-index over all Gorenstein* lattices of rank $n+1$ is coefficientwise minimized on the n-dimensional simplex [37]. In the light of our results for zonotopes in Section 4, it is natural to conjecture the following strengthening of Stanley's conjecture.

Conjecture 5.1. *Let L be a Gorenstein* lattice of rank $n+1$. Then the* **cd***-index $\Psi(L)$ satisfies the inequality $\Psi(L) \succeq \Psi(\Delta_n)$. That is, for all* **cd***-monomials u and v we have*

$$[u\mathbf{d}v]\Psi(L) - [u\mathbf{c}^2v]\Psi(L) \ \geq\ [u\mathbf{d}v]\Psi(\Delta_n) - [u\mathbf{c}^2v]\Psi(\Delta_n).$$

One possible method to prove this conjecture for polytopes is to use the following proposition and conjecture.

Proposition 5.2. *If the inequality $\Psi(\Delta_n) \preceq' \Psi(P)$ holds for all n-dimensional polytopes P then for all n-dimensional polytopes P we have $\Psi(\Delta_n) \preceq \Psi(P)$.*

Proof. This proof follows the exact same lines as the argument given for the implication $(b) \Longrightarrow (a)$ in the proof of Theorem 3.4. $\square$

Conjecture 5.3. *Let P be an n-dimensional polytope and F a face of dimension k of P. Let G be a k-dimensional face of the simplex Δ_n. Let $F_1, \ldots, F_r$ be the facets of P that contain the face F and let $G_1, \ldots, G_{n-1-k}$ be the facets of Δ_n containing the face G. Then*

$$\Psi((G_1 \cup \cdots \cup G_{n-1-k})') \preceq' \Psi((F_1 \cup \cdots \cup F_r)').$$

When $k = 0$ this conjecture states that $\Psi(\Delta_n) \preceq' \Psi((F_1 \cup \cdots \cup F_r)')$. Thus Conjecture 5.1 follows from Theorem 3.4, Proposition 5.2 and Conjecture 5.3.

An even more daring conjecture is the following:

Conjecture 5.4. *Let L be a Gorenstein* lattice of rank $n+1$. Then the* **cd***-index $\Psi(L)$ satisfies the inequality*

$$\frac{[u\mathbf{d}v]\Psi(L)}{[u\mathbf{c}^2v]\Psi(L)} \geq \frac{[u\mathbf{d}v]\Psi(\Delta_n)}{[u\mathbf{c}^2v]\Psi(\Delta_n)}.$$

If this conjecture fails, it would be desirable to obtain a lower bound for $[u\mathbf{d}v]\Psi(L)/[u\mathbf{c}^2v]\Psi(L)$. That is, find an estimate for the value

$$c_{u,v} = \inf_L \frac{[u\mathbf{d}v]\Psi(L)}{[u\mathbf{c}^2v]\Psi(L)},$$

where u and v are two **cd**-monomials such that the sum of their degrees is $n-2$ and L ranges over all Gorenstein* lattices of rank $n+1$.

Another question is to determine for which **cd**-monomials u and v is it possible to find two polytopes P and Q such that

$$\begin{aligned} [u]\Psi(P) &< [v]\Psi(P)\,, \\ [v]\Psi(Q) &< [u]\Psi(Q)\,? \end{aligned}$$

For instance, consider $u = \mathbf{d}^2$ and $v = \mathbf{cdc}$. We have that $[\mathbf{d}^2]\Psi(\Delta_4) = 4 < 5 = [\mathbf{cdc}]\Psi(\Delta_4)$ and $[\mathbf{cdc}]\Psi(\Box_4) = 16 < 20 = [\mathbf{d}^2]\Psi(\Box_4)$. However, by considering the known inequalities among the entries of the flag f-vector for four dimensional polytopes [4], one has that $[\mathbf{dc}^2]\Psi(P) \leq [\mathbf{cdc}]\Psi(P)$ and by duality $[\mathbf{c}^2\mathbf{d}]\Psi(P) \leq [\mathbf{cdc}]\Psi(P)$. Observe that these two inequalities do not follow from Theorem 3.4.

Meisinger, Kleinschmidt and Kalai proved that every 9-dimensional rational polytope has a three-dimensional face with less than 78 vertices or 78 facets [29, Theorem 5]. Their proof used the non-negativity of the toric g-vector of rational polytopes and convolutions of these inequalities. They also prove that every 9-dimensional polytope has the three-dimensional simplex as a quotient. It would be very interesting if one could improve their result using convolutions of the linear inequalities in Theorem 3.4.

Let H be a homogeneous **cd**-polynomial H of degree k, that is, we write $H = \sum_w \alpha_w \cdot w$, where the sum is over **cd**-monomials w of degree k. We call H *inequality generating* if the following inequality holds true:

$$\sum_w \alpha_w \cdot [uwv]\Psi(P) \geq 0,$$

for all polytopes P and **cd**-monomials u and v such that the sum of k and the degrees of u and v is the dimension of P. Using this language Theorem 3.4 states that the polynomial $\mathbf{d} - \mathbf{c}^2$ is inequality generating. Stanley's result that the **cd**-index of polytopes are non-negative amounts to saying that each **cd**-monomial is inequality generating. Can we find other examples of inequality generating polynomials? More generally, can we classify the set of inequality generating polynomials?

Now returning to zonotopes, the natural conjecture is the following.

Conjecture 5.5. *Let L be the dual of the lattice of regions of an oriented matroid. Then the* **cd***-index $\Psi(L)$ satisfies the inequality*

$$\frac{[u\mathbf{d}v]\Psi(L)}{[u\mathbf{c}^2v]\Psi(L)} \geq \frac{[u\mathbf{d}v]\Psi(\square_n)}{[u\mathbf{c}^2v]\Psi(\square_n)},$$

where $\square_n$ is the n-dimensional cube.

Other linear inequalities for the flag f-vector of zonotopes have been obtained by Varchenko and Liu; see [26, 28, 41].

It would interesting to continue the work of Readdy [31], who studied the question of determining the largest coefficient of the **ab**-index of certain polytopes. Thus to continue Corollary 3.5 it would be interesting to determine which coefficient of the **cd**-index is the largest for the simplex and the cube.

It can be shown that the rising property is preserved under the two linear operators $\mathrm{Pyr}, \mathrm{Prism}$ corresponding to the geometric pyramid and prism operations; see [24]. Is the rising property preserved under the two bilinear operators $M(\cdot, \cdot), N(\cdot, \cdot)$ that occur in the work of [22, 24]? Moreover, in [22] it is proved that three polytopes P, Q and R satisfy

$$\Psi(P \veebar (Q \times R)) \leq \Psi((P \veebar Q) \times R),$$

where $\veebar$ denotes the free join of polytopes and $\times$ the Cartesian product. Can this inequality be sharpened by replacing $\leq$ with $\preceq$?

Acknowledgements

I would like to thank the MIT Mathematics Department for their kind support where this research was done while the author was a Visiting Scholar. The author also thanks Margaret Readdy for many helpful discussions and Marge Bayer for suggesting references.

References

[1] G. Barthel, J.-P. Brasselet, K.-H. Fieseler and L. Kaup, Equivariant intersection cohomology of toric varieties, arXiv: math.AG/9904159.

[2] G. Barthel, J.-P. Brasselet, K.-H. Fieseler and L. Kaup, Combinatorial intersection cohomology for fans, arXiv: math.AG/0002181.

[3] F. Battaglia and E. Prato, Generalized toric varieties for simple non-rational convex polytopes, arXiv: math.CV/0004066.

[4] M. Bayer, The extended f-vectors of 4-polytopes, *J. Combin. Theory Ser. A* **44** (1987), 141–151.

[5] M. Bayer, Signs in the *cd*-index of Eulerian partially ordered sets, *Proc. Amer. Math. Soc.* **129** (2001), 2219–2225.

[6] M. Bayer and L. Billera, Generalized Dehn-Sommerville relations for polytopes, spheres and Eulerian partially ordered sets, *Invent. Math.* **79** (1985), 143–157.

[7] M. Bayer and R. Ehrenborg, The toric h-vectors of partially ordered sets, *Trans. Amer. Math. Soc.* **352** (2000), 4515–4531.

[8] M. Bayer and G. Hetyei, Flag vectors of Eulerian partially ordered sets, *European J. Combin.* **22** (2001), 5–26.

[9] M. Bayer and A. Klapper, A new index for polytopes, *Discrete Comput. Geom.* **6** (1991), 33–47.

[10] L. J. Billera and R. Ehrenborg, Monotonicity properties of the **cd**-index for polytopes, *Math. Z.* **233** (2000), 421–441.

[11] L. J. Billera, R. Ehrenborg, and M. Readdy, The **c**-2**d**-index of oriented matroids, *J. Combin. Theory Ser. A* **80** (1997), 79–105.

[12] L. J. Billera, R. Ehrenborg, and M. Readdy, The **cd**-index of zonotopes and arrangements, *Mathematical essays in honor of Gian-Carlo Rota* (B. E. Sagan and R. P. Stanley, eds.), Birkhäuser, Boston, 1998, pp. 23–40.

[13] L. J. Billera and G. Hetyei, Linear inequalities for flags in graded partially ordered sets, *J. Combin. Theory Ser. A* **89** (2000), 77–104.

[14] L. J. Billera and C. W. Lee, A proof of the sufficiency of McMullen's conditions for f-vectors of simplicial polytopes, *J. Combin. Theory Ser. A* **31** (1981), 237–255.
[15] L. J. Billera and N. Liu, Noncommutative enumeration in graded posets, *J. Algebraic Combin.* **12** (2000), 7–24.
[16] L. J. Billera and A. N. Myers, Shellability of interval orders, *Order* **15** (1999), 113–117.
[17] A. Björner, Shellable and Cohen-Macaulay partially ordered sets, *Trans. Amer. Math. Soc.* **260** (1980), 159–183.
[18] T. C. Braden and R. MacPherson, Intersection homology of toric varieties and a conjecture of Kalai, *Comment. Math. Helv.* **74** (1999), 442–455.
[19] P. Bressler and V. Lunts, Intersection Cohomology of Rational Polytopes, MPI-2000-87, `http://www.mpim-bonn.mpg.de/html/preprints/preprints.html`.
[20] H. Bruggesser and P. Mani, Shellable decompositions of spheres and cells, *Math. Scand.* **29** (1971), 197–205.
[21] R. Ehrenborg, k-Eulerian posets, *Order* **18** (2001), 227–236.
[22] R. Ehrenborg and H. Fox, Inequalities for **cd**-indices of joins and products of polytopes, to appear in *Combinatorica.*
[23] R. Ehrenborg and M. Readdy, The **r**-cubical lattice and a generalization of the **cd**-index, *European J. Combin.* **17** (1996), 709–725.
[24] R. Ehrenborg and M. Readdy, Coproducts and the **cd**-index, *J. Algebraic Combin.* **8** (1998), 273–299.
[25] R. Ehrenborg and M. Readdy, A homological proof for the existence of the **cd**-index, preprint 2001.
[26] K. Fukuda, S. Saito, and A. Tamura, Combinatorial face enumeration in arrangements of hyperplanes and oriented matroids, *Discrete Appl. Math.* **31** (1991), 141–149.
[27] G. Kalai, A new basis of polytopes, *J. Combin. Theory Ser. A* **49** (1988), 191–209.
[28] N. Liu, "Algebraic and combinatorial methods for face enumeration in polytopes," Doctoral dissertation, Cornell University, Ithaca, New York, 1995.
[29] G. Meisinger, P. Kleinschmidt and G. Kalai, Three theorems, with computer-aided proofs, on three-dimensional faces and quotients of polytopes, *Discrete Comput. Geom.* **24** (2000), 413–420.
[30] M. Purtill, André permutations, lexicographic shellability and the *cd*-index of a convex polytope, *Trans. Amer. Math. Soc.* **338** (1993), 77–104.
[31] M. A. Readdy, Extremal problems for the Möbius function in the face lattice of the n-octahedron, *Discrete Math., Special issue on Algebraic Combinatorics* **139** (1995), 361–380.
[32] R. P. Stanley, Finite lattices and Jordan-Hölder sets, *Algebra Universalis* **4** (1974), 361–371.
[33] R. P. Stanley, The number of faces of simplicial convex polytopes, *Adv. Math.* **35** (1980), 236–238.
[34] R. P. Stanley, "Enumerative Combinatorics, Vol. I," Wadsworth and Brooks/Cole, Pacific Grove, 1986.
[35] R. P. Stanley, Generalized h-vectors, intersection cohomology of toric varieties, and related results, in: "Commutative Algebra and Combinatorics" (M. Nagata and H. Matsumura, eds.), *Advanced Studies in Pure Mathematics* **11**, Kinokuniya, Tokyo and North-Holland, Amsterdam/New York 1987, pp. 187–213.
[36] R. P. Stanley, Flag f-vectors and the *cd*-index, *Math. Z.* **216** (1994), 483–499.
[37] R. P. Stanley, A survey of Eulerian posets, in: "Polytopes: Abstract, Convex, and Computational," (T. Bisztriczky, P. McMullen, R. Schneider, A. I. Weiss, eds.), NATO ASI Series C, vol. 440, Kluwer Academic Publishers, 1994.
[38] E. Steinitz, Über die Eulerischen Polyderrelationen, *Arch. Math. Phys.* **11** (1906), 86–88.
[39] C. Stenson, Relationships among flag f-vector inequalities for polytopes, preprint 2001.
[40] V. Timorin, On polytopes simple in edges (version 3), arXiv: math.AG/0010213.
[41] A. Varchenko, On the numbers of faces of a configuration of hyperplanes, *Soviet Math. Dokl.* **38** (1989), 291–295.

Department of Mathematics, University of Kentucky, Lexington, KY 40506-0027, USA
E-mail address: `jrge@ms.uky.edu`

EXPECTED NUMBER OF INVERSIONS AFTER A SEQUENCE OF RANDOM ADJACENT TRANSPOSITIONS — AN EXACT EXPRESSION

NIKLAS ERIKSEN

Abstract. A formula for calculating the expected number of inversions after t random adjacent transpositions has been presented by Eriksson et al. We have improved their result by determining a formula for the unknown integer sequence d_r that was used in their formula and also made the formula valid for large t.

Résumé. Une formule pour calculer le nombre attendu d'inversions après t transpositions adjacentes aléatoires a été présentée par Eriksson et al. Nous avons amélioré ce résultat en déterminant une formule pour la séquence inconnue d'entiers d_r, qui était utilisée dans leur formule et qui rendait la formule valide lorsque t prend une grande valeur.

1. Introduction

In a recent article [1], the Eriksson-Sjöstrand family calculated the expected number of inversions in a permutation, given the number of adjacent transpositions applied to it. Problems of this type have applications in computational biology, where the genome may be regarded as a permutation of genes. Consider two such genomes π and ρ, in which we have named the genes such that $\rho = id$. The evolutionary distance between π and ρ is assumed to be proportional to the number of evolutionary operations that have changed the gene order since the two genomes diverged. To calculate this number of operations, we can either calculate the least number of operations needed to transform π into $\rho = id$ (this corresponds to sorting π), which gives a lower bound of the true number of operations, or we can calculate the expected number of operations, given some measure on the difference between the two genomes. One such common measure is the number of breakpoints, that is the number of adjacencent pairs in π that are not consecutive.

In the paper by Eriksson et al., they calculated the inverse of the second alternative: they found the expected measure of difference given a certain number of operations. With this information, we may determine this measure of difference between two given genomes and then extract the number of operations that is expected to produce this difference. The same approach has been taken by Wang [2], for breakpoints and the long range inversions and transpostions usually considered in computational biology.

As mentioned, Eriksson et al. considered inversions and adjacent transpositions. Their result is the following

Theorem 1.1. *The expected number of inversions in a permutation in S_{n+1} after t random adjacent transpositions is, for $n \geq t$,*

$$E_{nt} = \sum_{r=0}^{t} \frac{(-1)^r}{n^r} \left[\binom{t}{r+1} 2^r C_r + 4 d_r \binom{t}{r} \right],$$

where d_r is an integer sequence that begins with $0, 0, 0, 1, 9, 69, 510$ and C_r are the Catalan numbers.

Supported by a grant from the Swedish Research Council.

There are a couple of things that can be improved in the result of Eriksson et al. First, their formula includes some numbers d_r that they have no expression formula for. Second, the formula is only valid for $n \geq t$.

In this paper, we will present an improved formula, where both these flaws have been eliminated. The theorem is given directly below, and the proof will appear in the following sections.

Theorem 1.2. *The expected number of inversions in a permutation in S_{n+1} after t random adjacent transpositions is*

$$E_{nt} = \sum_{r=1}^{t} \frac{1}{n^r} \binom{t}{r} \sum_{s=1}^{r} \binom{r-1}{s-1} (-1)^{r-s} 4^{r-s} g_{s,n}.$$

The integer sequence $g_{s,n}$ is given by

$$g_{s,n} = \sum_{l=0}^{n} \sum_{k \in \mathbb{N}} (-1)^k (n-2l) \binom{2\lceil \frac{s}{2} \rceil - 1}{\lceil \frac{s}{2} \rceil + l + k(n+1)} \sum_{j \in \mathbb{Z}} (-1)^j \binom{2\lfloor \frac{s}{2} \rfloor}{\lfloor \frac{s}{2} \rfloor + j(n+1)}$$

For $n \geq t$, we get

$$E_{nt} = \sum_{r=0}^{t} \frac{(-1)^r}{n^r} \left[2^r C_r \binom{t}{r+1} + 2\binom{t}{r} \sum_{s=3}^{r} \binom{r-1}{s-1} (-1)^{s-1} 4^{r-s} \binom{2\lfloor \frac{s}{2} \rfloor}{\lfloor \frac{s}{2} \rfloor} \sum_{l=0}^{\lfloor \frac{s-1}{2} \rfloor} l \binom{2\lceil \frac{s}{2} \rceil - 1}{\lceil \frac{s}{2} \rceil + l} \right]$$

where C_r are the Catalan numbers. Thus, the sequence d_r is given by

$$d_r = \frac{1}{2} \sum_{s=3}^{r} \binom{r-1}{s-1} (-1)^{s-1} 4^{r-s} \binom{2\lfloor \frac{s}{2} \rfloor}{\lfloor \frac{s}{2} \rfloor} \sum_{l=0}^{\lfloor \frac{s-1}{2} \rfloor} l \binom{2\lceil \frac{s}{2} \rceil - 1}{\lceil \frac{s}{2} \rceil + l}.$$

2. The heat flow model

To prove Theorem 1.2, we have used the heat flow model proposed by Eriksson et al. Before we state this model, we need a few definitions.

We look at the symmetric group S_{n+1}. The transposition that changes the elements π_i and π_{i+1} is denoted s_i. We let

$$\mathcal{P}_{nt} = \{s_{i_1} s_{i_2} \dots s_{i_t} : 1 \leq i_1, i_2, \dots, i_t \leq n\},$$

that is the set of sequences of exactly t adjacent transpositions.

Fix n. We define the matrix $(p_{ij})(t)$, where

$$p_{ij}(t) = \mathrm{Prob}(\pi_i < \pi_j)$$

for a permutation $\pi \in \mathcal{P}_{nt}$, where the adjacent transpositions $s_k, 1 \leq k \leq t$ have been chosen randomly from a uniform distribution. From this, it follows that

$$E_{nt} = \sum_{i>j} p_{ij}(t).$$

We now define a discrete heat flow process as follows. On a (finite or infinite) graph, every vertex has at time zero some heat associated to itself. In each time step, all vertices sends a fraction x of its heat to each of its neighbours. At the same time, it will receive the same fraction of each neighbours' heat. The following proposition is proven in [1].

Proposition 2.1. *(Eriksson et al. [1]) The sequence of (p_{ij})-matrices for $t = 0, 1, 2, \dots$ describes a discrete heat flow process with conductivity $x = 1/n$ on the grid graph depicted in Figure 1 (left).*

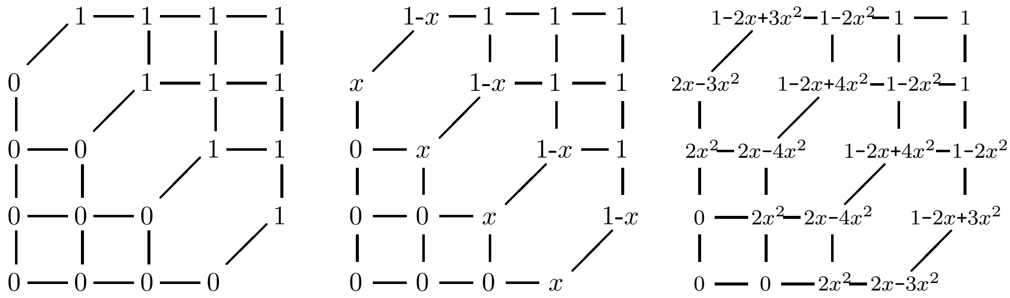


FIGURE 1. The matrices $(p_{ij})(0)$, $(p_{ij})(1)$ and $(p_{ij})(2)$ for $n = 4$.

In the same paper, they also show that we can replace the graph in Figure 1 by the grid in Figure 2. The sequence of (p_{ij})-matrices for $t = 0, 1, 2, \ldots$ describes a heat flow process on this grid graph. In this process, the heat on the diagonal will never change. Furthermore, we are only interested in the part below the diagonal. We thus get a model with two insulated boundaries (below and to the left) and one hot boundary (the diagonal). This is depicted in Figure 3.

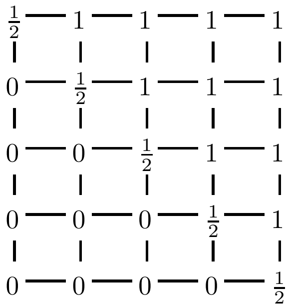


FIGURE 2. Grid graph with initial values

By reflection, we can extend this graph to a graph with no insulated boundaries (as in Figure 3). We will now calculate the amount of heat that flows from one of the borders (say the northeast one) onto this grid. This will equal the amount of heat in the upper right quarter of the grid, which is what we are trying to calculate. Remember that this heat equals E_{nt}.

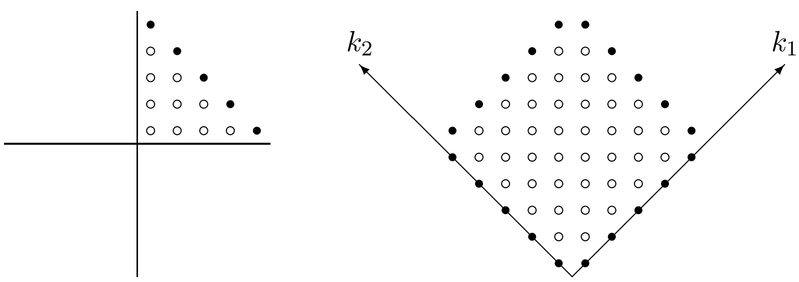


FIGURE 3. By reflection, the graph with one hot and two insulated boundaries is extended to a diamond shaped graph with no insulated boundaries. The new set of coordinates (k_1, k_2) is introduced.

The amazing thing about the heat flow model is that we can calculate the contribution from every heat packet separately, and then add them all together. In the model proposed by Eriksson et al., the vertices at the hot boundary send out heat packets with value $\frac{1}{2n}$ to their neighbours at each time step. These packets are then sent back and forth between the inner vertices. There are three possible travel steps for a packet [1]:

- It stays on the vertex unchanged.
- It travels to a neighbouring vertex, getting multiplied by $\frac{1}{n}$.
- It travels halfway to a neighbouring vertex, gets multiplied by $\frac{-1}{n}$ and returns to the vertex it came from.

Now, in order to calculate the total heat at a vertex, we sum, over all travel routes from the boundary, the heat packets that have traveled these routes. We define new coordinates k_1 and k_2 on this grid as in Figure 3 (the origin is at the bottom of the graph). If a packet has traveled from the northeast border to (i, j) in t days, we know the following.

- Out of the t days, there are r travel days. They can be chosen in $\binom{t}{r}$ ways.
- From these travel days, we must choose s true travel days, in which the packet changes vertex. This can be done in $\binom{r-1}{s-1}$ ways, since the packet must change vertex the first travel day.
- If the packet does not change vertex on a travel day, it has four direction to choose from. This gives the factor 4^{r-s}.
- The heat that reaches the destination is $\frac{(-1)^{r-s}}{2}\frac{1}{n^r}$.
- For each of the true travel days, both coordinates k_1 and k_2 change. Only paths that do not touch the boundary are valid. We will enumerate these walks, which we call **two-sided Catalan walks**.

It should be noted that Eriksson et al. used a similar approach, but only on a semi-infinite model, which gave a lower bound for E_{nt}.

We are now able to prove the first part of Theorem 1.2. We will sum over all vertices in the diamond graph, and for each vertex over all paths from the northeast border. These paths will display two-sided Catalan walks from $(0, a)$ to (s, b) (with a odd), where the y-coordinate corresponds to k_2, and two-sided Catalan walks from $(0, 1)$ to $(s-1, b)$ where the y-coordinate corresponds to $2n + 2 - k_1$. Let $b_{s,n}$ and $c_{s-1,n}$ be the number of such two-sided Catalan walks, respectively. This yields, with $x = 1/n$,

$$E_{nt} = \frac{1}{2}\sum_{r=1}^{t}\frac{1}{n^r}\binom{t}{r}\sum_{s=1}^{r}\binom{r-1}{s-1}(-1)^{r-s}4^{r-s}b_{s,n}c_{s-1,n}.$$

Thus, the first part of the theorem is proven (we have, of course, $g_{s,n} = \frac{b_{s,n}c_{s-1,n}}{2}$).

3. Two-sided Catalan walks

We start by formally defining two-sided Catalan walks and then proceed to enumerate them.

Definition 3.1. *A* **two-sided Catalan walk of height** n *is a walk on the integer grid from* $(0, a)$ *to* (s, b)*, where* $a, b \in \{1, 2, \ldots n-1\}$ *and* $s > 0$*, allowing only the steps* $(1, 1)$ *and* $(1, -1)$*, such that* $0 < y < n$ *at all positions along the way.*

We see that the number of two-sided Catalan walks from $(0, 1)$ to $(2k, 1)$ is C_k (ordinary Catalan numbers) if the height is larger than $k + 1$ (we can never hit the ceiling then).

Proposition 3.2. *The number of two-sided Catalan walks of height n from $(0,a)$ to (s,b) is given by*

$$\sum_{k\in\mathbb{Z}}\left(\binom{s}{\frac{s+b-a+2kn}{2}}-\binom{s}{\frac{s-b-a+2kn}{2}}\right)$$

or 0, if $s+b-a$ is an odd number.

This proposition can be proven using the standard reflection argument, in combination with the principle of inclusion-exclusion.

With this proposition, we are able to determine $b_{s,n}$ and $c_{s,n}$. We start with the latter.

Lemma 3.3. *The number of two-sided Catalan walks of height $2n+2$ from $(0,1)$ to (s,b) for all $0<b<2n+2$ is given by*

$$c_{s,n}=\sum_{k\in\mathbb{Z}}(-1)^k\binom{s}{\frac{s+2k(n+1)}{2}}$$

if s is an even number, and

$$c_{s,n}=\frac{1}{2}c_{s+1,n}$$

if s is an odd number.

Proof. We get, for even s,

$$\begin{aligned}c_{s,n}&=\sum_{m=0}^{n}\sum_{k\in\mathbb{Z}}\left(\binom{s}{\frac{s}{2}+m+2k(n+1)}-\binom{s}{\frac{s}{2}-m-1+2k(n+1)}\right)\\&=\sum_{k\in\mathbb{Z}}(-1)^k\binom{s}{\frac{s+2k(n+1)}{2}}.\end{aligned}$$

Most terms cancel by symmetry of the binomial coefficients. For odd s, we see that for each two-sided Catalan walk to $x=s$ we get two such walks to $x=s+1$. $\square$

Lemma 3.4. *The number of two-sided Catalan walks of height $2n+2$ from $(0,a)$ to (s,b) for all $0<a,b<2n+2$, a odd, is given by*

$$\begin{aligned}b_{s,n}&=2\sum_{l=0}^{n}\sum_{k\in\mathbb{N}}(-1)^k(n-2l)\binom{s}{\frac{s+1}{2}+l+k(n+1)}\\&=n2^s-2\sum_{l=0}^{n}2l\sum_{k\in\mathbb{N}}(-1)^k\binom{s}{\frac{s+1}{2}+l+k(n+1)}\\&=n2^s-4\ \beta_{s,n}\end{aligned}$$

if s is an odd number, and

$$b_{s,n}=2b_{s-1,n}=n2^s-8\ \beta_{s-1,n}$$

if s is an even number.

Proof. Assume s is an odd number. For all odd a but $n+1$, we get a term $\binom{s}{\frac{s+1}{2}}$. Hence, there are n such terms. Similarily, we get $n-2$ ($n-1$ positive and 1 negative) $\binom{s}{\frac{s+1}{2}+1}$ and $(n-4)\binom{s}{\frac{s+1}{2}+2}$, etc. to $(n-2n)\binom{s}{\frac{s+1}{2}+n}$. We then get $(n-2n)\binom{s}{\frac{s+1}{2}+n+1}$, $(n-2(n-1))\binom{s}{\frac{s+1}{2}+n+2}$, etc. Continuing in this fashion gives the first equality in the lemma. The leading 2 comes from symmetry, adding all paths going down.

For the second equality, we use that the row sums in Pascal's triangle are 2^n.

For even s, there are b_{s-1} paths to $x = s-1$. For each of these paths, there are two valid options (up or down) for the last step. $\square$

We have now proved the second part of our main theorem. What remains is the simplifications for $n \geq t$. Assuming this, we can simplify our formula using the following lemma.

Lemma 3.5.

$$\sum_{s=0}^{r}(-1)^s 2^{r-s}\binom{r}{s}\binom{s}{\lceil \frac{s}{2}\rceil} = C_r,$$

where C_r is the r:th Catalan number.

Proof. Consider vectors v of length $2r+1$, containing $r+1$ zeroes and r ones. The number $T(r,s)$ of such vectors that contain exactly $2s+1$ palindrome positions, i.e. positions i such that $v_i = v_{2r+2-i}$, can be found as follows. We concentrate on the first r positions. First choose which of these should be palindrome positions. Fill in the others arbitrarily. We then fill in the palindrome positions using $\lceil \frac{s}{2}\rceil$ zeroes and $\lfloor \frac{s}{2}\rfloor$ ones. All other positions can then be filled in so that the chosen palindrome positions really are palindrome positions and the other positions are not. It is easy to check that we get a valid palindrome vector, and that we do not miss any valid vectors. From this analysis, we find that

$$T(r,s) = 2^{r-s}\binom{r}{s}\binom{s}{\lceil \frac{s}{2}\rceil}.$$

It turns out that the element at position $r+1$ is 0 if s is even and 1 otherwise. If we remove this position, we get vectors of length $2r$ with r zeroes and r ones, for even s, and $r+1$ zeroes and $r-1$ ones for odd s. The number of such vectors are $\binom{2r}{r}$ and $\binom{2r}{r+1}$, respectively. We thus get

$$\sum_{s=0}^{r}(-1)^s T(r,s) = \binom{2r}{r} - \binom{2r}{r+1} = C_r.$$

$\square$

Now, for $n \geq t \geq r \geq s$, we get

$$g_{s,n} = b_{s,n}c_{s-1,n} = n2^{s-1}\binom{s-1}{\lceil \frac{s-1}{2}\rceil} - 2\binom{2\lfloor \frac{s}{2}\rfloor}{\lfloor \frac{s}{2}\rfloor}\sum_{l=0}^{\lfloor \frac{s-1}{2}\rfloor} l\binom{2\lceil \frac{s}{2}\rceil - 1}{\lceil \frac{s}{2}\rceil + l}.$$

This yields

$$E_{nt} = \sum_{r=0}^{t}\frac{(-1)^r}{n^r}\left[2^r C_r\binom{t}{r+1} + 2\binom{t}{r}\sum_{s=3}^{r}\binom{r-1}{s-1}(-1)^{s-1}4^{r-s}\binom{2\lfloor \frac{s}{2}\rfloor}{\lfloor \frac{s}{2}\rfloor}\sum_{l=0}^{\lfloor \frac{s-1}{2}\rfloor} l\binom{2\lceil \frac{s}{2}\rceil - 1}{\lceil \frac{s}{2}\rceil + l}\right].$$

4. An alternative formula

There is another way of writing E_{nt} that can be obtained using a similar model. We start with the same heat flow model, but instead of the three possible travel steps previously described, we merge two of them, giving these options:

- The packet changes vertex. It will then get multiplied with $x = \frac{1}{n}$.
- The packet does not change vertex. If it has not changed vertex before, nothing happens. Otherwise, it gets multiplied with $(1-4x)$.

We need no longer keep track of the true travel days (there will be no other travel days). We must, however, keep track of the first day (q) of travel. With this in mind, we easily find this expression valid:

$$E_{nt} = \frac{1}{2}\sum_{q=1}^{t}\sum_{r=0}^{t-q}\binom{t-q}{r}\left(1-\frac{4}{n}\right)^{t-q-r}\frac{1}{n^{r+1}}b_{r+1,n}c_{r,n}.$$

This gives the following theorem.

Theorem 4.1. *The expected number of inversions in a permutation in S_{n+1} after t random permutations is given by*

$$E_{nt} = \sum_{u=0}^{t-1}\left(\frac{n-4}{n}\right)^{u}\sum_{r=0}^{u}\binom{u}{r}\frac{1}{(n-4)^r}(2^r+\frac{2\beta_{r+1,n}}{n})c_{r,n}.$$

Proof. Trivial calculations give

$$\begin{aligned}
E_{nt} &= \frac{1}{2}\sum_{q=1}^{t}\sum_{r=0}^{t-q}\binom{t-q}{r}\left(1-\frac{4}{n}\right)^{t-q-r}\frac{1}{n^{r+1}}b_{r+1,n}c_{r,n}\\
&= \frac{1}{2}\sum_{u=0}^{t-1}\sum_{r=0}^{u}\binom{u}{r}\left(1-\frac{4}{n}\right)^{u-r}\frac{1}{n^{r+1}}b_{r+1,n}c_{r,n}\\
&= \sum_{u=0}^{t-1}\left(\frac{n-4}{n}\right)^{u}\sum_{r=0}^{u}\binom{u}{r}\frac{1}{(n-4)^r}(2^r+\frac{2\beta_{r+1,n}}{n})c_{r,n}.
\end{aligned}$$

□

This expression seems particularly useful for fixed n (try for instance $n = 4$). Also, it is easy to find out how much E_{nt} increases when we increase t one step. This is given by

$$\Delta_t E_{nt} = E_{n,t+1} - E_{nt} = \sum_{r=0}^{t}\binom{t}{r}\left(1-\frac{4}{n}\right)^{t-r}\frac{1}{n^{r+1}}b_{r+1,n}c_{r,n}.$$

It is easy to see that $\Delta_t E_{nt}$ is always positive for $n \geq 4$. This means that E_{nt} is monotonically increasing for almost all n. It should be pointed out that although this may seem trivial, for $n = 1$ (permutations of length 2), $E_{1,t}$ takes the values $0, 1, 0, 1, 0, 1\ldots$, which is not a monotone sequence.

To be able to apply this in the biological context, where we wish to estimate the number of transpositions given the inversion number of a permutation, we need this monotonicity property. The reason is that when we have found an expectation value E_{nt} which is close to our number of inversions, we must be sure that we will not find a better expectation value for a much larger t. If the sequence if monotone, this can never happen.

Acknowledgments

For help with the proof of Lemma 3.5, the author is indebted to Axel Hultman and Sloane's On-Line Encyclopedia of Integer Sequences.

References

[1] Henrik Eriksson, Kimmo Eriksson, Jonas Sjöstrand, Expected inversion number after k adjacent transpositions, in Formal Power Series and Algebraic Combinatorics, Krob, D., Mikhalev, A.A., Mikhalev, A.V. (Eds.), Springer Verlag (2000), 677–685.

[2] Li-San Wang, Exact-IEBP: A New Technique for Estimating Evolutionary Distances between Whole Genomes. Algorithms in Bioinformatics, Proceedings of WABI 2001, LNCS 2149, 175–188

Department of Mathematics, Royal Institute of Technology, S-100 44 Stockholm, Sweden
E-mail address: niklas@math.kth.se

A SIMPLE BIJECTION BETWEEN LECTURE HALL PARTITIONS AND PARTITIONS INTO ODD INTEGERS

NIKLAS ERIKSEN

Abstract. We give a simple bijection between lecture hall partitions, having at most n parts, and integer partitions, using only odd numbers less than or equal to $2n-1$. This reproves the lecture hall theorem of Bousquet-Mélou and Eriksson. We also use this bijection to find a recursion for the generating functions of the number of lecture hall partitions which use only the k back rows of a lecture hall of length n and some other special cases.

Résumé. Nous considérons une simple bijection entre les partitions d'un amphithéatre (lecture hall partitions), ayant au maximum n éléments, et les partitions d'entiers, utilisant seulement les nombres impairs inférieurs ou égaux a $2n-1$. Ceci redemontre le théorème de l'amphithéatre de Bousquet-Mélou et Eriksson. Nous utilisons également cette bijection pour trouver une récursivité dans les fonctions génératrices du nombre de partitions de l'amphithéatre qui utilisent seulement les k derniers rangs de l'amphithéatre de longueur n, et d'autres cas particuliers.

1. Introduction

Lecture hall partitions were presented for the first time by Bousquet-Mélou and Eriksson in 1997 [2]. Their definition is the following.

Definition 1.1. *A lecture hall partition into n parts is a partition $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ such that*

$$0 \leq \frac{\lambda_1}{1} \leq \frac{\lambda_2}{2} \leq \cdots \leq \frac{\lambda_n}{n}.$$

The name stems from the interpretation of the partition as a design for lecture halls (see Figure 1). The conditions imposed on the parts are sufficient to ensure that each student can see the teacher. We will henceforth use the name **rows** for the parts.

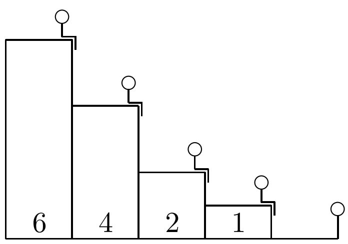


Figure 1. Students with teacher in a lecture hall. Thanks to the clever design, any sleeping student will be spotted by the teacher

There is a remarkable connection between lecture hall partitions into n parts and integer partitions with odd parts less than $2n$.

Supported by a grant from the Swedish Research Council.

Theorem 1.2 (Bousquet-Mélou and Eriksson [2]). *For fixed n, the generating function for the number of lecture hall partitions of N into n parts, $LH(N,n)$ is given by*

$$\sum_{N=0}^{\infty} LH(N,n)\; q^N = \prod_{i=1}^{n} \frac{1}{1-q^{2i-1}},$$

that is it equals the generating function of integer partitions into odd parts less than $2n$.

This theorem may be seen as a finite analogue of Euler's famous theorem stating that the number of partitions of k with odd parts equals the number of partitions of k with distinct parts; indeed, if we let n approach infinity, Theorem 1.2 becomes Euler's theorem.

Some notation: We will denote the set of lecture hall partitions into n parts $\mathcal{L}_n$ and the set of integer partitions into odd parts less than or equal to $2n-1$ will be denoted $\mathcal{O}_n$. We will prese= nt functions $\Phi_n : \mathcal{L}_n \to \mathcal{O}_n$ and $\Psi_n : \mathcal{O}_n \to \mathcal{L}_n$ such that $\Phi_n \circ \Psi_n = id_{\mathcal{O}_n}$ and $\Psi_n \circ \Phi_n = id_{\mathcal{L}_n}$.

Over the past years, several proofs of this theorem have been presented. The first paper [2] gave two proofs — one via Bott's formula for the Poincaré series of the affine Coxeter group $\widetilde{C}_n$ and one direct proof. The latter proof actually proved a refined version of the formula, keeping track of odd and even weights of the lecture hall partition as well.

In the sequel [3], Bousquet-Mélou and Eriksson introduced generalised lecture hall partitions as partitions $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ satisfying

$$0 \leq \frac{\lambda_1}{a_1} \leq \frac{\lambda_2}{a_2} \leq \ldots \leq \frac{\lambda_n}{a_n}$$

for some non-decreasing sequence of positive integers $a = (a_1, a_2, \ldots, a_n)$. The set of such sequences is denoted $\mathcal{L}_a$. Some results are derived for these partitions, but quite a few conjectures are also presented.

In a third paper [4], these authors provide another refinement of the theorem and two more proofs. The second one is claimed to be the first truely bijective proof of the lecture hall theorem. In fact, the proof is a bijection between lecture hall partitions with n parts and integer partitions with distinct parts between 1 and n and arbitrarily many parts between $n+1$ and $2n$. This is of course easily extendible to a bijection between lecture hall partitions and partitions with odd parts.

Apart from the original authors', there have been two contributions. The first one is by Andrews [1], who proves the theorem using MacMahon's 'partition analysis'. The latest contribution to this subject has been made by Yee [5], who also presents a combinatorial bijection.

With all these proofs, is there any *raison d'être* for another one? We claim, of course, that there is. The first reason is that the functions Φ_n and Ψ_n are, in all essentials, *independent* of n. For instance, the partition $\mu = (5,3,3)$ will give the lecture hall partition $\Psi_n(\mu) = (0, \ldots, 0, 4, 7)$ for any $n \geq 3$ (of course, the number of initial zeroes will differ). This contrasts with both previous bijections, although Yee's bijection has the property that for each μ there exists an N such that $\Psi_n(\mu)$ is independent of n if $n \geq N$.

Another reason to present this new proof is that the bijection gives new information on generalised lecture hall partitions. In fact, we are able to present a recursion for the generating function for the number = of lecture hall partitions of n with at least k empty front rows. We also present a recursion for the generating functions in the special case where the sequence $a = (a_1, a_2, \ldots, a_n)$ is increasing and $a_{n-2k} - a_{n-2k-1} = 1$ for $k \geq 0$.

We will start this paper by presenting the bijection in an intuitively clear way. We then proceed to state some technical definitions, which are followed by a proof that the map is bijective. Finally, we apply the bijection to the generalised lecture hall partitions as described above.

2. The bijection

We will describe the bijection by presenting the function Ψ_n that given a partition into odd integers will produce a lecture hall partition. This is done by building a lecture hall using components, which are determined by the odd number partition.

Here follows a short description of the steps involved.

- To each odd number n we associate a building block, B_n, consisting of n bricks. These are the basic parts of our lecture hall.
- The building blocks will be grouped into modules, M_k, according to rules defined below.
- Starting with the "smallest" module, we now add the modules, one at a time. Each time we add a module, we get a lecture hall partition, LH_k. When all modules have been added, we are done.

2.1. **The building blocks.** To each odd number n, we associate a building block of n bricks, B_n. First, put $\frac{n+1}{2}$ bricks at positions $(i, \frac{n+1}{2} - i + 1), 1 \leq i \leq \frac{n+1}{2}$, and then the remaining $\frac{n-1}{2}$ bricks at positions $(i, \frac{n-1}{2} - i + 1), 1 \leq i \leq \frac{n-1}{2}$. We thus obtain building blocks as in Figure 2.

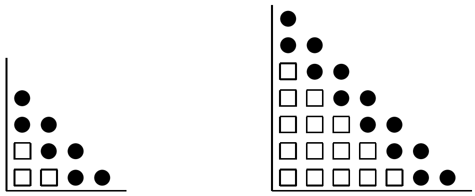


Figure 2. The building blocks B_7 and B_{13}. The dots are bricks and the boxes are empty spaces

2.2. **The modules.** First, we need to explain which building blocks go into which module. We will then show how to build the modules.

Let O be a partition into odd parts and let $2n-1$ be the largest part of O. We associate with O a matrix $A = (a_{i,j})_{i,j}$ with n rows and $\lceil n/2 \rceil$ columns. The top right half of this matrix is easy to fill: $a_{i,j} = 0$ if $2j > i + 1$. We shall describe below how to construct the other half. In the resulting matrix, the sum of the entries in the ith row will be the number of parts of O equal to $2(n-i)+1$. For $1 \leq k \leq \lceil n/2 \rceil$, we shall define the module M_k as the sum of the blocks encoded by the kth column of A: more precisely, $M_k = \sum_i a_{i,k} B_{2(n-i)+1}$.

We define the **sequence** $\operatorname{seq}(M_k)$ of a module M_k as $\{a_{i+2(k-1),k}\}_{i=1}^{n-2(k-1)}$. Order the parts of O in decreasing order. A part $2l-1$ should go into row $n-l+1$. We do this by increasing one element in the designated row of the first part in the decreasing order. In doing this, we should choose the rightmost column such that the sequences $\operatorname{seq}(M_k)$ are always lexocographically ordered. Then we remove the first element and iterate. An example will clarify this explanation.

Example 2.1. *The partition* $(17, 11, 11, 9, 7, 3, 3, 1)$ *corresponds to the building blocks* $\{B_{17}, B_{11}, B_{11}, B_9, B_7, B_3, B_3, B_1\}$*, which we will now divide into modules.* B_{17} *should go into row* $n - l + 1 = 9 - 9 + 1 = 1$ *and column* 1*, since it is the leftmost column allowed. Next, we have* B_{11}*, which goes into row* $9 - 6 + 1 = 4$*. We may add one to the element in column*

2, since we then get seq(M_2) $= (0,1,0\ldots) \leq (1,0,\ldots) =$ seq(M_1). Column 3, however, is forbidden, so column 2 is chosen. The second B_{11} goes into the same position.

As for the next block, B_9, putting it in column 3 would give seq(M_3) $>$ seq(M_2), but column 2 will do. We then continue in this fashion until we get the matrix in Figure 3. From this we conclude that $M_1 = B_{17}$, $M_2 = B_{11} + B_{11} + B_9$, $M_3 = B_7 + B_3 + B_1$ and $M_4 = B_3$.

$$
\begin{array}{r c}
 & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\
\begin{matrix} 17 \\ 15 \\ 13 \\ 11 \\ 9 \\ 7 \\ 5 \\ 3 \\ 1 \end{matrix} &
\begin{bmatrix}
1 & & & & \\
0 & & & & \\
0 & 0 & & & \\
0 & 2 & & & \\
0 & 1 & 0 & & \\
0 & 0 & 1 & & \\
0 & 0 & 0 & 0 & \\
0 & 0 & 1 & 1 & \\
0 & 0 & 1 & 0 & 0
\end{bmatrix}
\end{array}
$$

FIGURE 3. Putting each building block into the right module. Observe that the sequences are in lexicographic order.

Given the building blocks of a module, assembling the modules is easy. Just put the building block on top of each other. The empty spaces are heavier than the bricks and will sink to the bottom. This is shown in Figure 4. Since this operation has all the nice properties like associativity, commutativity and such, we will use the + sign for this operation.

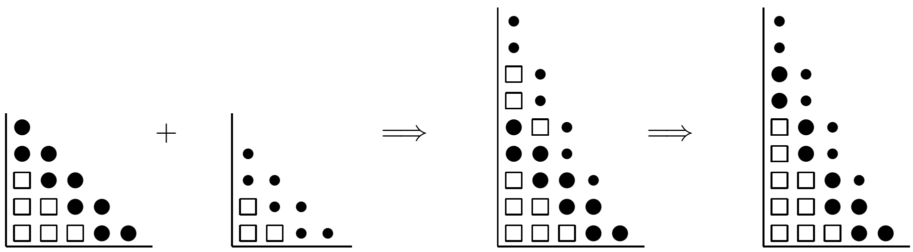


FIGURE 4. Adding B_9 and B_7

2.3. **The lecture hall partitions.** We assemble the modules, one by one, starting with the module with the highest index, k. The lecture hall partitions obtained from modules M_i to M_k, $i \leq k$, will be named LH_i. This will be denoted $LH_i = LH_{i+1} \oplus M_i$.

Given LH_{i+1} and M_i, we construct LH_i as follows. First insert LH_{i+1} into M_i from below. If any bricks in LH_{i+1} collide with bricks in M_i, the bricks in M_i will slide up. Then we push from the right, until there are no holes in the lecture hall partition.

Example 2.2. *An example of such an assembly can be viewed in Figure 5. The spaces in the module are no longer drawn. We are given the set $\{B_{11}, B_5, B_5, B_1\}$ of building blocks and get the modules $M_1 = B_{11}, M_2 = B_5 + B_5$ and $M_3 = B_1$. We get $LH_3 = \{0,0,0,0,0,0\} \oplus M_3 = \{0,0,0,0,0,1\}$ and $LH_=2 = LH_3 \oplus M_2 = \{0,0,0,1,4,6\}$. These assemblies were quite trivial, but here comes the tricky part, which can be viewed in the figure. We start with LH_2 (black) and $M_1 = B_{11}$ (white) (a). LH_2 is pushed into M_1 (b) and then we compress from the right to obtain (c).*

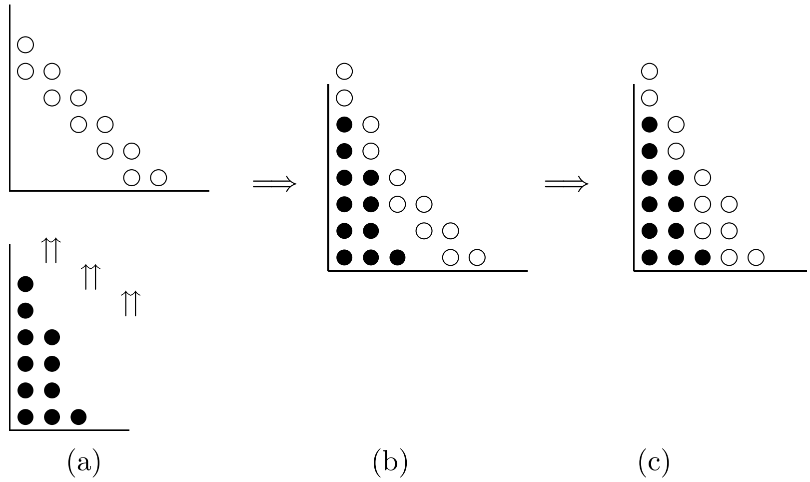


FIGURE 5. Assembling $M_1 = B_{11}$ (white) with $LH_2 = (B_5 + B_5) \oplus B_1$ (black).

3. Mathematical tools for lecture hall partitions

Some useful measures on the building blocks, as well as on the modules and the lecture halls, are the height, the ceiling and the range.

Definition 3.1. *The* **height** *of an object (building block, module, lecture hall) at a certain row, is the y-coordinate of the highest brick in that row. The* **ceiling** *of an object at a certain row is one less than the y-coordinate of the lowest brick in that row. If the row contains no bricks, both maps are zero. These are written as $h(A, k)$ and $c(A, k)$, respectively, for an object A and row k.*

Observation 3.2. *We observe that*

$$h(B_n, k) = \begin{cases} \frac{n+3}{2} - k, & 1 \le k \le \frac{n+3}{2}; \\ 0, & \textit{otherwise} \end{cases}$$

and

$$c(B_n, k) = \begin{cases} h(B_n, k) - 2, & 1 \le k \le \frac{n-1}{2}; \\ 0, & \textit{otherwise} \end{cases}$$

Definition 3.3. *The* **range** *of an object A and a row k is given by*

$$r(A, k) = \frac{h(A, k)}{h(A, k) - h(A, k+1)} + k,$$

provided that $h(A, k) > 0$, if A is a building block, and $h(A, k+1) > 0$ otherwise. For larger k, $h(A, k) = 0$. The range of an object A is the maximal range over all rows.

The range of a lecture hall partition tells at which position the teacher must be placed if the partition should be a valid lecture hall partition (the teacher must stand to the right of the range).

Example 3.4. *In Figure 5, the ranges are $r(LH, 1) = 5, r(LH, 2) = 5$, $r(LH, 3) = 7, r(LH, 4) = 5.5$ and $r(LH, 5) == 0$. We then get $r(LH) = 7$. The teacher must be placed at $x = 7$ to be visible to all students.*

For modules, we can easily calculate the height and the ceiling:

$$h(M,k) = \sum_{j=1}^{s} a_j h(B_{n_j},k)$$

and

$$c(M,k) = \sum_{j=1}^{s} a_j c(B_{n_j},k)$$

for all rows k.

We find that the range of a row in a module is the arithmetical mean of the ranges of the building blocks in the module, except for possibly some of the smallest ones.

Lemma 3.5. *Consider a module*

$$M = a_1 B_{n_1} + a_2 B_{n_2} + \ldots + a_s B_{n_s},$$

where $n_1 > n_2 < \ldots n_s$*, and let* $A_k = \{i : n_i \geq k\}$*. Then, for* $k < \frac{n_1+1}{2}$*, the range is given by*

$$r(M,k) = \frac{\sum\limits_{A_{2k-1}} a_i r(B_{n_i},k)}{\sum\limits_{A_{2k-1}} a_i}.$$

Proof. This follows from Observation 3.2 and the definition of range. We have

$$r(M,k) = \frac{h(M,k)}{h(M,k) - h(M,k+1)} + k$$

$$= \frac{\sum\limits_{A_{2k-3}} a_i \left(\frac{n_i+3}{2} - k\right)}{\sum\limits_{A_{2k-3}} a_i \left(\frac{n_i+3}{2} - k\right) - \sum\limits_{A_{2k-1}} a_i \left(\frac{n_i+3}{2} - k - 1\right)} + k$$

$$= \frac{\sum\limits_{A_{2k-1}} a_i \left(\frac{n_i+3}{2} - k\right)}{\sum\limits_{A_{2k-1}} a_i} + k$$

$$= \frac{\sum\limits_{A_{2k-1}} a_i \left(\frac{n_i+3}{2}\right)}{\sum\limits_{A_{2k-1}} a_i} = \frac{\sum\limits_{A_{2k-1}} a_i r(B_{n_i},k)}{\sum\limits_{A_{2k-1}} = a_i}.$$

□

We will call a module M_k **simple** if seq(M_k) contains no non-zero elements that are not adjacent. Otherwise, the module is **complex**. In Example 2.1, M_1, M_2 and M_4 are simple, but M_3 is complex.

Lemma 3.6. *For each lecture hall partition* LH *and row* k *such that the height* $h(LH,k)$ *is positive, there exists a unique set of numbers* $a > 0$*,* $b > 0$*,* $n = 2m+1 \geq 1$ *such that height of the module* $M = aB_n + bB_{n-2}$ *equals the height of* LH *at rows* k *and* $k+1$*. For a fixed* $n = 2m+1 \geq 1$*, if there exists number* $a \geq 0$*,* $b > 0$*, these are unique as well.*

Proof. We have three unknowns but only two equations:

$$a\left(\frac{n+3}{2}-k\right)+b\left(\frac{n+1}{2}-k\right)=h(LH,k)$$

and

$$a\left(\frac{n+3}{2}-k-1\right)+b\left(\frac{n+1}{2}-k-1\right)=h(LH,k+1).$$

This can, however, be rewritten into

$$a+b=h(LH,k)-h(LH,k+1)$$

and

$$(h(LH,k)-h(LH,k+1))\left(\frac{n+3}{2}-k\right)-b=h(LH,k).$$

We may vary b freely between 0 and $h(LH,k)-h(LH,k+1)-1$, so we see that there is exactly one solution to the last equation. This also gives a uniquely.

For fixed n, it is still clear that if such numbers exist, they are unique. □

This paves the way for the following definition.

Definition 3.7. *Given a lecture hall partition LH and a row k, we define the* **characteristic triple** (a,b,n), *which is a set of numbers such that $a>0, b\geq 0, n=2m+1\geq 1$ and such that the height of the module $M=aB_n+bB_{n-2}$ equals the heights of LH at rows k and $k+1$. The* **relaxed characteristic triple** (a,b,n) *is defined similarily, except for that we allow $a=0$.*

Definition 3.8. *We introduce the* **step operator** $^+$ *that operates on characteristic triples. We have $(a,b,n)^+=(a+1,b-1,n)$ if $b\geq 1$ and $(a,0,n)^+=(1,a-1,n+2)$. If a row k has characteristic triple (a,b,n), applying the step operator corresponds to increasing the height of rows k and $k+1$ by one.*

Definition 3.9. *The* **top range row** *of a lecture hall partition LH is the rightmost row k such that its characteristic triple (a,b,n) has $n=2\lceil r(LH)\rceil-3$, maximal a, and maximal b, given the value of a.*

Example 3.10. *The top range row of LH in Figure 5 is of course 3, since it has the highest range. The characteristic triple associated with row 3 is* $(1,0,11)$.

Observation 3.11. *Assume that we are constructing a lecture hall partition using the procedure described above. The height of the rows in LH_i is then given as follows.*

- *If $h(LH_{i+1},k)>c(M_i,k)$, then $h(LH_i,k)=h(M_i,k)-c(M_i,k)+h(LH_{i+1},k)$.*
- *If $h(LH_{i+1},k)\leq c(M_i,k)$ and either $k\in\{1,2\}$ or $h(LH_{i+1},k-j)>c(M_i,k-j)$ for $j\in\{1,2\}$, then $h(LH_i,k)=h(M_i,k)$.*
- *Otherwise, $h(LH_i,k)=h(LH_{i+1},k-2)$.*

Definition 3.12. *If $h(LH_{i+1},k)>c(M_i,k)$ for any row k, we say that the M_i is* **disturbed**.

Remark 3.13. *Each time we add a module, we increase the number of rows by two, except for the case with only B_1s in the module. This fact will be used in the last section to calculate the number of lecture hall partitions with the k front rows empty.*

4. Getting back

We have described how to generate a well-defined partition from a partition into odd integers. What remains is to verify that we indeed obtain a lecture hall partition, and that two different partitions of odd integers do not produce the same lecture hall partition. The first question will be answered implicitly, since we will show that there is a limit on the range of the partition obtained, thus implying that the partition really fulfills the conditions imposed on a lecture hall partition. The other will be adressed by producing a way to obtain the partition into odd numbers from any lecture hall partition. Then the LHP generator is clearly bijective.

We will show that given any partition $LH_1 = \Psi_n(O)$, we can reveal the contents of M_1. We can then remove M_1 and iterate, to find the contents of all modules. This will give the odd part partition $O = \Phi_n(LH_1)$ and also show that LH_1 is indeed a lecture hall partition, since the range is limited from above by $n + 1$.

We must, at this stage, warn sensitive readers that the next page contains some ugly mathematics. The beauty of this bijection lies in its simple construction, not in the ease with which we show that it is bijective.

Theorem 4.1. *Given a partition $LH_1 = \Psi_n(O)$, where $O \in \mathcal{O}_n$, we are able to read off M_1 and remove it. The module M_1 is read as follows. First, find the top range row k and let (a, b, n) be its characteristic triple. Set $M_1 = aB_n + bB_{n-2}$. We then continue with row $k - 2l$, for $l = 1, 2, \ldots$, in that order. In each step, we determine the smallest index p of any building block in M_1. Then, if there exists a relaxed characteristic triple $(a_l, b_l, p - 2)$ or (a_l, b_l, p) on row $k - 2l$ using heights $h(LH, k - 2l) - h(M_1, k - 2l)$ and $h(LH, k - 2l + 1) - h(M_1, k - 2l + 1)$, then we add $a_l B_{p-2} + b_l B_{p-4}$ (or $a_l B_p + b_l B_{p-2}$, respectively), to M_1 and continue. Otherwise, we are done and we can remove M_1. As a special rule when we get to the leftmost rows, we should note that we will never accept the relaxed characteristic triple $(0, s, 3)$ at row 0 and we also note that if $LH = (0, \ldots, 0, m)$, this corresponds to the module mB_1.*

Proof. This theorem will be shown by induction. We will always assume that M_2 can be read from LH_2 as described, and look at how LH_1 appears. It should be noted that since $\mathrm{seq}(M_1) > \mathrm{seq}(M_2)$, $r(LH_1) \geq r(LH_2) + 2$.

It is easy to see that the first module added will be a lecture hall partition with only one well-defined range. We then find the characteristic triple and are done.

Now assume that LH_2 is non-empty. We will, in turn, cover the four possible cases: M_1 is simple or complex, disturbed or not disturbed.

- M_1 **is simple and undisturbed:** It is trivial to see that $h(LH_1, k) = h(M_1, k)$ for $k = 1, 2$ and $h(LH_1, k) = h(LH_2, k - 2)$ otherwise. Since M_1 is simple, we will find the top range at row 1, and the characteristic triple will tell the tale on M_1.
- M_1 **is complex and undisturbed:** Again, the heights are distributed as in the simple case. However, the top range is no longer found at row 1. Rather, if the top range of LH_2 is found at row $k - 2$, we now find it at row k. Since we got valid characteristic triples for LH_2, we also get them for LH_1. Finally, we reach $k = 1$ or $k = 0$, and we then read the same heights as in M_1. These are also valid and we get the right M_1.
- M_1 **is simple and disturbed:** We assume that the first k' rows of M_1 are disturbed (there can not be any other disturbed rows, since this would give too large range for LH_2). We then have $h(LH_1, k) = h(LH_2, k) + h(M_1, k) - c(M_1, k) = h(LH_2, k) + 2(a + b)$ for $k \leq k'$, $h(LH_1, k) = h(M_1, k)$ for $k = k' + 1, k' + 2$, and $h(LH_1, k) = h(LH_2, k - 2)$ otherwise. We need to show that the ranges of rows $k < k'$ are to

small to affect our construction. We start by showing that the top range is given by row $k'+1$ or $k'+2$.

It is clear that the top range row can not be found among the rows k such that $k > k'+2$. We must show that it can not be found among the first k' rows either. Let the characteristic triple of row $k'+1$ be (a, b, n). We then know that the characteristic triple of a row $k < k'$ in LH_2 either has the form $(a', b', n-l)$ for $l > 4$ or $(a', b', n-4)$ with $a' < a$ (otherwise, the modules would not be sorted alphabetically). We also know that $a'+b' > a+b$ (or row k would not be disturbed). Now, the corresponding row in LH_1 has height that is $2(a+b)$ larger. We apply the step operator $2(a+b)$ times to the characteristic triple. Since $a'+b' > a+b$, we never get a characteristic triple that is greater than (a, b, n).

We have shown that the top range can be found only on rows $k'+1$ and $k'+2$. We must now show that we can not continue reading another characteristic triple. We assume that the top range row is $k'+1$ and that the characteristic triple of $k'-1$ in LH_2 is $(a', b', n-4)$, with $a' < a$ and $d = a'+b'-(a+b) > 0$. If we apply the step operator $2(a+b)$ times, we get the characteristic triple $(a'-2d, b'+2d, n)$ or, if d is sufficiently high, $(a'-2d+a'+b', b'+2d-a'-b', n-2)$ or $(a'+2(a+b), b'-2(a+b), n-4)$. From these triples, we must break loose a triple (a, b, n). If we do this to the first, we get $(a'-2d, b'+2d, n) = (a'-2d, b+2d+a-a', n) + (d, 0, n-2) = (a, b, n) + (d-(a-a'), a-a', n-6)$ (in the last equality, we have increased the first term by $2d + (a-a')$ and reduced the second term by an equal amount). We see that the characteristic triple $(d-(a-a'), a-a', n-6)$ is not valid, since $a-a' > 0$. Similar calculations show that the other cases do not produce valid triples either.

Had the top range been row $k'+2$, we should have look at the characteristic triple of k' instead. But then we still have that the heights of rows k' and $k'+1$ in LH_1 are given by the corresponding height in LH_2 increased by $2(a+b)$, and the result follows from the analysis above.

M_1 **is complex and disturbed:** As for the undisturbed case, we are able to read off all valid building blocks. The only case where we can go wrong is when we enter the disturbed zone. However, since we then have removed (in thought) from the heights all building blocks that belong to M_1, we have a case similar to the simple case. From that analysis, we find that we can not continue to read valid characteristic triples.

□

Example 4.2. *Let us exemplify this by looking at the lecture hall partition* $(0, 1, 3, 4, 6, 8)$ *found in Figure 5. According to the theorem, we should start by finding the top range row and its characteristic triple. We lready now that this is row* 3 *and triple* $(1, 0, 11)$*. We thus set* $M_1 = B_{11}$*. We then look at row* $3-2 = 1$*. We should use the heights* $h(LH, 1) - h(M_1, 1) = 8 - 6 = 2$ *and* $h(LH, 2) - h(M_1, 2) = 6 - 5 = 1$*. The corresponding characteristic triple would be* $(1, 0, 3)$ *or, if we allow the relaxed version,* $(0, 1, 5)$*. In neither case, the last coordinate is* 11 *or* $11 - 2 = 9$*, which we need in order to add more building blocks to* M_1*. Since we found something that we wished to add, but could not,* M_1 *will not contain any more building blocks, and we can remove* M_1 *from* LH*.*

We now get $LH_2 = (0, 0, 0, 1, 4, 6)$*. It is easy to see that the top range row is* 1 *and that the corresponding chracteristic triple is* $(2, 0, 5)$*. We then set* $M_2 = B_5 + B_5$*. Since we are already at row* 1*, we can not continue to the left and are done. Removing* M_2 *will give* $LH_3 = (0, 0, 0, 0, 0, 1)$*, and it is not hard to see that this gives* $M_3 = B_1$*. From this we conclude that* $\Phi_n(0, 1, 3, 4, 6, 8) = (11, 5, 5, 1)$ *for* $n \geq 6$*.*

5. The distant teacher and other generalisations

We now turn to generalised lecture hall partitions. For starters, we take a closer look at real world lecture halls. Usually, there is some distance between the teacher and the students. The following theorem will give the generating function for this case. The function will be defined recursively.

Theorem 5.1. *Let $P_{LH}(N, n, k)$ be the number of ways to partition N into $\lambda = [\lambda_1, \ldots, \lambda_k)$ such that we have*

$$0 \leq \frac{\lambda_1}{n-k+1} \leq \frac{\lambda_2}{n-k+2} \leq \ldots \leq \frac{\lambda_k}{n}.$$

Then the generating function is

$$\sum_N P_{LH}(N, n, k)\ q^N = Q(n, k)$$

where $Q(n, k)$ is given recursively by

$$Q(n, k) = Q(n-1, k) + \frac{q^{2n-1}}{(1-q^{2n-1})(1-q^{2n-3})} Q(n-2, k-2), \quad \text{for } 2 \leq k < n,$$

$$Q(n, 0) = 1,\ Q(n, 1) = \frac{1}{1-q},\ Q(n, n) = \frac{1}{\prod_{i=1}^{n}(1-q^{2i-1})}.$$

Proof. First look a the boundary conditions. For $k = n$ we have the lecture hall theorem and for $k = 1$ and $k = 0$, the results are trivial.

Let us now take a closer look at the proof of the bijection above. For each module we create, we will use two more rows in the lecture hall (unless the module contains only ones). Thus, if we wish to partition λ, we can either use the number $2n-1$, thereby creating a new module, or not use the number $2n-1$. In the first case, we add the factor $\frac{q^{2n-1}}{(1-q^{2n-1})(1-q^{2n-3})}$ to acknowledge the fact that $2n-1$ is used at least once and that $2n-3$ may be used freely, and to this we multiply $Q(n-2, k-2)$ for the rest of the modules, which can only use number strictly less then $2n-3$ and may only use the remaining $k-2$ rows. On the other hand, not using $2n-1$ will not reduce the available number of rows. □

Using the same line of thinking, the following theorem follows naturally.

Theorem 5.2. *Let a be an increasing sequence of n positive integers such that $a_{n-2i} - a_{n-2i-1} = 1, 0 \leq i < \lfloor \frac{n}{2} \rfloor$ and $P_{LH}(N, a)$ be the number of ways to partition N into $\lambda = (\lambda_1, \ldots, \lambda_n)$ such that we have*

$$0 \leq \frac{\lambda_1}{a_1} \leq \ldots \leq \frac{\lambda_n}{a_n}.$$

Then the generating function is

$$\sum_N P_{LH}(N, a)\ q^N = Q(a_n, a)$$

where $Q(m, a)$ is given recursively by

$$Q(m, a) = Q(m-1, a) + \frac{q^{2m-1} Q(m-(a_n - a_{n-2}), (a_1, \ldots, a_{n-2}))}{(1-q^{2m-1})(1-q^{2m-3})}, \quad \text{for } n > 1,\ m > 1,$$

$$Q(m, ()) = 1,\ Q(m, (a_1)) = \frac{1}{1-q},$$

$$Q(1, a) = \frac{1}{1-q},\ Q(2, a) = \frac{1}{(1-q)(1-q^3)}.$$

It should be noted that the building blocks and the way the modules are put together change somewhat in this case. An example will clarify this better than any formal definition.

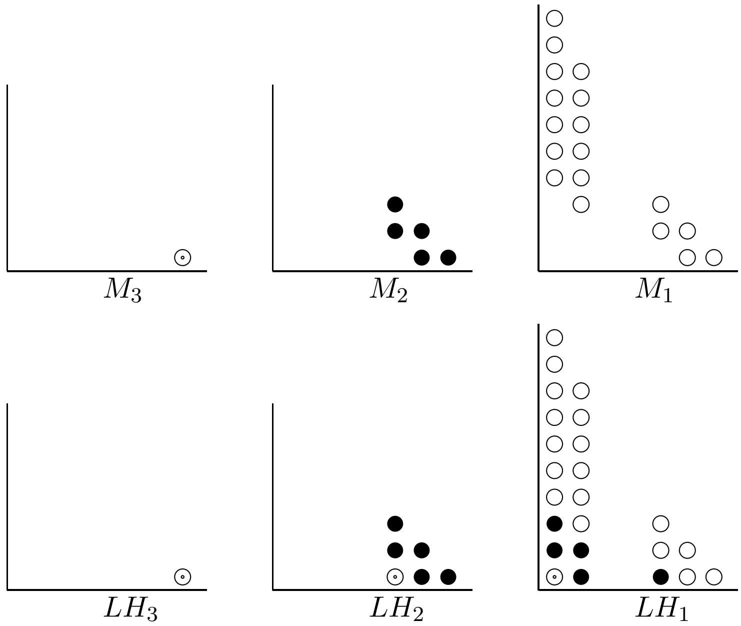


FIGURE 6. We have marked the bricks from M_3 with a dot and the bricks from M_2 are marked black, to make it easier to follow what is happening. We see that $M_3 = B_1$ and $M_2 = B_5$ have their usual appearances, although they are shifted somewhat to the right, but $M_1 = B_{13} + B_5$ looks differently from the standard appearance, since some rows are kept empty.

Example 5.3. *We wish to put* $\{13, 5, 5, 1\}$ *in an* $(1, 2, 3, 6, 7)$*-lecture hall partition. This gives the matrix*

$$\begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 13 \\ 11 \\ 9 \\ 7 \\ 5 \\ 3 \\ 1 \end{array} & \left[\begin{array}{ccc} 1 & & \\ 0 & & \\ 0 & & \\ 0 & & \\ 1 & 1 & \\ 0 & 0 & \\ 0 & 0 & 1 \end{array}\right] \end{array}$$

The second module now starts 4 *steps lower than the first module in the matrix. The reason is that* $a_n - a_{n-2} = a_5 - a_3 = 7 - 3 = 4$ *and not* 2*, as we get in the standard case. However, we still demand that the sequences are ordered lexicographically.*

We get $M_1 = B_{13} + B_5, M_2 = B_5$ *and* $M_3 = B_1$*. The modules and the lecture hall partitions can be found in Figure 6. In essence, we may say that the we get the same modules as before, but those bricks that occupy rows that should be empty are moved to the closest rows to the left. By this procedure, the building blocks will not occupy any rows that must be empty. We also build* LH_k *at row* a_{n-2k} *instead of row* a_n*. It is then slided to the left before we add the next module.*

Discussion

In [3], Bousquet-Mélou and Eriksson conjecture that all sequences $a = (a_1, \dots, a_n)$ that give generating functions of the form

$$\frac{1}{(1-q^{e_1})(1-q^{e_2})\cdots(1-q^{e_n})}$$

have $a_1|a_i$ for all a_i. The results obtained in the previous section allows for calculations that could falsify their conjecture, but so far, no counterexample has been found. In the same article, several other conjectures were made, and we intend to take a closer look at them to see if we can use this new bijection to verify or falsify them.

Acknowledgments

I am deeply indebted to Jakob von Döbeln, who not only introduced me to Lecture Hall partitions, but also worked with me through the first attempts to formulate the bijection a couple of years ago. I also wish to thank my supervisor Kimmo Eriksson for all his support. Finally, an anonymous referee has put down an enormous effort on this paper to clear out ambiguities and make it a lot easier to read. Thank you!

References

[1] Andrews, George E., MacMahon's partitions analysis. I. The Lecture hall partition theorem, *in* Mathematical Essays in Honor of G.-C. Rota.", 1–22, Birkh=E4user, Cambridge, MA, 1998.
[2] Bousquet-Mélou, Mireille; Eriksson, Kimmo, Lecture hall partitions, *Ramanujan J.* **1** (1997), 101–111.
[3] Bousquet-Mélou, Mireille; Eriksson, Kimmo, Lecture hall partitions II, *Ramanujan J.* **1** (1997), 165–185.
[4] Bousquet-Mélou, Mireille; Eriksson, Kimmo, A refinement of the Lecture hall theorem, *J. Combinatorial Theory A* **86** (1999), 63–84.
[5] Yee, Ae Ja, On combinatorics of Lecture hall Partitions, *Ramanujan J.* **5** (2001), 247–262.

Department of Mathematics, Royal Institute of Technology, S-100 44 Stockholm, Sweden
E-mail address: `niklas@math.kth.se`

DENSE PACKING OF PATTERNS IN A PERMUTATION

HENRIK ERIKSSON, KIMMO ERIKSSON, SVANTE LINUSSON, AND JOHAN WÄSTLUND

Abstract. We study the problem of finding the shortest permutation containing all patterns of a given length k. An upper bound of $3k^2/4$ is established. We also prove that as $k \to \infty$, there are permutations of length $k^2/4 + o(k^2)$ containing almost all patterns of length k.

Résumé. Nous étudions le problème de trouver la permutation la plus courte contenant tous les motifs d'une longueur donnée k. Une borne supérieure de $3k^2/4$ est établie. Nous montrons aussi qu'il y a des permutations de longueur $k^2/4 + o(k^2)$ contenant presque tous les motifs de longueur k.

This paper was born at FPSAC'01 in Arizona, where Richard Stanley presented the mathematical work of the late Rodica Simion. Among many other things, she was a pioneer in the field of pattern avoiding permutations. In his talk, Stanley mentioned that Rodica Simion and Frank Schmidt [2] gave a formula for the number of permutations of length n that do not avoid *any* pattern of length three. There is little hope of finding a similar formula for arbitrary pattern length k, considering how complex the theory of pattern avoiding permutations becomes for larger k. But listening to Stanley, the four present authors became curious about a related question:

The pattern packing problem: What is the length L_k of the shortest permutation containing all patterns of length k?

We later found that this pattern packing problem was posed already in 1999 by Arratia [1], but only trivial bounds on L_k have been given so far. The pattern packing problem is of course reminiscent of other dense packing problems, such as that of finding a shortest bit sequence containing every binary k-word as a contiguous subsequence. For the latter problem there is a well-known solution, the so called *de Bruijn sequences*, which contain every k-word exactly once. We cannot hope for such an efficient solution to the pattern packing problem. For instance, the shortest possible permutation containing all patterns of length $k = 2$ is evidently of length $L_2 = 3$ (there are four of them: 132, 213, 231, 312). But such a permutation contains $\binom{3}{2} = 3$ subsequences of length 2, of which at most two can have different patterns.

In spite of our efforts, the problem is still unsolved. We offer the conjecture that $L_k \sim k^2/2$ asymptotically.

Our story is one of partial results, to be developed in nine sections as follows.

(1) The minimal length L_k of a permutation containing all k-patterns lies between k^2/e^2 and k^2.
(2) We represent patterns by dot configurations on a square grid and we characterize compact representations in terms of ascents and inverse descents.
(3) The pedestrian game is introduced to describe equivalent representations of a pattern.
(4) Strong convergence of the game implies a unique terminal position.
(5) The terminal configuration is in fact the compact representation of the pattern.

(6) By a probabilistic argument, we show that any dot configuration can be played onto white squares of a k by $3k/2$ chessboard. This improves the upper bound to $L_k \leq 3k^2/4$, our main result.
(7) *A variation*: If we relax our ambition to finding a permutation containing *almost all* k-patterns, we show that a length of $k^2/4 + o(k^2)$ suffices.
(8) *A second variation*: If we are satisfied with a permutation containing each k-pattern *or its inverse*, then a length of $k^2/2$ is sufficient.
(9) *Final twist*: Returning to the terminal positions of the pedestrian game, a simple counting argument gives a new proof of an identity of Carlitz.

1. Elementary bounds on L_k

In this section, we shall see that L_k is asymptotically proportional to k^2 and obtain the elementary bounds

$$\frac{k^2}{e^2} \leq L_k \leq k^2.$$

The lower bound is derived from the observation that the total number of k-subsequences in the permutation must be at least as big as the number of all possible k-patterns. In other words,

$$\binom{L_k}{k} \geq k!,$$

or equivalently,

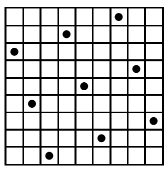
$$L_k(L_k - 1) \cdots (L_k - k + 1) \geq (k!)^2. \tag{1}$$

The left-hand side is less than L_k^k. By Stirling's formula, the right-hand side is approximately $2\pi k(k/e)^{2k}$. Hence we obtain a lower bound

$$\frac{k^2}{e^2} < L_k.$$

This bound is mentioned in [1], where it is also conjectured to be sharp (contradicting our own conjecture).

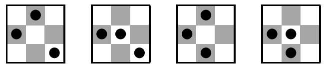
The simple upper bound of $L_k \leq k^2$ is obtained as follows. Define a k^2-permutation by arranging the numbers between 1 and k^2 into k sequences of length k, each decreasing by k in every step, the sequences taken in increasing order. For example, for $k = 3$ the permutation of length 9 is 741852963. The dot matrix of such a permutation (with values increasing upwards) will look like a slightly *tilted square*:



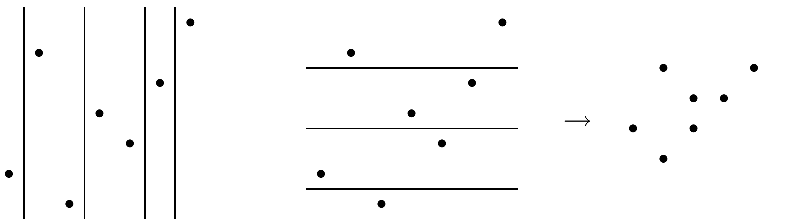
Clearly every k-pattern can be found in this permutation; simply take one dot in each row and column of the tilted square. For example, the pattern 231 is found in the subsequence 483.

2. More compact packings

Since the rows in the tilted square actually correspond to increasing subsequences, while columns correspond to decreasing subsequences, it is possible to realize a pattern by a more compact subset of dots in the tilted square. For example, the permutation pattern 231 is realized by all the following dot patterns in the tilted square (and a few others).



There is a direct way of figuring out the most compact such dot representation of a pattern: Start with the dot matrix of the pattern. Draw a vertical line at every ascent, and a horizontal line at every descent in the inverse permutation (obtained by reflection in the line $y = x$). Adjust the dots so that they lie straight in the rows and columns induced by the lines. For example, the compact representation of `2614357` is obtained as follows:



vertical lines horizontal lines compact dots

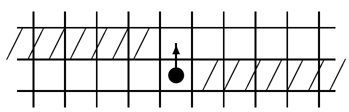
A direct description of the compact position of a given permutation can be given in terms of descents of the permutation and its inverse. We write the permutation π as $(\pi_1\ \pi_2\ \pi_3\ \ldots\ \pi_k)$.

A pair π_i, π_{i+1} of adjacent symbols is a *descent* if $\pi_i > \pi_{i+1}$, otherwise it is an *ascent*. We let $d(\pi)$ be the number of descents in π, and we let $a(\pi)$ be the number of ascents. A descent in the inverse π^{-1} is a pair of symbols $x, x{+}1$ such that $x{+}1$ comes before (but not necessarily adjacent to) x in the permutation π. Such a pair is therefore called an *inverse descent*.

For each symbol π_i in π, we count the ascents *before* π_i, that is, pairs π_j, π_{j+1} with $j < i$ and $\pi_j < \pi_{j+1}$. We also count the inverse descents *below* π_i, by which we mean pairs of symbols $x, x{+}1$ such that $x < \pi_i$ and $x{+}1$ comes before x. The compact dot configuration in the tilted square is constructed by putting, for each symbol π_i in π, a dot in the site that has as many columns east of it as there are ascents before π_i, and as many rows south of it as there are inverse descents below π_i.

3. The Melbourne pedestrian game

We will now introduce a game for transforming dot patterns. The setting is k pedestrians walking on a square grid in Melbourne. In Australian traffic, as a pedestrian you are supposed to look for any traffic to your right before you cross a street, and also watch out for any traffic to your left on the other side of the street. Thus, in our game a pedestrian is allowed to take a step forward to an unoccupied site on the grid if and only if there is nobody anywhere to her right before she takes the step, and will be nobody anywhere to her left after she has taken the step. "Forward" may be any direction on the grid (north, south, east or west).



Observe that moves are reversible. A *position* in the game is a placement of the k pedestrians on k distinct sites on the grid. Two positions are *game equivalent* if one can be reached from the other by a sequence of moves. Since moves can be reversed, this is an equivalence relation.

Proposition 3.1. *Two positions are game equivalent if and only if they realize the same pattern in the tilted square.*

Proof. The pattern realized by a position is invariant under the game: When a dot moves one step it changes its relative position only to dots that are to the right of the old site or to the left of the new site, but a move can be made only when no such dots are present. Hence, positions that are game equivalent realize the same pattern.

To prove the converse, we shall prove that any position can be played to a permutation (one dot in each row and column of a k by k square). By the game invariance of patterns, this permutation must be the unique permutation realizing this pattern. Hence any two positions that realize the same pattern can be played to each other via the permutation.

For any position, we can play the eastmost dot of the top row upwards as far as we want. Among the remaining dots we can again play the eastmost dot of the top row upwards as far as it does not reach the row of the first dot. Continuing in this way, we obtain a position with the dots on distinct rows. We then perform the same operation on columns, playing one dot at a time eastwards. Since this can be done without changing the row of any dot, we reach a position where the dots occupy sites on k distinct rows and k distinct columns.

Any empty rows or columns can be filled by playing in the dot from the row or column next to it. Hence we obtain a permutation. □

Corollary 3.2. *There are $k!$ game equivalence classes with k pedestrians.*

The following theorem establishes a connection between the pedestrian game and the pattern packing problem:

Theorem 3.3. *L_k is the minimal size of a union of one representative from each game equivalence class in the pedestrian game with k pedestrians.*

Proof. A set S of sites which is a union of one representative from each game-equivalence class has the property that every position can be played so that all dots are moved into S. Hence the permutation that is represented by S contains each pattern of length k. Conversely, a permutation that contains each pattern of length k can be represented as a set of sites with the property that any position with k dots can be played into it. This set of sites therefore contains a representative of each game-equivalence class. □

4. Strong convergence of the pedestrian game

In this section we consider the pedestrian game played on a k by k board. Proposition 3.1 still holds in this setting, for a position with k dots which contains several dots in the same rows or columns can be untangled within the board limits. The modified procedure is as follows.

First move the top dot to the top row, if necessary. Then make sure that there are at least two dots in the top two rows, possibly by moving a second dot to the north. In general, for every $m \leq k$, make sure that there are at least m dots in the top m rows. This also means that the bottom $k-m$ rows will contain at most $k-m$ dots. In particular, the bottom row contains at most one dot, and by moving the bottom dot to the south, we can make sure that there is exactly one. Repeating this for each row, we get exactly one dot in each row, and to obtain this, we have not made any horizontal moves. This means that if the same operation is performed on columns, we reach a permutation, i.e. a position where each row and each column contains exactly one dot.

Hence each game-equivalence class on the k by k square has a unique representative with one dot in each row and one dot in each column. We now show that there is another natural representation of game-equivalence classes.

Proposition 4.1. *Each game-equivalence class on the k by k board has a unique representative from which it is impossible to make a move in the south or west directions.*

Proof. By the *directed* pedestrian game, we mean the pedestrian game played on the positive quadrant, with the restriction that only moves in the south and west directions are permitted. A position where no such move is possible is called a *terminal* position. As we will show, the directed pedestrian game is *strongly convergent*, which means that from any given initial position, all move sequences lead to the same terminal position in the same number of moves. It is clear that this also proves the proposition. □

A simple criterion for strong convergence is the *polygon property* [3]: In any position where two different moves, x and y, are legal, there are two play sequences of the same length and beginning with x and y respectively leading to the same position.

Lemma 4.2. *The directed pedestrian game on the k by k board is strongly convergent.*

Proof. We have to verify the polygon property. In this case, we show that two different moves from the same position actually *commute*, that is xy and yx are both legal and give the same result.

If one and the same dot can make a move in either of the directions south and west, then it is clear that after a move south, the move to the west will still be legal, and vice versa.

If two different dots can move in any of the two legal directions, it is easy to see that these moves cannot interfere. It follows that the directed pedestrian game is strongly convergent. □

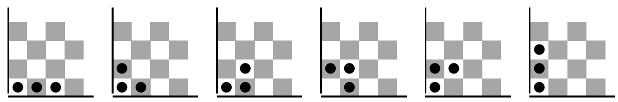


FIGURE 1. The six terminal positions in the directed pedestrian game for $k = 3$.

5. The terminal position is the compact representation

The unique terminal position of the directed pedestrian game provides a canonical dot representation for every k-pattern. We shall now see that it is the same compact configuration that we constructed from the ascents and inverse descents of the permutation.

Theorem 5.1. *The compact dot representation of π defined in Sec. 2 is the terminal position corresponding to π.*

Proof. First, we show that this position represents the permutation π. Consider any pair of symbols $x < y$ that occur in order, that is

$$\pi = (\ \ldots\ x\ \ldots\ y\ \ldots\ ).$$

Then there is at least one ascent between x and y, so the dot corresponding to y is in a column east of the dot corresponding to x. Since $x < y$, there are at least as many inverse descents below y as there are below x, so either the dots are in the same row, or the one corresponding to y is higher.

If the pair occurs in the wrong order, that is $\pi = (\ \ldots\ y\ \ldots\ x\ \ldots\ )$ then there are at least as many ascents before x as there are before y, and at least one more inverse descent below y than below x. Hence in this case the dot corresponding to x will be in a row south of the dot corresponding to y, and in the same column or a column east of it.

This shows that every pair of dots has the correct relative location. Hence the position described represents π.

To show that the position we have defined is a terminal position, we first prove that no dot can move west. The dot corresponding to the symbol π_1 obviously cannot. If a pair π_{i-1}, π_i is a descent, then the dot corresponding to π_{i-1} will be above the one corresponding to π_i, in the same column. Hence the dot corresponding to π_i cannot move west. If on the other hand π_{i-1}, π_i is an ascent, then the dot corresponding to π_{i-1} will be in the column immediately to the west of the dot corresponding to π_i, and above it or in the same row. Hence the dot corresponding to π_i still cannot move west.

To show that no dot can move south either, we note that the symbol `1` is already in the bottom row. If a pair $x-1 \dots x$ is an inverse ascent, the corresponding dots will be in the same row so the x-dot cannot move south. And for an inverse descent $x \dots x-1$, the x-dot will have the other dot one row below, in the same column or to the east. In either case, it cannot move south. □

6. The main result

Our original motive for studying compact configurations was a desire to improve the upper bound $L_k \le k^2$. We computed L_k for small k and the following table points to the existence of a tighter bound.

k	1	2	3	4	5
L_k	1	3	5	9	13

To our disappointment, the union of all compact configurations covers all of the $k \times k$-board except a thin slice along the border which is asymptotically insignificant. A closer study of the optimal permutations found by the computer revealed that it may be easier to pack sparse configuration in a chessboard full of holes!

For $k = 3$ there exist only two permutations of minimal length ($L_k = 5$) containing all k-patterns. One of them is `41352`, which can be interpreted in terms of chessboards. As shown in the figure, `41352` is given by the white squares of the 3 by 3 chessboard in our standard way of associating patterns with subsets of squares on the grid.

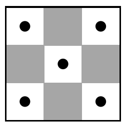


Figure 2. The white squares permutation `41352`.

The same thing is true for $k = 5$; the thirteen white squares of a 5×5 chessboard define a permutation containing all 5-patterns. It is tempting to conjecture that the white squares permutation of any $k \times k$-board contains all k-patterns, at least if k is odd, and in the first draft of this paper we gave in to that temptation. From our failed attempts to prove the false conjecture a probabilistic argument emerged that nonconstructively demonstrates the existence of a counterexample. Fortunately, a similar argument proves that a fifty percent wider rectangle *can* accommodate all patterns on its white squares.

Theorem 6.1. *The permutation given by the white squares of the $k \times \lfloor 3k/2 \rfloor$ chessboard contains all k-patterns, therefore*

$$L_k \le \frac{3}{4}k^2 \quad \text{for } k > 1$$

Proof. For any given pattern, we start with the standard dot configuration in the k by k square to the west. If all dots are already on white squares, we are done; otherwise, some horizontal moves need to be made. Going from west to east, each time we encounter a dot on a black square, we move it together with all dots east of it one step. If the black dot is the second dot in a descent, this mass move goes west, otherwise it goes east. If we are lucky, the dots will stay within the stipulated rectangle, but we may as well have bad luck, as shown in the figure.

Returning to the starting position, we note that some vertical moves are possible. At every inverse ascent, we have the option of moving the upper dot, along with all dots above it, one step to the south. We now use a probabilistic argument. For every inverse ascent, we flip a coin. On heads, we do nothing; on tails, we move the upper dot and every dot above it to the south. The dots obviously stay within the original k by k square. As before, we then make the necessary moves in the horizontal direction to put the dots on white squares, and we will show that there is now a nonzero probability of the dots staying inside the stipulated rectangle.

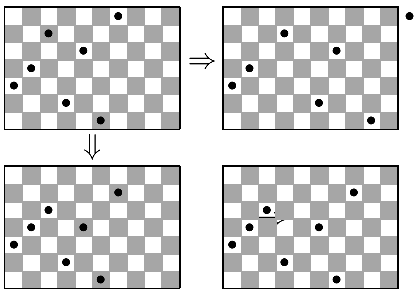


FIGURE 3. Two ways to move 3462517 to white squares

We estimate the expected number of moves to the east, counting a move to the west as -1. The first dot will cause a move to the east if it is on a black square after the southward moves. Every other dot will cause a move if after the southward moves it is on a square of the color opposite to that of the previous dot. Otherwise it will automatically be pushed to a white square when the previous dot is moved to a white square. The potential move will be to the west if the dot is the second dot in a descent, and otherwise to the east.

The key point is that the probability of a dot being on a square of color opposite to that of the previous dot is $1/2$ as soon as there is at least one potential southward move that will change the color of the upper dot but not of the lower one. Since the southward moves are decided by independent coin flips, the probability that the two dots end up on squares of different colors is exactly $1/2$. Now, it is easy to see that two adjacent symbols in a permutation must have an inverse ascent in the interval between them unless it is a descent by one, like 54. In that special case, the probability of the dots occupying squares of different colors drops to zero.

To sum up, the first dot will cause at most one move to the east, and for the others, a descent by one will contribute zero moves to the east, while a descent by more than one will contribute $-1/2$ and an ascent $1/2$ to the expected number of moves to the east. Therefore the expected number of necessary moves to the east is at most

$$1 + \frac{1}{2}a(\pi).$$

Hence with one exception, the identity permutation $[123\ldots k]$ with $k-1$ ascents (which is trivial), the expected number of moves to the east is at most $k/2$, so some lucky outcome of the coin flips will need at most $\lfloor k/2 \rfloor$ extra columns in order to play all the dots to white squares. □

7. Variation 1: Permutations containing almost all k-patterns

We show that a permutation need not be much longer than $k^2/4$ in order to contain *almost* all patterns of size k.

Theorem 7.1. *Let f be a function such that $f(k)$ tends to infinity as k does. Then there is a sequence $\sigma(k)$ of permutations of length less than $k^2/4 + f(k)k^{3/2}$ that contains almost all patterns of size k in the sense that the fraction of patterns of length k that are not contained in $\sigma(k)$ tends to zero as k tends to infinity.*

Proof. Let π be a random permutation of length k. We can write $d(\pi) = d_{odd} + d_{even}$, where d_{odd} and d_{even} are the number of descents in odd and even position, respectively, that is, d_{odd} is the number of descents (π_i, π_{i+1}) for which i is odd, and d_{even} is the number of such descents for which i is even. The both d_{odd} and d_{even} are binomial distributed. Their distributions can therefore be approximated by normal distributions with mean $k/4$ and standard deviation $O(\sqrt{k})$.

Now let g be a function such that $g(k) \to \infty$ as $k \to \infty$. Then the probability that d_{even} or d_{odd} deviates by more than $g(k)\sqrt{k}/2$ from $k/4$ tends to zero as k tends to infinity. Hence the probability that $d(\pi) < k/2 - g(k)\sqrt{k}$ tends to zero. Similarly, the probability that $a(\pi^{-1}) < k/2 - g(k)\sqrt{k}$ tends to zero.

It follows that with high probability, the permutation π can be played into a square of side $k/2 + g(k)\sqrt{k}$. Hence there is a permutation of size $\left(k/2 + g(k)\sqrt{k}\right)^2$ that contains almost all patterns of size k.

If we put

$$\left(k/2 + g(k)\sqrt{k}\right)^2 = k^2/4 + f(k)k^{3/2},$$

then it is clear that g tends to infinity if and only if f does. The theorem follows. □

8. Variation 2: Permutations containing pattern or inverse

We can construct a permutation of length $\binom{k}{2}$ which in a slightly weaker sense contains every pattern of length k. This permutation consists of the elements on and below the diagonal in the tilted square. For example, $T_2 =$`312` and $T_3 =$`641523`.

Theorem 8.1. *For every pattern τ of length k, the triangular permutation T_k contains either τ or τ^{-1}.*

Proof. A dot in the terminal position of a permutation will be outside T_k only if the sum of the number of ascents before the symbol and the number of inverse descents below it is at least k. An inverse descent in τ is of course a descent in τ^{-1}, so the total number of ascents and inverse descents is $a(\tau) + d(\tau^{-1})$. If that total is less than k, the permutation T_k will certainly contain τ. Analogously, if the total $a(\tau^{-1}) + d(\tau)$ is less than k, the permutation T_k will certainly contain τ^{-1}.

But there are $k-1$ positions which are either descents or ascents, so we have

$$d(\tau) + a(\tau^{-1}) + d(\tau^{-1}) + a(\tau) = 2(k-1).$$

The theorem follows from the fact that either τ or its inverse must have the property that the total number of ascents plus the total number of inverse descents is at most $k-1$. □

9. Final twist: Enumeration of game positions

We consider the pedestrian game played on an m by n board. We wish to count the representations of a given permutation, that is, the number of positions in a given game-equivalence class.

If a certain permutation π can be represented at all on an m by n board (m rows and n columns), then there is a terminal position P_0 of π in the south-west directed pedestrian game. A different representation P of the same permutation π can now be described by labeling each dot with the two coordinates of the distance it has to move from P_0 to P. Since P_0 is the terminal position, each dot has to move north and east to reach P. Moreover, if a dot x is west of a dot y, then y has to move at least as many steps east as x, since any move to the east by x will bump all other dots one step east.

The terminal position P_0 occupies $a(\pi)+1$ columns and $d(\pi^{-1})+1$ rows. Therefore, the number of times we can allow the eastmost dot to be bumped without being bumped off the m by n board is $n-a(\pi)-1$. Hence the sequence of horizontal distances that the dots, taken from west to east, have to travel to get from P_0 to P forms a weakly increasing sequence of k nonnegative integers that is bounded from above by $n-a(\pi)-1$. An elementary theorem in enumerative combinatorics tells us that the number of such sequences is

$$\binom{n-a(\pi)-1+k}{k} = \binom{n+d(\pi)}{k}.$$

Similarly, the number of sequences of vertical distances from P_0 to P of the dots taken from bottom to top is

$$\binom{m+a(\pi^{-1})}{k}.$$

Since we get a representation of π for each pair of such integer sequences, we see that the number of representations of a permutation π on an m by n board is

$$\binom{m+a(\pi^{-1})}{k}\binom{n+d(\pi)}{k}.$$

From this we obtain, by summing over all permutations in S_k, a classical identity, probably due to Carlitz.

$$\sum_{\pi\in S_k}\binom{m+a(\pi^{-1})}{k}\binom{n+d(\pi)}{k} = \binom{mn}{k},$$

since the total number of positions of k dots on an m by n board is $\binom{mn}{k}$.

References

[1] R. Arratia, On the Stanley-Wilf conjecture for the number of permutations avoiding a given pattern, *The Electronic Journal of Combinatorics* **6** (1999), #N1.

[2] R. Simion and F. W. Schmidt, Restricted permutations, *European Journal of Combinatorics* **6** (1985), no. 4, 383–406.

[3] K. Eriksson, Strong convergence and a game of numbers, Eur. J. Combinatorics 17 (1996), 379–390.

NADA, KTH, SE-100 44 Stockholm, Sweden
E-mail address: henrik@nada.kth.se

IMA, Mälardalens högskola, SE-721 23 Västerås, Sweden
E-mail address: kimmo.eriksson@mdh.se

MaI, Linköpings universitet, SE-581 83 Linköping, Sweden
E-mail address: {linusson, jowas}@mai.liu.se

ALGEBRAIC SUCCESSION RULES

JEAN-MARC FÉDOU AND CHRISTINE GARCIA

Abstract. In this paper, we show that succession rules

$$(k) \rightsquigarrow (1)^{\alpha_{k-1}} \dots (k-1)^{\alpha_1} (k)^{\lambda_0} \dots (k+p)^{\lambda_p}$$

have an algebraic generating function when the sequence (α_i) is rational. We decompose algebraically the paths in the corresponding generating tree and deduce an algebraic equation satisfied by the noncommutative generating function.

Résumé. Nous montrons dans cet article que les règles de succession

$$(k) \rightsquigarrow (1)^{\alpha_{k-1}} \dots (k-1)^{\alpha_1} (k)^{\lambda_0} \dots (k+p)^{\lambda_p}$$

ont une série génératrice algébrique quand la suite (α_i) est rationnelle. Nous décomposons algébriquement les chemins dans l'arbre de génération correspondant et nous en déduisons une équation algébrique satisfaite par la série génératrice non commutative.

1. Introduction

The succession rules were first introduced by Chung, Graham, Hoggatt and Kleimann in [3] to study Baxter permutations. The method was later successfully used by West [11], Dulucq, Gire and Guibert [5, 6, 7] for the enumeration of permutations with forbidden sequences. The concept has more recently been exploited by Barcucci, Del Lungo, Pergola and Pinzani [2] as the ECO method for the enumeration and recursive construction of various classes of combinatorial objects. The succession rule approach has several equivalent interpretations, ECO rules, random paths, infinite automata or Riordan arrays, and deals with different kinds of generating functions (rational, algebraic or exponential). The problem of classifying successions rules according to the type of their generating functions has been proposed by R.Pinzani [2] in the area of ECO systems. A classical and easy result is that finite succession rules have rational generating function since they correspond to a regular language. It is shown in [1] that every finite transformation of Catalan succession rule

$$(k) \rightsquigarrow (1)(2) \dots (k)(k+1)$$

is algebraic. In the same paper are also described succession rules leading to exponential generating functions which have been more extensively studied by S.Corteel in [4].

Our paper is devoted to the study of algebraic system of succession rules having algebraic generating function. Our approach is closely related with the Schutzenberger methodology, which consists in finding first a bijection between the objects and the words of an algebraic language, second a non ambiguous grammar for the language and finally take the commutative image and deduce an algebraic system for the generating function. We first explore the basic Catalan example and explain its algebraicity using a non ambiguous decomposition of the paths in the Catalan generating tree. For that, we define its noncommutative formal power series using the infinite alphabet of positive integers. We use a new operation $\oplus$ which allows us to get a non ambigous decomposition of the formal power series associated to the generating tree. We deduce the classical Catalan algebraic equation by taking the commutative image of the formal power series. This method is then extended to the more general succession rule

$$(k) \rightsquigarrow (1)\dots(k-1)(k)^{\lambda_0}\dots(k+p)^{\lambda_p}$$

for any finite sequence (λ_i).

Next we describe an algebraic decomposition for the succession rule

$$(k) \rightsquigarrow (1)^{\alpha_{k-1}}\dots(k-1)^{\alpha_1}(k)^{\lambda_0}\dots(k+p)^{\lambda_p}$$

for any *constant* sequence (α_i) proving thereby that their generating function is algebraic when the sequence (α_i) is *rational*.

2. Definitions

A succession rule is a function $(k) \rightsquigarrow (e_1(k))(e_2(k))\dots(e_{s_k}(k))$ which associates to each positive integer k a finite multiset of integers, called successors of k. A generating tree is a succession rule with a particular integer a, called axiom. We suppose in the following that a equals 1. A generating tree can be viewed as the infinite tree constructed with a root labelled by the axiom and where each node labelled k has sons labelled according to the succession rule. Hence, ECO systems are those generating trees where each integer has exactly k successors.

For a generating tree Ω,

$$\Omega \left\{ \begin{array}{l} (1) \\ (k) \rightsquigarrow (e_1(k))(e_2(k))\dots(e_{s_k}(k)) \end{array} \right. \tag{1}$$

we define the language L_Ω as the set of words over $\mathbb{N}$, begining by the axiom 1 and satisfying the succession rule, $(k) \rightsquigarrow (e_1(k))(e_2(k))\dots(e_{s_k}(k))$. Each word w of L_Ω corresponds to at least one path of Ω. We note $m(w)$ the number of paths w in the generating tree Ω. We denote by S_Ω the noncommutative formal power series,

$$S_\Omega = \sum_{w\in L_\Omega} m(w)w.$$

By construction, the generating tree Ω and the noncommutative formal power series S_Ω have the same generating function,

$$F_\Omega(z) = \sum_n f_n z^n$$

where

$$f_n = \sum_{w\in L_\Omega, |w|=n} m(w).$$

For simplifying the notation, we write F_Ω for $F_\Omega(z)$. We use standard external product and concatenation, by an integer n, over the noncommutative formal power series S_Ω. We write:

$$nS_\Omega = \sum_{w\in L_\Omega} (nm(w))w$$

$$(n).S_\Omega = \sum_{w\in L_\Omega} m(w)(n.w).$$

We define a new operation $\oplus$ as follows,

Definition 1. *For $i \in \mathbb{N}^+$, we define by $i^\oplus = i+1$. By extension if $w = w_1w_2\dots w_i$ is a word of L_Ω then $w^\oplus = w_1^\oplus w_2^\oplus \dots w_i^\oplus$ and $S_\Omega^\oplus = \sum_{w\in L_\Omega} m(w)w^\oplus$*

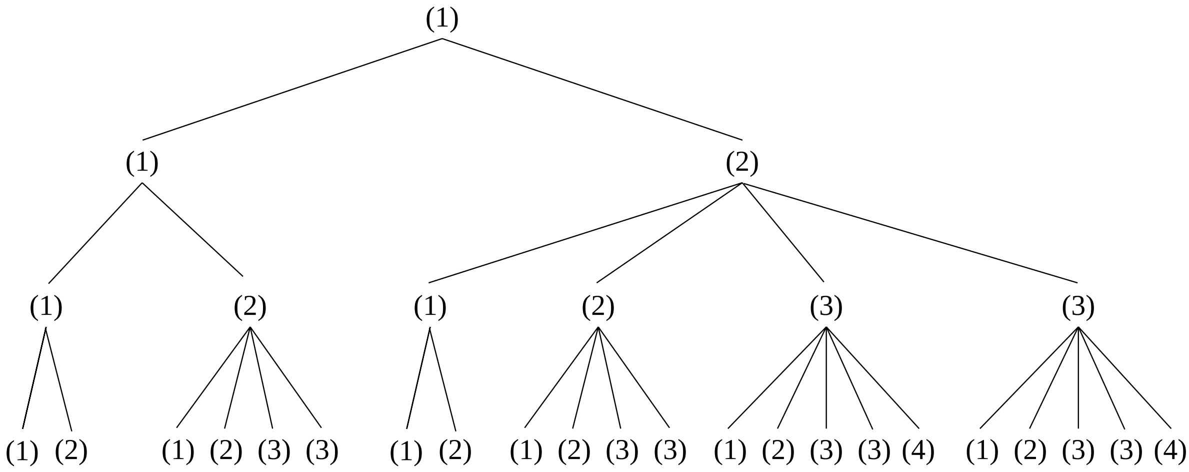


FIGURE 1. Partial generating tree of the rule Θ

Example 2. *The Figure 1 give the partial generating tree of the rule* Θ *with axiom* 1.

$$\Theta \left\{ \begin{array}{l} (1) \\ (k) \rightsquigarrow (1)(2)(3)(3)(4)\dots(k)(k+1) \end{array} \right.$$

$$\begin{array}{rcl} S_\Theta & = & (1) + (11) + (12) + (111) + (112) + (121) + (122) + 2(123) + \dots \\ F_\Theta & = & z + 2z^2 + 6z^3 + 22z^4 + \dots \end{array}$$

The noncommutative formal power series approach allows us to interpret finite transformations and show that they do not change the algebraicity of the generating function. We are interested only with *total* algebraic succession rules, that is succession rules where algebraicity is acquired for any choice of the axiom, and for all the restrictions according to the last letter. More precisely, we suppose that for any integer i, the generating function F_i of the paths ending with i are algebraic.

Definition 3. *The transformation* T_1 *is the addition of a constant* c *for one succession rule,*

$$T_1(\Omega) \left\{ \begin{array}{l} (1) \\ (k_0) \rightsquigarrow (e_1(k_0))(e_2(k_0))\dots(e_{s_{k_0}}(k_0))(c) \\ (k) \rightsquigarrow (e_1(k))(e_2(k))\dots(e_{s_k}(k)), for \;\; k \neq k_0 \end{array} \right.$$

Definition 4. *The transformation* T_2 *is the addition of a constant* c *uniformly,*

$$T_2(\Omega) \left\{ \begin{array}{l} (1) \\ (k) \rightsquigarrow (e_1(k))(e_2(k))\dots(e_{s_k}(k))(c) \end{array} \right.$$

Proposition 5. *All finite modifications by* T_1 *and* T_2 *of the succession rule* Ω *are algebraic.*

Proof. Let S_{k_0} be the formal sum of the words corresponding to the paths of the generating tree Ω ending with k_0, let F_{k_0} the generating function of S_{k_0}. We have $S_{T_1(\Omega)} = S_\Omega + S_{k_0}S_{T_1(\Omega)}$, and deduce $F_{T_1(\Omega)} = F_\Omega + F_{k_0}F_{T_1(\Omega)}$, so $F_{T_1(\Omega)}$ is algebraic when F_Ω and F_{k_0} are algebraic.
Let now ${}_cS$ be the formal power series of

$$\left\{ \begin{array}{l} (c) \\ (k) \rightsquigarrow (e_1(k))(e_2(k))\dots(e_{s_k}(k)) \end{array} \right.$$

that is the succession rule Ω where the axiom 1 has been replaced by c. Let ${}_cF$ the generating function of ${}_cS$. We have $S_{T_2(\Omega)} = S_\Omega({}_cS)^* = \frac{S_\Omega}{1-({}_cS)}$, which concludes the proof. □

Example 6. *Let Γ be the Catalan generating tree defined by,*

$$\Gamma \left\{ \begin{array}{l} (1) \\ (k) \rightsquigarrow (1)(2)\dots(k)(k+1) \end{array} \right.$$

the Θ generating tree of Example2 is obtained applying the T_2 transformation over Γ with $c = 3$.

Remark Proposition 5 will be usefull in Theorem 12.

3. Algebraicity of Catalan generating tree Γ

In this section, we first describe a non ambiguous decomposition of the Catalan generating tree, which explains its algebraicity.

Theorem 7. *The formal power series S_Γ satisfies the equation*

$$S_\Gamma = (1) + (1).S_\Gamma + (1).S_\Gamma^{\oplus}.(\epsilon + S_\Gamma)$$

Proof. The proof can be easily deduced from the following non ambiguous inductive description of the set of the paths in the generating tree. Let $w \neq 1$ be a non trivial path of the generating tree Γ. Then w can be written $w = 1u$ where either u begins with 1 and therefore is a path of Γ or u begins with 2 and we have two cases,

- if each letter of u is > 1 then $u = v^{\oplus}$ where v is a path of Γ,
- if not, u can be writen $v_1^{\oplus}v_2$ where v_1 and v_2 are paths of Γ, v_2 being the longest suffix of u beginning by 1.

We deduce that $S_\Gamma = (1) + (1).S_\Gamma + (1).S_\Gamma^{\oplus} + (1).S_\Gamma^{\oplus}.S_\Gamma$ □

Corollary 8. *The generating function F_Γ of the succession rule Γ satisfies :*

$$F_\Gamma = z + zF_\Gamma + zF_\Gamma(1 + F_\Gamma)$$

Theorem 7 can be easily generalized to the case of more general succession rule,

$$\Lambda \left\{ \begin{array}{l} (1) \\ (k) \rightsquigarrow (1)\dots(k-1)(k)^{\lambda_0}\dots(k+p)^{\lambda_p} \end{array} \right.$$

A similar inductive description (see Figure 2) of the associated language leads to the following theorem.

Theorem 9.

$$\begin{aligned} S_\Lambda \;=\; & (1) + \\ & \lambda_0(1).S_\Lambda + \\ & \lambda_1(1).S_\Lambda^{\oplus}.(\epsilon + S_\Lambda) + \\ & \lambda_2(1).S_\Lambda^{\oplus\oplus}.(\epsilon + S_\Lambda^{\oplus}).(\epsilon + S_\Lambda) + \\ & \vdots \\ & \lambda_p(1).S_\Lambda^{p\oplus}.(\epsilon + S_\Lambda^{(p-1)\oplus}).\dots.(\epsilon + S_\Lambda^{\oplus})(\epsilon + S_\Lambda), \end{aligned}$$

where $S_\Lambda^{i\oplus} = (S_\Lambda^{(i-1)\oplus})^{\oplus}$ and $S_\Lambda^{1\oplus} = S_\Lambda^{\oplus}$.

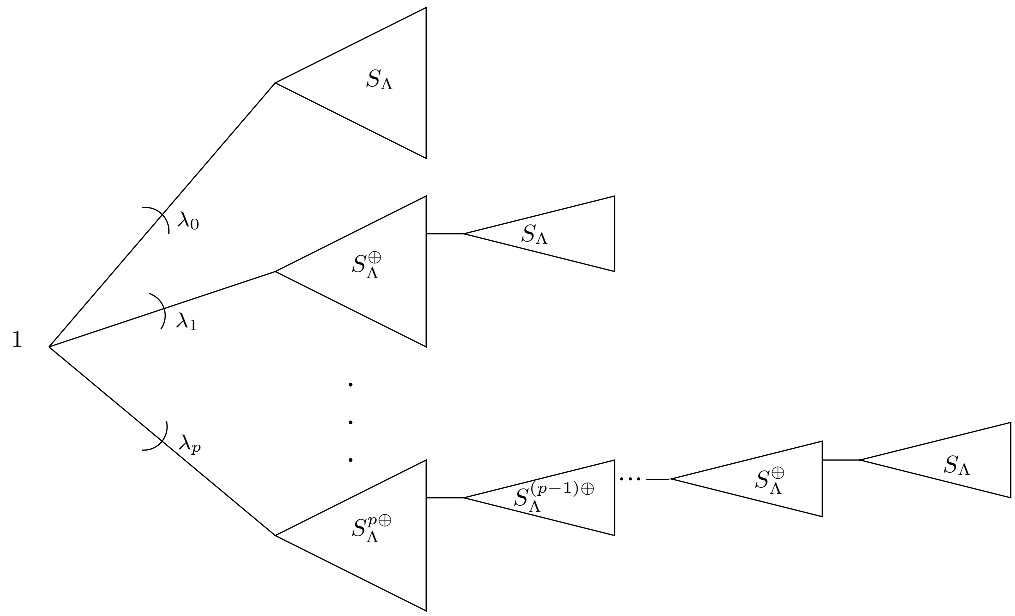


Figure 2. Generalized Catalan generating tree Λ

Corollary 10. *The generating function F_Λ of the succession rule Λ is algebraic and satisfies :*

$$\begin{aligned} F_\Lambda &= z + \\ & \lambda_0 z F_\Lambda + \\ & \lambda_1 z F_\Lambda(1+F_\Lambda) + \\ & \lambda_2 z F_\Lambda(1+F_\Lambda)^2 + \\ & \vdots \\ & \lambda_p z F_\Lambda(1+F_\Lambda)^p \\ &= z(1+\sum_{i=0}^{p} \lambda_i F_\Lambda(1+F_\Lambda)^i) \end{aligned}$$

All the algebraic generating function given in the small catalog of ECO-systems of [1] can be deduced from the previous theorem. For instance, Motzkin numbers correspond to the sequence $\lambda_i = (0,1,0,\ldots)$, Schröder numbers correspond to the sequence $\lambda_i = (1,2,0,\ldots)$ and Ternary trees correspond to the sequence $\lambda_i = (1,1,1,0,\ldots)$.

4. Algebraicity and rationality

In this section, we study more general rules having the following general form,

$$\Upsilon \left\{ \begin{array}{l} (1) \\ (k) \rightsquigarrow (1)^{\alpha_{k-1}} \ldots (k-1)^{\alpha_1} (k)^{\lambda_0} \ldots (k+p)^{\lambda_p} \end{array} \right.$$

As in the previous section, the words occuring in S_Υ have λ_i rises from (k) to $(k+i)$. The difficulty is to deal here with the α_i kind of descents from (k) to $(k-i)$.

Theorem 11. *The rationality of the sequence (α_i) implies the algebraicity of F_Υ.*

Proof. We will give the proof in the particular case of $p = 2$ and (α_i) a rational sequence of degree 2. The general case follows the same scheme. We begin by giving the different equations obtained from the recursive decomposition of the paths in the generating tree Υ. We need to define S_i as the formal sum of the paths in Υ ending by i:

$$
\begin{aligned}
S &= (1) + \lambda_0(1)S + \lambda_1(1)S^{\oplus} + \lambda_1(1)\sum_{j\geq 1}\alpha_j S_j^{\oplus} S \\
&\quad \lambda_2(1)S^{\oplus\oplus} + \lambda_2(1)\sum_{j\geq 1}\alpha_{j+1}S_j^{\oplus\oplus}S + \lambda_2(1)\sum_{j\geq 1}\alpha_j S_j^{\oplus\oplus}S^{\oplus} \\
&\quad +\lambda_2(1)\sum_{j\geq 1}\sum_{k\geq 1}\alpha_j\alpha_k S_j^{\oplus\oplus}S_k^{\oplus}S \\
S_1 &= (1) + \lambda_0(1)S_1 + \lambda_1(1)\sum_{j\geq 1}\alpha_j S_j^{\oplus}S_1 \\
&\quad +\lambda_2(1)\sum_{j\geq 1}\alpha_{j+1}S_j^{\oplus\oplus}S_1 \\
&\quad +\lambda_2(1)\sum_{j\geq 1}\sum_{k\geq 1}\alpha_j\alpha_k S_j^{\oplus\oplus}S_k^{\oplus}S_1 \\
S_2 &= \lambda_0(1)S_2 + \lambda_1(1)S_1^{\oplus} + \lambda_1(1)\sum_{j\geq 1}\alpha_j S_j^{\oplus}S_2 \\
&\quad +\lambda_2(1)\sum_{j\geq 1}\alpha_{j+1}S_j^{\oplus\oplus}S_2 + \lambda_2(1)\sum_{j\geq 1}\alpha_j S_j^{\oplus\oplus}S_1^{\oplus} \\
&\quad +\lambda_2(1)\sum_{j\geq 1}\sum_{k\geq 1}\alpha_j\alpha_k S_j^{\oplus\oplus}S_k^{\oplus}S_2
\end{aligned}
$$

and for $i \geq 3$,

$$
\begin{aligned}
S_i &= \lambda_0(1)S_i + \lambda_1(1)S_{i-1}^{\oplus} + \lambda_1(1)\sum_{j\geq 1}\alpha_j S_j^{\oplus}S_i \\
&\quad \lambda_2(1)S_{i-2}^{\oplus\oplus} + \lambda_2(1)\sum_{j\geq 1}\alpha_{j+1}S_j^{\oplus\oplus}S_i + \lambda_2(1)\sum_{j\geq 1}\alpha_j S_j^{\oplus\oplus}S_{i-1}^{\oplus} \\
&\quad +\lambda_2(1)\sum_{j\geq 1}\sum_{k\geq 1}\alpha_j\alpha_k S_j^{\oplus\oplus}S_k^{\oplus}S_i
\end{aligned}
$$

We note $F = F_\Upsilon$ for short, the generating function of Υ, F_i the generating functions of S_i, and

$$F = \sum_{i\geq 1} F_i$$

$$G = \sum_{i\geq 1} \alpha_i F_i$$

$$H = \sum_{i\geq 1} \alpha_{i+1} F_i$$

Thus we have,

$$\begin{aligned}
F &= z + \lambda_0 zF + \lambda_1 zF + \lambda_1 zGF + \lambda_2 zF + \lambda_2 zHF + \lambda_2 zGF + \lambda_2 zG^2 F \\
F_1 &= z + \lambda_0 zF_1 + \lambda_1 zGF_1 + \lambda_2 zHF_1 + \lambda_2 zG^2 F_1 \\
F_2 &= \lambda_0 zF_2 + \lambda_1 zF_1 + \lambda_1 zGF_2 + \lambda_2 zHF_2 + \lambda_2 zGF_1 + \lambda_2 zG^2 F_2
\end{aligned}$$

and for $i \geq 3$,

$$F_i = \lambda_0 zF_i + \lambda_1 zF_{i-1} + \lambda_1 zGF_i + \lambda_2 zF_{i-2} + \lambda_2 zHF_i + \lambda_2 zGF_{i-1} + \lambda_2 zG^2 F_i$$

The generating functions F_i satisfies a linear recurence relation of degree 2, $F_{i+2} = uF_{i+1} + vF_i$ for $i \geq 1$, where u and v are rational functions depending on $\lambda_0, \lambda_1, \lambda_2, G$ and H.

Moreover, supposing that the sequence (α_i) is rational of degree 2 means that it satisfies a recurence relation $\alpha_{i+2} = a\alpha_{i+1} + b\alpha_i + c$, for some a, b, c coefficients. Developing $\sum_{i\geq 1} \alpha_{i+2} M^i$, we get,

$$\begin{aligned}
\sum_{i\geq 1} \alpha_{i+2} M^i &= a \sum_{i\geq 1} \alpha_{i+1} M^i + b \sum_{i\geq 1} \alpha_i M^i + c \sum_{i\geq 1} M^i \\
&= a\alpha_2 M + b\alpha_1 M + b\alpha_2 M^2 + +c \sum_{i\geq 1} M^i + (aM + bM^2) \sum_{i\geq 1} \alpha_{i+2} M^i
\end{aligned}$$

where in particular, M can be any square matrix.

Let now

$$M = \begin{pmatrix} u & v \\ 1 & 0 \end{pmatrix},$$

where u and v are the coefficients of the linear recurrence satisfied by F_i. Classicaly, we have

$$\begin{pmatrix} F_{i+2} \\ F_{i+1} \end{pmatrix} = M \begin{pmatrix} F_{i+1} \\ F_i \end{pmatrix} = M^2 \begin{pmatrix} F_i \\ F_{i-1} \end{pmatrix} = \ldots = M^i \begin{pmatrix} F_2 \\ F_1 \end{pmatrix},$$

and we obtain

$$\begin{aligned}
\sum_{i\geq 1} \alpha_{i+2} \begin{pmatrix} F_{i+2} \\ F_{i+1} \end{pmatrix} &= (a\alpha_2 M + b\alpha_1 M + b\alpha_2 M^2) \begin{pmatrix} F_2 \\ F_1 \end{pmatrix} + c \sum_{i\geq 1} M^i \begin{pmatrix} F_2 \\ F_1 \end{pmatrix} \\
&\quad + (aM + bM^2) \sum_{i\geq 1} \alpha_{i+2} \begin{pmatrix} F_{i+2} \\ F_{i+1} \end{pmatrix}
\end{aligned}$$

As $\sum_{i\geq 1} \alpha_{i+2} F_{i+2} = G - \alpha_1 F_1 - \alpha_2 F_2$ and $\sum_{i\geq 1} \alpha_{i+2} F_{i+1} = H - \alpha_2 F_1$, we deduce an algebraic system of equations satisfied by G and H,

$$\begin{aligned}
\begin{pmatrix} G \\ H \end{pmatrix} &= (a\alpha_2 M + b\alpha_1 M + b\alpha_2 M^2) \begin{pmatrix} F_2 \\ F_1 \end{pmatrix} + c \begin{pmatrix} F - F_1 - F_2 \\ F - F_1 \end{pmatrix} \\
&\quad + (aM + bM^2) \begin{pmatrix} G - \alpha_1 F_1 - \alpha_2 F_2 \\ H - \alpha_2 F_1 \end{pmatrix}
\end{aligned}$$

□

5. Conclusion

Theorem 11 and Proposition 5 allow us to generalize the results of Flajolet and al [1]:

Theorem 12. *All finite transformations of the succession rule*

$$(k) \rightsquigarrow (1)^{\alpha_{k-1}} \dots (k-1)^{\alpha_1}(k)^{\lambda_0} \dots (k+p)^{\lambda_p}$$

are algebraic when (α_i) *is rational.*

A conjecture is to have a similar result when the sequence (α_i) is algebraic as discussed with Cyril Banderier during GASCOM'01.

We give some examples where such succesion rules naturally appear (see Figure 3). Diagonally directed convex polyominoes [14] (or fully directed compact animals) are known to be counted according to their number of diagonals by $\frac{1}{2n+1}\begin{pmatrix} 3n \\ n \end{pmatrix}$. They naturally satisfy the succession rule

$$\begin{cases} (1) \\ (k) \rightsquigarrow (1)^{k+1}(2)^k \dots (k-1)^3(k)^2(k+1) \end{cases}$$

Another example concerns a new succesion rule for Catalan numbers.

$$\begin{cases} (1) \\ (k) \rightsquigarrow (1)^{2^{k-2}}(2)^{2^{k-3}} \dots (k-2)^2(k-1)(k+1) \end{cases}$$

This succession rule generates the partition $\{B_1, \dots B_p\}$ of $[n]$ such that the numbers $1, 2, \dots, n$ are arranged in order around a circle, then the convex hulls of the blocks $B_1, \dots, B_p$ are pairwise disjoint [13]. Indeed, let k be the number of isolated points around 1. The 2^{k-1} successors of this configuration are obtained by taking all the subset of $\{\alpha_1 = 1, \alpha_2, \dots, \alpha_k, n+1\}$ containing $n+1$.

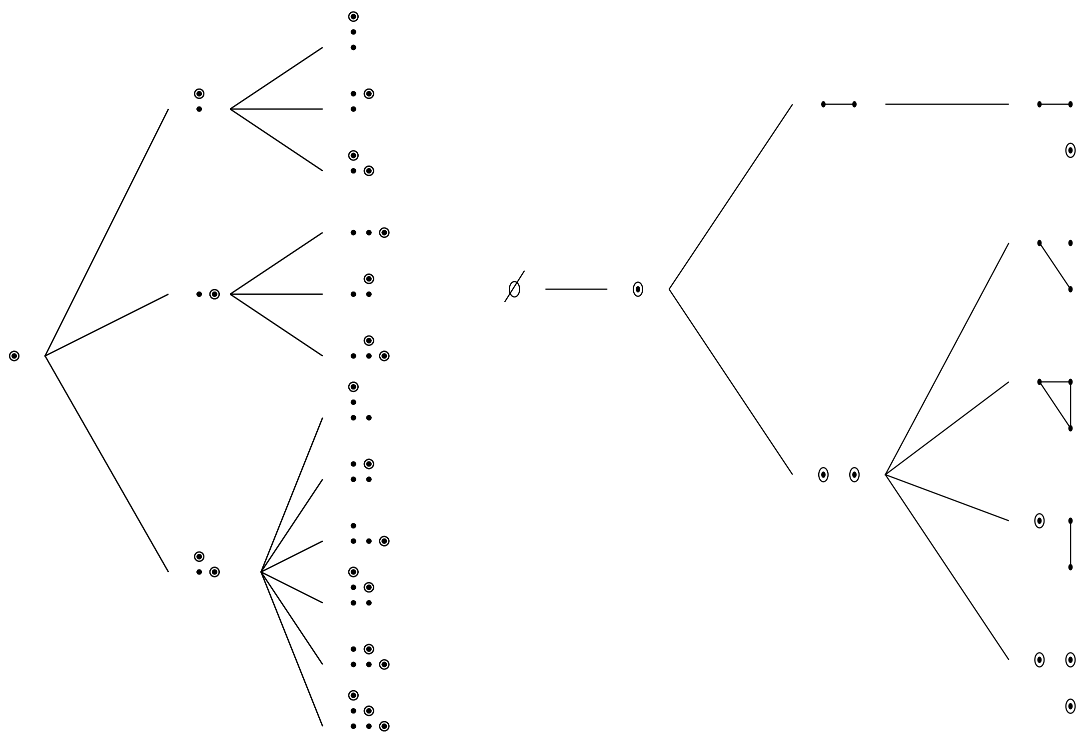


FIGURE 3. Generating trees for FDC animals and Catalan blocks

REFERENCES

[1] C. Banderier, M. Bousquet-Mélou, A. Denise, P. Flajolet, D. Gardy, and D. Gouyou-Beauchamps. Generating functions for generating trees. *Discrete Mathematics*, 2000. An earlier version "On Generating Functions of Generating Trees" was in the proceedings of FPSAC'99 and as INRIA Report 3661.

[2] E. Barcucci, A. Del Lungo, E. Pergola, R. Pinzani, ECO: A methodology for the Enumeration of Combinatorial Objects, *Journal of Difference Equations and Applications,* 5 (1999) 435-490.

[3] F. R. K. Chung, R. L. Graham, V. E. Hoggatt, M. Kleimann, The number of Baxter permutations, *J. Combin. Theory Ser. A,* 24 (1978) 382-394.

[4] S. Corteel. *Séries génératrices exponentielles pour les eco-systèmes signés,* Proceedings of the 12-th International Conference on Formal Power Series and Algebraic Combinatorics, Moscow, 2000.

[5] S. Dulucq, S. Gire, J. West Permutations with forbidden subsequences and nonseparable planar maps, *Discrete Mathematics,* 153 (1996), 85-103.

[6] S. Dulucq, S. Gire, O. Guibert, A combinatorial proof of J.West's conjecture, *Discrete Mathematics,* 187 (1998), 71-96.

[7] S. Dulucq, S. Gire, O. Guibert, Enumération de permutations à motifs exclus, *Actes du 30ième séminaire Lotharingien de Combinatoire, Hesselberg,* Janvier 1993.

[8] D. Merlini and M.C. Verri, Generating trees and proper Riordan Arrays, *Discrete Mathematics,* 218 (2000), 167-183.

[9] D. Merlini, R. Sprugnoli and M.C. Verri, An algebra for Proper Generating Trees, *Proceedings of CMCS, Versailles* (2000), 127-139

[10] E. Pergola, R. Pinzani and S. Rinaldi, A set of well-defined operations on succession rules *Proceedings of CMCS, Versailles* (2000), 141-152.

[11] J. West, Permutations with forbidden subsequences and stack-sortable permutations, *Ph.D. Thesis, MIT, Cambridge, MA* (1990).

[12] J. West, Generating trees and the Catalan and Schröder numbers, *Discrete Mathematics,* 146 (1995), 247-262.

[13] R P. Stanley, Exercises on Catalan and Related Numbers, *Enumerative Combinatorics,* 2 (1999), 221-247.

[14] S. Feretić, A q-Enumeration of Directed Diagonally Convex Polyominoes, *FPSAC'99,* (1999), 195-206.

I3S, URA 1376 du CNRS Sophia-Antipolis, FRANCE
E-mail address: `cgarcia,fedou@unice.fr`

ACTIVITY PRESERVING BIJECTIONS BETWEEN SPANNING TREES AND ORIENTATIONS IN GRAPHS

EMERIC GIOAN AND MICHEL LAS VERGNAS

Abstract. The main results of the paper are two dual algorithms which map in a 1-1 way the set of spanning trees with internal activity 1 and external activity 0 of an ordered graph onto the set of acyclic orientations with adjacent unique source and sink. More generally, these algorithms extend to an activity preserving correspondence between spanning trees and orientations. For certain linear orderings of the edges, they also provide a bijection between spanning trees with external activity 0 and acyclic orientations with a given unique sink.

Résumé. Les principaux résultats de ce papier sont deux algorithmes duaux qui définissent une bijection entre l'ensemble des arbres couvrants d'activité interne 1 et d'activité externe 0 d'un graphe ordonné et l'ensemble de ses orientations acycliques avec un unique puits et une unique source adjacents. Plus généralement ces algorithmes s'étendent à une correspondance préservant les activités entre arbres couvrants et orientations. Pour certains ordres totaux de l'ensemble des arêtes, ils fournissent également une bijection entre arbres couvrants d'activité externe 0 et orientations acycliques ayant un unique puits donné.

1. Introduction

The Tutte polynomial $t(G;x,y)$ of a graph G is a two variable polynomial containing as specializations several fundamental numerical invariants of G such as the numbers of spanning trees, q-colorings, acyclic orientations, etc. We refer the reader to [1] for a comprehensive survey of properties and applications of Tutte polynomials of graphs, and, more generally, matroids.

Suppose the edge-set of G is linearly ordered. W.T. Tutte has shown that

$$t(G;x,y)=\sum_{i,j} t_{i,j}x^iy^j$$

where $t_{i,j}$ is the number of spanning trees of G such that i edges are smallest in their fundamental cocycle and j edges are smallest in their fundamental cycle [15]. On the other hand, M. Las Vergnas has shown that

$$t(G;x,y)=\sum_{i,j} o_{i,j}2^{-i-j}x^iy^j$$

where $o_{i,j}$ is the number of orientations of G such that i edges are smallest in some directed cocycle and j edges are smallest in some directed cycle [11]. This last formula generalizes a well-known result of R. Stanley : the number of acyclic orientations of G is equal to $t(G;2,0)$ [14].

Comparing these two expressions for $t(G;x,y)$ we get $o_{i,j}=2^{i+j}t_{i,j}$ for all i,j. A natural question arises of a bijective proof for this formula [11]. The problem is to define a correspondence between spanning trees and orientations, preserving parameters (i,j), called

1991 *Mathematics Subject Classification.* Primary : 05C99. Secondary : 05B35 52C40.

Key words and phrases. graph, spanning tree, activity, directed graph, acyclic, orientation, source, sink, algorithm, bijection, Tutte polynomial, matroid, oriented matroid.

activities in the literature, and compatible with the above formula. More precisely, the desired correspondence should associate with a (i,j)-*active* spanning tree of G, a set of 2^{i+j} (i,j)-active orientations of G, in such a way that each orientation of G is the image of a unique spanning tree. The main object of the present paper is to describe such a correspondence, which we call the *active correspondence*.

Spanning trees and orientations with $(1,0)$ activities (or, dually, $(0,1)$ activities) constitute the main case of our construction. Several papers of the literature deal with $(1,0)$-orientations of graphs, i.e. acyclic orientations with adjacent unique source and sink. Enumeration problems on $(1,0)$-orientations are considered by C. Greene and T. Zaslavsky in [10] in the contexts of graphs and hyperplane arrangements. In particular, they prove that the number of acyclic orientations of a graph with adjacent unique source and sink is $2\beta(G)$, or, equivalently, we have $o_{1,0} = 2t_{1,0}$ (implying that this number does not depend on the particular source and sink). In [6] bijective proofs are given to a result of [10] on acyclic orientations with unique sink (see below, and Section 6). Orientations with $(1,0)$ activities are studied in [5] for their relevance in several graph algorithms. On the other hand, the external activity of a spanning tree has recently retained some attention in relation with the chip-firing game and the sandpile model [3] (see also [2] for the particular case of K_n and parking functions).

Section 3 contains the main results : two dual algorithms establish a bijection between spanning trees and orientations with $(1,0)$ activities. In Section 4, we obtain as a corollary, a bijection for $(0,1)$ activities. In Section 5, these bijections are extended to a correspondence between spanning trees and orientations consistent with the formula $o_{i,j} = 2^{i+j}t_{i,j}$, thus answering the above query. We point out that this correspondence not only preserves activities but also active elements. The construction uses reductions from general activities to the $(1,0)$ case. Finally, in Section 6, we show that the correspondence of Section 5 produces a bijection between internal spanning trees and acyclic orientations with a unique sink at a given vertex.

A bijection between acyclic orientations with a unique sink at a given vertex and internal spanning trees has recently appeared in the literature [6]. The present one is different, as it preserves activities, whereas the bijection of [6] does not (see Section 6). The bijection of Section 3 constitutes an answer to a question of [6] (see (a) p.145). It should be mentioned that one of the present authors has already published an activity preserving correspondence between spanning trees and orientations in graphs [12], an extended abstract apparently overlooked in [6]. However, the correspondence in [12][13] is different from the present one (see Section 4). It has not been - and maybe cannot be - generalized beyond regular matroids, whereas the correspondence of this paper generalizes to oriented matroids [7][8] [9], and thus, is probably more natural. We point out that the graphical case deserved the present specific treatment, as stronger orthogonality properties permit significative simplifications, and also for properties involving vertices (Section 6).

2. Notation and terminology

The present paper deals exclusively with graphs. We point out that all definitions and results of this section, except Minty's Lemma, have extensions to the wider context of matroid and oriented matroid theories. Throughout the paper, we will implicitly assume that graphs under consideration are connected, and that cycles and cocycles are *elementary* (i. e. minimal for inclusion).

Let G be a graph with edge-set E, and $T \subseteq E$ be a spanning tree of G. For $e \in E \setminus T$, we denote by $C(T;e)$ the *fundamental cycle* of e with respect to T, i.e. the unique cycle contained in $T \cup \{e\}$, obtained from the unique path of T joining the two vertices of e.

For $e \in T$, we denote by $C^*(T;e)$ the *fundamental cocycle* of e with respect to T, i.e. the unique cocycle contained in $(E \setminus T) \cup \{e\}$. The cocycle $C^*(T;e)$ is the set of edges of G joining the two connected components of $T \setminus \{e\}$. For $e \in E \setminus T$ and $f \in T$, we have clearly $f \in C(T;e)$ if and only if $e \in C^*(T;f)$, and then $C(T;e) \cap C^*(T;f) = \{e,f\}$.

We say that a graph G is *ordered* if its edge-set E is linearly ordered. The notion of *activities* of a spanning trees T in an ordered graph G is due to W.T. Tutte [15]. The *internal activity* $\iota(T)$ is the number of edges $e \in T$ smallest in their fundamental cocycle $C^*(T;e)$, and the *external activity* $\varepsilon(T)$ is the number of edges $e \in E \setminus T$ smallest in their fundamental cycle $C(T;e)$. We denote by $t_{i,j}(G)$ the number of spanning trees of G such that $\iota(T) = i$ and $\varepsilon(T) = j$.

Spanning tree activities have been introduced by Tutte to generalize, in a self- dual way, the chromatic polynomial of a graph. The *dichromate*, now called the *Tutte polynomial*, has been originally defined as

$$t(G;x,y) = \sum_{i,j} t_{i,j} x^i y^j.$$

The main point in [15] is to prove that the coefficients $t_{i,j}$ are independent from the ordering of E. Nowadays, the simplest definition of the Tutte polynomial of a graph (or, more generally, a matroid) is by means of a closed formula in terms of subsets of edges, and the above formula is a theorem proved by deletion/contraction of the greatest edge (see [1]).

A cycle resp. cocycle in a directed graph is *directed* if all its edges are directed consistently. The *(primal) orientation activity* of an ordered directed graph G, or *O-activity*, denoted by $o(G)$, is the number of edges smallest in some directed cycle. The *dual orientation activity* of G, or *O^*-activity*, denoted by $o^*(G)$, is the number of edges smallest in some directed cocycle. We denote by $o_{i,j}(G)$ the number of orientations $\overrightarrow{G}$ of G such that $o^*(\overrightarrow{G}) = i$ and $o(\overrightarrow{G}) = j$. The definitions of O- and O^*-activities have been introduced in [11] in view of the formula

$$t(G;x,y) = \sum_{i,j} o_{i,j} 2^{-i-j} x^i y^j$$

This formula implies that $o_{i,j}$ does not depend on the ordering, and that $o_{i,j} = 2^{i+j} t_{i,j}$. The proof in [11] is by deletion/contraction of the greatest edge.

Internal and external activities of spanning trees, and also the two types of orientation activities, are dual notions from the point of view of graph duality. If G is a planar graph imbedded in the plane, and G^* is a dual of G, we have $\varepsilon_{G^*}(T) = \iota_G(E \setminus T)$. If G is directed, a *directed dual* of G is a planar dual G^* directed such that all directions of corresponding edges in G and G^* define rotations of the same type, clockwise or counterclockwise. Then, we have $o^*(G) = o(G^*)$. The graph G is said *acyclic* if there is no directed cycle, i.e. if $o(G) = 0$, and, dually, is said *totally cyclic* (or *strongly connected*) if $o^*(G) = 0$.

In a directed graph, the two possible *cycle directions* along an elementary cycle C can be distinguished by defining C^+ resp. C^- as the set of edges of C directed consistently resp. in the opposite direction. An elementary cocycle D is the set of edges joining two subsets partitioning the vertex-set of G into two connected subgraphs. There are two *cocycle directions* defined by an ordering of these two subsets. The two possible cocycle directions can be distinguished by defining D^+ resp. D^- as the set of edges of D directed from the first subset to the second resp. from the second to the first. In a directed graph, the notation $C(T;e)$ for $e \in E \setminus T$ resp. $C^*(T;e)$ for $e \in T$ can be precised by choosing the cycle direction resp. cocycle direction consistent with the direction of e, i.e. such that e is in the positive part.

We make a crucial use in the proof of Theorem 4 (9) of the (directed) *graphical orthogonality property*

$$|C^+ \cap D^+| + |C^- \cap D^-| = |C^- \cap D^+| + |C^+ \cap D^-|$$

between a cycle C and a cocycle D. In all other places, the weaker (directed) *orthogonality property* $C \cap D \neq \emptyset$ implies $(C^+ \cap D^+) \cup (C^- \cap D^-) \neq \emptyset$ and $(C^- \cap D^+) \cup (C^+ \cap D^-) \neq \emptyset$ suffices for our purpose. We mention that the graphical orthogonality property characterizes regular matroids (Minty 1975), whereas the orthogonality property characterizes oriented matroids (Bland-Las Vergnas 1978).

3. The bijection for $(1,0)$-activities

We recall that $t_{1,0}(G) \neq 0$ if and only if the graph G is 2-connected and has no loop [1].

Proposition 1. *Let G be an ordered directed graph, with smallest edge $e_1 = s's''$ directed from s' to s''. Then $o^*(G) = 1$ and $o(G) = 0$ if and only if G is acyclic, with unique source s' and unique sink s''.*

Proof. A directed graph has external activity 0 if and only if it is acyclic by definition. In an acyclic graph, e_1 belongs to a cocycle, so it is the smallest element of a cocycle. An acyclic graph has a source (otherwise one could construct easily a directed cycle). The set of edges having this source as an extremity is then a directed cocycle.

If the graph has dual activity 1 then this source must be an extremity of e_1 (because e_1 is the only possible minimal element of a cocycle). The same properties holding for the opposite orientation, the graph has a sink and any sink must be an extremity of e_1. This proves that the graph has unique source s' and unique sink s''.

Conversely, suppose G has a unique source s' and a unique sink s''. The two connected subgraphs induced by the partition of V defined by a cocycle are also acyclic. Hence, so they must have a source and a sink. If the cocycle is directed, there exist a source of G in one component and a sink of G in the other. Necessarily these two vertices are s' and s'', and so e_1 belong to the directed cocycle. □

Proposition 2. *Let G be an ordered graph, and T be a spanning tree of G. Set $T = \{b_1 < b_2 < \ldots < b_r\}$ and $E \setminus T = \{a_1 < a_2 < \ldots < a_{n-r}\}$.*

(i) $\varepsilon(T) = 0$ if and only if $b_j = \mathrm{Min}\ (E \setminus \bigcup_{1 \leq i < j} C^*(T; b_i))$ for $j = 1, 2, \ldots, r$.

(ii) $\iota(T) = 1$ if and only if $a_j = \mathrm{Min}\ ((E \setminus \{e_1\}) \setminus \bigcup_{1 \leq i < j} C(T; a_i))$ for $j = 1, 2, \ldots, n - r$.

Proof. (i) Let $e = \mathrm{Min}\ (E \setminus \bigcup_{1 \leq i < j} C^*(T; b_i))$, and suppose $e < b_j$. We have $e \notin T$, since $e \notin \{b_1, \ldots, b_{j-1}\}$ by definition. Set $C = C(T; e)$. If $b_i \in C$, we have $e \in C^*(T; b_i)$, therefore $C \cap \{b_1, \ldots, b_{j-1}\} = \emptyset$. It follows that $C \cap T \subseteq \{b_j, \ldots, b_r\}$, then $e = \mathrm{Min}\ C$, hence $\varepsilon(T) > 0$.

Conversely, suppose $b_j = \mathrm{Min}\ (E \setminus \bigcup_{1 \leq i < j} C^*(T; b_i))$ for $j = 1, 2, \ldots, r$. Let $e \in E \setminus T$. Set $C = C(T; e)$, and let $b_j = \mathrm{Min}\ C \cap T$. We have $e \notin \bigcup_{1 \leq i < j} C^*(T; b_i))$, otherwise $b_i \in C$ for some $i < j$. Hence $b_j < e$, and e is not externally active.

(ii) Let $e = \mathrm{Min}\ ((E \setminus \{e_1\}) \setminus \bigcup_{1 \leq i < j} C(T; a_i))$, and suppose $e < a_j$. We have $e \in T$, since $e \notin \{a_1, \ldots, a_{j-1}\}$ by definition. Set $D = C^*(T; e)$. If $a_i \in D$, we have $e \in C(T; a_i)$, therefore $D \cap \{a_1, \ldots, a_{j-1}\} = \emptyset$. It follows that $D \cap (E \setminus T) \subseteq \{a_j, \ldots, a_{n-r}\}$, then $e = \mathrm{Min}\ D$, hence $\iota(T) > 1$.

Conversely, suppose $a_j = \mathrm{Min}\ ((E \setminus \{e_1\}) \setminus \bigcup_{1 \leq i < j} C(T; a_i))$ for $j = 1, 2, \ldots, n - r$. Let $e \in T \setminus \{e_1\}$. Set $D = C^*(T; e)$, and let $a_j = \mathrm{Min}\ D \setminus T$. We have $e \notin \bigcup_{1 \leq i < j} C(T; a_i))$, otherwise $a_i \in D$ for some $i < j$. Hence $a_j < e$, and e is not internally active. □

The following proposition defines the active correspondence for $(1,0)$- activities.

Proposition 3. *Let G be an ordered graph, with edge-set $E = \{e_1 = s's'' < e_2 < \ldots < e_n\}$, and T be a spanning tree of G with internal activity 1 and external activity 0. The following two algorithms produce the same acyclic orientation of G, with unique source s' and unique sink s''.*

Step 0 (in both algorithms) : direct the smallest edge e_1 from s' to s''.

(i) Algorithm 1

Let $E \setminus T = \{a_1 = e_2 < a_2 < \ldots < a_{n-r}\}$.

Step $i = 1, 2, \ldots, n-r$: direct the undirected edges of $C(T; a_i)$ in the cycle direction opposite to the direction of its smallest edge.

(ii) Algorithm 2.

Let $T = \{b_1 = e_1 < b_2 < \ldots < b_r\}$.

Step 1 : direct all edges $\neq e_1$ of $C^*(T; b_1)$ in the cocycle direction defined by e_1

Step $i = 2, \ldots, r$: direct the undirected edges of $C^*(T; b_i)$ in the cocycle direction opposite to the direction of its smallest edge.

An example for Algorithms 1 and 2 applied to the 4-wheel W_4 is given by Figure 1.

Proof. Since G has a spanning tree T with $(1, 0)$ activities, it has no isthmus or loop.

(1) *Algorithm 1 directs all edges of G, and* $(1')$ *Algorithm 2 directs all edges of G*

We show inductively that all edges a_i $i = 1, 2, \ldots, n-r$ are directed by Algorithm 1. We have to check that after Step $i-1$ the edge $b = \text{Min } C(T; a_i)$ is directed. This is clear for $i = 1$ since then $b = e_1$, so suppose $i \geq 2$. We have $b \in T$, otherwise $b = a_i$ would be externally active. If $b = e_1$, then a_i is directed at Step i of Algorithm 1. If $b \neq e_1$, then b is not the smallest element of its fundamental cocycle since $\iota(T) = 1$. Set $a_j = \text{Min } C^*(T; b)$. We have $a_j < b < a_i$, hence a_j is directed after Step $i-1$ by induction. Since $b \in C(T; a_j)$, the edge b has been directed by Algorithm 1 at a Step $\leq j < i$, hence a_i is directed at Step i. On the other hand, since G has no isthmus, we have $\bigcup_i (C(T; a_i) = E$, hence all edges of T are directed by Algorithm 1.

The proof of $(1')$ is dual.

(2) *Algorithm 1 and Algorithm 2 produce the same orientation of G*

The proof is by induction on the rank in the ordering. Let $a \in E \setminus T$, and set $b = \text{Min } C(T; a)$, $a' = \text{Min } C^*(T; b)$. We have $b \in T$, otherwise $b = a$ is externally active, contradicting $\varepsilon(T) = 0$. If $a' \in T$, the edge $b = a'$ is internally active, hence $b = e_1$ since $\iota(T) = 1$. In this case a and e_1 have opposite directions in $C(T; a)$ for Algorithm 1. We have $a \in C^*(T; e_1)$, hence a and e_1 have the same direction in $C^*(T; e_1)$ by Step 1 of Algorithm 2. These directions agree by orthogonality. Otherwise, we have $b \neq e_1$ and $a' \in E \setminus T$. We have $a \in C^*(T; b)$ and $a' < b < a$. By Algorithm 1, the edges b and a have opposite directions in $C(T; a)$. We have $C(T; a_i) \cap C^*(T; b) = \{a, b\}$, hence by orthogonality a and b have the same direction in $C^*(T; b)$. As b is the smallest edge in T such that $a \in C^*(T; b)$, it follows that a is undirected when b is directed by Algorithm 2. Therefore a and b have same direction in $C^*(T; b)$ for Algorithm 2, opposite to the direction of a'. Since by induction, the directions of b agree in Algorithms 1 and 2, the same conclusion holds for a.

The proof for $b \in T$ is similar and left to the reader.

Let $\overrightarrow{G}$ be the orientation of G constructed by Algorithms 1 and 2.

(3) $o^*(\overrightarrow{G}) = 1$ *and* $(3')$ $o(\overrightarrow{G}) = 0$

Suppose there is a directed cocycle D in $\overrightarrow{G}$ with $\text{Min } D \neq e_1$, contradicting (3). Since G has no isthmus, we have $\bigcup_i C(T; a_i) = E$. Let i be the smallest integer such that $D \cap C(T; a_i) \neq \emptyset$. Let $b \in D \cap C(T; a_i) \setminus \{a_i\}$. Since $b \in C(T; a_i) \setminus \{a_i\}$, we have $b \in T$. By the choice of i, the edge b is directed at step i of Algorithm 1. Set $e = \text{Min } C(T; a_i)$. We have $e \neq a_i$ otherwise a_i would be externally active, contradicting $\varepsilon(T) = 0$. If $i = 1$, we have $a_i = e_2$ and $e = b = e_1$, contradicting our assumption. Hence $i \geq 2$. By (1), the edge e is directed after Step $i-1$ of Algorithm 1 and since b is not we have $e \neq b$. Hence, by definition of Algorithm 1, both b and a_i are directed in the same direction of $C(T; a_i)$, opposite to the direction of e. It follows that all edges in $D \cap C(T; a_i)$ have the same direction in both D and $C(T; a_i)$, contradicting orthogonality.

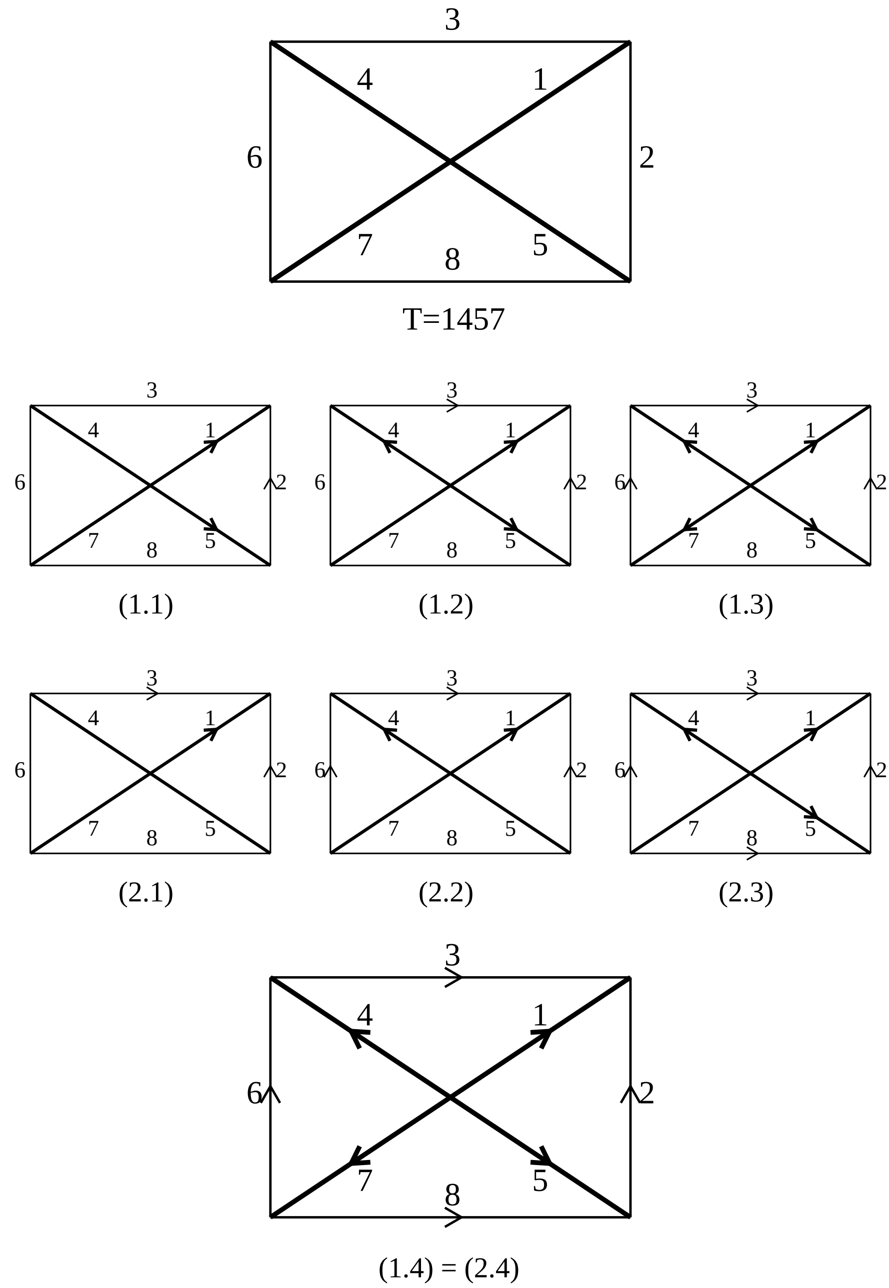


Figure 1

Suppose there is a directed cycle C in $\overrightarrow{G}$, contradicting $(3')$. Since G has no loop, we have $\bigcup_i C^*(T; b_i) = E$. Let i be the smallest integer such that $C \cap C^*(T; b_i) \neq \emptyset$. Let $a \in C \cap C^*(T; b_i) \setminus \{b_i\}$. By the choice of i, the edge a is directed at step i of Algorithm 2. If $i = 1$, i.e. $b_1 = e_1$, then a and b_i have the same direction in $C^*(T; b_i)$ by definition of Step 1 of Algorithm 2. Suppose $i \leq 2$. Set $e = \text{Min } C^*(T; b_i)$. By $(1')$, the edge e is directed after Step $i-1$ of Algorithm 2 and since a is not, we have $e \neq a$. On the other hand, $e \neq b_i$ otherwise b_i would be internally active, implying $i = 1$ since $\iota(T) = 1$. Hence, by definition of Step $i \geq 2$ in Algorithm 2, both a and b_i are directed in the same direction of $C^*(T; b_i)$, opposite to the direction of e. It follows that all edges in $C \cap C^*(T; b_i)$ have the same direction in both C and $C^*(T; b_i)$, contradicting orthogonality. □

Theorem 4. *Let G be an ordered graph. The application defined by Algorithms 1 and 2 is a bijection from the set of spanning trees of G with $(1,0)$ activities onto the set of orientations of G with $(1,0)$ activities such that the direction of the first edge is fixed.*

Proof. Since $2t_{1,0} = o_{1,0}$ by [10], it suffices to show that the application is injective. Suppose there exist two different spanning trees $T = \{b_1 < b_2 < ... < b_r\}$ and $T' = \{b'_1 < ... < b'r\}$ with $(1,0)$ activities such that Algorithms 1 and 2 produce the same directed graph.
(1) Let k be the smallest integer such that $C^*(T; b_k) \neq C^*(T'; b'_k)$. By Proposition 2, we have $b_i = b'_i$ for all $i \leq k$. Set $b = b_k = b'_k$, $D = C^*(T; b)$ and $D' = C^*(T'; b)$. We have $b \in D^+ \cap D'^+$.
(2) $T \cap D' \subseteq \{b = b_k, \ldots, b_r\}$, *and* (1′) $T' \cap D \subseteq \{b = b'_k, \ldots, b'_r\}$
If $i < k$, by (1) we have $b_i = b'_i \notin C^*(T'; b'_k) = D'$.
(3) $T \cap D' \subseteq D'^+$, *and* (3′) $T' \cap D \subseteq D^+$
Let $b_i \in T \cap D'$. By (2), we have $i \geq k$. If $i = k$, then $b_i = b_k = b'_k = b \in D'^+$. Suppose $i > k$. Since $b_i \in D' = C^*(T'; b'_k)$, the edge b_i is directed at a step $j \leq k$ of Algorithm 2 applied to T'. If $j < k$, we have $b'_j = b_j \in T$, hence $b_i \notin C^*(T; b_j) = C^*(T'; b'_j)$, so that b_i cannot be directed at Step j.
Therefore $j = k$. If $k > 1$, the edges $b = b'_k$ and b_i are directed by Algorithm 2 in the same cocycle direction of D' (opposite to the direction of the smallest edge of D'), hence $b_i \in D'^+$. If $k = 1$, then, by definition of Step 1 in Algorithm 2, we have $D' = D'^+$.
(4) $|T \cap D'| \geq 2$ *and* (4′) $|T' \cap D| \geq 2$
Since T is a spanning tree and D' a cocycle, we have $|T \cap D'| \geq 1$. If $|T \cap D'| = 1$, then D' is a fundamental cocycle of T, and necessarily, since $b = b_k \in T$, we have $D' = C^*(T; b) = D$, contradicting the definition of k. Therefore $|T \cap D'| \geq 2$.
(5) Let a be the smallest element of the set

$$\bigcup_{e \in (T \cap D') \setminus \{b\}} C^*(T; e) \cup \bigcup_{e \in (T' \cap D) \setminus \{b\}} C^*(T'; e)$$

which is not empty by (4). By symmetry, we may suppose that $a = \mathrm{Min}\ C^*(T; e)$ for some $e \in (T \cap D') \setminus \{b\}$. We have $e = b_\ell$ for some $\ell > k$ by (2). In particular $\ell > 1$.
(6) $a \notin T$
If $a \in T$, then $a = e$ and $a = \mathrm{Min}\ C^*(T; a)$ is internally active. Hence $a = e_1 = b_1$, contradicting $\ell > 1$ (5).
Set $C = C(T; a)$.
(7) $a \notin T'$
Suppose $a \in T'$.
We have $a > b$ by (6). If $a \in D$, we have $a \in (T' \cap D) \setminus \{b\}$, hence $a \leq \mathrm{Min}\ C^*(T'; a)$ by (5). Therefore a is internally active, hence $a = e_1$, contradicting (6). So $a \notin D$. Since $a > b$, we have also $a \notin D'$.
Let $x \in C \cap D'$. We have $x \neq b$ since $a \notin D$, and $x \neq a$ since $a \notin D'$. Therefore, $x \in (C \setminus \{a\}) \cap D') \setminus \{b\} \subseteq (T \cap D') \setminus \{b\}$. Hence $a \leq \mathrm{Min}\ (C^*(T; x)$, and in fact $a = \mathrm{Min}\ (C^*(T; x)$ since $x \in C = C(T; a)$ implies $a \in C^*(T; x)$. By Algorithm 2 applied to T, the edge x is directed in the cocycle direction opposite to a in the cocycle $C^*(T; x)$, hence by orthogonality a and x have the same cycle direction on C, i.e. $x \in C^+$. On the other hand, we have $x \in D' = C^*(T'; b_k)$ and $x \notin C^*(T'; b'_i) = C^*(T; b_i)$ for $i < k$, since x in T.

Hence, the edge x is directed at Step k of Algorithm 2 applied to T'. Since $x > b_k = b$, the edges b and x have the same cocycle direction in D', i.e. $x \in D'^+$. It follows that $C \cap D' \subseteq C^+ \cap D'^+$.

By (5), we have $a \in C^*(T;e)$, hence $e \in C(T;a) = C$, and also $e \in D'$. We have $e \in C \cap D'$ and $C \cap D' \subseteq C^+ \cap D'^+$, contradicting the orthogonality property.

Set $C' = C(T';a)$. We have $a \in C^+ \cap C'^+$.

(8) $(C \cap D') \setminus \{a,b\} \subseteq C^+ \cap D'^+$ *and* $(8')$ $(C' \cap D) \setminus \{a,b\} \subseteq C'^+ \cap D^+$

We have $C \setminus \{a\} \subseteq T$, hence $(C \cap D') \setminus \{a,b\} \subseteq T \cap D' \subseteq D'^+$ by (3). Let $x \in (C \cap D') \setminus \{a,b\}$ We have $x \in (T \cap D') \setminus \{b\}$, hence $a \leq \text{Min } C^*(T;x)$ by (5). On the other hand $x \in C = C(T;a)$, hence $a \in C^*(T;x)$. It follows that $a = \text{Min } C^*(T;x)$. We have $x = b_i$ with $i > k$. By Algorithm 2 applied to T, at Step i the edge $x = b_i$ is directed in the cocycle direction of $C^*(T;x)$ opposite to the direction of a. Now $C = C(T;a) \cap C^*(T;x) = \{x,a\}$, hence by orthogonality the edges x and a have the same cycle direction in the cycle C, i.e. $x \in C^+$.

(9) $C \cap D' \subseteq \{a,b\}$ *and* $(9')$ $C' \cap D \subseteq \{a,b\}$

Suppose $C \cap D' \subseteq \{a,b\} \neq \emptyset$. By (8) and graphical orthogonality, we have $a \in D'^-$ or $b \in C^-$, and both hold if $\{a,b\} \subseteq C \cap D'$.

Suppose $a \in D'^-$. Then $a \in C' \cap D' \subseteq \{a,b\}$, hence by orthogonality, we have $C' \cap D' = \{a,b\}$ and $b \in C'^+$. By (8') and graphical orthogonality applied to $C' \cap D$, we have $a \in D^-$. Then $a \in C \cap D \subseteq \{a,b\}$, hence by orthogonality, we have $C \cap D = \{a,b\}$ and $b \in C^+$. Therefore $\{a,b\} \subseteq C \cap D'$, both $a \in D'^-$ and $b \in C^-$ should hold : contradiction.

The case $b \in C^-$ is similar, and left to the reader.

(10) By (5), we have $a = \text{Min } C^*(T;e)$, with $e = b_\ell \in (T \cap D') \setminus \{b\}$ and $\ell > k$. We have $e \in C = C(T;a)$, hence $e \in C \cap ((T \cap D') \setminus \{b\}) \subseteq (C \cap D') \setminus \{b\} \subseteq \{a\}$ by (9). Therefore $a = e$. Hence $a = \text{Min } C^*(T;a)$, i.e. a is internally active. Then, necessarily, $a = e_1 = b_1$, since T and T' have internal activity 1, contradicting $e = b_\ell$ with $\ell > 1$ (5). □

4. The bijection for $(0,1)$-activities

The case of $(0,1)$-activities can be reduced to $(1,0)$-activities by duality.

Proposition 5. *Let G be an ordered graph with edge-set $\{e_1 < e_2 \dots\}$.*

(i) If T is a spanning tree with $(1,0)$ activities, then $T \setminus \{e_1\} \cup \{e_2\}$ is a spanning tree with $(0,1)$ activities. The mapping defined by $T \mapsto T \setminus \{e_1\} \cup \{e_2\}$ is a bijection between the sets of spanning trees of G with $(1,0)$ resp. $(0,1)$ activities.

(ii) If $\overrightarrow{G}$ is an orientation of G with $(1,0)$ orientation activties, then $-_{e_1}\overrightarrow{G}$ has $(1,0)$ orientation activities. The mapping defined by $G \mapsto -_{e_1}\overrightarrow{G}$ is a bijection between the sets of orientations of G with $(1,0)$ resp. $(0,1)$ activities. □

The proof of Proposition 5 is straightforward.

Figure 2 shows an application of Proposition 5 to the planar graph W_4 considered in Figure 1. We observe that the $(0,1)$-orientation associated with the spanning tree $T = 2368$ is different from the orientation associated with the same tree by the algorithm of [12] : the edge 8 of [12 Fig.4] is reversed in Figure 2.

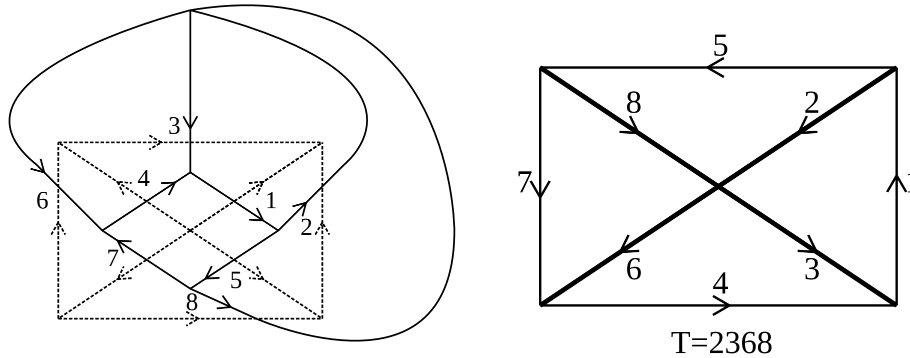


Figure 2

5. The general correspondence

For short, we call *active correspondence*, the particular correspondence we construct in this section, associating with a general spanning tree of activities (i, j) a set of 2^{i+j} orientations with the same activities, in such a way that each orientation is the image of a unique spanning tree.

The main content of this section is that the construction of the active correspondence can be reduced to the $(1, 0)$ case by means of *active partitions* of the edge-set. It turns out that, contrasting with Sections 3,4,6, where specific properties of graphs are used, Section 5 is a mere specialization to graphs of properties holding in matroids and oriented matroids. In consequence, we will only sketch the main results, and refer the reader to [7][8][9] for details and proofs.

Active partitions can be described either in terms of spanning trees in an ordered graph, or of orientations in an ordered directed graph. One main point is that if a spanning tree and an orientation are related by the active correspondence, then either definition produce the same active partition.

The definition of an active partition in terms of spanning trees is much more involved than its definition in terms of orientations. Let G be an ordered graph with edge-set E, and T be a spanning tree of G with activities (i, j). The first step is to construct a set F such that $E \setminus F$ is the set of *internal* elements of T and F its set of *external* elements. For the convenience of the reader, we sketch the construction of F (see [4] for more details and proofs).

Set $F_0 = \emptyset$, and define

$$F_{i+1} = F_i \cup \bigcup_{\substack{e \in E \setminus T \\ C_<(T;e) \subseteq F_i}} C(T; e)$$

where $C_<(T; e)$ is the set of elements of $C(T; e)$ strictly smaller than e. Clearly F_1 is the set of externally active elements of T, and we have $F_1 \subseteq F_2 \subseteq \cdots \subseteq F_i \subseteq \ldots$. Set $F = \cup_i F_i$.

Then F separates the internal and external activities : $T \setminus F$ is a spanning tree with $(i, 0)$ activities of the contraction G/F of G by F, and $T \cap F$ is a spanning tree with $(0, j)$ activities of the subgraph $G(F)$ [4].

Let $\overrightarrow{G}$ be an orientation associated with T by the active correspondence. By a result of G. Minty (1960), in a directed graph an edge belongs either to a directed cycle or to directed cocycle, but not to both. Then F is the *totally cyclic part* of $\overrightarrow{G}$, i.e. the union of all directed cycles of $\overrightarrow{G}$, and $E \setminus F$ is the *acyclic part* of G, i.e. the union of all directed cocycles of $\overrightarrow{G}$.

It follows from this first reduction that without loss of generality, we may restrict the construction to $(i,0)$ or $(0,j)$ activities. Furthermore, internal and external elements, and also totally cyclic parts and acyclic parts, being related by duality (cycles and cocycles play dual parts), we may restrict ourselves to internal spanning trees and acyclic orientations.

The second step reduces the construction to $(1,0)$ activities. The reduction in terms of spanning trees is very similar to the above one, starting with active elements instead of the empty set (see [7][9]). For the sake of simplicity, we describe the reduction in terms of orientations. Of course, by so doing, we rather describe the reciprocal of the active correspondence. But this is clearly equivalent to the direct construction.

Let $\overrightarrow{G}$ be an acyclic orientations of the ordered graph G with $\o^*(\overrightarrow{G}) = i$, and let $a_1 < \dots < a_i$ be its O^*-active edges. Then for $j = 1, 2, \dots, i$ set

$$A_j = \bigcup_{\substack{D \text{ directed cocycle} \\ \text{Min } D = a_j}} D \setminus \bigcup_{\substack{D \text{ directed cocycle} \\ \text{Min } D > a_j}} D$$

Then the *active partition* of $\overrightarrow{G}$ for the orientation is the partition

$$E = A_1 + A_2 + \cdots + A_i$$

Set

$$\overrightarrow{G}_j = \overrightarrow{G}/(A_1 \cup A_2 \cup \ldots \cup A_{j-1}) \setminus (A_{j+1} \cup A_{j+2} \cup \ldots \cup A_i)$$

where, as usual $\setminus$ denotes the deletion, and / denotes the contraction.

Theorem 6. *The graph $\overrightarrow{G}_j$ on the edge-set A_j has $(1,0)$ orientation activities. By reversing the bijection of Section 3 on each graph $\overrightarrow{G}_j$ for $j = 1, 2, \dots, i$, we associate with each $\overrightarrow{G}_j$ a spanning tree T_j with $(1,0)$ activities. Then $T = T_1 + T_2 + \ldots + T_i$ is a spanning tree of $\overrightarrow{G}$ with $(i,0)$ activities. We define the active correspondence by associating the spanning tree T with $\overrightarrow{G}$. This active correspondence has the desired properties, and moreover preserves active elements.*

The proof of Theorem 6, and the statement and proof of the mixed case, when both F and $E \setminus F$ are not empty, can be found in [7][8] in the more general context of oriented matroids. We will illustrate its content in Section 6 on an example (Figures 3 and 4).

The fact that the bijection of Section 3 is actually (1-2) has an immediate consequence. By reversing in all possible ways, independently, all edge directions in each A_j, we get a set of 2^i orientations, called the *activity class* of $\overrightarrow{G}$.

All orientations in an activity class have the same active elements, the same active partition, and are associated with the same spanning tree.

The activitiy classes constitute a partition of the set of orientations of a graph. The active correspondence induces an activity preserving bijection between spanning trees and activity classes of orientations.

6. Application to acyclic orientations with a unique sink

C. Greene and T. Zaslavsky have shown in [10] that the number of acyclic orientations of a graph G with a unique sink at a given vertex is equal to $t(G;1,0)$. In [6], D.D. Gebhard and B.E. Sagan give three bijective proofs of this result. The third one [6 Th.4.1] is by means of an explicit bijection between acyclic orientations with a given unique sink and spanning trees with external activity 0, or *internal* spanning trees, as suggested by the relation $t(G;1,0) = \sum_i t_{i,0}$.

It turns out that the correspondence defined in Section 5 provides another bijection between internal spanning trees and acyclic orientations with a given unique sink, which moreover preserves active edges. The internally active edges of an internal tree becomes O^*-active edges of the orientation.

Lemma 1. *In an ordered graph, the smallest edge of any cocycle belongs to the lexicographically smallest spanning tree.*

Proof. Let G be an ordered graph, T_0 be its lexicographically smallest spanning tree, and D be any cocycle. We have $D \cap T_0 \neq \emptyset$. Set $a = \text{Min } D$ and $b = \text{Min } D \cap T_0$. Then, $a \notin T_0$ implies $a < b$, hence $T_0 \cup \{a\} \setminus \{b\}$ is a spanning tree of G, lexicographically smaller than T_0, a contradiction. □

We say that a spanning tree T in an ordered graph is *increasing with respect to a vertex* s if the edges increase for the ordering along any path of T beginning at s.

Proposition 7. *Let G be an ordered graph such that the lexicographically smallest spanning tree is increasing with respect to a vertex s.*

Then there is exactly one acyclic orientation with a unique sink at s in each activity class of acyclic orientations of G.

Note that the hypothesis implies s is an extremity of the smallest (non loop) edge of G.

Proof. Let T_0 denote the lexicographically smallest spanning tree of $G = (V, E)$. By hypothesis T_0 is increasing.
(1) *The edges of a directed (elementary) cocycle D defined by a 2- partition $V = V_1 + V_2$ in an acyclic orientation $\overrightarrow{G}$ of G with a unique sink at $s \in V_1$ are directed from V_2 to V_1.*

Since $\overrightarrow{G}$ is acyclic, $\overrightarrow{G}(V_2)$ contains at least one sink s'. If the edges of D were directed from V_1 to V_2, then s' would be a sink of G with $s \neq s'$, contradicting the unicity.
(2) *If $\overrightarrow{G}$ is an acyclic orientation of G with a unique sink at s, then the O^*-active edges of T_0 are directed toward s.*

Let a be a O^*-active edge of $\overrightarrow{G}$, and D be a directed cocycle with smallest edge a. By Lemma 1, we have $a \in T_0$. Since T_0 is increasing and a smallest in D, there is no edge of D on the path of T_0 from s to the closest vertex of a. Hence, with notation of (1), this path is in V_1, and by (1) a is directed towards s.

Conversely, let $\overrightarrow{G}$ be the (unique) graph in a given activity class of acyclic orientations of G such that the O^*-active edges of this class are directed towards s on T_0. The graph $\overrightarrow{G}$ exists and is unique by the properties stated in Section 5.
(3) *The graph $\overrightarrow{G}$ has a unique sink at s.*

Since $\overrightarrow{G}$ is acyclic, it has at least one sink s'. The smallest edge a of $\overrightarrow{G}$ incident to s' is in T_0 by Lemma 1. Since the edge a is directed towards s in T_0 by construction of $\overrightarrow{G}$, and T_0 is increasing with respect to s, if $s \neq s'$ then there exists another edge $b < a$ on T_0 incident to s', contradicting the minimality of a. □

Theorem 8. *Let G be an ordered graph, such that the lexicographically smallest spanning tree is increasing with respect to a vertex s.*

Then the mapping sending an internal spanning tree T of G to the unique acyclic orientation with a unique sink at s belonging to the activity class of orientations associated with T by the correspondence of Theorem 6, is an activity prerving bijection from the set of internal spanning trees of G onto the set of acyclic orientations of G with a unique sink at s. □

Theorem 8 is a straightforward corollary of Theorem 6 and Proposition 7. Note that given any spanning tree T in a graph G, and a vertex s, it is always possible - and easy - to linearly order the edges of G so that T is the lexicographically smallest spanning tree and is increasing with respect to s. Label the edges of T by consecutive integers $1, 2, \dots$ in succesive layers defined by their distance to s. After T has been labelled, label arbitrarily the edges not in T.

The bijections provided by Theorem 8 are different from the Gebhard-Sagan bijections. We observe that these bijections are activity preserving by construction, whereas Gebbard-Sagan bijections are not in general. The orientation in Figure 1 of [10 p.139] has O^*-activity 2, but the spanning tree constructed by the algorithm has internal activity 3.

Figure 3 uses the graph W_4, also used in Figures 1 and 2, to give an example for Theorem 8. The Tutte polynomial of W_4 is

$$t(W_4; x, y) = x^4 + y^4 + 4x^3 + 4x^2y + 4xy^2 + 4y^3 + 6x^2 + 9xy + 6y^2 + 3x + 3y$$

The graph W_4 has $t(W_4; 1, 0) = 14$ internal spanning trees.

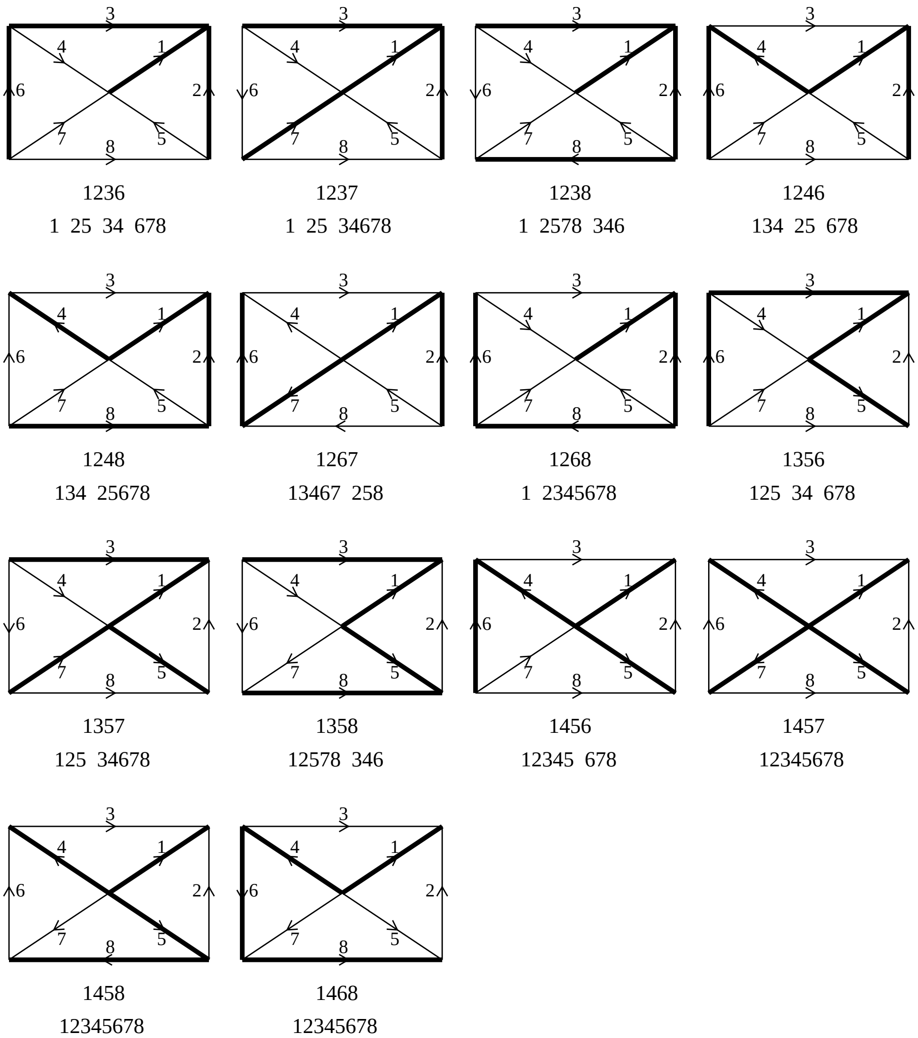


Figure 3

The lexicographically smallest spanning tree 1236 is increasing with respect to the NE (North-East) vertex. For each acyclic orientation with unique sink at the NE vertex, we have indicated the internal spanning tree T given by Theorem 8 (its edges are drawn in heavy lines). We have also indicated the active partition. The internal activity is the number of parts of the active partitions, and the active edges are the first element of each part. By reversing all edge directions in arbitrarily chosen parts of the active partition, we get the activity class associated T. By Proposition 7, in each activity class exactly one acyclic orientation has a unique sink at the NE vertex : this orientation is shown on Figure 3.

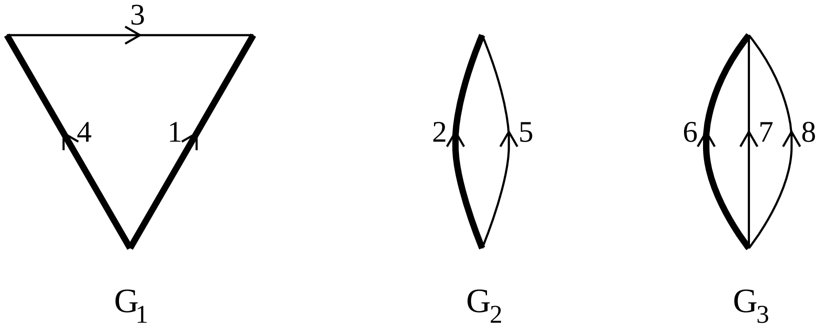


Figure 4

Hence Figure 3 also illustrates the bijection from internal spanning trees to activity classes of acyclic orientation (a restriction of the active correspondence) defined in Section 5.

Figure 4 gives details of the construction of Section 5 for the spanning tree $T = 1246$. The active partition is 134+25+678. The graphs of Theorem 6 are $G_1 = G \setminus 25678$, $G_2 = G/134 \setminus 678$, $G_3 = G/12345$. The spanning trees with $(1, 0)$ activities being unique in these very simple graphs one can check easily that we have $T_1 = 14$, $T_2 = 2$, $T_3 = 6$, and, of course, $1246 = 14 + 2 + 6$.

References

[1] T. Brylawski, J. Oxley, The Tutte polynomial and its applications, in Matroid applications, N. White ed., Cambridge University Press 1992, 123-225

[2] J.S. Beissinger, U.N. Peled, A note on major sequences and external activity in trees, The Wilf Festschrift (Philadelphia, PA, 1996), Electron. J. Combin. 4 (1997), no. 2, Research Paper 4, approx. 10 pp. (electronic).

[3] R. Cori, Y. Le Borgne, Sandpile model and the Tutte polynomial, preprint.

[4] G. Etienne, M. Las Vergnas, External and internal elements of a matroid basis, Discrete Math. 179 (1998), 111-119.

[5] H. de Fraysseix, P. Ossona de Mendez, P. Rosenstiehl, Bipolar orientations revisited, Discrete Applied Mathematics 56 (1995), 157-179.

[6] D.D. Gebhard, B.E. Sagan, Sinks in acyclic orientations of graphs, J. Combinatorial Theory Ser. B 80 (2000), 130-146.

[7] E. Gioan, Correspondance naturelle entre les bases et les réorientations des matroïdes orientés, Thèse de Doctorat, ready soon.

[8] E. Gioan, M. Las Vergnas, Natural activity preserving correspondence between bases and reorientations of oriented matroids, in preparation.

[9] E. Gioan, M. Las Vergnas, On bases, reorientations and linear programming in oriented matroids, Proceedings of the Workshop on Tutte polynomials (Barcelona 2001), Advances in Applied Mathematics, submitted.

[10] C. Greene, T. Zaslavsky, On the interpretation of Whitney numbers through arrangements of hyperplanes, zonotopes, non-Radon partitions and orientations of graphs, Trans. Amer. Math. Soc. 280 (1983), 97-126.
[11] M. Las Vergnas, The Tutte polynomial of a morphism of matroids II. Activities of orientations, in Progress in Graph Theory, J.A. Bondy & U.S.R. Murty eds., Proc. Waterloo Silver Jubilee Combinatorial Conference 1982, Academic Press, Toronto 1984, 367-380.
[12] M. Las Vergnas, A correspondence between spanning trees and orientations in graphs, in Graph Theory and Combinatorics, B. Bollobás ed., Proc. Cambridge Combinatorial Conference in Honour P. Erdös (Cambridge 1983), Academic Press, London 1984, 233-238.
[13] E. Gioan, M. Las Vergnas, A correspondence between spanning trees and orientations in graphs, in preparation.
[14] R.P. Stanley, Acyclic orientations of graphs, Discrete Math. 5 (1973), 171-178.
[15] W.T. Tutte, A contribution to the theory of chromatic polynomials, Canadian J. Mathematics 6 (1954), 80-91.

Université Bordeaux I, LaBRI, 351 cours de la Libération, 33405 Talence, France
E-mail address: `gioan@labri.fr`

Université Pierre et Marie Curie (Paris 6), case 189 - Combinatoire, 4 place Jussieu, 75005 Paris, France
E-mail address: `mlv@ccr.jussieu.fr`

MINUSCULE HEAPS OVER SIMPLY-LACED, STAR-SHAPED DYNKIN DIAGRAMS

HAGIWARA MANABU

Abstract. A minuscule heap is a partially ordered set, together with a labeling of its elements by the nodes of a Dynkin diagram, satisfying certain conditions derived by J. Stembridge. The minuscule heaps over a given Dynkin diagram associated with a symmetrizable Kac-Moody Lie algebra correspond bijectively with the elements of its Weyl group called λ-minuscule by D.Peterson for some integral weight λ, where λ is not necessarily dominant. This paper classifies the minuscule heaps over a class of Dynkin diagrams which we call star-shaped.

Résumé. Un empilement minuscule est un ensemble partiellement ordonné avec un étiquetage de ses éléments par les nœuds d'un diagramme de Dynkin, satisfaisant certaines conditions formulées par J. Stembridge. Les empilements minuscules sur un diagramme de Dynkin associé à une algèbre de Kac–Moody symétrisable correspondent biunivoquement aux éléments minuscules de son groupe de Weyl W. Un élément de W est appelé minuscule s'il est λ-minuscule (dans la terminologie de D. Peterson) pour un poids λ entier quelconque, où λ n'est pas nécessairement dominant. Cet article classifie les empilements minuscules sur une classe de diagrammes de Dynkin que nous appelons de forme étoilée.

1. Introduction

The aim of this paper is to classify the minuscule heaps over simply-laced, star-shaped Dynkin diagrams.

A simply-laced, star-shaped Dynkin diagram Γ is a simple graph (without loops or multiple edges) like the one in Figure 1. It has a node o, and several **branches** $R^1, R^2, \ldots, R^l$ emanating from o. We call o the **center** of Γ, and the number of nodes on R^i (not including o) the **length** of the branch R^i. If $l \geq 3$, then o is uniquely determined by Γ. We mainly deal with such cases. If the length of R^i is l_i, then we say that Γ is of **type** $S(l_1, l_2, \ldots, l_r)$.

Γ is an example of a Dynkin diagram, namely an encoding of a generalized Cartan matrix $A = (a_{i,j})_{i,j\in I}$, associated to which is a Kac-Moody Lie algebra $\mathfrak{g} = \mathfrak{g}(A)$(see [1]). The set I indexing the rows and columns of A is the node set of Γ, which we denote by $N(\Gamma)$. $\mathfrak{g}(A)$ is a generalization of a finite dimensional semi-simple Lie algebra, say over $\mathbb{C}$, and defined by a certain presentation determined by A. All simple finite-dimensional cases (types $A_n (n \geq 1), D_n (n \geq 4)$ and $E_n (n = 6, 7, 8)$) are included in our class.

Minuscule heaps arose in connection with the λ-minuscule elements of the Weyl group W of $\mathfrak{g}$. According to R. Proctor [6] and J. Stembridge [9] the notion of λ-minuscule elements of W was defined by D. Peterson in his unpublished work in the 1980's. Let λ be an integral weight for $\mathfrak{g}$. An element w of W is called **λ-minuscule** if it has a reduced decomposition $s_{i_1}, s_{i_2}, \ldots, s_{i_p}$ such that

$$s_{i_l}(s_{i_{k+1}} \ldots s_{i_p}\lambda) = s_{i_{k+1}} \ldots s_{i_p}\lambda - \alpha_{i_k} \text{ for all } 1 \leq k \leq p,$$

and is called minuscule if w is λ-minuscule for some integral weight λ. Here α_{i_k} is the simple root corresponding to s_{i_k}. It is known that a minuscule element is fully commutative ,

The author would like to thank Professors I.Terada and K.Koike for helpful discussions.

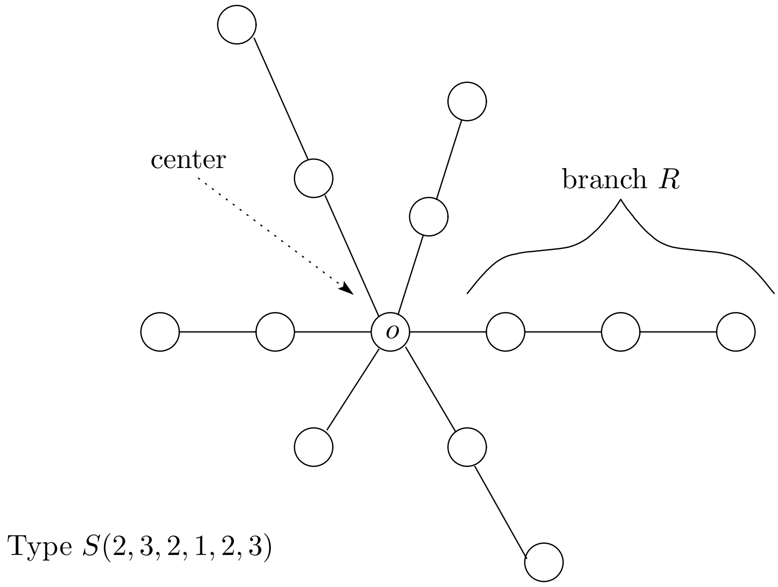


FIGURE 1. A star-shaped Dynkin diagram

namely any reduced decomposition can be converted into any other by exchanging adjacent commuting generators several times (see [6, §15], [7, Theorem A] and [8, Theorem 2.2], or [9, Proposition 2.1]). (We note that there is an element which is fully commutative but not minuscule.) To such an element w, one can associate a Γ-labeled poset called its **heap**. A **Γ-labeled poset** is a triple $(P, \leq \phi)$ in which $(P, \leq)$ is a poset and $\phi : P \to N(\Gamma)$ is any map (called the **labeling map**). A linear extension of a Γ-labeled poset naturally gives a word in the generators of W. The heap of a fully commutative element w is a Γ-labeled poset whose linear extensions give all reduced decompositions of w. A **minuscule heap** is the heap of a minuscule element of W. Stembridge obtained the following structural conditions for a finite Γ-labeled poset to be a minuscule heap ([8, Proposition 3.1])

- **(H1):** If $p \to q$ in P, then $\phi(p)$ and $\phi(q)$ are either equal or adjacent in Γ. Moreover, if $p, q \in P$ are incomparable, then $\phi(p)$ and $\phi(q)$ are not equal, and not adjacent in Γ.
- **(H2):** If $p, q \in P, p < q, \phi(p) = \phi(q) = v$ and no element in $[p, q]$ except p, q are labeled v, then exactly two elements in $[p, q]$ have labels adjacent to v. (This is a simplified version accommodated to the simply-laced cases only.)

The interval appearing in (H2) is important in minuscule heaps, and will be called a **v-interval**. We start from this characterization, namely we define a **minuscule heap** over Γ to be a **finite** Γ-labeled poset $(P, \leq, \phi)$ satisfying (H1) and (H2). The isomorphism classes of minuscule heaps over Γ corresponds bijectively with the minuscule element of W, where an isomorphism is defined to be a poset isomorphism commuting with the labeling maps. R. Proctor showed that, if Γ is simply-laced and λ is dominant, then the minuscule heap constructed from a λ-minuscule element is a d-complete poset, a notion defined by himself. d-complete posets enjoy nice properties such as the hook length formula and jeu de taquin, and are expected to be a nice class of posets that generalize Young diagrams. He introduced the operation of slant sum, and enumerated all 15 types of "slant-irreducible" d-complete posets, namely the ones irreducible with respect to the slant sum decomposition. Then J. Stembridge classified the slant-irreducible minuscule heaps over multiply-laced Dynkin diagrams Γ, where λ was still assumed to be dominant. In this paper, we assume that

Γ is simply-laced and star-shaped, but remove the assumption that λ is dominant. Our motivation to investigate the minuscule heaps without the assumption is connection to skew Young diagrams. In fact a minuscule heap and a dominant minuscule heap over Γ are a skew-Young diagram and a Young diagram if Γ is of type A. Thus we can regard a minuscule heap over Γ as a skew d-complete poset.

As an intermediary for classifying these minuscule heaps over such Γ, we introduce the notion of D-matrices (see §4). They represent the structure of the intervals $[b_o, t_o]$ of minuscule heaps, where b_o and t_o are the smallest and largest elements labeled by o, the "central node" of Γ, respectively. We characterize the D-matrices for any fixed such Γ, and then give a complete description of the set of all minuscule heaps which share the structure of $[b_o, t_o]$ represented by each D-matrix. To describe these minuscule heaps, we introduce the notion of slant lattice over Γ (see §4). It plays the role of a "universal holder" to embed all minuscule heaps over Γ, and provide a "standard coordinate system" to compare them up to isomorphism. Our main results are Theorems 4.8 and 5.6. We also show that a simply-laced Dynkin diagram Γ (not necessarily star-shaped) has only a finite number of minuscule heaps if and only if Γ is of type A_n ($n \geq 1$), D_n ($n \geq 4$) or E_n ($n \geq 6$) (see §6).

The paper is organized as follows. §4, 5 form the main part of this paper, where we classify the minuscule heaps over simply-laced, star-shaped Dynkin diagrams. To reach there, we collect some basic facts in §2, and introduce the notion of the slant lattice in §3. §6 is devoted to the above question on the finiteness of minuscule heaps.

2. Preliminaries

First note that all poset appearing in this paper, including infinite ones, satisfy the following condition:

(*)If $p, q \in P$ and $p \leq q$, then there exists a finite sequence of elements of P, say $p_0, p_1, \ldots, p_l$, such that $p_0 = p, p_l = q$ and p_i covers p_{i-1} for $1 \leq i \leq l$.

We call such a sequence $p_0, p_1, \ldots, p_l$ a **saturated chain from p to q**.

Let $(P, \leq, \phi)$ be a Γ-labeled poset. For each $v \in N(\Gamma)$, we denote by P_v the set of all elements in P labeled v. For $\Gamma' \subset \Gamma$, we denote $\cup_{v \in N(\Gamma')} P_v$ by $P_{\Gamma'}$. It is each to see the following.

Proposition 2.1. *Let Γ be any Dynkin diagram. Let $(P, \leq, \phi)$ be a Γ-labeled poset satisfying (H1), and v a node of Γ. Then P_v is totally ordered.*

Now let $(P, \leq, \phi)$ be a minuscule heap over Γ. By the **support** of P we mean the image of ϕ, which is denoted by supp$P$. Minuscule heaps with acyclic support has additional nice properties. Following [8], we denote this condition by (H4), namely

- **(H4):** supp$P$ is acyclic. ((H3) is used in [8] for another condition for dominant minuscule heaps.)

Note that (H4) is always satisfied if Γ is star-shaped (see §4), since such Γ are acyclic.

Proposition 2.2. *Let $(P, \leq, \phi)$ be a minuscule heap over Γ.*

(1) *If C is a convex subset of P, then $(C, \leq_{|C}, \phi_{|C})$ is a minuscule heap over Γ, where $\leq_{|C}$ and $\phi_{|C}$ are the restrictions of the ordering $\leq$ and over Γ. In particular, all order ideals, order filters, intervals, open intervals, and connected components of P are minuscule heaps over Γ.*

(2) *The dual poset of P is a minuscule heap over Γ. Namely, $(P, \leq^*, \phi)$ is a minuscule heap over Γ.*

It is also easy to see the following.

Proposition 2.3. *Let $(P, \leq, \phi)$ be a Γ-labeled poset satisfying (H1). Then P is connected if and only if* supp$P$ *is connected.*

We say that two subdiagrams Γ_1 and Γ_2 of Γ are **strongly disjoint** if their node sets are disjoint and if no node of Γ_1 is adjacent to any node of Γ_2 in Γ.

Remark 2.4. Let $P_1, P_2, \ldots, P_c$ be the connected components of P. Proposition 2.3 implies that the subdiagrams Γ_i of Γ with node sets $\phi(P_i), i = 1, 2, \ldots, c$, are connected and pairwise strongly disjoint. Hence $\Gamma_1, \Gamma_2, \ldots, \Gamma_c$ are the connected components of supp$P$. This establishes a one-to-one correspondence between the connected components of P and those of supp$P$. A Γ-labeled poset is a minuscule heap over Γ if and only if its connected components are minuscule heaps over Γ and their supports are pairwise strongly disjoint.

Our aim is to classify the minuscule heaps P over simply-laced, star-shaped Γ up to isomorphism of Γ-labeled posets. By Remark 2.4, it is sufficient to study each connected component. At most one of the connected components contains o in its support, and the rest have supports of type A. The type A minuscule heaps turn out to be all of the labeled posets described in [1, 1] (isomorphic to skew Young diagrams). So we concentrate on the case where supp$P$ is connected and $o \in$ supp$P$.

Let Γ and Γ' be Dynkin diagrams. We say that a Γ-labeled poset $(P, \leq, \phi)$ and a Γ'-labeled poset $(P', \leq', \phi')$ are **abstractly isomorphic** (or isomorphic if no confusion would arise) if there is a poset isomorphism $\alpha : P \to P'$ and an isomorphism of subdiagrams $\beta :$ supp$P \to$ suppP' such that β maps the label of p to the label of $\alpha(p)$ for every $p \in P$. For an integer $k \geq 3$, we denote by $d_k(1)$, as was done by Proctor [6], the labeled poset illustrated in Figure 2. An interval$[p, q]$ abstractly isomorphic to $d_k(1)$ will be called a **double-tailed diamond**, with the special case where $k = 3$ being called a **diamond**.

Let Γ be any Dynkin diagram. The following proposition is due to Stembridge.

Proposition 2.5. [8, Propositin 3.3] *Let $(P, \leq \phi)$ be a minuscule heap satisfying $(H4)$. Let v be a node in $N(\Gamma)$ and let $[p, q]$ be a v-interval. Then $[p, q]$ is a double-tailed diamond. In particular, if q covers two distinct elements, then $[p, q]$ is a diamond.*

Remark 2.6. In [8], Stembridge calls $d_k(1)$ a subinterval of type D_k.

The following proposition is also due to Stembridge.

Proposition 2.7. [8, Corollary 3.4] *If a minuscule heap $(P, \leq, \phi)$ satisfies $(H4)$, then P is a ranked poset, i.e. there exists a function $f : P \to \mathbb{Z}$, called a rank function, such that $f(q) = f(p) + 1$ for any covering pair $p \to q$, which we mean $p < q$ and $(p, q) = \emptyset$.*

3. The slant lattice

In this section we define the notion of the slant lattice over an acyclic Dynkin diagram Γ, and show that every minuscule heap over Γ can be "cover-embedded", which we define below, into this poset. For the moment, we do not assume that Γ is acyclic. We only assume (H4).

Now we define the slant lattice. From this point, we assume that Γ itself is connected and acyclic, so that any minuscule heap over Γ satisfies (H4).

For $(u, i), (v, j) \in N(\Gamma) \times \mathbb{Z}$, we write $(u, i) \to (v, j)$ if and only if $j = i + 1$ and v, u are adjacent nodes of Γ. We write $\leq$ for the reflective and transitive closure of $\to$.

Lemma 3.1. *Suppose that Γ is connected and acyclic, and let $\to, \leq$ be the relations on $N(\Gamma) \times \mathbb{Z}$ defined above.*

(1): *$\leq$ is a partial ordering in $N(\Gamma) \times \mathbb{Z}$.*

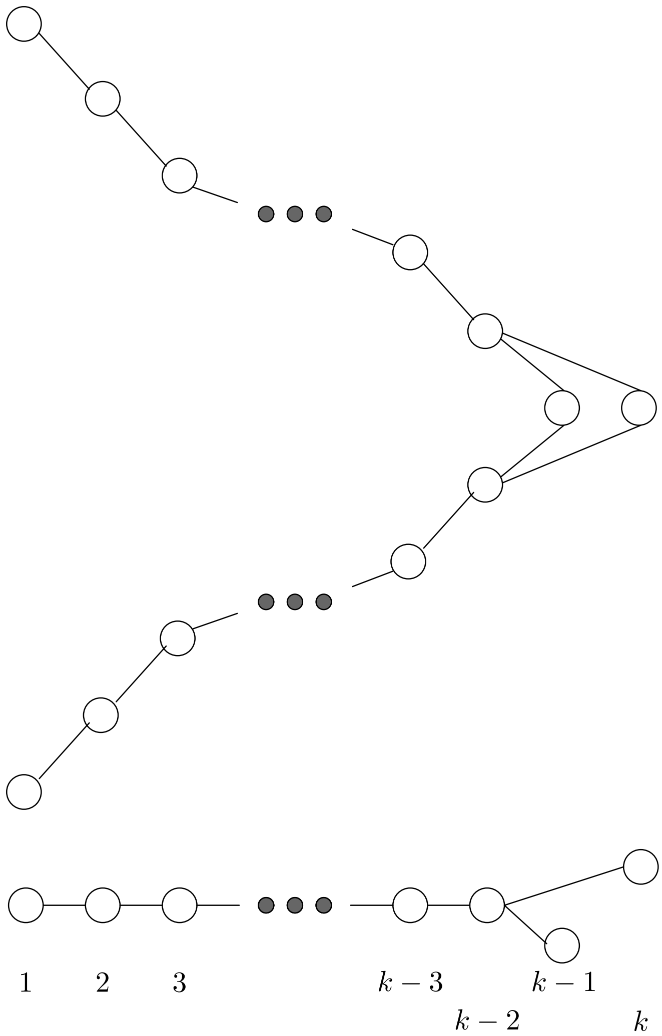


FIGURE 2. The double-tailed diamond $d_k(1)$

Let (u, i) and (v, j) be elements of $N(\Gamma) \times \mathbb{Z}$.

- **(2):** *If Γ contains at least 2 nodes, then we have $(u, i) \leq (v, j)$ if and only if $i \leq j, d(u, v) \leq j - i$, and $d(u, v) \equiv j - i \pmod 2$. Here $d(u, v)$ denotes the distance between u and v in Γ, namely the smallest $l \in \mathbb{Z}_{\geq 0}$ such that there exists a sequence $u = u_0, u_1, \ldots, u_l = v$ of nodes, or ∞ if no such l exists.*
- **(3):** *(u, i) is covered by (v, j) in $N(\Gamma) \times \mathbb{Z}$ if and only if $(u, i) \to (v, j)$.*

If S is a subset of a poset P, we consider two orderings in S induced from P. One is just the restriction of the ordering P. The subset S equipped with this ordering will be simply called a **subposet** of P (**in the ordinary sense** if it is ambiguous). The other ordering, generally weaker than the one above, is obtained by first taking the covering relation in P, restricting it to S, and then taking its reflexive-and-transitive closure. It is straightforward to check that this is in fact a partial order. In this ordering, two elements $p, p' \in S$ are in order if and only if there is a (finite) saturated chain $p = p_0, p_1, \ldots, p_l = p'$ of P consisting solely of elements of S. It can be checked that $p \in S$ is covered by $p' \in S$ in this ordering if and only if p is covered by p' in P. (This is not the case with the restriction of the ordering of P.) We call it the ordering **cover-induced** from P, and we call S together with this ordering a **cover-subposet** of P. Note that, for a general P, the ordering cover-induced

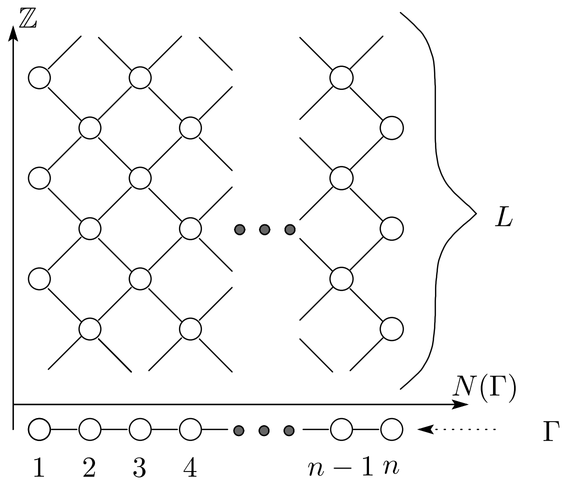


FIGURE 3. A slant lattice of type A

on P itself may be strictly weaker than the original ordering, but our assumption (*) on P assures that this does not happen. Now suppose P and Q are posets. We say that a map $\phi : P \to Q$ is a **cover-embedding** if it gives a poset isomorphism of P with the cover-subposet $\phi(P)$ of Q, namely if p is covered by p' in P if and only if $\phi(p)$ is covered by $\phi(p')$ in Q.

For a minuscule heap P over Γ, there is a unique rank function f on P up to an additive constant for each connected component. Naturally f induces the following injection ν from P to $N(\Gamma) \times \mathbb{Z}$,

$$\nu : p \mapsto (\phi(p), f(p)).$$

We regard $N(\Gamma) \times \mathbb{Z}$ as a Γ-labeled poset by defining the label of each element (v, i) to be v.

Proposition 3.2. *Assume that Γ is acyclic. Let $(P, \leq, \phi)$ be a connected minuscule heap over Γ, let f be a rank function on P, and let ν be the map defined above, Then ν is a cover-embedding that commutes with the labeling maps.*

From now on, assume that Γ is connected. If we fix an element p of P, we can choose a rank function f such that $f(p) = 0$. We define a **slant lattice L over Γ**

$$\text{by } L = \{(q, u) \in N(\Gamma) \times \mathbb{Z} | f(q) - d(v, u) \equiv 0 \pmod 2\}$$

(see Figure 3). If Γ contains at least two nodes, then L coincides with the connected component of the poset $N(\Gamma) \times \mathbb{Z}$ containing $(\phi(p), 0)$. Our definition of L depends on the choice of (p, v), but it is unique up to a shift along the $\mathbb{Z}$ axis. Namely, Suppose we have another slant lattice L' constructed from another element $p' \in P$ and a rank function f'. If $f'(p) \equiv 0 \pmod 2$, then we have $L' = L$. If $f'(p) \equiv 1 \pmod 2$, then we have $L' = \{(v, i) \in N(\Gamma) \times \mathbb{Z} | (v, i-1) \in L\}$. If P is not connected, then we may choose f so as to embed $\mathrm{Im}\nu \subset L$.

The ordering in L induced from $N(\Gamma) \times \mathbb{Z}$ in the usual sense coincides with the ordering cover-induced from $N(\Gamma) \times \mathbb{Z}$. The following is clear.

Corollary 3.3. *Let Γ be a connected acyclic Dynkin diagram, and let $(P, \leq \phi)$ be a connected minuscule heap over Γ. Let f be as above, and let Γ be the slant lattice over Γ defined by f. Then the corresponding ν is a cover-embedding of P into L.*

Let Q be a Γ-labeled poset such that there exist a cover-embedding $\nu : Q \to L$ and $\#Q < \infty$. Let f be the restriction of the second projection $\nu(Q)(\subset N(\Gamma) \times \mathbb{Z}) \to \mathbb{Z}$. For

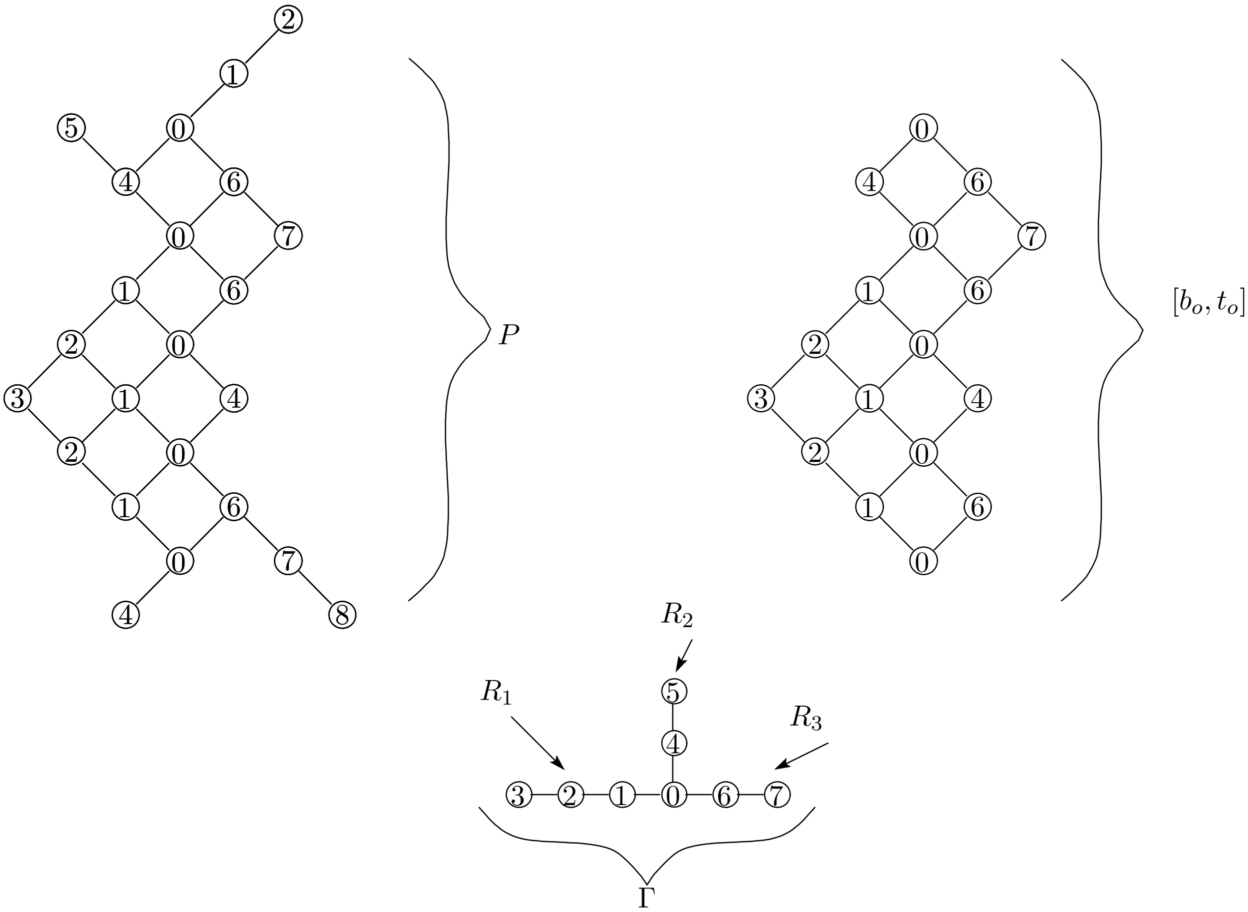


FIGURE 4. A minuscule heap and its core

each $v \in N(\Gamma)$, we can set t_v (resp. b_v) to be the unique maximal (resp. minimal) element of Q_v since L_v is totally ordered. We say that Q_v is **full** if $f(t_v) - f(b_v) = 2r$, where $r+1$ is the number of elements of Q_v.

Proposition 3.4. *Let $(P, \leq, \phi)$ be a minuscule heap satisfying (H4), and let v be a node of Γ. P_v is full if and only if all v-intervals are diamonds.*

4. THE STAR-SHAPED CASE: CORES AND D-MATRICES

Let Γ be a star-shaped Dynkin diagram, and let R be a branch of Γ. We denote the nodes of R by $R_1, R_2, \ldots, R_l$ in the increasing order of the distances from o. We denote by $\bar{R}$ the subdiagram with node set $N(R) \cup \{o\}$, and we sometimes denote o by R_0.

Let P be a minuscule heap over Γ with connected support containing o. By Proposition 2.1, P_o has a unique maximal (resp. minimal) element t_o (resp. b_o). Since $[b_o, t_o]$ is convex, it is also a minuscule heap. We call $[b_o, t_o]$ the **core** of P, and we say that P is **unadorned** if $P = [b_o, t_o]$(see Fig 4). We proceed in two steps. In this section we classify the **unadorned** minuscule heaps over Γ by associating then with what we call D-matrices. In §5, we determine what adornments can be added to the core.

We can determine the possibilities of the R_i-intervals as follow.

Proposition 4.1. *Let $(P, \leq, \phi)$ be a minuscule heap over Γ. Let R be a branch of Γ, and let l be its length. Then*

- *Any o-interval in P is a diamond, namely P_o is full.*

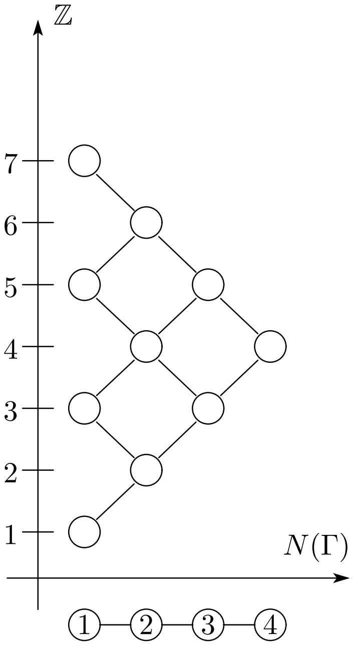


FIGURE 5. A wing over the Dynkin diagram of type A_4

- *If $1 \leq h < l$, then any R_h-interval in P is either a diamond or isomorphic to $d_{h+3}(3)$.*
- *Any R_l-interval in P in isomorphic to $d_{l+3}(3)$.*

(In particular, if Γ is of type A then P_v is full for each $v \in N(\Gamma)$.)

From now on, choose a rank function on f with $f(b_o) = 0$ and choose a slant lattice L which contains $(o, 0)$, namely which contains Imν. We may identify P with Imν.

Now fix a branch R and determine the shape of $[b_o, t_o \cap P_{\bar{B}}$. We distinguish between two kinds of o-intervals, namely the ones containing an element labeled R_1 (which we call R-diamonds) and the rest (non-R-diamonds).

Let Γ' be the Dynkin diagram of type A_n with node set $\{1, 2, \ldots, n\}$ and L' be a slant lattice over Γ' containing $(1, 1)$. We define a subset Q of L' by

$$Q := \{(v, q) | 1 \leq v \leq n, v \leq q \leq 2n - v\}.$$

We regard Q as a cover-subposet of L, and call a Γ'-labeled poset isomorphic to Q a **wing** over Γ'(see Figure 5) of **width** n.

Proposition 4.2. **(1):** *In the above notation, $[o_k, o_{k+s}] \cap P_{\bar{R}}$ is a wind over $\bar{R}$.*

(2): *$[b_o, t_o] \cap P_{\bar{R}}$ is contained in the union of all wings over $\bar{R}$ in P.*

(3): *Two adjacent R-diamond blocks are separated by exactly one non-R-diamond.*

(4): *$[b_o, t_o] \cap P_R$ is contained in the union of all wings over R. If two R-diamonds in P are separated by non-R-diamonds only, then the number of such non-R-diamonds must be one.*

Let $b_0 = o_0, o_1, \ldots, o_c = t_o$ be the elements of P_o in the increasing order. Then $[o_0, o_1]$,$[o_1, o_2]$, ..., $[o_{c-1}, o_c]$ give all o-intervals of P. We call a sequence o-intervals $[o_k, o_{k+1}]$, $[o_{k+1}, o_{k+2}]$, ..., $[o_{k+s-1}, o_{k+s}]$ an R-**diamond block** if $[o_k, o_{k+1}]$, $[o_{k+1}, o_{k+2}]$, ..., $[o_{k+s-1}, o_{k+s}]$ are R-diamonds and $[o_{k-1}, o_k]$, $[o_{k+s}, o_{k+s+1}]$ are non-R-diamonds (or $k = 0$ or $k + s = c$). We call s the **length** of this R-diamond block.

Proposition 4.3. *Let $a_1, \dots, a_r$ be the lengths of the R-diamond blocks in $[b_o, t_o]$ arranged from bottom to top. Then the sequence $a_1, \dots, a_r$ is unimodal: i.e. we have $a_1 \le a_2 \le \cdots \le a_i \ge \cdots \ge a_{r-1} \ge a_r$, for some $1 \le i \le r$.*

Let Γ be the Dynkin diagram of type $S(l_1, \dots, l_r)$ and let $R^1, \dots, R^r$ be the branches of Γ of length $l_1, \dots, l_r$ respectively. We call an $r \times m$ integer matrix $B = (b_{i,j})$, where m is any nonnegative integer, a D-matrix for Γ if it satisfies the following conditions:

(1): $b_{i,j} = 0$ or 1 for all i and j.
(2): For each j, we have $\sum_{i=1}^{r} b_{i,j} = 2$.
(3): For each i, the ith row has the form

$$(\#) \qquad (\overbrace{0, \dots, 0}^{c_i}, \overbrace{1, \dots, 1}^{a_{i,1}}, 0, \overbrace{1, \dots, 1}^{a_{i,2}}, 0, \dots, 0, \overbrace{1, \dots, 1}^{a_{i,s_i-1}}, 0, \overbrace{1, \dots, 1}^{a_{i,s_i}}, \overbrace{0, \dots, 0}^{d_i})$$

for some $s_i \in \mathbb{Z}_{\ge 0}, a_{i,1}, a_{i,2}, \dots, a_{i,s_i} \in [1, l_i]_{\mathbb{Z}}$ and $c_i, d_i \in \mathbb{Z}_{\ge 0}$, and the sequence $a_{i,1}, a_{i,2}, \dots, a_{i,s_i}$ is unimodal. If $s_i = 0$, then this means that all entries in row i are 0.

We include an empty matrix as a special case where $m = 0$. What we saw above and the shape of o-intervals lead us to the following.

Lemma 4.4. *Let P, b_o, t_o, f as above. Define an $r \times (f(t_o)/2)$-matrix $B = (b_{i,j})$ by*

$$b_{i,j} = \begin{cases} 1 & \text{if } [o_{j-1}, o_j] \text{ is an } R^i\text{-diamond} \\ 0 & \text{otherwise,} \end{cases}$$

where o_j is the element of P_o with rank $2j$. Then B is a D-matrix for Γ. We call this B the D-matrix of P.

Example 4.5. *The D-matrix B for Γ constructed from P of Fig. 4 is*

$$B = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

Conversely, we can construct an unadorned minuscule heap for each D-matrix for Γ as follows. Recall that we have fixed a slant lattice L over Γ containing $(o, 0)$.

Lemma 4.6. *Let B be a D-matrix for Γ with m columns. Put*

$$Q = \{(o,0), (o,2), \dots, (o,2m)\} \\ \cup \bigcup_{i=1}^{r} \{(R^i_h, j) \in L \mid 1 \le h \le l_i, b_{i,k} = 1 \text{ for all } k \in [j-h, j+h]\}.$$

Let $\le$ denote the ordering in Q cover-induced from L, and let $\phi : Q \to N(\Gamma)$ denote the restriction of the first projection $N(\Gamma) \times \mathbb{Z} \to N(\Gamma)$. Then $(Q, \le, \phi)$ is a unadorned minuscule heap over Γ.

Example 4.7. *Let us construct the minuscule heap Q from B in Example 4.5. By the definition of Q, we have $Q_0 = \{(0,0), (0,2), (0,4), (0,6), (0,8)\}$. P_{R_1} consists of wing of width 3. P_{R_2} consists of 2 wings of width 1. P_{R_3} consists of 2 wings, and each widths are 2 and 1 from bottom (see Fig. 6). In fact, Q is isomorphic to the core of P.*

Let $\mathcal{H}_0$ denote the set of isomorphic classes of unadorned minuscule heaps over Γ. We can summarize the results of this section as follows. This is the first part of our main result.

Theorem 4.8. *There is a one-to-one correspondence between $\mathcal{H}_0$ and the D-matrices for Γ.*

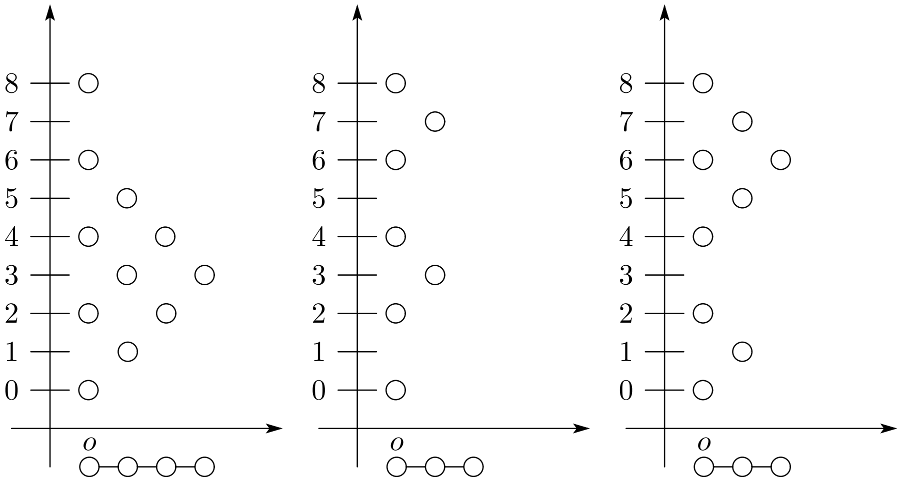


FIGURE 6. The core constructed from the D-matrix B as in Example 4.5

5. The star-shaped case: adornments

In this section, we determine what adornments can be added to the cores.

Let Γ be a simply-laced, star-shaped Dynkin diagram of type $S(l_1, l_2, \ldots, l_r)$, and L be a slant lattice over Γ which contains $(o, 0)$. Let $(P, \preceq) \subset L$ be a connected minuscule heap over Γ with $o \in \mathrm{supp}P$ and $f(b_o) = 0$, where f is the restriction of the second projection $N(\Gamma) \times \mathbb{Z} \to \mathbb{Z}$ and $\preceq$ is the ordering cover-induced from L.

Let R be a branch. We note that $P_{\bar{R}}$ is may not be a minuscule heap.

For every $p, q \in P$, a sequence $p = p'_0, p'_1 \ldots, p'_{l'} = q$ in P such that either $p'_{i-1} \to p'_i$ or $p'_i \to p'_{i-1}$ holds for each $i, 1 \le i \le l$ is called a **Hasse walk from** p **to** q. The following is a key lemma in the proof we omitted below.

Lemma 5.1. *Let p be an element of $P_{\bar{R}}$. Put $h = d(o, \phi(p))$, where $d(,)$ is the distance of two nodes as we have set in §4. Then there exists a unique Hasse walk $p_0, p_1, \ldots, p_h$ in $P_{\bar{R}}$ such that*

- **(1):** $\phi(p_0) = o$, $\phi(p_i) = R_i (1 \le^{\forall} i \le h)$ *and* $p_h = p$.
- **(2):** *If p_{j-1} is covered by p_j, then no element of $P_{R_{j-1}}$ covers p_j in P.*

(These conditions say that, if we regard the sequence as a walk from p to p_0, we keep moving closer to P_o incessantly, and we go up instead of down whenever possible.)

We call such a Hasse walk in Lemma 5.1 the **approach to** p **from above**. We call a Hasse walk which is the approach to p from above in the dual poset the **approach to** p **from below**. In the sequel, we investigate the form of a connected component Q of the cover-subposet $P_{\bar{R}}$ (resp. P_R) of P (and hence of L). We simply call such a subset a connected component of $P_{\bar{R}}$.

By Lemma 5.1, we have $\phi(Q) = \{o, R_1, R_2, \ldots, R_m\}$ for some $m \ge 0$. The following three Propositions determine the possible shapes of Q.

Proposition 5.2. *Let Q be a connected component of $P_{\bar{R}}$. If $\phi(Q)$ is of type A_{m+1}, then* $\#Q_{R_m} = 1$.

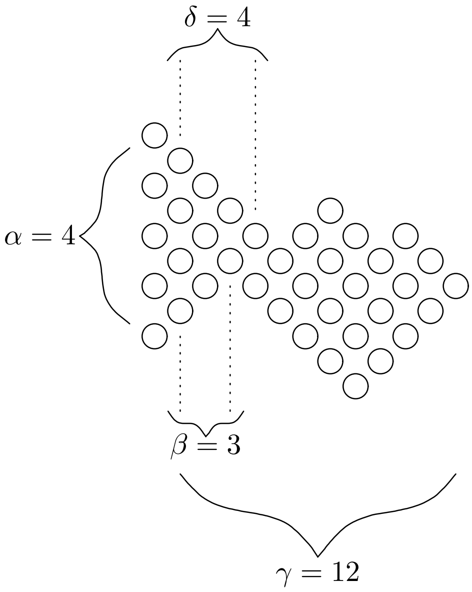


FIGURE 7. An element of $\mathcal{B}^{4,3}_{12,4}$

Proposition 5.3. *Let* (R_0, j_0),(R_1, j_1), $\dots$, (R_m, j_m) *(resp.* (R_0, i_0),(R_1, i_1), $\dots$, (R_m, i_m)*) be the approach to the unique element of* Q_{R_m} *from above (resp. below). Then we have*

$$Q = \bigcup_{0 \le k \le m} \{(R_k, h) \in L | i_k \le h \le j_k\}. \tag{1a}$$

We call the approach to the unique element of Q_{R_m} from above (resp. below) **the upper (resp. the lower) boundary of** Q.

Let $\alpha, \beta, \gamma, \delta$ be nonnegative integers. We define $\mathcal{B}^{\alpha,\beta}_{\gamma,\delta}$ (see Figure 7) to be the set of all subsets N of L such that

$$N = \{(R_k, h) | i_k \le h \le j_k, 0 \le k \le \gamma\}$$

for some Hasse walks (R_0, i_0),(R_1, i_1), $\dots$, (R_γ, i_γ) and (R_o, j_0),(R_1, j_1), $\dots$, (R_γ, j_γ) in L such that

- **(2a):** $j_0 - i_0 = 2\alpha$,
- **(2b):** $(R_0, i_0) \to (R_1, i_1) \to \cdots \to (R_\beta, i_\beta)$,
- **(2c):** $(R_\beta, i_\beta) \leftarrow (R_{\beta+1}, i_{\beta+1})$ if $\beta \ne \gamma$,
- **(2d):** $(R_0, j_0) \leftarrow (R_1, j_1) \leftarrow \cdots \leftarrow (R_\delta, j_\delta)$,
- **(2e):** $(R_\delta, j_\delta) \to (R_{\delta+1}, j_{\delta+1})$ if $\delta \ne \gamma$, and
- **(2f):** $i_k \le j_k$ for each $0 \le k \le \gamma - 1$, and $i_\gamma = j_\gamma$.

Proposition 5.4. *Let* $\alpha, \beta, \gamma, \delta$ *be nonnegative integers. Then* $\mathcal{B}^{\alpha,\beta}_{\gamma,\delta} \ne \emptyset$ *if and only if (3a)-(3c) hold.*

- **(3a):** $\alpha, \beta, \delta \le \gamma \le l$
- **(3b):** $\beta \le \alpha$ *or* $\delta \le \alpha$
- **(3c):** *If* $\beta, \delta < \gamma$, *then* $\gamma \ge \alpha + 2$.
  If $\beta < \gamma$ *or* $\delta < \gamma$, *then* $\gamma \ge \alpha + 1$.

Let B be a D-matrix for Γ. Let P be a minuscule heap over Γ, not necessarily unadorned, whose $[b_o, t_o]$ is represented by B. Let R be a branch of Γ and let

$(\overbrace{0,\dots,0}^{c_R},\overbrace{1,\dots,1}^{a_{R,1}},0,\overbrace{1,\dots,1}^{a_{R,2}},0,\dots,0,\overbrace{1,\dots,1}^{a_{R,e}},\overbrace{0,\dots,0}^{d_R})$ be the row of B corresponding to R. Put

$$\alpha_R = (\alpha_{R,1}, \alpha_{R,2}, \dots, \alpha_{R,h_R}) = (\overbrace{0,0,\dots,0}^{c_R}, a_{R,1}, a_{R,2}, \dots, a_{R,e}, \overbrace{0,0,\dots,0}^{d_R}).$$

Let $Q_R^1, Q_R^2, \dots, Q_R^{h_R}$ be the connected components of $P_{\bar{R}}$ from bottom to top. Then there are unique nonnegative integers $\beta_{R,i}, \gamma_{R,i}, \delta_{R,i}$ such that $Q_R^i \in \mathcal{B}^{\alpha_{R,i},\beta_{R,i}}_{\gamma_{R,i},\delta_{R,i}}$. Like α_R, we put $\beta_R = (\beta_{R,1}.\beta_{R,2}, \dots, \beta_{R,h_R})$, $\gamma_R = (\gamma_{R,1}, \gamma_{R,2}, \dots, \gamma_{R,h_R})$ and $\delta_R = (\delta_{R,1}, \delta_{R,2}, \dots, \delta_{R,h_R})$.

Example 5.5. *Let us calculate $\alpha, \beta, \gamma, \delta$ corresponding to Fig. 4.*

$$\begin{aligned}
\alpha_{R_1} &= (3,0), \alpha_{R_2} = (0,1,1), \alpha_{R_3} = (1,2)\\
\beta_{R_1} &= (3,2), \beta_{R_2} = (0,1,2), \beta_{R_3} = (1,2)\\
\gamma_{R_1} &= (3,2), \gamma_{R_2} = (1,1,2), \gamma_{R_3} = (3,2)\\
\delta_{R_1} &= (3,0), \delta_{R_2} = (1,1,1), \delta_{R_3} = (3,2)
\end{aligned}$$

Conversely, we can construct minuscule heaps from a collection of such elements of $\mathcal{B}^{\alpha,\beta}_{\gamma,\delta}$ as follows. The following theorem gives a complete parameterization of the (isomorphism classes) of minuscule heaps having a fixed D-matrix B. This is the second part of our main result. We omit the arguments to check that the resulting subsets of L are actually minuscule heaps.

Theorem 5.6. [3] *Let B be a D-matrix for Γ, and let P denote the unadorned minuscule heap over Γ corresponding to B constructed in Lemma 4.6. For each branch R, define an integer sequence $\alpha_R = (\alpha_{R,i})_{i=1}^{h_R}$ from B as above. Let $\beta_R = (\beta_{R,i})_{i=1}^{h_R}$, $\gamma_R = (\gamma_{R,i})_{i=1}^{h_R}$, $\delta_R = (\delta_{R,i})_{i=1}^{h_R}$ be integer sequences satisfying the following conditions:*

- **(1):** *For each R and $1 \le i \le h_R$, the quadruple $\alpha_{R,i}, \beta_{R,i}, \gamma_{R,i}, \delta_{R,i}$ satisfy the conditions in Proposition 5.4.*
- **(2):** *For each R, the sequence γ_R is unimodal.*
- **(3):** *If $\delta_{R,i} < \gamma_{R,i}$, then $\beta_{R,i+1} = \gamma_{R,i+1} < \delta_{R,i}$ $(1 \le i < h_R)$.*
- **(4):** *If $\beta_{R,i} < \gamma_{R,i}$, then $\delta_{R,i-1} = \gamma_{R,i-1} < \beta_{R,i}$ $(1 < i \le h_R)$.*

For each R and $1 \le i \le h_R$, choose $Q^{R,i} \in \mathcal{B}^{\alpha_{R,i},\beta_{R,i}}_{\gamma_{R,i},\delta_{R,i}}$, and replace the ith wing over R (counted from the bottom) in P by $Q^{R,i}$. These $Q^{R,i}$ do not overlap with one another, and the resulting Γ-labeled poset P' is a minuscule heap over Γ with connected support containing o, having B as the D-matrix, the $Q^{R,i}, i = 1, 2, \dots, h_R$, being the connected components of $P'_{\bar{R}}$ for each R. Moreover, all minuscule heaps over Γ with connected support containing o are obtained in this manner.

6. The MH-finite Dynkin diagrams

We say that a Dynkin diagram Γ is **MH-finite** if the number in the isomorphism classes of minuscule heaps over Γ is finite (or equivalently the number of minuscule elements of the corresponding Weyl group is finite). Our goal in this section is the following theorem,

Theorem 6.1. *A connected simply-laced Dynkin diagram is MH-finite, if and only if Γ is of type $A_n (n \ge 1)$, $D_n (n \ge 4)$, and $E_n (n \ge 6)$.*

Let W be the Weyl group of the Kac-Moody Lie algebra associated with Γ. Stembridge and R. Proctor showed (see [6, §15], [7, Theorem A] and [8, Theorem 2.2], or [9, Proposition 2.1]) that any minuscule element $w \in W$ is fully commutative, namely any reduced expression for w can be obtained from any other by using only the Coxeter relations that

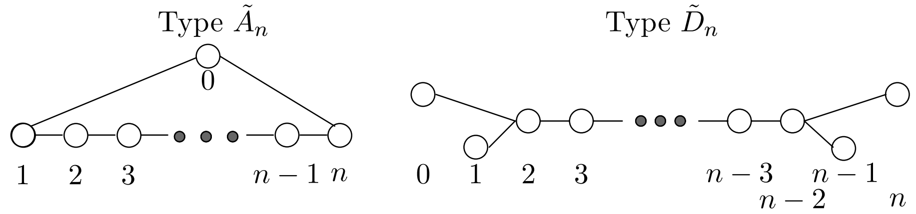


FIGURE 8. The Dynkin diagrams of type $\tilde{A}_n$ and $\tilde{D}_n$

involve commuting generators. On the other hand, Stembridge also showed in [8] that the number of fully commutative elements is finite if and only if Γ is either of type A, D or E if Γ is simply-laced. Thus the "if" part of the theorem is proved.

Lemma 6.2. *Dynkin diagrams of type $\tilde{A}_n (n \geq 2), \tilde{D}_n$ are not MH-finite. (See Fig 8 for the definition of the Dynkin diagram of type $\tilde{A}_n$.)*

By Lemma 6.2, a connected MH-finite Dynkin diagram cannot contain a subdiagram of type $\tilde{A}$ or $\tilde{D}$. Thus it must be star-shaped. We prove this by constructing an infinite series of D-matrices for Γ.

Lemma 6.3. *The Dynkin diagrams of type $S(1,1,1,1), S(2,2,2), S(3,3,1)$ are not MH-finite.*

We can prove Theorem 6.1 by Lemmas above and a few argument.

REFERENCES

[1] S.Billey, W.Jockusch and R.Stanley, Some combinatorial properties of Schubert polynomials, J. Algebraic Combinatorics, 2, 1993, 345-374
[2] C. K. Fan, A Hecke Algebra Quotient and Properties of Commutative Elements of a Weyl Group, Ph.D. thesis, MIT, 1995.
[3] Hagiwara.M, Minuscule Heaps Over Simply-laced, Star-shaped Dynkin Diagrams, preprint.
[4] J. Humphreys, "Reflection groups and Coxeter groups," Cambridge Univ. Press, 1990.
[5] V. G. Kac, "Infinite Dimensional Lie Algebras," Cambridge Univ. Press, Cambridge, 1990.
[6] Robert A. Proctor, Dynkin Diagram Classification of λ-Minuscule Bruhat Lattices and of d-Complete Posets, Journal of Algebraic Combinatorics, 9, 1999, 61-94.
[7] Robert A. Proctor, Minuscule elements of Weyl groups, the numbers game, and d-complete posets, J.Algebra, 213, 1999, 272-303
[8] John R. Stembridge, Minuscule Elements of Weyl Groups, J. Algebra, 235, 2001, 722-743.
[9] John R. Stembridge, On the Fully Commutative Elements of Coxeter Groups, J. Algebraic Combinatorics, 5, 1996, 353-385.

GRADUATE SCHOOL OF MATHEMATICAL SCIENCES, UNIVERSITY OF TOKYO, KOMABA 3-8-1, MEGURO-KU, TOKYO 153-8914, JAPAN
E-mail address: manau@ms.u-tokyo.ac.jp

A CONJECTURED COMBINATORIAL FORMULA FOR THE HILBERT SERIES FOR DIAGONAL HARMONICS

J. HAGLUND AND N. LOEHR

Abstract. We introduce a conjectured way of expressing the Hilbert Series of diagonal harmonics as a weighted sum over parking functions. Our conjecture is based on a pair of statistics for the q,t-Catalan sequence discovered by M. Haiman and proven by the first author and A. Garsia. We show how our q,t-parking function formula for the Hilbert Series can be expressed more compactly as a sum over permutations. We also derive two equivalent forms of our conjecture, one of which is based on the original pair of statistics for the q,t-Catalan introduced by the first author and the other of which is expressed in terms of rooted, labelled trees.

Résumé. Nous présentons une façon conjecturelle d'exprimer les séries de Hilbert des harmoniques diagonales en tant que sommes pondérées de fonctions de stationnement. Notre conjecture repose sur une paire de statistiques (découverte par M. Haiman et ensuite prouvée par le premier auteur et A. Garsia) associé e à la séquence de q,t-Catalan. Nous montrons comment notre formule de q,t-stationnement pour la série de Hilbert peut s'exprimer de façon compacte en utilisant une somme sur les permutations. Nous dérivons aussi deux formes équivalentes de notre conjecture. La première, introduite par le premier auteur, est basée sur la paire originale de statistiques associée à la séquence de q,t-Catalan, alors que la seconde s'exprime en termes d'arbres enracinés et étiquetés.

1. Background

In the early 1990's Garsia and Haiman introduced the following conjecture [7].

Conjecture 1.1. *For each positive integer n define a rational function $C_n(q,t)$ by*

$$C_n(q,t) = \sum_{\mu\vdash n} \frac{t^{2\Sigma l} q^{2\Sigma a}(1-t)(1-q)(\sum q^{a'}t^{l'}) \prod^{0,0}(1-q^{a'}t^{l'})}{\prod (q^a - t^{l+1})(t^l - q^{a+1})}, \tag{1}$$

where the outer sum is over all partitions μ of n, the products and sums in the inner summand are over the squares of the Ferrers diagram of μ, and the arm a, coarm a', leg l, and coleg l' of a square are as in Fig. 1. The $0,0$ above the product indicates we leave out the upper-left corner square of the diagram of μ. Then $C_n(q,t)$ is a polynomial in q and t with nonnegative integer coefficients.

Conjecture 1.1 is a special case of a more general conjecture in [7], that the Frobenius Series $\mathcal{F}_n(q,t)$ of the space H_n of "Diagonal Harmonics"(an S_n-module first introduced by Haiman in [9]) could be written as

$$\mathcal{F}_n(q,t) = \sum_{\mu\vdash n} \frac{\tilde{H}_\mu[X;q,t] t^{\Sigma l} q^{\Sigma a}(1-t)(1-q)(\sum q^{a'}t^{l'}) \prod^{0,0}(1-q^{a'}t^{l'})}{\prod (q^a - t^{l+1})(t^l - q^{a+1})}. \tag{2}$$

Here $\tilde{H}_\mu[X;q,t] = \sum_\lambda \tilde{K}_{\lambda,\mu}(q,t) s_\lambda$ is the "modified" Macdonald polynomial, with $s_\lambda[X]$ the Schur function, $\tilde{K}_{\lambda,\mu}(q,t) = t^{\Sigma l'} K_{\lambda,\mu}(q,1/t)$, and $K_{\lambda,\mu}(q,t)$ is Macdonald's q,t-Kostka

1991 *Mathematics Subject Classification.* Primary 05A15, 05A30 Secondary 05E05.

Key words and phrases. Catalan Number, Parking Functions, Labelled Trees, Hilbert Series.

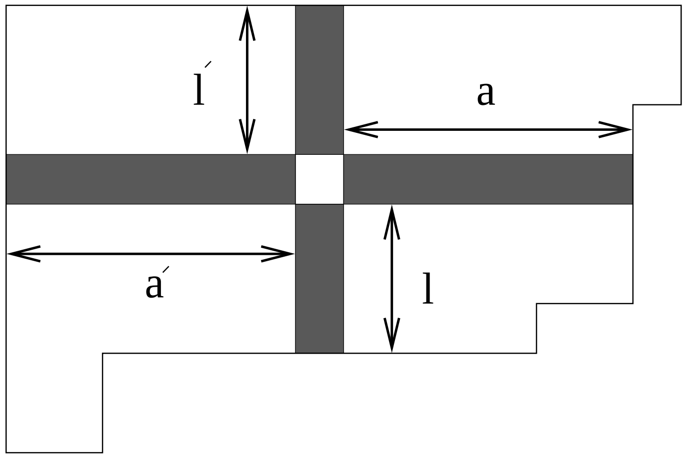


FIGURE 1. The arm, coarm, leg and coleg of a square

number, defined in [14]. They also showed that the right-hand-side of (2) equals $\nabla e_n[X]$, where ∇ is a linear operator defined on the modified Macdonald basis $\tilde{H}_\mu[X;q,t]$ by

$$\nabla \tilde{H}_\mu[X;q,t] = t^{\Sigma l} q^{\Sigma a} \tilde{H}_\mu[X;q,t],$$

and $e_n[X]$ is the nth elementary symmetric function. A special case of this conjecture is that the rational function expression for $C_n(q,t)$, which can be obtained by taking the coefficient of $s_{1^n}[X]$ in (2), equals the component of $\mathcal{F}_n(q,t)$ corresponding to the sign character χ^{1^n}, and hence has nonnegative coefficients.

A Dyck path is a lattice path from $(0,0)$ to (n,n) that never goes below the main diagonal $(i,i), 0 \le i \le n$. It is well-known that the number of such paths is the nth Catalan number $C_n = \frac{1}{n+1}\binom{2n}{n}$. We call the number of squares below the path but above the main diagonal the *area* of the Dyck path. Garsia and Haiman proved that

$$q^{\binom{n}{2}} C_n(q,1/q) = \frac{1}{[n+1]} \begin{bmatrix} 2n \\ n \end{bmatrix}$$

where $[k] = (1-q^k)/(1-q)$ and $\begin{bmatrix} n \\ k \end{bmatrix}$ is the q-binomial coefficient $(q;q)_n/((q;q)_k(q;q)_{n-k})$. They also showed that

$$C_n(q,1) = \sum_{D \in \mathcal{D}_n} q^{\text{area}(D)},$$

where the sum is over all Dyck paths from $(0,0)$ to (n,n). Based in part on these results, they called $C_n(q,t)$ the *q,t-Catalan sequence* [7].

In conjunction with Conjecture 1.1, they posed the question of finding a pair of statistics (qstat, tstat) so that $C_n(q,t)$ could be written in the form

$$\sum_{D \in \mathcal{D}_n} q^{\text{qstat}(D)} t^{\text{tstat}(D)}.$$

This problem was solved by the first author [8], who introduced a new statistic we here call *dmaj*. To define it, start by placing a ball at the upper right-hand corner (n,n) of a Dyck path D, then push the ball straight left. Once the ball intersects a vertical step of the path, it "ricochets" straight down until it intersects the diagonal, after which the process is iterated; the ball goes left until it hits another vertical step of the path, then down to the

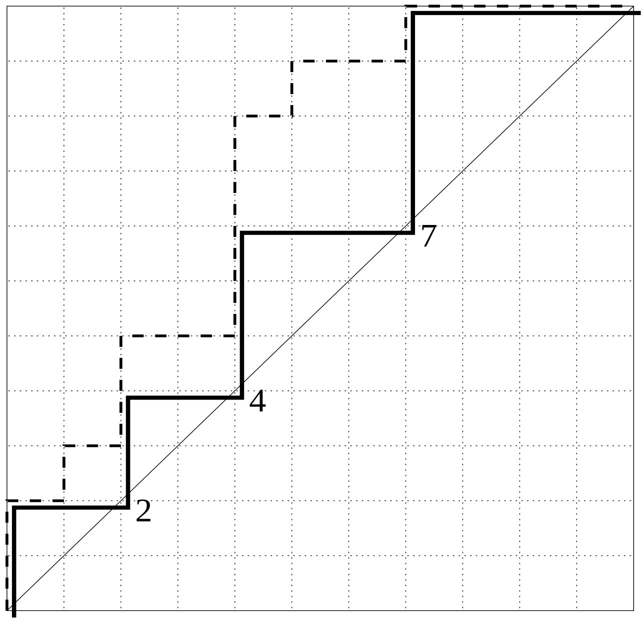


FIGURE 2. The statistic dmaj for a path. The Dyck path is the dashed line, the solid line is the bouncing ball. Here dmaj = 2+4+7 = 13 and area = 22.

diagonal, etc. On the way from (n, n) to $(0, 0)$ the ball will strike the diagonal at various points (i_j, i_j). We define $\mathrm{dmaj}(D)$ to be the sum of these i_j, and set

$$F_n(q,t) = \sum_{D \in \mathcal{D}_n} q^{\mathrm{area}(D)} t^{\mathrm{dmaj}(D)}.$$

See Fig. 1.

In [8], the first author conjectured that $F_n(q,t) = C_n(q,t)$, and also introduced a stratified function $F_{n,s}(q,t)$, defined as the sum, over all Dyck paths which end in exactly s horizontal steps, of $q^{\mathrm{area}(D)} t^{\mathrm{dmaj}(D)}$. He showed this function satisfies the recurrence relation

$$F_{n,s}(q,t) = \sum_{r=0}^{n-s} \begin{bmatrix} r+s-1 \\ r \end{bmatrix} q^{\binom{s}{2}} t^{n-s} F_{n-s,r}(q,t),$$

and by iterating this recurrence obtained the explicit formula

$$\begin{aligned} F_n(q,t) \quad = \quad & \sum_{k=1}^{n} \sum_{\substack{\alpha_1+\ldots+\alpha_k \\ \alpha_i>0}} t^{(k-1)\alpha_1+(k-2)\alpha_2+\ldots+\alpha_{k-1}} \\ & q^{\binom{\alpha_1}{2}+\ldots+\binom{\alpha_k}{2}} \begin{bmatrix} \alpha_1+\alpha_2-1 \\ \alpha_1 \end{bmatrix} \cdots \begin{bmatrix} \alpha_{k-1}+\alpha_k-1 \\ \alpha_{k-1} \end{bmatrix}. \end{aligned} \tag{3}$$

Garsia and the first author found a conjectured expression for $F_{n,s}(q,t)$ in terms of the nabla operator, namely

$$F_{n,s}(q,t) = t^{n-s} q^{\binom{s}{2}} \nabla e_{n-s} \left[X \frac{1-q^s}{1-q} \right]. \tag{4}$$

They were then able to prove (4) by using extended versions of summation formulas for generalized Pieri coefficients and other plethystic identities that Garsia and a number of coauthors have developed over the last ten years [5, 6]. As a corollary they proved Haglund's conjecture that $F_n(q,t) = C_n(q,t)$.

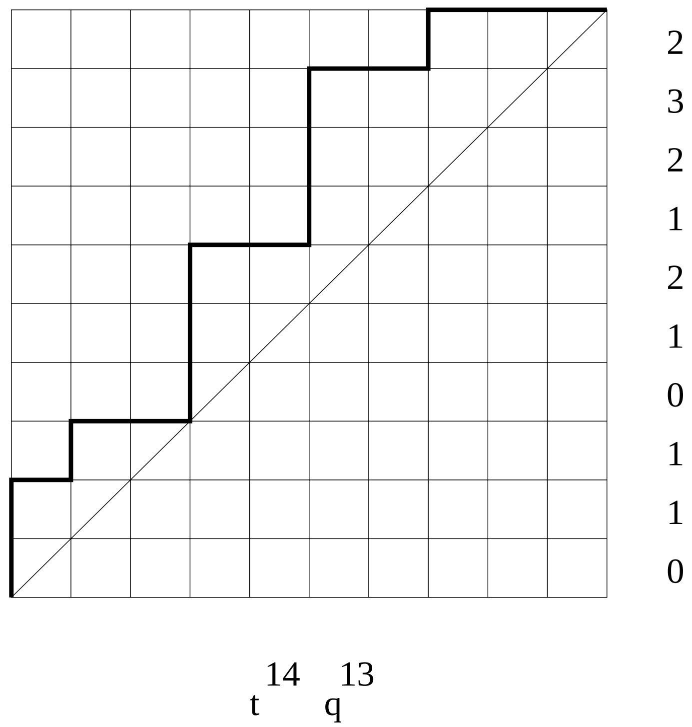


FIGURE 3. A Dyck path, with row lengths on the right

There is another pair of statistics for $C_n(q,t)$, discovered by M. Haiman while Garsia and the first author were still trying to prove (4). Given a Dyck path D, let $a_i(D)$ be the number of squares in the ith row, from the top, that are below D and strictly above the main diagonal. Note that the sum of the $a_i(D)$ equals area(D), that D ends in $a_1(D)+1$ horizontal steps, and that $a_n(D) = 0$ for all D. Call the sequence $a_1(D), a_2(D), \ldots a_n(D)$ the *area sequence* of D. We then define a statistic dinv(D) to be the sum of the cardinalities of the two sets

$$\{(i,j) : i < j \text{ and } a_i(D) = a_j(D)\}$$

and

$$\{(i,j) : i < j \text{ and } a_i(D) + 1 = a_j(D)\}.$$

For example, for the path of Fig. 1, dinv $= 14$.

Haiman conjectured that

$$C_n(q,t) = \sum_{D \in \mathcal{D}_n} q^{\mathrm{area}(D)} t^{\mathrm{dinv}(D)}. \tag{5}$$

The discovery of (5) was evidently motivated by elements of his celebrated proof of the "$n!$" conjecture [11], which says that certain S_n-submodules $\mathbf{M}_\mu$ of H_n have dimension $n!$. Since he had previously shown [10] that this conjecture implied that $\tilde{H}_\mu[X;q,t]$ is the Frobenius Series for M_μ, this resolves Macdonald's famous conjecture that the $K_{\lambda,\mu}(q,t)$ are polynomials in q,t with nonnegative integer coefficients. More recently Haiman has also proven (2), the explicit formula for the Frobenius Series of H_n [12].

At first it seemed that (5) was quite different than Haglund's conjecture, but before long Garsia, Haiman, and the first author realized they are closely related. To see why, note that a sequence B of n nonnegative integers $b_1 \cdots b_n$ is the area sequence of a Dyck path if and only if $b_n = 0$ and B contains no "two descents", i.e. values of i, $1 \le i \le n-1$, with

$b_i - b_{i+1} \geq 2$. Given a multiset $A = \{0^{\alpha_k} 1^{\alpha_{k-1}} \cdots (k-1)^{\alpha_1}\}$ of α_k copies of 0, etc., consider the sum of $q^{\text{area}(D)} t^{\text{dinv}(D)}$ over all Dyck paths D whose area sequence is some multiset permutation of A. Note that for any such D, $\text{area}(D) = \alpha_1(k-1) + \ldots + \alpha_{k-1}$. Note also that the contribution to dinv coming from $|\{(i,j) : i < j \text{ and } a_i(D) = a_j(D)\}|$ will equal $\binom{\alpha_1}{2} + \ldots + \binom{\alpha_k}{2}$ for all these D.

We can construct these area sequences by first permuting the α_k 0's and α_{k-1} 1's in any fashion, with a 0 at the end, which can be done in $\binom{\alpha_{k-1}+\alpha_k-1}{\alpha_{k-1}}$ ways. When we take into account the contribution to dinv from these various permutations, we generate the term

$$\begin{bmatrix} \alpha_{k-1} + \alpha_k - 1 \\ \alpha_{k-1} \end{bmatrix}_t .$$

Next we insert the α_{k-2} 2's into the sequence, which cannot be placed in front of any of the 0's since we must avoid 2-descents. This generates the factor

$$\begin{bmatrix} \alpha_{k-2} + \alpha_{k-1} - 1 \\ \alpha_{k-2} \end{bmatrix}_t .$$

It is now clear from (3) that

$$F_n(t,q) = \sum_{D \in \mathcal{D}_n} q^{\text{area}(D)} t^{\text{dinv}(D)}.$$

It follows easily from (1) that $C_n(q,t) = C_n(t,q)$, since the terms for conjugate partitions interchange q and t. Thus Garsia and the first author's result that $F_n(q,t) = C_n(q,t)$ also implies (5).

2. Statistics for the Hilbert Series

Haiman's proof of (2) [12] implies an earlier conjecture of his, that the space H_n has dimension $(n+1)^{n-1}$. It also implies that an explicit expression, as a sum of rational functions, for the Hilbert Series of H_n can be obtained by substituting, for each partition λ, f^λ, the number of standard tableaux of shape λ, in for $s_\lambda[X]$ in the right-hand-side of (2).

The number $(n+1)^{n-1}$ also counts the number of *parking functions* on n cars. A parking function P is obtained by starting with a Dyck path $D(P)$ and placing n "cars", denoted by the integers 1 through n, in the squares immediately to the right of the vertical segments of D, with the restriction that if car i is placed immediately on top of car j, then $i < j$. An example of a parking function is given in Fig. 2.

In this section we introduce a conjectured combinatorial formula for the Hilbert Series, which involves pairing the area statistic with a natural extension of the statistic dinv to parking functions. Given a parking function P, we define $r_i(P)$ to be the number (car) in the ith row (from the top) of P. We then let $\text{dinv}(P)$ be the sum of the cardinalities of the two sets

$$\{(i,j) : i < j, r_i(P) > r_j(P) \text{ and } a_i(D) = a_j(D)\}$$

and

$$\{(i,j) : i < j, r_i(P) < r_j(P) \text{ and } a_i(D) + 1 = a_j(D)\}.$$

For the parking function of Fig. 2, dinv $= 6$, since "inversions" occur for pairs (i,j) of rows $(1,3), (1,5), (4,5), (7,8), (7,9),$ and $(8,9)$.

Conjecture 2.1. *Define*

$$R_n(q,t) = \sum_P q^{\text{area}(D(P))} t^{\text{dinv}(P)}, \tag{6}$$

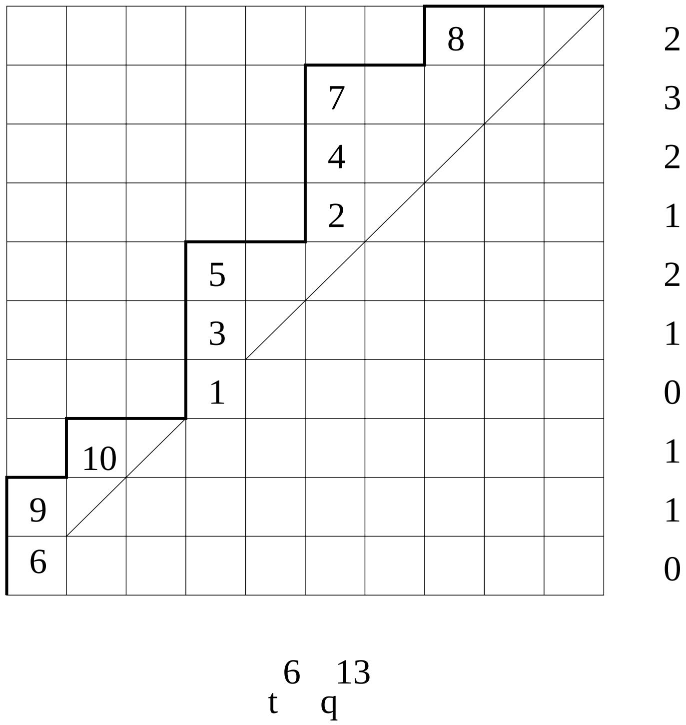


FIGURE 4. A parking function, with row lengths on the right

where the sum is over all parking functions on n cars. Then $R_n(q,t) = \mathcal{H}_n(q,t)$, where $\mathcal{H}_n(q,t)$ denotes the Hilbert Series of H_n.

Conjecture 2.1 has been verified using Maple by the first author and A. Garsia for $n \leq 7$.

We have a more compact way of writing $R_n(q,t)$ as a sum over permutations, which is based on the following lemma.

Lemma 2.2. *Given sets $A = \{a_1, a_2, \ldots, a_s\}$, $B = \{b_1, \ldots, b_{n-s}\}$ with $a_1 < a_2 < \cdots < a_s$, $b_1 < b_2 < \cdots < b_{n-s}$, $A \cap B = \emptyset$ and $A \cup B = \{1, 2, \ldots, n\}$, define*

$$F(A,B) = \sum_P t^{\mathrm{dinv}(P)},$$

where the sum is over all parking functions P whose set of cars on the main diagonal (rows of length 0) consist of the elements of A, in any order, and whose cars on the diagonal just above the main diagonal (rows of length 1) consist of the elements of B, in any order. Then

$$F(A,B) = [s]!_t [b_{n-s} - (n-s)]_t [b_{n-s-1} - (n-s) + 2]_t \cdots [b_1 + n - s - 2]_t.$$

Proof. By reading the cars in a parking function starting with the car in the top row and moving down, a parking function considered in the sum above can be identified with a permutation (linear list) of elements of $\{a_1, \ldots, a_s, b_1, \ldots, b_{n-s}\}$ where b_j immediately precedes a_k implies $b_j > a_k$. We will construct such a sequence recursively by first placing the a's in any order, say $a_{\alpha_1} a_{\alpha_2} \cdots a_{\alpha_s}$. Now b_{n-s} can be placed into this sequence in any of $b_{n-s} - 1 - (n-s-1)$ places, since there are $b_{n-s} - 1$ numbers less than b_{n-s}, but $n-s-1$ are in $\{b_1, \ldots, b_{n-s-1}\}$ and thus can't be in $a_{\alpha_1} \cdots a_{\alpha_s}$. If we insert b_{n-s} in front of the leftmost a_{α_j} satisfying $a_{\alpha_j} < b_{n-s}$, then any a_i's to the left of b_{n-s} will be greater than b_{n-s}, and will not generate any inversions. If we insert b_{n-s} in front of the next-to-leftmost a_{α_j} satisfying $a_{\alpha_j} < b_{n-s}$, then there is one a_i to the left of b_{n-s} less than b_{n-s} and we get a contribution of t. Hence the various possible placements of b_{n-s} generate a factor of $[b_{n-s} - (n-s)]_t$, independent of the permutation $\alpha_1 \cdots \alpha_n$. Next we insert b_{n-s-1} in front of b_{n-s} or any a_j satisfying $a_j < b_{n-s-1}$. If we place b_{n-s-1} in front of the leftmost of these possibilities we do not generate any inversions. If we place b_{n-s-1} in front of the next-to-leftmost choice we will either have b_{n-s} or a_{α_j} (with $a_{\alpha_j} < b_{n-s-1}$) to the left of b_{n-s-1}, and in either case this pair contributes t. Thus we see the insertion of b_{n-s-1} will generate a factor of $[b_{n-s-1} - (n-s-1) + 1]_t$ and by induction all the b's together will generate $[b_{n-s}-(n-s)]_t[b_{n-s-1}-(n-s)+2]_t \cdots [b_1+n-s-2]_t$. Since this was independent of $\alpha_1, \ldots, \alpha_s$, when we sum over all permutations of the a's, counting inversions amongst the a's only, we get the remaining $[s]!_t$ factor. □

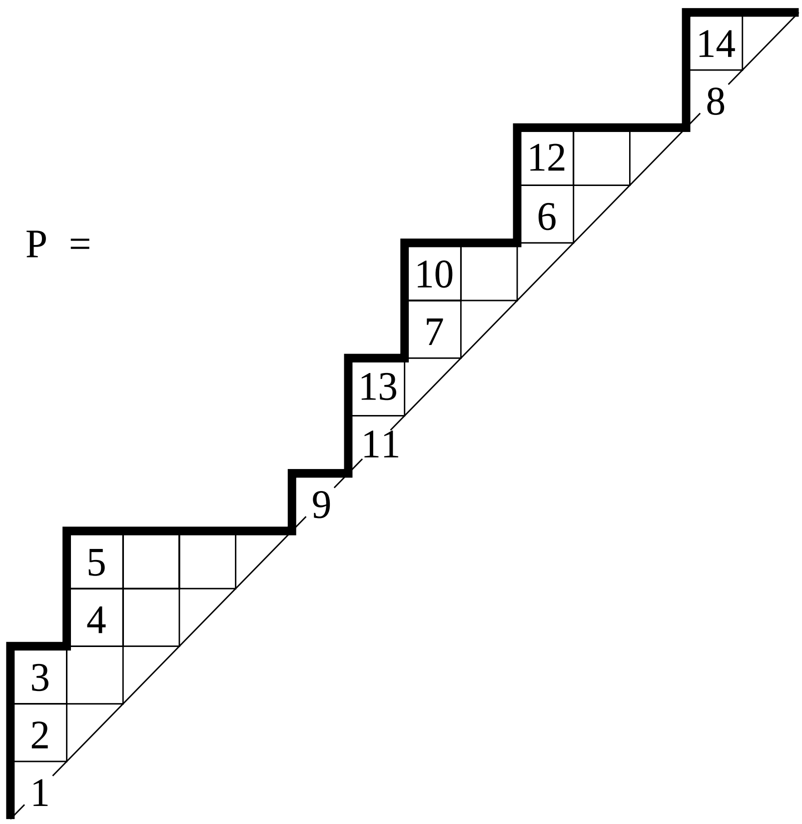
Using Lemma 2.2we can derive a product formula for the sum of t^{dinv} over all parking functions whose cars on the ith diagonal are from the set A_i, $i = 0, \ldots, k$ for general k. Say for example $k = 2$ so we are considering parking functions with rows of length 0, 1, and 2 with cars from disjoint subsets A, B, and C respectively, of cardinalities u, v, w, with $A \cup B \cup C = \{1, \ldots, n\}$. We start with a permutation of the elements of A, then insert the largest of the elements of B. Since the number of inversions depends only on the relative order of the elements of A and B, we will get a factor of $[\tilde{b}_v - v]_t$, where $\tilde{b}_v$ is what b_v would become if we reduced the elements of A and B to the set $\{1, 2, \ldots u+v\}$, keeping the relative order of each element to the others intact. Then when inserting c_w into the sequence after inserting all the elements of B, we would get the factor $[\tilde{c}_w - w]_t$, where $\tilde{c}_w$ is what c_w would become if we reduced the elements of B and C to the set $\{1, 2, \ldots v+w\}$, keeping the relative order of each element to the others intact. Thus we end up with the final term of

$$[u]!_t[\tilde{b}_v - v] \cdots [\tilde{b}_1 + v - 2]_t[\tilde{c}_w - w]_t \cdots [\tilde{c}_1 + w - 2]_t. \tag{7}$$

For which set partitions $A, B, C, \ldots$ of $\{1, 2, \ldots, n\}$ is there at least one parking function with cars on the main diagonal from A, the next diagonal from B, and so on? A necessary condition is that the largest element of B be larger than the smallest element of A, the largest element of C be larger than the smallest element of B, and so on, since some b_i must be on top of some a_i, some c_i must be on top of some b_i, etc. This condition is also sufficient, since we could put the a_i in columns 1 through $|A|$, with the smallest of these (a_1) in column $|A|$, then put the b_i in columns $|A|$ through $|A \cup B| - 1$, with the largest b_i on top of a_1 in column $|A|$, and b_1 in column $|A \cup B| - 1$, etc. Call such a sequence of sets "valid". Assume for the moment $k = 2$, so we have sets A, B, and C of cardinalities u, v, w, respectively. In the permutation $\sigma = c_1 c_2 \cdots c_w b_1 \cdots b_v a_1 \cdots a_u$, there are descents at spots w (since $c_w > b_1$) and $w + v$. This argument shows that there is a bijection between valid sequences of $k+1$ sets and permutations of $\{1, \ldots, n\}$ with k descents. Note that the area of all the corresponding parking functions in our example is $2w + v$, which is also the major index of the permutation σ. It is easy to see this holds in general. Furthermore, M. Haiman has pointed out that the numbers $\tilde{c}_i$ somewhat awkwardly described above can be easily defined in terms of the elements of σ. To do so, define a sequence σ' by $\sigma'_i = \sigma_i$, $1 \le i \le n$, and $\sigma'_{n+1} = 0$. Then for each $i, 1 \le i \le n$, let $u_i(\sigma)$ be the length of the longest consecutive sequence $\sigma'_i \sigma'_{i+1} \cdots \sigma'_j$ that starts at σ'_i and has either no descents, or exactly one descent and $\sigma'_i > \sigma'_j$. For example, if $\sigma = 47358126 \in S_8$, then $(u_1, u_2, \ldots, u_8) = (3, 3, 5, 4, 4, 4, 3, 2)$. It is easy to check that $\tilde{c}_1 = u_1 - 1$, $\tilde{c}_2 = u_2 - 1$, etc. We finally arrive at the following.

Theorem 2.3.

$$R_n(q, t) = \sum_{\sigma \in S_n} q^{\mathrm{maj}(\sigma)} \prod_{i=1}^{n} [u_i(\sigma) - 1]_t.$$



i	a_i	r_i
1	1	14
2	0	8
3	2	12
4	1	6
5	2	10
6	1	7
7	1	13
8	0	11
9	0	9
10	3	5
11	2	4
12	2	3
13	1	2
14	0	1

area(P) = 16 dinv(P)=19 dinv(D(P)) = 41

FIGURE 5. A labelled Dyck path (version 1).

3. Parking Functions and the dmaj Statistic

The relation

$$\sum_{D\in\mathcal{D}_n} q^{area(D)}t^{dinv(D)} = \sum_{D\in\mathcal{D}_n} q^{dmaj(D)}t^{area(D)}$$

suggests there should also be a way of extending the (area, dmaj) pair of statistics to get an alternate form of $R_n(q,t)$. We discuss one such extension in this section.

Let $\mathcal{P}_n$ denote the set of parking functions on n cars, as defined previously. We now view $\mathcal{P}_n$ as a collection of *labelled* Dyck paths. Fig. 5 shows a typical element of $\mathcal{P}_{14}$.

It is convenient to introduce another set of labelled Dyck paths, which we call $\mathcal{Q}_n$. To construct a typical object $Q \in \mathcal{Q}_n$, we attach labels to a path $D \in \mathcal{D}_n$ according to the following rules. Let $q_1 q_2 \cdots q_n$ be a permutation of the labels $\{1, 2, \ldots, n\}$. Place each label q_i in the ith row of the diagram for D, in the *main diagonal cell.* There is one restriction: For each "left-turn" in the Dyck path (i.e., an EAST step followed immediately by a NORTH step, reading from southwest to northeast), the label q_i appearing due east of the NORTH step must be less than the label q_j appearing due south of the EAST step. See Fig. 3 for an example. In the figure, capital letters mark the left-turns in the Dyck path. Since $4 < 5$,

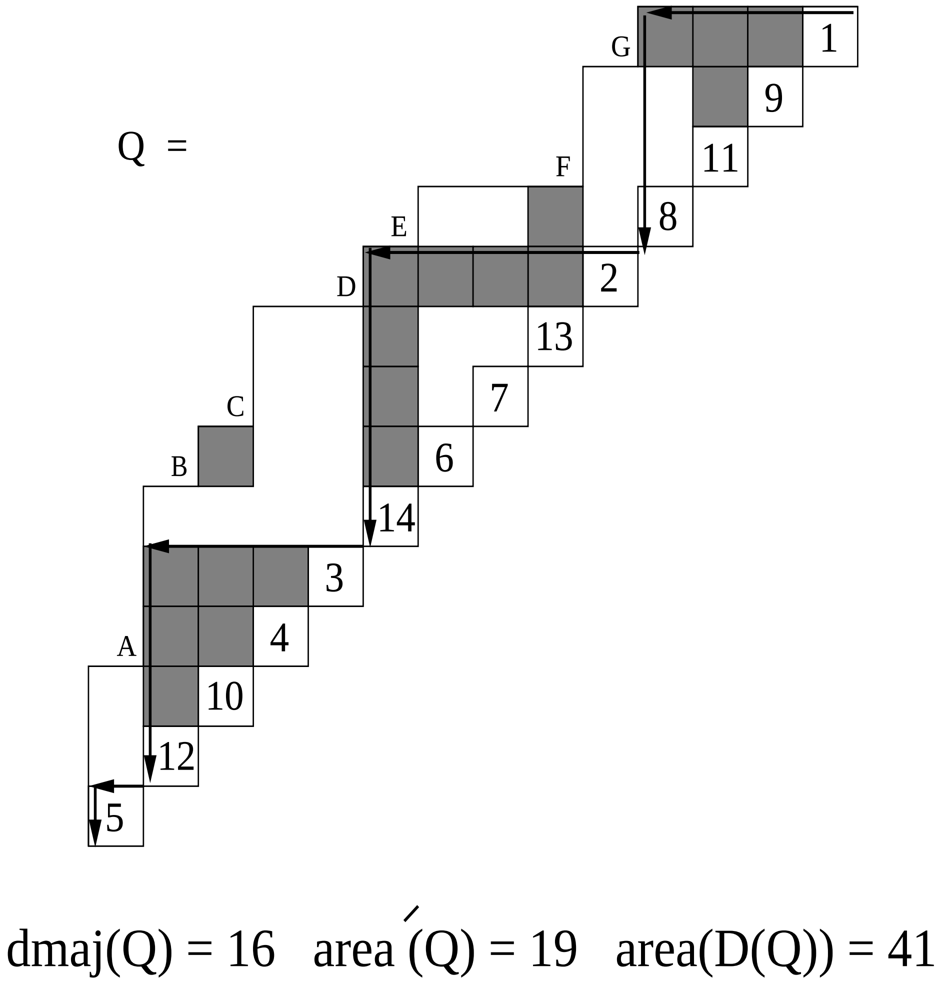


FIGURE 6. A labelled Dyck path (version 2).

$6 < 12$, $7 < 10$, $2 < 3$, $8 < 14$, $11 < 13$, and $1 < 2$, the labelled path shown does belong to $\mathcal{Q}_{14}$.

Given a labelled path Q constructed from the ordinary Dyck path $D = D(Q)$, define $\mathrm{dmaj}(Q)$ to be $\mathrm{dmaj}(D(Q))$, which was defined earlier. Also define $\mathrm{area}'(Q)$ to be the number of cells c in the diagram for Q such that:

(1) Cell c is strictly between the Dyck path D and the main diagonal; AND
(2) The label on the main diagonal due east of c is less than the label on the main diagonal due south of c.

In Fig. 3, only the shaded cells satisfy both conditions and hence contribute to $\mathrm{area}'(Q)$. Evidently, $\mathrm{area}'(Q) \leq \mathrm{area}(D(Q))$ for all Q, and strict inequality can occur.

We conjecture that

$$S_n(q,t) = \sum_{Q \in \mathcal{Q}_n} q^{dmaj(Q)} t^{area'(Q)} \tag{8}$$

also gives the Hilbert series for H_n. We will show this conjecture is equivalent to the previous one by giving a bijective proof that $R_n(q,t) = S_n(q,t)$.

Bijections. We begin with the case of unlabelled Dyck paths. Fix a path $D \in \mathcal{D}_n$. We will construct a new path $E \in \mathcal{D}_n$ such that $\mathrm{area}(D) = \mathrm{dmaj}(E)$ and $\mathrm{dinv}(D) = \mathrm{area}(E)$, which proves that

$$\sum_{D \in \mathcal{D}_n} q^{area(D)} t^{dinv(D)} = \sum_{E \in \mathcal{D}_n} q^{dmaj(E)} t^{area(E)}.$$

The bijection is essentially a combinatorial version of the proof of this formula given in Section 1.

Consider the area sequence $a(D) = (a_1(D), \ldots, a_n(D))$. It is easy to see that such a list of numbers corresponds to a valid Dyck path iff $a_i \geq 0$ for all i, $a_n = 0$, and $a_i \leq a_{i+1} + 1$ for all $i < n$. Set $s = \max_{1 \leq i \leq n} a_i$. For $0 \leq j \leq s$, let b_j be the number of occurrences of j in $a(D)$. It follows from the above conditions on $a(D)$ that $b_j > 0$ for all j; moreover, $b_0 + \cdots + b_s = n$.

To construct E, we first draw a bounce path B whose successive horizontal moves (starting from (n, n)) have lengths $b_0, \ldots, b_s$. This bounce path, together with the main diagonal line $y = x$, creates a sequence of $s+1$ triangles which we shall call $T_0, \ldots, T_s$. For $1 \leq i \leq s$, there is an empty rectangular region R_i located north of triangle T_i and west of triangle T_{i-1}. Note that rectangle R_i has width b_i and height b_{i-1}.

We now describe how to construct the portion of the path E located in rectangle R_i. Fix i, and let w_i be the word obtained from $a(D)$ by deleting all symbols other than $i-1$ and i. Then w_i consists of b_{i-1} occurrences of $i-1$ and b_i occurrences of i; also, by the conditions on $a(D)$, the last symbol in w_i must be $i-1$. Read the symbols in w_i from left to right. Starting at the northwest tip of triangle T_i, draw an EAST step when the symbol i is read; draw a NORTH step when the symbol $i-1$ is read. Note that this partial path must terminate in a NORTH step. For later use, we remark that the "left-turns" of E in the region R_i correspond precisely to the *descents* in the word w_i. Because of the condition $a_i \leq a_{i+1} + 1$, the set of descents in all the words w_i corresponds bijectively with the set of descents in the full word $a(D)$.

After filling all the rectangular regions in this way, we obtain the Dyck path E. Observe that, because the paths within each R_i ended in NORTH steps, B is the bounce path derived from E. Therefore,

$$\begin{aligned}
\mathrm{dmaj}(E) &= (n - b_0) + (n - b_0 - b_1) + \cdots + (n - b_0 - b_1 - \cdots - b_s) \\
&= n(s+1) - (s+1)b_0 - sb_1 - \cdots - (s+1-j)b_j - \cdots - 1b_s \\
&= (s+1)(n - b_0 - \cdots - b_s) + \sum_{j=0}^{s} j b_j \\
&= \sum_{j=0}^{s} j b_j = \sum_{i=1}^{n} a_i \\
&= \mathrm{area}(D).
\end{aligned}$$

Furthermore, from the definitions of b_j and w_i, it is easy to see that the formula for dinv can be rewritten as

$$\mathrm{dinv}(D) = \sum_{j=0}^{s} \binom{b_j}{2} + \sum_{i=1}^{s} \mathrm{coinv}(w_i),$$

where $\mathrm{coinv}(w_i)$ is the number of coinversions in the word w_i. Now $\binom{b_j}{2}$ is the number of area cells in the triangle T_j, and $\mathrm{coinv}(w_i)$ is the number of cells beneath the path E in the rectangle R_i. Hence, $\mathrm{dinv}(D) = \mathrm{area}(E)$.

The process used to construct E from D is reversible. First, we obtain $b_0, \ldots, b_s$ by examining the bounce path of E. Next, we recover $a(D)$ from E by starting with b_0 zeroes and successively inserting b_1 ones, then b_2 twos, etc., according to the partial paths in R_1, R_2, etc. The condition that $a_i \leq a_{i+1}+1$ ensures that there will be a unique way to perform this insertion procedure. Hence, we obtain the desired bijection.

As an example, if we take D to be the path shown in Fig. 5 (ignoring labels), then E will be the path shown in Fig. 3 (ignoring labels).

Next, we consider the case of labelled Dyck paths. We will give a bijection from $\mathcal{P}_n$ to $\mathcal{Q}_n$ that sends area to dmaj and sends dinv to area$'$. This bijection proves that

$$R_n(q,t) = \sum_{P\in\mathcal{P}_n} q^{area(P)}t^{dinv(P)} = \sum_{Q\in\mathcal{Q}_n} q^{dmaj(Q)}t^{area'(Q)} = S_n(q,t).$$

Fix $P \in \mathcal{P}_n$. We shall construct $Q \in \mathcal{Q}_n$ with $\mathrm{dmaj}(Q) = \mathrm{area}(P)$ and $\mathrm{area}'(Q) = \mathrm{dinv}(P)$. As an example, the labelled path P in Fig. 5 will map to the labelled path Q in Fig. 3.

Let $D = D(P)$ denote the underlying unlabelled Dyck path of P. Let E be the unlabelled Dyck path produced by the above bijection, with $\mathrm{dmaj}(E) = \mathrm{area}(D)$ and $\mathrm{area}(E) = \mathrm{dinv}(D)$. E will be the underlying unlabelled path for Q (i.e., $D(Q) = E$).

We obtain Q by attaching labels to E, as follows. Scan each of the diagonals of P, from southwest to northeast, starting with the main diagonal and proceeding upward. Enter the labels of P, in the order in which they are encountered, on the main diagonal of Q going from northeast to southwest. For instance, in Fig. 5, P has the labels 1, 9, 11, 8 on the main diagonal, followed by the labels 2, 13, 7, 6, 14 on the first superdiagonal, etc. Hence (see Fig. 3), the labels on the main diagonal of Q are 1, 9, 11, 8, 2, 13, 7, 6, 14, … starting from (n,n). Clearly, we can recover the labelling of P from the labelling of Q.

Here is an equivalent way of describing the relation between the labels in P and Q. Recall that $E = D(Q)$ can be dissected into triangles $T_0, \ldots, T_s$ and rectangles $R_1, \ldots, R_s$. For $0 \leq j \leq s$, the b_j labels on the main diagonal of Q inside triangle T_j (read from top to bottom) are the labels appearing in the leftmost cells of the b_j rows of $D = D(P)$ for which $a_i(D) = j$ (read from bottom to top).

Recall that the labels of P in a given column must increase from bottom to top. To check the validity of a given labelling, it clearly suffices to check that *adjacent* labels in the same column are always properly ordered. Suppose that the labels r_i and r_{i+1} in rows i and $i+1$ both occur in column j. This occurs iff $a_i = a_{i+1}+1$ iff there is a *descent* of $a(D)$ at position i (recall that $a_i \leq a_{i+1}+1$). We observed earlier that the descents of $a(D)$ correspond bijectively to the left-turns of $E = D(Q)$. From here, it is easy to verify that label r_{i+1} appears in Q due east of the left-turn corresponding to the descent $a_{i+1} > a_i$, and the label r_i appears in Q due south of this left-turn. Hence, the labelling restrictions on P imply the corresponding labelling restrictions on Q, and conversely.

Clearly, $\mathrm{dmaj}(Q) = \mathrm{dmaj}(E) = \mathrm{area}(D) = \mathrm{area}(P)$. We now show that $\mathrm{area}'(Q) = \mathrm{dinv}(P)$. Consider a typical area cell c of the path $E = D(Q)$. Suppose first that c is inside triangle T_k. Let $x_1, x_2, \ldots, x_{a_k}$ be the labels on the diagonal of Q inside T_k, from top to bottom. As noted above, the labels $x_1, x_2, \ldots, x_{a_k}$ are just the numbers r_i in all positions i for which $a_i = k$, *written in reverse order.* Thus, the cells in T_k that contribute to $\mathrm{area}'(Q)$ correspond precisely to inversions in the word $x_{a_k}, \ldots, x_2, x_1$. We obtain a bijection between the contributing cells in T_k and the set

$$\{(i,j) : i < j, r_i(P) > r_j(P) \text{ and } a_i(D) = a_j(D) = k\} \quad (0 \leq k \leq s).$$

A similar argument applies to a cell c in rectangle R_k. The horizontal position of the cell determines a unique r_j such that $a_j = k$, and the vertical position of the cell determines a unique r_i such that $a_i = k-1$. All pairs (i,j) for which $a_i + 1 = a_j$ are accounted

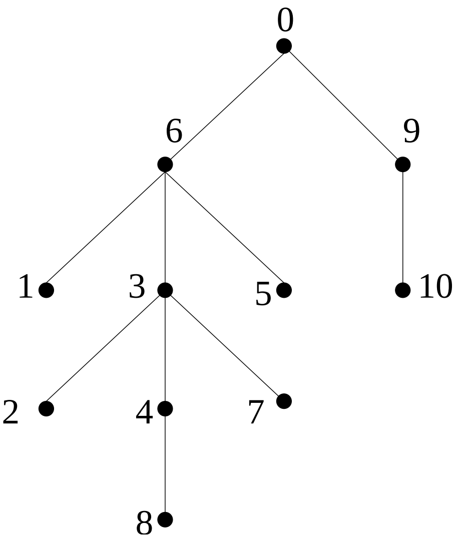


FIGURE 7. The tree for the parking function of Figure 2

for exactly once in this fashion (as k ranges from 1 to s). Now, we get a contribution to $\mathrm{dinv}(P)$ iff $i < j$ and $r_i < r_j$; this occurs precisely when the associated cell c satisfies the two conditions for contributing to $\mathrm{area}'(Q)$. We conclude that the number of contributing cells in all the rectangular regions R_k is exactly the cardinality of the set

$$\{(i,j) : i < j, r_i(P) < r_j(P) \text{ and } a_i(D) + 1 = a_j(D)\}.$$

Combining this result with the one in the preceding paragraph, we conclude that $\mathrm{area}'(Q) = \mathrm{dinv}(P)$. This completes the proof.

4. Labelled Trees

There are a number of known bijections between parking functions on n cars and forests of labelled trees on n vertices [2, 4, 13]; see also [16, p.140]. These typically use an alternate definition of a parking function, as a function $f : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ with the property that the number of values of j with $f(j) \leq i$ is at least i for all $1 \leq i \leq n$. Such a function can be obtained from the geometric representation of a parking function from section 2 by letting $f(i)$ be the column containing the car i. It isn't easy to translate Conjecture 2.1 into a statement about trees using these bijections since it is hard to keep track of what happens to the area and dinv statistics. In this section we describe a simple bijection between forests of rooted labelled trees, and parking functions in the geometric form of section 2 which makes it easy to describe versions of these statistics for trees.

Given a parking function P as in section 2 and a given car i, travel northeast, staying in the same diagonal, until we either leave the $n \times n$ square or run into another car. If we run into a car at the bottom of a column, then we say all the cars in that column are "children" of car i. If we leave the square or run into a car which isn't at the bottom of a column, car i has no children. We define a rooted, labelled tree $T(P)$ with root labelled 0 by the condition that the node with label i (call this node i) is a child of the root node if and only if car i is in the first column of P, and for a non-root node node i, i has j as a child if and only if car j is a child of car i in P. Also, when we draw a tree T we put the children of node i below node i, increasing from left to right. For example, if P is the parking function from Fig. 2, $T(P)$ is represented in Fig. 4.

We recursively define the *preference order* of such a tree as the sequence of n numbers, beginning with the children of the root node, listed left to right (smallest to largest) followed by the preference order of the descendants of the largest child of the root node, followed by the preference order of the descendants of the next-to-largest child of the root node, etc. We also define a function $d_i(T)$ recursively by $d_0(T) = 0$, and if j is the kth-smallest child of i, then $d_j(T) = d_i(T) + k - 1$. For example, the tree of Fig. 4 has preference order $6, 9, 10, 1, 3, 5, 2, 4, 7, 8$, and the d values of these (non-root) nodes are $0, 1, 1, 0, 1, 2, 1, 2, 3, 2$, respectively. Given a tree T, we construct a Dyck path $D(T)$ by the condition that the length of the k row, from the bottom, of $D(T)$ is the d value of the kth element of the preference order of T. We then construct a parking function $P(T)$ by placing the kth number from the preference sequence of T in the kth row, from the bottom, immediately to the right of $D(T)$. Next define $\mathrm{area}(T)$ to be the sum of the d-values of the nodes of T, and $\mathrm{dinv}(T)$ to be the sum of the cardinalities of the two sets

$$\{(i,j) : i < j, d(i) = d(j) \text{ and } i \text{ occurs before } j \text{ in preference order}\}$$

and

$$\{(i,j) : i < j, d(i) + 1 = d(j), \text{ and } i \text{ occurs after } j \text{ in preference order}\}.$$

It is not hard to see that $P(T(P)) = P$, $T(P(T)) = T$ and furthermore that $\mathrm{area}(T(P)) = \mathrm{area}(P)$ and $\mathrm{dinv}(T(P)) = \mathrm{dinv}(P)$. Thus Conjecture 2 is equivalent to the following.

Conjecture 4.1.

$$\mathcal{H}_n(q,t) = \sum_T q^{\mathrm{area}(T)} t^{\mathrm{dinv}(T)}$$

where the sum is over all labelled, rooted trees T with $n+1$ vertices and root node labelled 0.

5. Open Problems

The main obstacle to proving Conjecture 2.1 by the methods of [5, 6] is the lack of a recurrence relation for $R_n(q,t)$. We have also been unable to resolve a number of interesting bijective questions related to Conjecture 2.1, most notably to prove $R_n(q,t)$ is symmetric in q and t by finding an involution on parking functions which interchanges area and dinv. In fact, we can't even prove the much weaker statement that $R_n(q,1) = R_n(1,q)$. We have also been unable to show $q^{\binom{n}{2}} R_n(q, 1/q) = (1 + q + \ldots + q^n)^{n-1}$, which is the value for $q^{\binom{n}{2}} \mathcal{H}_n(q, 1/q)$ conjectured by Stanley [9]. One could also hope to refine Conjecture 2.1 to find a pair of statistics on parking functions that generate $\mathcal{F}_n(q,t)$.

The statistics and bijections discussed in section 3 are specializations of more general constructs that apply to variations of the q,t-Catalan numbers and parking functions. In particular, the second author has discovered combinatorial results concerning the "extended family" of q,t-Catlalan sequences $C_n^m(q,t)$ introduced by Garsia and Haiman [7]. These generalizations are discussed in the second author's thesis [to appear].

References

[1] F. Bergeron, A. Garsia, M. Haiman, and G. Tesler, Identities and positivity conjectures for some remarkable operators in the theory of symmetric functions, *Asian J. Math.* **6** (1999), 363–420.

[2] J.S. Beissinger and U.N. Peled, A note on major sequences and external activity in trees, The Wilf Festschrift (Philadelphia, PA 1996) *Electron. J. Combin.* **4** (1997), Research Paper 4, approx. 10 pp.

[3] Ö. Eğecioğlu and J. Remmel, Bijections for Cayley trees, spanning trees, and their q-analogues, *J. Combin. Theory Ser. A* **42** (1986), 15–30.

[4] D. Foata and J. Riordan, Mappings of acyclic and parking functions, *Aequationes Math.* **10** (1974), 10–22.

[5] A. Garsia and J. Haglund, A positivity result in the theory of Macdonald polynomials, *Proc. Nat. Acad. Sci. U.S.A* **98** (2001), 4313–4316.
[6] A. Garsia and J. Haglund, A proof of the (q, t)-Catalan positivity conjectur , *Discrete Math.*, to appear (the LACIM 2000 issue).
[7] A. Garsia and M. Haiman, A remarkable q, t-Catalan sequence and q-Lagrange inversion, *J. Algebraic Combin.* **5** (1996), 191–244.
[8] J. Haglund, Conjectured statistics for the (q, t)-Catalan numbers, *Adv. in Math.* (To appear).
[9] M. Haiman, Conjectures on the quotient ring by diagonal invariants, *J. Algebraic Combin.* **3** (1994), 17–76.
[10] M. Haiman, Macdonald polynomials and geometry, in *New Perspectives in Algebraic Combinatorics* 37, Billers, et. al. (eds.), MSRI Publication, 1999, 207–254.
[11] M. Haiman, Hilbert Schemes, polygraphs, and the Macdonald positivity conjecture, *J. Amer. Math. Soc.* **14** (2001), 941–1006
[12] M. Haiman, Vanishing theorems and character formulas for the Hilbert scheme of points in the plane, (preprint, 2001), available from `http://www.math.berkeley.edu/ mhaiman/`.
[13] G. Krewerau, Une famille de polynômes ayant plusieurs proprietes enumeratives, **11** (1980), 309–320.
[14] I.G. Macdonald, *Symmetric Functions and Hall Polynomials*, The Clarendon Press, Oxford University Press, New York, 1995.
[15] J. Riordan, Ballots and trees, *C. Combinatorial Theory* **6** (1969), 408–411.
[16] R.P. Stanley, *Enumerative Combinatorics Volume 2* Cambridge University Press, Cambridge, United Kingdom, 1999.

Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104-6395
E-mail address: `jhaglund@math.upenn.edu`

Department of Mathematics, University of California at San Diego, La Jolla, CA 92093-0112
E-mail address: `nloehr@math.ucsd.edu`

THE NUMBER OF REPRESENTATIONS OF A NUMBER BY VARIOUS FORMS

MICHAEL D. HIRSCHHORN

ABSTRACT. We find formulae for the number of representations of the integer n as, for example, the sum of two triangles and two squares, or of four triangles, in terms of divisor functions. Indeed, we find sixteen formulae of this type, a few of which are known, the remainder apparently new.

RÉSUMÉ. Nous trouvons des formules pour le nombre des représentations d'un entier n comme, par example, somme de deux nombres triangulaires et de deux carrés, ou de quatres nombres triangulaires, en termes des fonctions des diviseurs de n. En effet, nous trouvons seize formules de ce type, dont quelques unes sont connues, les autres apparemment nouvelles.

1. Introduction

There are several classical results which give the number of representations of a number by a quadratic form in terms of a divisor function. The object of this note is to consider four such results and from them to derive many more of the same sort.

With $d_{r,m}(n)$ denoting the number of divisors d of n with $d \equiv r(\text{mod } m)$ amd $\sigma(n)$ the sum of the divisors of n, the results we consider are the following. Proofs of all four can be found in [2].

Theorem 1. *(Jacobi, 1828). The number of representations of $n \geq 1$ as the sum of two squares is*

$$r\{\Box + \Box\}(n) = 4\left(d_{1,4}(n) - d_{3,4}(n)\right). \tag{J1}$$

Theorem 2. *(Dirichlet, 1840). The number of representations of $n \geq 1$ as the sum of a square and twice a square is*

$$r\{\Box + 2\Box\}(n) = 2\left(d_{1,8}(n) + d_{3,8}(n) - d_{5,8}(n) - d_{7,8}(n)\right). \tag{D}$$

Theorem 3. *(L.Lorenz, 1871). The number of representations of $n \geq 1$ as the sum of a square and three times a square is*

$$r\{\Box + 3\Box\}(n) = 2\left(d_{1,3}(n) - d_{2,3}(n)\right) + 4\left(d_{4,12}(n) - d_{8,12}(n)\right). \tag{L}$$

Theorem 4. *(Jacobi, 1829). The number of representations of $n \geq 1$ as the sum of four squares is*

$$r\{\Box + \Box + \Box + \Box\}(n) = 8 \sum_{d|n, 4\nmid d} d = 8\left(\sigma(n) - 4\sigma\left(\frac{n}{4}\right)\right). \tag{J2}$$

We shall prove the following sixteen results of the same sort.

(1.1) $r\{\Delta+\Delta\}(n) = d_{1,4}(4n+1) - d_{3,4}(4n+1),$

(1.2) $r\{\Box+2\Delta\}(n) = d_{1,4}(4n+1) - d_{3,4}(4n+1),$

(1.3) $r\{2\Box+\Delta\}(n) = d_{1,4}(8n+1) - d_{3,4}(8n+1),$

(1.4) $r\{\Delta+4\Delta\}(n) = \frac{1}{2}\left(d_{1,4}(8n+5) - d_{3,4}(8n+5)\right),$

(1.5) $r\{\Delta+2\Delta\}(n) = \frac{1}{2}\left(d_{1,8}(8n+3) + d_{3,8}(8n+3) - d_{5,8}(8n+3) - d_{7,8}(8n+3)\right),$

(1.6) $r\{\Box+\Delta\}(n) = d_{1,8}(8n+1) + d_{3,8}(8n+1) - d_{5,8}(8n+1) - d_{7,8}(8n+1),$

(1.7) $r\{\Box+4\Delta\}(n) = d_{1,8}(2n+1) + d_{3,8}(2n+1) - d_{5,8}(2n+1) - d_{7,8}(2n+1),$

(1.8) $r\{\Delta+3\Delta\}(n) = d_{1,3}(2n+1) - d_{2,3}(2n+1),$

(1.9) $r\{3\Box+2\Delta\}(n) = d_{1,3}(4n+1) - d_{2,3}(4n+1),$

(1.10) $r\{\Box+6\Delta\}(n) = d_{1,3}(4n+3) - d_{2,3}(4n+3),$

(1.11) $r\{6\Box+\Delta\}(n) = d_{1,3}(8n+1) - d_{2,3}(8n+1),$

(1.12) $r\{\Delta+12\Delta\}(n) = \frac{1}{2}\left(d_{1,3}(8n+13) - d_{2,3}(8n+13)\right),$

(1.13) $r\{2\Box+3\Delta\}(n) = d_{1,3}(8n+3) - d_{2,3}(8n+3),$

(1.14) $r\{3\Delta+4\Delta\}(n) = \frac{1}{2}\left(d_{1,3}(8n+7) - d_{2,3}(8n+7)\right),$

(1.15) $r\{\Delta+\Delta+\Delta+\Delta\}(n) = \sigma(2n+1),$

(1.16) $r\{\Box+\Box+\Box+\Box\}(n) = \sigma(4n+1).$

It should be noted that not all these results are new. For instance, (1.8) is equivalent to a result of Ramanujan ([1, pp.223-224, 3, 4, p.229]).

2. Preliminary Results

As usual, let

$$\phi(q) = \sum_{-\infty}^{\infty} q^{n^2}, \quad \psi(q) = \sum_{n\geq 0} q^{(n^2+n)/2}.$$

We shall require the easy lemmas

(2.1) $$\phi(q)\psi(q^2) = \psi(q)^2,$$

(2.2) $$\phi(q) = \phi(q^4) + 2q\psi(q^8),$$

as well as the (apparently new) result

(2.3) $$\psi(q)\psi(q^3) = \phi(q^6)\psi(q^4) + q\phi(q^2)\psi(q^{12}).$$

Proofs of lemmas.

(2.1) $$\phi(q)\psi(q^2) = \frac{(q^2)_\infty^5}{(q)_\infty^2(q^4)_\infty^2} \cdot \frac{(q^4)_\infty^2}{(q^2)_\infty} = \frac{(q^2)_\infty^4}{(q)_\infty^2} = \psi(q)^2.$$

(2.2) $$\phi(q) = \sum_{-\infty}^{\infty} q^{n^2} = \sum_{n \text{ even}} q^{n^2} + \sum_{n \text{ odd}} q^{n^2} = \sum_{-\infty}^{\infty} q^{4n^2} + \sum_{-\infty}^{\infty} q^{4n^2+4n+1} = \phi(q^4) + 2q\psi(q^8).$$

$$q^4\psi(q^8)\psi(q^{24}) = \sum_{k,l=-\infty}^{\infty} q^{(4k+1)^2+3(4l+1)^2} = \sum_{k,l=-\infty}^{\infty} q^{4(k+3l+1)^2+12(k-l)^2} = \sum_{u-v\equiv 1 \pmod 4} q^{4u^2+12v^2}. \tag{2.3}$$

We now consider the two cases v even, u even. If v is even, $v = 2k$, $u = 4l+1$ or $-4l-1$, according as k is even or odd, while if u is even, $u = 2k$, $v = 4l+1$ or $-4l-1$, according as k is odd or even. Thus

$$\begin{aligned}\sum_{u-v\equiv 1 \,(\mathrm{mod} 4)} q^{4u^2+12v^2} &= \sum_{k,l=-\infty}^{\infty} q^{4(4l+1)^2+12(2k)^2} + \sum_{k,l=-\infty}^{\infty} q^{4(2k)^2+12(4l+1)^2} \\ &= q^4\psi(q^{32})\phi(q^{48}) + q^{12}\phi(q^{16})\psi(q^{96}),\end{aligned}$$

as required.

Proofs of theorems.

(J1) is equivalent to

$$\phi(q)^2 = 1 + 4\sum_{n\geq 1}(d_{1,4}(n) - d_{3,4}(n))q^n. \tag{3.1}$$

That is, by (2.2) and (2.1),

$$\big(\phi(q^4) + 2q\psi(q^8)\big)^2 = \big(\phi(q^4)^2 + 4q^2\psi(q^8)^2\big) + 4q\psi(q^4)^2 = 1 + 4\sum_{n\geq 1}\left(d_{1,4}(n) - d_{3,4}(n)\right)q^n. \tag{3.2}$$

If we extract those terms in which the power of q is $1 \pmod 4$, divide by $4q$ and replace q^4 by q, we find

$$\psi(q)^2 = \sum_{n\geq 0}\left(d_{1,4}(4n+1) - d_{3,4}(4n+1)\right)q^n, \tag{3.3}$$

from which we obtain (1.1).

By (2.1), (3.3) can be written

$$\phi(q)\psi(q^2) = \sum_{n\geq 0}(d_{1,4}(4n+1) - d_{3,4}(4n+1))q^n, \tag{3.4}$$

from which (1.2) follows.

Using (2.2), (3.4) can be written

$$\psi(q^2)\left(\phi(q^4) + 2q\psi(q^8)\right) = \sum_{n\geq 0}\left(d_{1,4}(4n+1) - d_{3,4}(4n+1)\right)q^n, \tag{3.5}$$

from which (1.3) and (1.4) follow.

(D) is equivalent to

$$\phi(q)\phi(q^2) = 1 + 2\sum_{n\geq 1}(d_{1,8}(n) + d_{3,8}(n) - d_{5,8}(n) - d_{7,8}(n))q^n, \tag{3.6}$$

or, by (2.2),

$$(\phi(q^4)+2q\psi(q^8))(\phi(q^8)+2q^2\psi(q^{16})) = 1+2\sum_{n\geq 1}(d_{1,8}(n)+d_{3,8}(n)-d_{5,8}(n)-d_{7,8}(n))q^n, \tag{3.7}$$

from which (1.5), (1.6) and (1.7) follow.

(L) is equivalent to

$$\begin{aligned}\phi(q)\phi(q^3) &= 1+2\sum_{n\geq1}(d_{1,3}(n)-d_{2,3}(n))q^n+4\sum_{n\geq1}(d_{4,12}(n)-d_{8,12}(n))q^n \\ &= 1+2\sum_{n\geq1}(d_{1,3}(n)-d_{2,3}(n))q^n+4\sum_{n\geq1}(d_{1,3}(n)-d_{2,3}(n))q^{4n},\end{aligned} \tag{3.8}$$

or

$$\begin{aligned}&\left(\phi(q^4)+2q\psi(q^8)\right)\left(\phi(q^{12})+2q^3\psi(q^{24})\right) \\ &\qquad = 1+2\sum_{n\geq1}\left(d_{1,3}(n)-d_{2,3}(n)\right)q^n+4\sum_{n\geq1}\left(d_{1,3}(n)-d_{2,3}(n)\right)q^{4n}.\end{aligned} \tag{3.9}$$

If from (3.9) we extract those terms in which the power of q is 0 (mod 4) and replace q^4 by q, we find

$$\begin{aligned}&\phi(q)\phi(q^3)+4q\psi(q^2)\psi(q^6) \\ &\qquad = 1+2\sum_{n\geq1}(d_{1,3}(4n)-d_{2,3}(4n))q^n+4\sum_{n\geq1}(d_{1,3}(n)-d_{2,3}(n))q^n \\ &\qquad = 1+6\sum_{n\geq1}(d_{1,3}(n)-d_{2,3}(n))q^n\end{aligned} \tag{3.10}$$

where we have used the fact, which needs a little consideration, that

$$d_{1,3}(4n)-d_{2,3}(4n)=d_{1,3}(n)-d_{2,3}(n).$$

If we subtract (3.8) from (3.10), we find

$$4q\psi(q^2)\psi(q^6)=4\sum_{n\geq1}(d_{1,3}(n)-d_{2,3}(n))q^n-4\sum_{n\geq1}\left(d_{1,3}(n)-d_{2,3}(n)\right)q^{4n}. \tag{3.11}$$

The right hand side of (3.11) is an odd function of q; if we divide by $4q$ and replace q^2 by q, we find

$$\psi(q)\psi(q^3)=\sum_{n\geq0}\left(d_{1,3}(2n+1)-d_{2,3}(2n+1)\right)q^n, \tag{3.12}$$

from which (1.8) follows.
If we invoke (2.3), (3.12) becomes

$$\phi(q^6)\psi(q^4)+q\phi(q^2)\psi(q^{12})=\sum_{n\geq0}\left(d_{1,3}(2n+1)-d_{2,3}(2n+1)\right)q^n, \tag{3.13}$$

so

$$\phi(q^3)\psi(q^2)=\sum_{n\geq0}(d_{1,3}(4n+1)-d_{2,3}(4n+1))q^n \tag{3.14}$$

and

$$\phi(q)\psi(q^6)=\sum_{n\geq0}(d_{1,3}(4n+3)-d_{2,3}(4n+3))q^n. \tag{3.15}$$

(1.9) and (1.10) follow.
(3.14) and (3.15) can be written respectively

$$\psi(q^2)\left(\phi(q^{12})+2q^3\psi(q^{24})\right)=\sum_{n\geq0}\left(d_{1,3}(4n+1)-d_{2,3}(4n+1)\right)q^n \tag{3.16}$$

and

$$\psi(q^6)\left(\phi(q^4)+2q\psi(q^8)\right)=\sum_{n\geq 0}(d_{1,3}(4n+3)-d_{2,3}(4n+3))q^n. \tag{3.17}$$

(1.11), (1.12), (1.13) and (1.14) follow.

(J2) is equivalent to

$$\phi(q)^4=1+8\sum_{n\geq 1}\left(\sum_{d|n,\,4\nmid d} d\right)q^n. \tag{3.18}$$

Now, the left hand side is

(3.19)

$$\begin{aligned}\phi(q)^4&=\left(\phi(q^4)+2q\psi(q^8)\right)^4\\&=\left(\phi(q^4)^4+16q^4\psi(q^8)^4\right)+8q\phi(q^4)^3\psi(q^8)+24q^2\phi(q^4)^2\psi(q^8)^2+32q^3\phi(q^4)\psi(q^8)^3\\&=\left(\phi(q^4)^4+16q^4\psi(q^8)^4\right)+8q\phi(q^4)^2\psi(q^4)^2+24q^2\psi(q^4)^4+32q^3\psi(q^4)^2\psi(q^8)^2.\end{aligned}$$

So (3.18) becomes

$$\begin{aligned}&\left(\phi(q^4)^4+16q^4\psi(q^8)^4\right)+8q\phi(q^4)^2\psi(q^4)^2+24q^2\psi(q^4)^4+32q^3\psi(q^4)^2\psi(q^8)^2\\&=1+8\sum_{n\geq 1}\left(\sum_{d|n,\,4\nmid d} d\right)q^n.\end{aligned} \tag{3.20}$$

We deduce that

$$24\psi(q)^4=8\sum_{n\geq 0}\left(\sum_{d|4n+2} d\right)q^n=24\sum_{n\geq 0}\sigma(2n+1)q^n \tag{3.21}$$

and

$$8\phi(q)^2\psi(q)^2=8\sum_{n\geq 0}\left(\sum_{d|4n+1} d\right)q^n=8\sum_{n\geq 0}\sigma(4n+1)q^n, \tag{3.22}$$

which are (1.15) and (1.16).

References

[1] Berndt, B.C., "Ramanujan's Notebooks", Part III, Springer-Verlag, New York, 1991.

[2] Hirschhorn, M.D., "Partial fractions and four classical theorems of number theory", *Amer. Math, Monthly*, **106**(2000), 268-272.

[3] Hirschhorn, M.D., "Three classical results on representations of a number", *The Andrews Festschrift, Seventeen Papers on Classical Number Theory and Combinatorics*, D.Foata, G-N.Han, eds., Springer-Verlag, (2001),159-165.

[4] Ramanujan, S., *Notebooks*, **Vol. 2**, Tata Institute of Fundamental Research, Mumbai, 1957.

Department of Pure Mathematics, UNSW, Sydney NSW, 2052, Australia

E-mail address: `m.hirschhorn@unsw.edu.au`

POLYGRAPH ARRANGEMENTS
(EXTENDED ABSTRACT)

AXEL HULTMAN

Abstract. A class of subspace arrangements, $Z(n,m)$, known as polygraph arrangements was exploited by Haiman in order to prove the $n!$ theorem. By showing that their intersection lattices, $L(Z(n,m))$, are EL-shellable, we determine the cohomology groups of the complements of the arrangements. Moreover, we generalize the shellability results to a class of lattices which deserve to be called Dowling generalizations of $L(Z(n,m))$. As a consequence, we obtain the cohomology groups of the complements of certain Dowling analogies of polygraph arrangements.

Résumé. Une classe d'arrangements de sous-espaces de Z(n, m), connue sous le nom des arrangements de polygraphes, a été exploitée par Haiman afin de prouver le théorme $n!$. En prouvant que leurs treillis d'intersections, $L(Z(n,m))$, sont EL-shellable, nous déterminons les groupes de cohomologies des compléments d'arrangements. De plus, nous généralisons les résultats de shellabilité à une classe de treillis que nous appelons les $L(Z(n,m))$-généralisations de Dowling. Entre autres, nous obtenons les groupes de cohomologies des compléments de certains analogues de Dowling des arrangements de polygraphe.

1. Introduction

Macdonald [11] introduced a family of polynomials known as *Macdonald polynomials*. They constitute a basis of the algebra of symmetric functions in the variables $x_1, x_2, \ldots$ with coefficients in the field of fractions of $\mathbb{Q}[y,z]$. Transformation to the basis of Schur functions gives rise to transition coefficients that are called *Kostka-Macdonald coefficients*. Until recently, the conjecture that the Kostka-Macdonald coefficients in fact are polynomials in y and z with nonnegative integer coefficients was open. This conjecture was known as the *Macdonald positivity conjecture*.

Garsia and Haiman [6] conjectured that the Kostka-Macdonald coefficients are multiplicities of graded characters of certain S_n-modules. An equivalent (see Haiman [9]) formulation of this has become known as the $n!$ *conjecture*, since it asserts that the said modules are of dimension $n!$. This implies the positivity conjecture.

Recently, Haiman [8] proved the $n!$ conjecture. The proof relies on the fact that a class of subspace arrangements in $(\mathbb{C}^2)^{n+m}$, called *polygraph arrangements*, have coordinate rings that are free modules over the polynomial ring in one coordinate set on $(\mathbb{C}^2)^n$.

In this paper, we show that certain lattices, which deserve to be called Dowling generalizations of the intersection lattices of the polygraph arrangements, are EL-shellable. Via the Goresky-MacPherson formula, this allows us to determine the cohomology groups of the complements of the polygraph arrangements as well as of Dowling analogies of these arrangements. In particular, it turns out that the cohomology is torsion-free and vanishing in "most" dimensions.

The structure of this paper is as follows. After briefly reviewing basic definitions and tools in Section 2, we deal with the case of ordinary polygraph arrangements in Sections 3 and 4. In Section 5, we give Dowling generalizations of the results in Section 4.

The proofs of Section 5 certainly specialize to proofs of the theorems in Section 4. However, they do not quite specialize to the proofs given in Section 4; the latter are simpler and more transparent. This is the reason why we treat ordinary polygraph arrangements and their Dowling generalizations separately.

Acknowledgement. The author is grateful to his advisor Anders Björner who suggested the study of polygraph arrangements. Moreover, his careful reading and remarks have led to substantial improvements in the paper.

2. Tools for investigation of subspace arrangements

We give a brief survey of the techniques that are used in this paper. For basic combinatorial and topological concepts, the reader is referred to the textbooks by Stanley [13] and Munkres [12]. For more on subspace arrangements, see e.g. Björner's survey article [2].

A *subspace arrangement* is a collection $\mathcal{A} = \{A_1, \dots, A_n\}$ of affine subspaces of $\mathbf{k}^m$, where $\mathbf{k}$ is some field. In case $\mathbf{k} \in \{\mathbb{R}, \mathbb{C}\}$, one is often interested in the topological features of the complement $\mathcal{M}_{\mathcal{A}} := \mathbf{k}^m \setminus (\bigcup_{i=1}^n A_i)$.

2.1. The Goresky-MacPherson formula.

To any poset P, we associate the *order complex* $\Delta(P)$. This is the simplicial complex having the chains of P as simplices. If P has a minimal and/or a maximal element, then the symbols $\hat{0}$ and $\hat{1}$ will be used to denote them, respectively. The *proper part* $\overline{P}$ is the poset $P \setminus \{\hat{0}, \hat{1}\}$.

The *intersection semi-lattice* $L(\mathcal{A})$ of $\mathcal{A}$ is the meet semi-lattice of all nonempty intersections of subsets of $\mathcal{A}$ ordered by reverse inclusion. It is a lattice iff $\bigcap_{i=1}^n A_i \neq \emptyset$. In case $\mathbf{k} \in \{\mathbb{R}, \mathbb{C}\}$, the following result by Goresky and MacPherson [7] relates the reduced cohomology groups of $\mathcal{M}_{\mathcal{A}}$ and the reduced homology of the lower intervals of $L(\mathcal{A})$:

Theorem 2.1. (The Goresky-MacPherson formula) *Let $\mathcal{A}$ be a real subspace arrangement (i.e. $\mathbf{k} = \mathbb{R}$). Then, for all i,*

$$\widetilde{H}^i(\mathcal{M}_{\mathcal{A}}; \mathbb{Z}) \cong \bigoplus_{x \in L(\mathcal{A}) \setminus \{\hat{0}\}} \widetilde{H}_{codim_{\mathbb{R}}(x)-i-2}(\Delta(\overline{[\hat{0}, x]}); \mathbb{Z}). \qquad \square$$

Note that we can apply Theorem 2.1 to complex arrangements by identifying $\mathbb{C}$ with $\mathbb{R}^2$. Then $codim_{\mathbb{R}}(\cdot)$ is replaced by $2codim_{\mathbb{C}}(\cdot)$.

2.2. Lexicographic shellings.

If we are interested in the cohomology of $\mathcal{M}_{\mathcal{A}}$, then Theorem 2.1 leaves us with the task of determining the homology of $L(\mathcal{A})$ and its lower intervals. To this end, the technique of EL-shellability described below will be useful to us. It was introduced for ranked posets by Björner [1] and later extended to arbitrary posets by Björner and Wachs [4].

For an arbitrary poset P, let $\hat{P}$ denote the poset obtained by adding an additional maximal element $\hat{1}$ and an additional minimal element $\hat{0}$ to P. Let $R(\hat{P}) \subset \hat{P}^2$ denote the covering relation of $\hat{P}$. We write $x \to y$ if y covers x. An *edge-labelling* of P is a map $\lambda : R(\hat{P}) \to \Lambda$, where Λ is some poset of labels. A saturated chain $c = \{c_1 \to \dots \to c_t\} \subseteq \hat{P}$ is *rising* if $\lambda(c_1 \to c_2) < \dots < \lambda(c_{t-1} \to c_t)$. The chain c is *falling* if, instead, $\lambda(c_1 \to c_2) \geq \dots \geq \lambda(c_{t-1} \to c_t)$. Given an interval $[x, y] \subseteq \hat{P}$, we compare two saturated chains $c = \{x = c_1 \to \dots \to c_{t_c} = y\}$ and $d = \{x = d_1 \to \dots \to d_{t_d} = y\}$ using the lexicographic order induced by Λ on the sequences $\lambda(c_1 \to c_2), \dots, \lambda(c_{t_c-1} \to c_{t_c})$ and $\lambda(d_1 \to d_2), \dots, \lambda(d_{t_d-1} \to d_{t_d})$.

Definition 2.2. *$\lambda : R(\hat{P}) \to \Lambda$ is an* EL-labelling *of P if every $\hat{P}$-interval contains a unique rising saturated chain and this chain is lexicographically least of all saturated chains in the interval. If P admits an EL-labelling, then P is called* EL-shellable.

Clearly, if P is EL-shellable, then so is every interval of P. Moreover, the homotopy type of $\Delta(P)$ can be read off the labelling:

Theorem 2.3. (see [4, Thm. 5.9]) *If λ is an EL-labelling of P, then $\Delta(P)$ is homotopy equivalent to a wedge of spheres. The spheres of dimension i in the wedge are indexed by the falling maximal chains of length $i+2$ in $\hat{P}$. In particular, if P is ranked, then $\Delta(P)$ is homotopy equivalent to a wedge of spheres in top dimension.* □

3. Polygraph arrangements

Now we describe our main objects of study. Let V be a d-dimensional vector space over $\mathbf{k}$. For $m, n \in \mathbb{N}$ and a function $f : [m] \to [n]$, we let

$$W_f = \{(x_{f(1)}, \dots, x_{f(m)}, x_1, \dots, x_n) \in V^{m+n} \mid x_i \in V \ \forall i \in [n]\}.$$

This is a linear subspace of V^{m+n}. The *polygraph arrangement*, $Z_V(n,m)$, is the collection of all such subspaces:

$$Z_V(n,m) := \{W_f \mid f : [m] \to [n]\}.$$

Often, the choice of V (and $\mathbf{k}$) is not important. Therefore, we frequently write $Z(n,m)$ instead of $Z_V(n,m)$. To avoid confusion, we mention that Haiman [8] lets $Z(n,m)$ denote the union, not just the collection, of all W_f.

We need a combinatorial description of the intersection lattice $L(Z(n,m))$. From now on, we let $P = \{p_1 < \cdots < p_m\}$ and $Q = \{q_1 < \cdots < q_n\}$ be fixed disjoint ordered sets. For a subset $S \subseteq P \cup Q$, we use the notation $S^P := S \cap P$ and $S^Q := S \cap Q$, so that $S = S^P \cup S^Q$. Consider the following lattice, which is a join-subsemilattice of the partition lattice $\Pi_{P \cup Q}$.

$$\begin{aligned} L(Q,P) := \ & \{\pi_1|\dots|\pi_t \in \Pi_{P\cup Q} \mid \pi_i^P = \emptyset \Rightarrow |\pi_i^Q| = 1 \\ & \ \text{and } \pi_i^P \neq \emptyset \Rightarrow \pi_i^Q \neq \emptyset \ \forall i \in [t], t \text{ arbitrary}\} \cup \{\hat{0}\}. \end{aligned}$$

Proposition 3.1. *$L(Z(n,m))$ and $L(Q,P)$ are isomorphic.*

Proof. Pick a subset $F \subseteq \{f : [m] \to [n]\}$. The element $\bigcap_{f \in F} W_f \in L(Z(n,m))$ can be represented by a bipartite graph $G_F = (P \cup Q, E)$, where $\{p_i, q_j\} \in E$ iff $f(i) = j$ for some $f \in F$. Two such graphs represent the same subspace of V^{m+n} precisely if they have the same connected components. Clearly, all bipartite graphs on $P \cup Q$ in which $deg(p_i) \geq 1$ for all $i \in [m]$ occur in this way. Thus, $L(Z(n,m))$ is isomorphic to the lattice of partitions of $P \cup Q$ that correspond to connected components in bipartite graphs on $P \cup Q$ with no isolated vertices in P. This is precisely $L(Q,P)$. □

Corollary 3.2. *The intersection lattice $L(Z(n,m))$ is ranked of length n.* □

4. An EL-labelling of $\overline{L(Q,P)}$

We will give an edge-labelling λ of $\overline{L(Q,P)}$. It will turn out to be an EL-labelling. Our poset Λ of labels is

$\Lambda = \{A_2 < \cdots < A_n < B_1 < \cdots < B_n < \underbrace{11\ldots11}_{m} < \underbrace{11\ldots12}_{m} < \cdots < \underbrace{nn\ldots n}_{m} < C_2 < \cdots < C_n\}$.

The labelling $\lambda : R(L(Q,P)) \to \Lambda$ is defined by:

- $\lambda(\pi \to \tau) = A_x$ if two non-singleton blocks, π_i and π_j, in π are merged in τ and $q_x = max(\pi_i^Q \cup \pi_j^Q)$.
- $\lambda(\pi \to \tau) = B_x$ if a singleton, q_x, and a non-singleton block, π_i, in π are merged in τ and $q_x < max(\pi_i^Q)$.
- $\lambda(\pi \to \tau) = C_x$ if a singleton, q_x, and a non-singleton block, π_i, in π are merged in τ and $q_x > max(\pi_i^Q)$.
- $\lambda(\hat{0} \to \tau) = f(1)f(2)\ldots f(m)$ (juxtapositioning) if τ corresponds to the subspace W_f.

Theorem 4.1. *The labelling λ is an EL-labelling of $\overline{L(Q,P)}$.*

Proof. Pick an interval $I = [\pi,\tau] \subseteq L(Q,P)$. We must verify that I contains a unique rising chain and that this chain is lexicographically least in I.

Suppose, to begin with, that $\pi \neq \hat{0}$. If there are non-singleton blocks in π which are merged in τ, then any rising chain must begin with the merging of these blocks. This gives rise to A-labels, and for the subscripts of these labels to form a rising sequence, the order in which to merge the blocks is unique. Next, all singletons q_i that are to be merged with non-singleton blocks containing some $q_j > q_i$ must be so. This gives rise to B-labels, and again there is a unique way to make their subscripts form a rising sequence. Finally, the remaining singletons that are larger than all Q-elements in their blocks in τ are to be merged, this giving rise to C-labels. Once again, there is a unique order in which to do this within a rising chain. Thus, we have constructed the unique rising chain in I. Note that if we had replaced the word *rising* with the phrase *lexicographically least*, then the above construction would give us the unique lexicographically least chain. Hence, it coincides with the rising chain.

Now, suppose that $\pi = \hat{0}$. Note that I contains exactly one atom a_f, corresponding to the subspace W_f, with the property that $[a_f,\tau]$ contains a chain with only C-labels. The function f sends $i \in [m]$ to the least $j \in [n]$ such that p_i and q_j are in the same block in τ. As before, the C-labels occur with rising subscripts in exactly one chain in $[a_f,\tau]$. Note that $\lambda(\hat{0} \to a_f) < \lambda(\hat{0} \to a)$ for all atoms $a \in I \setminus \{a_f\}$. Hence, the rising chain is again lexicographically least in I. □

Remark. Let $r = (r_1,\ldots,r_m) \in [n]^m$. Haiman [10] has considered the subarrangement $Z(n,m,r) \subseteq Z(n,m)$ which consists of those W_f that satisfy $f(i) \leq r_i$ for all $i \in [m]$. It is not difficult to see that, with straightforward modifications, the proof of Theorem 4.1 goes through for the appropriate subsemilattice $L(Q,P,r)$ of $L(Q,P)$.

Theorem 2.3 tells us that $\Delta(\overline{L(Q,P)})$ is homotopy equivalent to a wedge of spheres in top dimension, the spheres being indexed by the falling chains in $L(Q,P)$ under the labelling λ. In order to calculate the number of spheres in the wedge, we define an easily counted set of combinatorial objects which is in 1-1 correspondence with the set of falling chains in $L(Q,P)$.

Consider the set $C(Q,P)$ of ordered partitions $(\pi_1,\ldots,\pi_k)$ of $P \cup Q$ such that $q_n \in \pi_1$, $\pi_i^P \neq \emptyset$ and $\pi_i^Q \neq \emptyset$ for all $i \in [k]$, k arbitrary. Define $\Gamma(n,m) := |C(Q,P)|$. Clearly,

$\Gamma(n,m) = \sum_{k=1}^{min(n,m)} S(n,k)S(m,k)k!(k-1)!$,

where the $S(i,j)$ are Stirling numbers of the second kind.

Now we establish the bijection mentioned above.

Theorem 4.2. $\Delta(\overline{L(Q,P)})$ *is homotopy equivalent to a wedge of* $(n-2)$*-dimensional spheres. The number of spheres in the wedge is*

$\sum_{k=1}^{min(n,m)} S(n,k)S(m,k)k!(k-1)!$.

Proof. We construct a bijection $\phi : \{\text{falling chains in } L(Q,P)\} \to C(Q,P)$ as follows. Let $c = \{\hat{0} \to c_1 \to \ldots \to c_n = \hat{1}\} \subseteq L(Q,P)$ be a falling chain. Then, for some j, all c_i, $i < j$, contain singleton blocks whereas all c_i, $i \geq j$, do not. The blocks in c_j are the blocks in $\phi(c)$. Let π_1 be the block in c_j which contains q_n. Since c is falling, c_{j+1} is obtained by merging π_1 with some other block which we call π_2. Then c_{j+2} is obtained by merging $\pi_1 \cup \pi_2$ with a block which we denote π_3 and so on, until finally $\hat{1}$ is obtained from c_{n-1} by merging $\pi_1 \cup \cdots \cup \pi_{n-j}$ with the only other block, which is then given the name π_{n-j+1}. We define $\phi(c) := (\pi_1, \ldots, \pi_{n-j+1})$.

To check injectivity of ϕ, consider two distinct falling chains $c = \{\hat{0} \to c_1 \to \ldots \to c_n = \hat{1}\}$ and $d = \{\hat{0} \to d_1 \to \ldots \to d_n = \hat{1}\}$ in $L(Q,P)$. Let j be the smallest index such that $c_j \neq d_j$.

If $j = 1$, then c_1 and d_1 correspond to different functions $f_c, f_d : [m] \to [n]$. Note that $\{q_i \in Q \mid i \in f_c([m])\}$ is the set of maximal Q-elements in blocks in $\phi(c)$, since no falling chain possesses C-labels. An analogous statement holds for f_d. Therefore, if $f_c([m]) \neq f_d([m])$, then $\phi(c) \neq \phi(d)$. If, on the other hand, $f_c([m]) = f_d([m])$, then we can pick $i \in [m]$ such that $f_c(i) \neq f_d(i)$ and both $f_c(i)$ and $f_d(i)$ are maximal Q-elements in blocks in both $\phi(c)$ and $\phi(d)$. Therefore, p_i and $q_{f_c(i)}$ belong to the same block in $\phi(c)$ but to different blocks in $\phi(d)$. Hence, $\phi(c) \neq \phi(d)$.

Now suppose $j > 1$. If $\lambda(c_{j-1} \to c_j) = B_x$, for some x, then $\lambda(d_{j-1} \to d_j) = B_x$, too. Since $c_j \neq d_j$, this means that the block containing x in $\phi(c)$ is different from the block containing x in $\phi(d)$. This implies $\phi(c) \neq \phi(d)$.

The only case left is $\lambda(c_{j-1} \to c_j) = \lambda(d_{j-1} \to d_j) = A_n$. This implies that the set of blocks in $\phi(c)$ is equal to the set of blocks in $\phi(d)$. Since $c \neq d$, $\phi(c)$ must differ from $\phi(d)$ in the order of the blocks. Hence $\phi(c) \neq \phi(d)$ in this case too, and ϕ is injective.

To establish surjectivity, choose $\pi = (\pi_1, \ldots, \pi_k) \in C(Q,P)$. We will construct a falling chain $c \subseteq L(Q,P)$ such that $\phi(c) = \pi$. Let $\hat{f} : P \to Q$ be the function mapping all elements in π_i^P to $max(\pi_i^Q)$ for all $i \in [k]$. Define $f : [m] \to [n]$ to be the corresponding function on the indices, i.e. by requiring that $\hat{f}(p_i) = q_{f(i)}$ for all $i \in [m]$. The atom of $L(Q,P)$ corresponding to W_f is c_1. The chain $\{\hat{0} \to c_1 \to \ldots \to c_{n-k+1} = \pi_1|\ldots|\pi_k\}$ is produced by merging the singletons in c_1 one by one with their corresponding non-singleton blocks in the only possible way which ends with $\pi_1|\ldots|\pi_k$ while giving rise to a falling sequence of B-labels. For $l \in [k-1]$, let $c_{n-k+l} = \pi_1 \cup \cdots \cup \pi_l|\pi_{l+1}|\ldots|\pi_k$. Now, $c = \{\hat{0} \to c_1 \to \ldots \to c_{n-1} \to \hat{1}\}$ is mapped to π by ϕ, so ϕ is surjective. $\square$

Note that $\Gamma(n,m) = \Gamma(m,n)$. This implies an unsuspected numerical relationship between the combinatorially very distinct arrangements $Z(n,m)$ and $Z(m,n)$.

The cohomology groups of the complement $\mathcal{M}_{Z_{\mathbb{R}^d}(n,m)}$ are determined by the Goresky-MacPherson formula (Theorem 2.1) and the following corollary:

Corollary 4.3. *Let* $\pi = \pi_1|\ldots|\pi_k \in L(Q,P)$. *Then* $\Delta(\overline{[\hat{0},\pi]})$ *is homotopy equivalent to a wedge of* $(n-1-k)$*-dimensional spheres. The number of spheres in the wedge is* $\prod_{j=1}^{k}\Gamma(|\pi_j^Q|,|\pi_j^P|)$.

Proof. $\overline{[\hat{0},\pi]}$ is ranked, and it is EL-shellable since $\overline{L(Q,P)}$ is. Hence, by Theorem 2.3, $\Delta(\overline{[\hat{0},\pi]})$ is homotopy equivalent to a wedge of spheres in top dimension. The number of spheres in the wedge is the absolute value $|\mu(\hat{0},\pi)|$ of the Möbius function. Note that $[\hat{0},\pi] \cong L(\pi_1^Q,\pi_1^P)\times\cdots\times L(\pi_k^Q,\pi_k^P)$. The Möbius function is multiplicative, so $\mu(\hat{0},\pi) = \prod_{j=1}^{k}\mu_j(\hat{0},\pi_j)$, where μ_j is the Möbius function of $L(\pi_j^Q,\pi_j^P)$. The corollary now follows from Theorem 4.2. □

In general dimension, the expression for the cohomology of the complement, although determined by Corollary 4.3, is not pretty. In the following theorem we restrict ourselves to weaker, readable, information. As before, the complex case is obtained by identifying $\mathbb{C}$ and $\mathbb{R}^2$.

Theorem 4.4. *For all* i, $\widetilde{H}^i(\mathcal{M}_{Z_{\mathbb{R}^d}(n,m)};\mathbb{Z})$ *is torsion-free. Let* $\widetilde{\beta}^i$ *denote its rank. We have*

(1) $\widetilde{\beta}^i = 0$, *unless* $i = d(m-1)+j(d-1)$ *for some* $j \in [n]$.

(2) $\widetilde{\beta}^{dm-1} = n^m$, *if* $d \geq 2$.

(3) $\widetilde{\beta}^{d(m+n-1)-n} = \Gamma(n,m)$, *if* $d \geq 2$.

□

Remark. Unlike its complement, the union $\cup\mathcal{A}$ of an arrangement of linear subspaces is topologically not very exciting; it is a cone with apex in the origin. A more interesting object is the *link*, $lk(\mathcal{A}) := S^{l-1}\cap(\cup\mathcal{A})$, where l is the dimension of the space in which the arrangement is embedded. From Ziegler and Živaljević [14, Thm. 2.4], it follows that the link of a real linear subspace arrangement with shellable intersection lattice has the homotopy type of a wedge of spheres. In particular, this applies to the polygraph arrangements $Z_{\mathbb{R}^d}(n,m)$.

5. A Dowling generalization

5.1. **Dowling lattices.** Let G be a finite group and n a positive integer. G acts on the set $([n]\times G)\cup\{0\}$ by $0g := 0$ and $(i,h)g := (i,hg)$ for $i \in [n]$ and $g,h \in G$. For a subset $S \subseteq ([n]\times G)\cup\{0\}$, we define $Sg := \{xg \mid x \in S\}$. A partition $\pi = \pi_1|\ldots|\pi_t$ of $([n]\times G)\cup\{0\}$ is *G-symmetric* if $\pi_i g \in \pi$ for all $g \in G$ and $i \in [t]$. The block π_i is called *g-symmetric*, for $g \in G$, if $\pi_i g = \pi_i$. If the identity element is the only $g \in G$ which makes π_i g-symmetric, then π_i is called *simple*. Note that if π is G-symmetric, then the block containing 0 is necessarily g-symmetric for all $g \in G$.

Definition 5.1. *Let* G *be a finite group. The* Dowling lattice Π_n^G *is the lattice of all* G*-symmetric partitions* π *of* $([n]\times G)\cup\{0\}$ *such that all blocks not containing* 0 *are simple. The block containing* 0 *is called the* null block *of* π.

Note that $\Pi_n^{\{e\}} \cong \Pi_{n+1}$. Thus, Dowling lattices constitute a generalization of the partition lattice. They were first introduced by Dowling [5]. Two more special cases are worth mentioning. The lattice $\Pi_n^{\mathbb{Z}_2}$ is isomorphic to the partition lattice of type B, i.e. the intersection lattice of the arrangement of reflecting hyperplanes of the Coxeter group B_n. This is a special case of $\Pi_n^{\mathbb{Z}_r}$, which is isomorphic to the intersection lattice of the *Dowling* *arrangement*, i.e. the arrangement in $\mathbb{C}^n$ of complex hyperplanes given by the equations $x_i = \zeta^k x_j$ and $x_l = 0$, where $i < j \in [n]$, $k \in [r]$, $l \in [n]$ and ζ is a primitive r:th root of unity.

For obvious reasons, the notation tends to get horrible when dealing with Dowling lattices. We agree on some conventions to simplify it. We write $i^g := (i, g)$ for $i \in [n]$ and $g \in G$. The G-orbit of a simple block in $\pi \in \Pi_n^G$ has cardinality $|G|$ and is of course completely determined by any representative. When we write out π, we therefore often omit all but one (arbitrary) block in every orbit of a simple block. Thus, $\pi = \pi_1|\ldots|\pi_t \in \Pi_n^G$ should be interpreted as an element with t G-orbits of blocks; hence with $(t-1)|G|+1$ blocks (since the null block is alone in its orbit). When the G-elements in the superscripts are irrelevant, namely in the null block and in singletons, we omit them, too. For example, we write $02|4|1^0 3^1$ for the element $0(2,0)(2,1)(2,2)|(4,0)|(4,1)|(4,2)|(1,0)(3,1)|(1,1)(3,2)|(1,2)(3,0)$ in $\Pi_4^{\mathbb{Z}_3}$.

We view an element in Π_n^G as a "signed" partition of $[n] \cup \{0\}$, where G is the group of "signs". Sometimes we wish to disregard the "signs". Therefore, for $S \subseteq ([n] \times G) \cup \{0\}$, we define $\overline{S} := \{i \in [n] \mid i^g \in S \text{ for some } g \in G\} \cup S^0$, where $S^0 := \{0\}$ if $0 \in S$ and $S^0 := \emptyset$ otherwise. With this, we can define the *absolute value* $\overline{\pi} \in \Pi_{[n]\cup\{0\}}$ of $\pi = \pi_1|\ldots|\pi_t \in \Pi_n^G$ by $\overline{\pi} := \overline{\pi_1}|\ldots|\overline{\pi_t}$. If π_i is the null block of π, then we say that $\overline{\pi_i}$ is the null block of $\overline{\pi}$.

5.2. **Dowling analogies of $L(Q,P)$.** Recall that x is a *modular* element in a ranked lattice L if $rank(x) + rank(y) = rank(x \vee y) + rank(x \wedge y)$ for all $y \in L$. Björner [3] observed that the lattice $L(Q,P)$ can be constructed in the following way, which suggests possible generalizations of the results in Section 3. Consider the modular element $\pi = p_1|\ldots|p_m|Q$ in the partition lattice $\Pi_{P\cup Q}$. Note that the set of complements $Co(\pi) := \{\tau \in \Pi_{P\cup Q} \mid \tau \wedge \pi = \hat{0} \text{ and } \tau \vee \pi = \hat{1}\}$ is precisely the set of atoms in $L(Q,P)$, so that $L(Q,P)$ is the lattice join-generated by $Co(\pi)$.

Now, let G be a finite group and consider the element $\pi = 0Q|p_1|\ldots|p_m$ in the Dowling lattice $\Pi_{P\cup Q}^G$ (meaning that we replace $[n]$ with $P \cup Q$ in Definition 5.1). By [5, Thm. 4], π is modular. Let $L^G(Q,P)$ be the lattice which is join-generated by $Co(\pi)$. Note that $Co(\pi)$ consists of the elements in which every simple block contains exactly one $Q \times G$-element and the null block contains no $Q \times G$-elements. Therefore, $L^G(Q,P)$ consists of those elements in $\Pi_{P\cup Q}^G$ in which every singleton is either 0 or from $Q \times G$ and every non-singleton intersects $P \times G$. In other words, for $\pi \in \Pi_{P\cup Q}^G$, we have $\pi \in L^G(Q,P)$ iff $\overline{\pi} \in L(Q \cup \{0\}, P)$.

We have $L^{\{e\}}(Q,P) \cong L(Q \cup \{0\}, P)$. The cases $G = \mathbb{Z}_2$ and $G = \mathbb{Z}_r$ are also interesting. As before, let V be a vector space over $\mathbf{k}$, and let r be a positive integer.

Definition 5.2. *Suppose that $char(\mathbf{k}) \neq 2$. The* polygraph arrangement of type B, *$Z_V^B(n,m)$, is the collection of all subspaces of the form*

$$\{(\tau_1 x_{f(1)}, \ldots, \tau_m x_{f(m)}, x_1, \ldots, x_n) \in V^{m+n} \mid x_i \in V \ \forall i \in [n]\}$$

over all $f : [m] \to [n]$ and $(\tau_1, \ldots, \tau_m) \in \{-1, 0, 1\}^m$.

Definition 5.3. *Suppose that $\mathbf{k}$ contains a primitive r:th root of unity. The* Dowling polygraph arrangement, *$Z_V^r(n,m)$, is the collection of all subspaces of the form*

$$\{(\tau_1 x_{f(1)}, \ldots, \tau_m x_{f(m)}, x_1, \ldots, x_n) \in V^{m+n} \mid x_i \in V \ \forall i \in [n]\}$$

over all $f : [m] \to [n]$ and $(\tau_1, \ldots, \tau_m) \in \{0, \zeta, \zeta^2, \ldots, \zeta^r = 1\}^m$, where ζ is a primitive r:th root of unity.

As before, we frequently suppress the vector space in the subscript. It is clear that $L(Z^B(n,m)) \cong L^{\mathbb{Z}_2}(Q,P)$ and $L(Z^r(n,m)) \cong L^{\mathbb{Z}_r}(Q,P)$.

It turns out that $L^G(Q,P)$ is EL-shellable. We define an edge-labelling $\omega : R(L^G(Q,P)) \to \Omega$, where Ω is the following poset of labels:

$$\Omega = \{\alpha_1 < \cdots < \alpha_n < A_2 < \cdots < A_n < B_1 < \cdots < B_n < \underbrace{00\ldots00}_{m} < \underbrace{00\ldots01}_{m} < \cdots < \underbrace{nn\ldots n}_{m} < \beta_1 < \cdots < \beta_n < C_2 < \cdots < C_n\}.$$

To simplify notation, we agree that from now on, the term *block* means a block which is neither a singleton nor a null block. Bearing this in mind, we define ω as follows:

- $\omega(\pi \to \tau) = \alpha_x$ if a block, $\overline{\pi_i}$, and the null block in $\overline{\pi}$ are merged in $\overline{\tau}$ and $q_x = max(\overline{\pi_i}^Q)$.
- $\omega(\pi \to \tau) = \beta_x$ if a singleton, q_x, and the null block in $\overline{\pi}$ are merged in $\overline{\tau}$.
- $\omega(\pi \to \tau) = A_x$ if two blocks, $\overline{\pi_i}$ and $\overline{\pi_j}$, in $\overline{\pi}$ are merged in $\overline{\tau}$ and $q_x = max(\overline{\pi_i}^Q \cup \overline{\pi_j}^Q)$.
- $\omega(\pi \to \tau) = B_x$ if a singleton, q_x, and a block, $\overline{\pi_i}$, in $\overline{\pi}$ are merged in $\overline{\tau}$ and $q_x < max(\overline{\pi_i}^Q)$.
- $\omega(\pi \to \tau) = C_x$ if a singleton, q_x, and a block, $\overline{\pi_i}$, in $\overline{\pi}$ are merged in $\overline{\tau}$ and $q_x > max(\overline{\pi_i}^Q)$.
- $\omega(\hat{0} \to \tau) = f(1)f(2)\ldots f(m)$ (juxtapositioning), where f is the function $f : [m] \to [n] \cup \{0\}$ which satisfies that $q_{f(i)}$ is the unique element in $Q \cup \{q_0\}$ sharing block (or null block) with p_i in $\overline{\tau}$. (Here, $q_0 := 0$.)

Given an atom $a \in L^G(Q,P)$, we define $f_a : [m] \to [n] \cup \{0\}$ by requiring that $f_a(1)\ldots f_a(m) = \omega(\hat{0} \to a)$.

We omit the proof of the following theorem; it is along the same lines as the proof of Theorem 4.1. Note, however, that ω does not specialize to the labelling λ of Section 4 when $G = \{e\}$.

Theorem 5.4. *The labelling ω is an EL-labelling of $\overline{L^G(Q,P)}$.* □

As in Section 4, we may exploit the EL-labelling ω to calculate the homotopy type of $\Delta(\overline{L^G(Q,P)})$. The key facts that we need are expressed in Lemma 5.8 and Lemma 5.9. The proof of the former requires three further lemmata. We omit the technical, but reasonably straightforward, proofs of two of them.

Lemma 5.5. *Let R be the set of elements in $L^G(Q,P)$ that contain no non-zero singletons. Fix an atom $a \in L^G(Q,P)$. Define ϕ_a^G to be the number of ω-falling chains $c = \{\hat{0} \to a = c_1 \to \ldots \to c_{t+1}\}$ such that $c_{t+1} \in R$ and $\omega(c_j \to c_{j+1})$ is a B-label for all $j \in [t]$. Then*

$$\phi_a^G = |G|^{n-|f_a([m])\setminus\{0\}|} \prod_{i\in[n]\setminus f_a([m])} |\{j \in f_a([m]) \mid j > i\}|. \quad \square$$

We define a map $\psi : 2^{[n]} \to \mathbb{N}$ by $\psi(S) = \prod_{i\in[n]\setminus S} |\{j \in S \mid j > i\}|$, so that we obtain $\phi_a^G = |G|^{n-|f_a([m])\setminus\{0\}|}\psi(f_a([m]) \setminus \{0\})$. The Stirling numbers are related to ψ in the following way:

Lemma 5.6. $\sum_{S\in\binom{[n]}{k}} \psi(S) = S(n,k)$.

Proof. $\psi(S)$ counts all partitions $\pi = \pi_1|\ldots|\pi_{|S|}$ of $[n]$ with the property that $S = \{max(\pi_i) \mid i \in [|S|]\}$. □

Lemma 5.7. *Let $S \subseteq [n]$ be fixed. Let Λ_S be the set of atoms $a \in L^G(Q, P)$ such that $f_a([m]) \setminus \{0\} = S$. Then*

$$|\Lambda_S| = |S|! \sum_{j=0}^{m-|S|} \binom{m}{j} S(m-j, |S|)|G|^{m-j}. \qquad \square$$

Lemma 5.8. *Let R_k be the set of elements ρ in $L^G(Q, P)$ such that $\overline{\rho}$ consists of a null-block, k blocks and no non-zero singletons. Define $\phi^G_\downarrow(k)$ to be the number of ω-falling chains $c = \{\hat{0} \to c_1 \to \ldots \to c_{t+1}\}$ such that $c_{t+1} \in R_k$ and $\omega(c_i \to c_{i+1})$ is a B-label for all $i \in [t]$. Then*

$$\phi^G_\downarrow(k) = S(n, k)k! \sum_{j=0}^{m-k} \binom{m}{j} S(m-j, k)|G|^{m+n-j-k}.$$

In particular, this number only depends on k.

Proof. Let c be as in the statement of the lemma. Since all $\overline{c_i}$, $i \in [t]$, have the same number of blocks, k, we obtain

$$\phi^G_\downarrow(k) = \sum_{S\in\binom{[n]}{k}} \sum_{a\in\Lambda_S} \phi^G_a =$$

$$\overset{(1)}{=} \sum_{S\in\binom{[n]}{k}} \sum_{a\in\Lambda_S} |G|^{n-k}\psi(S) =$$

$$= \sum_{S\in\binom{[n]}{k}} |\Lambda_S| \cdot |G|^{n-k}\psi(S) =$$

$$\overset{(2)}{=} |G|^{n-k} S(n, k)|\Lambda_S| =$$

$$\overset{(3)}{=} |G|^{n-k} S(n, k)k! \sum_{j=0}^{m-k} \binom{m}{j} S(m-j, k)|G|^{m-j}.$$

Here, (1) follows from Lemma 5.5, (2) follows from Lemma 5.6 and (3) follows from Lemma 5.7. $\square$

Lemma 5.9. *Suppose $\rho \in R_k$. Then the number of ω-falling chains in $[\rho, \hat{1}]$ is $\phi^G_\uparrow(k)$, where*

$$\phi^G_\uparrow(k) := (1 + |G|)(1 + 2|G|) \ldots (1 + (k-1)|G|).$$

In particular, this number only depends on k.

Proof. There is a natural isomorphism between $[\rho, \hat{1}]$ and the Dowling lattice Π^G_k obtained by identifying the blocks of ρ with the set $[k] \times G$ and the null block of ρ with 0. Hence, ω induces an EL-labelling of $\overline{\Pi^G_k}$. As then follows from Dowling's [5] computation of the Möbius function of Π^G_k, $\Delta(\overline{\Pi^G_k})$ is homotopy equivalent to a wedge of $\phi^G_\uparrow(k)$ spheres (of top dimension). Therefore, by Theorem 2.3, $\phi_\uparrow(k)$ must be the number of ω-falling chains in $[\rho, \hat{1}]$. $\square$

Now, we are ready to count the falling chains in $L^G(Q, P)$.

Theorem 5.10. *$\Delta(\overline{L^G(Q, P)})$ is homotopy equivalent to a wedge of $(n-1)$-dimensional spheres. Let $\Gamma^G(n, m)$ denote the number of spheres in the wedge. Then,*

$$\Gamma^G(n, m) = \sum_{k=1}^{min(n,m)} S(n, k)k! \sum_{j=0}^{m-k} \binom{m}{j} S(m-j, k)|G|^{m+n-j-k} \prod_{i=1}^{k-1}(1 + i|G|).$$

Proof. It is clear that the number of falling chains in $L^G(Q,P)$ under ω is $\sum_{k=0}^{min(n,m)} \phi_{\downarrow}^G(k)\phi_{\uparrow}^G(k)$. The theorem now follows from Lemma 5.8, Lemma 5.9 and Theorem 2.3. □

Below, let $\alpha = min(n+1,m)$. We check that Theorem 5.10 indeed generalizes Theorem 4.2. Note that

$$\begin{aligned}
\Gamma^{\{e\}}(n,m) &= \textstyle\sum_{k=1}^{min(n,m)} S(n,k)k! \sum_{j=0}^{m-k} \binom{m}{j} S(m-j,k)k! = \\
&= \textstyle\sum_{k=1}^{min(n,m)} S(n,k)S(m+1,k+1)(k!)^2 = \\
&= \textstyle\sum_{k=1}^{\alpha} S(n,k)S(m+1,k+1)(k!)^2 = \\
&= \textstyle\sum_{k=1}^{\alpha} \frac{S(n+1,k)-S(n,k-1)}{k}(S(m,k)+(k+1)S(m,k+1))(k!)^2 = \\
&= \Gamma(n+1,m) - \textstyle\sum_{k=1}^{\alpha} S(n,k-1)S(m,k)k!(k-1)! + \\
&\quad + \textstyle\sum_{k=1}^{\alpha} S(n,k)S(m,k+1)(k+1)!k! = \\
&\overset{(\star)}{=} \Gamma(n+1,m) - S(n,0)S(m,1)1!0! + S(n,\alpha)S(m,\alpha+1)(\alpha+1)!\alpha! = \\
&= \Gamma(n+1,m),
\end{aligned}$$

as required. The identity $(\star)$ follows from substituting $j = k-1$ in the first sum on the left hand side.

Corollary 5.11. *Pick $\pi \in L^G(Q,P)$. Suppose that $\overline{\pi} = 0\pi_0|\pi_1|\pi_2|\ldots|\pi_t$. Then, $\Delta(\overline{[\hat{0},\pi]})$ is homotopy equivalent to a wedge of top-dimensional spheres. The number of spheres in the wedge is $\Gamma^G(|\pi_0^Q|,|\pi_0^P|) \cdot \prod_{i=1}^{t} \Gamma(|\pi_i^Q|,|\pi_i^P|)$.*

Proof. Note that $[\hat{0},\pi] \cong L^G(\pi_0^Q,\pi_0^P) \times L(\pi_1^Q,\pi_1^P) \times \cdots \times L(\pi_t^Q,\pi_t^P)$. The rest of the proof is analogous to the proof of Corollary 4.3. □

As in Section 4, Corollary 5.11 provides the information needed to calculate the cohomology groups of $\mathcal{M}_{Z^r(n,m)}$ with the Goresky-MacPherson formula, thereby obtaining a generalization of Theorem 4.4. We omit the details.

References

[1] A. Björner, Shellable and Cohen-Macaulay partially ordered sets, *Trans. Amer. Math. Soc.* **260** (1980), 159–183.
[2] A. Björner, Subspace arrangements, *First European congress of math.*, Paris, 1992 (A. Joseph et al., eds.), 321–370, *Progr. Math* **119**, *Birkhäuser, Basel*, 1994.
[3] A. Björner, personal communication, 2001.
[4] A. Björner and M. Wachs, Shellable nonpure complexes and posets I, *Trans. Amer. Math. Soc.* **348** (1996), 1299–1327.
[5] T. A. Dowling, A class of geometric lattices based on finite groups, *J. Combin. Theory Ser. B* **14** (1973), 61–86.
[6] A. M. Garsia and M. Haiman, A graded representation model for Macdonald's polynomials, *Proc. Nat. Acad. Sci. U.S.A.* **90** (1993), 3607–3610.
[7] M. Goresky and R. D. MacPherson, Stratified Morse theory, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 14, Springer-Verlag, Berlin/Heidelberg/New York, 1988.
[8] M. Haiman, Hilbert schemes, polygraphs, and the Macdonald positivity conjecture, *J. Amer. Math. Soc*, to appear.

[9] M. Haiman, Macdonald polynomials and geometry, in *New perspectives in algebraic combinatorics* (L. J. Billera et al., eds.), 207–254, Cambridge Univ. Press, 1999.
[10] M. Haiman, personal communication, 2001.
[11] I. G. Macdonald, A new class of symmetric functions, *Actes du 20e Séminaire Lotharingien*, vol. 372/S-20, Publications I.R.M.A., Strasbourg, 1988, 131–171.
[12] J. R. Munkres, *Elements of algebraic topology*, Addison-Wesley, Menlo Park, 1984.
[13] R. P. Stanley, *Enumerative combinatorics*, vol. 1, Cambridge Univ. Press, 1997.
[14] G. M. Ziegler and R. T. Živaljević, Homotopy types of subspace arrangements via diagrams of spaces, *Math. Ann.*, **295** (1993), 527–548.

Department of Mathematics, Royal Institute of Technology, S-100 44, Stockholm, Sweden
E-mail address: `axel@math.kth.se`

TRIBASIC INTEGRALS AND IDENTITIES OF ROGERS-RAMANUJAN TYPE

M. E. H. ISMAIL AND D. STANTON

Abstract. Using complex analysis, integrals involving three independent q's are evaluated as infinite products. This leads to identities of Rogers-Ramanujan type.

Résumé. En utilisant trois q independant, nous evaluons des integraux en analyse complexe et trouver des produits infinis. Ceci nous donnons des generalisations des identites de Rogers-Ramanujan.

1. Introduction

The purpose of this extended abstract (no proofs are given) is to show the extensive relationship between integrals and identities of Rogers-Ramanujan type. We concentrate on integral evaluations involving infinite products with three independent q's.

In [4, 6], it was shown that the Rogers-Ramanujan identities of modulus 5 follow from evaluating the bibasic integral

$$\frac{(qt^2;q)_\infty}{2\pi}\int_0^\pi \frac{(q^5,e^{2i\theta},e^{-2i\theta};q^5)_\infty}{(te^{i\theta},te^{-i\theta};q)_\infty}d\theta. \tag{1}$$

We generalize the above integral by replacing the three bases of the infinite products q^5, q and q by independent bases s, p and q. The resulting integral can be evaluated as an infinite product for special values of t. By specializing s, p, and q the product sides of identities of Rogers-Ramanujan type appear. The special values of t for which such identities exist can be found by considering the singularities of the integrals as functions of t.

2. Integrals

We use a variant of Schwarz's theorem [1] to evaluate limits of integrals.

Theorem 1. *Let $f(\theta,z)$ be continuous in θ for $\theta\in[0,2\pi]$, and for all z so that $r\geq|z|\geq r-\epsilon$ for some positive ϵ. Assume further that $f(\theta,z)$ converges to $f(\theta,re^{i\phi})$ as $z\to re^{i\phi}$ uniformly in θ, for $\theta\in[0,2\pi]$. Then*

$$\lim_{z\to re^{i\phi}}\int_0^{2\pi}\frac{(r^2-|z|^2)\,f(\theta,z)d\theta}{2\pi|re^{i\theta}-z|^2}=f(\phi,re^{i\phi}).$$

We next apply Theorem 1 to find limiting values of the tribasic integrals

$$G_1(t,p,q,s)=\frac{(-t;q)_\infty}{2\pi}\int_0^\pi\frac{(s,e^{2i\theta},e^{-2i\theta};s)_\infty}{(te^{2i\theta},te^{-2i\theta};p)_\infty}\,d\theta,\quad |t|<1,$$

and

$$G_2(t,p,q,s)=\frac{(pt^2;q)_\infty}{2\pi}\int_0^\pi\frac{(s,e^{2i\theta},e^{-2i\theta};s)_\infty}{(te^{i\theta},te^{-i\theta};p)_\infty}\,d\theta,\quad |t|<1.$$

Research partially supported by NSF grants DMS 99-70865, DMS 99-70627.

Note that $G_2(t,q,q,q^5)$ is the integral (1) of the introduction giving the Rogers-Ramanujan identities modulo 5.

Corollary 2. *We have an analytic continuation of $G_2(t,p,q,s)$ to $p^{1/2} < |t| < p^{-1/2}$ such that*

$$\begin{aligned}\lim_{t\to -1} G_1(t,p,q,s) &= \frac{(q;q)_\infty}{(p,p;p)_\infty}\,(s,-s,-s;s)_\infty,\\ \lim_{t\to p^{-1/2}} G_2(t,p,q,s) &= \frac{(s,p,s/p;s)_\infty(q;q)_\infty}{(p,p;p)_\infty}.\end{aligned}$$

In particular for the Rogers-Ramanujan identities modulo 5,

$$G_2(q^{-1/2},q,q,q^5) = \frac{1}{(q^2,q^3;q^5)_\infty}.$$

One can also extend the evaluation for G_2 using further analytic continuations. Let $F(z) = (s,z^2,1/z^2;s)_\infty$.

Theorem 3. *For $k = 1,2,\ldots$, $G_2(t,p,p,s)$ can be analytically continued to*

$$p^{1-k/2} < |t| < p^{-k/2}, \textit{ via}$$

$$G_2(t,p,p,s) = \frac{(pt^2;p)_\infty}{4\pi}\int_0^{2\pi} \frac{F(e^{i\theta}p^{-k/2})}{(te^{i\theta}p^{-k/2},tp^{k/2}e^{-i\theta};p)_\infty}\,d\theta$$

$$+\frac{(s;s)_\infty}{2(1-t^2)(p;p)_\infty}\sum_{j=0}^{k-1}\frac{(t^2;p)_j(t^2p^{2j},t^{-2}p^{-2j};s)_\infty}{(1/p;1/p)_j}.$$

Furthermore

$$\lim_{t\to p^{-k/2}} G_2(t,p,p,s) = \frac{1}{2(1-p^{-k})(p;p)_\infty}\sum_{j=0}^{k}\begin{bmatrix}k\\ j\end{bmatrix}_p (s,p^{2j-k},p^{k-2j};s)_\infty p^{j(j-k)}.$$

3. Sums

Using techniques from orthogonal polynomials, one can give power series representations in t for several special choices of p in the integrals

$$S_{p,q}(t) = G_2(t,q,q,p) = \frac{(qt^2;q)_\infty(p;p)_\infty}{2\pi}\int_0^{\pi}\frac{(e^{2i\theta},e^{-2i\theta};p)_\infty}{(te^{i\theta},te^{-i\theta};q)_\infty}d\theta,$$

$$H_{p,q}(t) = G_2(t,p,p^2,q^2) = \frac{(q^2;q^2)_\infty(pt^2;p^2)_\infty}{2\pi}\int_0^{\pi}\frac{(e^{2i\theta},e^{-2i\theta};q^2)_\infty}{(te^{i\theta},te^{-i\theta};p)_\infty}d\theta,$$

$$J_{p,q}(t) = G_1(t,p^2,p,q) = \frac{(q;q)_\infty(-t;p)_\infty}{2\pi}\int_0^{\pi}\frac{(e^{2i\theta},e^{-2i\theta};q)_\infty}{(te^{2i\theta},te^{-2i\theta};p^2)_\infty}d\theta.$$

For example

$$
\begin{aligned}
S_{q^5,q}(t) &= \sum_{n=0}^{\infty} \frac{q^{n^2+2n}t^{2n}}{(q;q)_n},\\
S_{q^5,q^2}(t) &= \sum_{n=0}^{\infty} \frac{(q;q^2)_n}{(q^2;q^2)_{2n}} q^{n(3n+2)}(-t^2)^n,\\
S_{q^7,q^2}(t) &= \sum_{n=0}^{\infty} \frac{(q;q^2)_n}{(q^2;q^2)_{2n}} q^{2n(n+2)}t^{2n},\\
S_{-q^3,q}(t) &= \sum_{n=0}^{\infty} \frac{(-1;q^2)_n}{(q;q)_{2n}} q^{n(n+2)}t^{2n}.\\
S_{\omega q^3,q}(t) &= 1+(1-\omega)\sum_{n=1}^{\infty} \frac{(q^3;q^3)_{n-1}}{(q;q)_{2n}(q;q)_{n-1}} q^{n(n+2)}t^{2n}.
\end{aligned}
$$

$$
\begin{aligned}
H_{q^2,q}(t) &= \sum_{n=0}^{\infty} \frac{q^{2n^2}(-t^2)^n}{(q^4;q^4)_n} = (q^2t^2;q^4)_\infty,\\
H_{iq,q}(t) &= \sum_{n=0}^{\infty} \frac{(-1;q^4)_n}{(iq;iq)_{2n}}(-qt)^{2n},\\
H_{q,q^2}(t) &= \sum_{n=0}^{\infty} \frac{(q^2;q^4)_n}{(q;q)_{2n}}(q^2t^2)^n = \frac{(-q^3t^2;q^2)_\infty}{(q^2t;q^2)_\infty},\\
H_{q,q^3}(t) &= \sum_{n=0}^{\infty} \frac{(-q;q)_{2n}}{(q^2;q^2)_n}(qt)^{2n},\\
H_{q^2,q^3}(t) &= \sum_{n=0}^{\infty} \frac{(q^4t^2)^n}{(q^4;q^4)_n} = \frac{1}{(q^4t^2;q^4)_\infty},
\end{aligned}
$$

$$
\begin{aligned}
J_{q^2,q}(t) &= \sum_{n=0}^{\infty} \frac{(q;q^2)_n}{(q^4;q^4)_n} q^{n^2}(-t)^n,\\
J_{-q,q}(t) &= \sum_{n=0}^{\infty} \frac{(-1;q^2)_n}{(q^2;q^2)_n}(-qt)^n = \frac{(qt;q^2)_\infty}{(-qt;q^2)_\infty},\\
J_{q,q^2}(t) &= \sum_{n=0}^{\infty} \frac{(q;q^2)_n}{(q^2;q^2)_n}(qt)^n = \frac{(q^2t;q^2)_\infty}{(qt;q^2)_\infty},\\
J_{q,q^3}(t) &= \sum_{n=0}^{\infty} \frac{(q;q)_{2n}}{(q;q)_n(q^2;q^2)_n}(qt)^n,\\
J_{q^2,q^3}(t) &= \sum_{n=0}^{\infty} \frac{(q;q^2)_n}{(q^4;q^4)_n}(q^2t)^n.
\end{aligned}
$$

Note that several of these series converge for all t, thus are analytic continuations of the integrals, allowing one to specialize t, and apply Corollary 2 and Theorem 3. These are Rogers-Ramanujan identities. For example, specializing $S_{\omega q^3,q}(t)$, $\omega = e^{2\pi i/3}$, yields

Theorem 4. *We have*

$$1 \;+\; (1-\omega)\sum_{n=1}^{\infty}\frac{(q^3;q^3)_{n-1}}{(q;q)_{2n}(q;q)_{n-1}}q^{n(n+1)} = \frac{(\omega q^3,\omega q^2,q;\omega q^3)_\infty}{(q;q)_\infty}$$

$$1 \;+\; (1-\omega)\sum_{n=1}^{\infty}\frac{(q^3;q^3)_{n-1}}{(q;q)_{2n}(q;q)_{n-1}}q^{n^2} = \frac{(\omega q^3,q^2,\omega q;\omega q^3)_\infty}{(q;q)_\infty}.$$

4. Integer partition interpretations

Several of the identities are equivalent to integer partition statements, here are three examples.

Corollary 5. *Let $A(n)$ be the number of integer partitions of n into parts congruent to 2,4, 10 or 12 mod 14 and distinct parts congruent to 1,5,7,9, or 13 mod 14. Let $B(n)$ be the number of integer partitions of n*
(1) whose odd parts are consecutive (starting with 1) and have multiplicity one or two,
(2) whose largest even part is at most two more than twice the largest odd part.
Then $A(n) = B(n)$.

Corollary 6. *Let $A(n)$ be the number of integer partitions of n into parts congruent to 2,6,8, or 12 mod 14 and distinct parts congruent to 3,7, or 11 mod 14. Let $B(n)$ be the number of integer partitions of n*
(1) whose odd parts are consecutive (starting with 1) and have multiplicity two or three,
(2) whose largest even part is at most two more than twice the largest odd part.
Then $A(n) = B(n)$.

Corollary 7. *Let $A(n)$ be the number of integer partitions of n into parts congruent to 4 or 6 mod 10 and distinct parts congruent to 3,5, or 7 mod 10. Let $B(n)$ be the number of integer partitions of n*
(1) whose odd parts are consecutive (starting with 1) and have multiplicity three or four,
(2) whose largest even part is at most two more than twice the largest odd part.
Then $A(n) = B(n)$.

5. m-versions

The Rogers-Ramanujan identities have the natural generalization [4]

$$\sum_{n=0}^{\infty}\frac{q^{n^2+mn}}{(q;q)_n} = \frac{a_m(q)}{(q,q^4;q)_\infty} + \frac{b_m(q)}{(q^2,q^3;q)_\infty}, \tag{2}$$

where $a_m(q)$ and $b_m(q)$ are Laurent polynomials in q which are explicitly known. We refer to (2) as an "m-version" of the Rogers-Ramanujan identities.

According to Theorem 3

$$\lim_{t\to q^{-m/2}} S_{p,q}(t) = \frac{1}{2(1-q^{-m})}\sum_{j=0}^{m}\begin{bmatrix} m \\ j\end{bmatrix}_q \frac{(p,q^{2j-m},q^{m-2j};p)_\infty}{(q;q)_\infty}q^{j(j-m)}, \tag{3}$$

for $m = 1, 2, \ldots$. Equation (3) generalizes (2) and gives an explicit form for the generalizations of the polynomials $a_m(q)$ and $b_m(q)$. This alternating form is a special case of the hook difference polynomials in [3]. However explicit positive forms may be found using the recurrence relations for the polynomials. In the mod 5 case, this recurrence is three-term and related to orthogonal polynomials. But in other cases higher order recurrences do occur.

For example

$$\begin{aligned}
S_{q^7,q^2}(q^{-m}) &= \sum_{n=0}^{\infty} \frac{(q;q^2)_n}{(q^2;q^2)_{2n}}\, q^{2n^2+4n-2mn} \\
&= c_1(m,7,2)\frac{(q^7,q^1,q^6;q^7)_\infty}{(q^2,q^2)_\infty} + c_2(m,7,2)\frac{(q^7,q^2,q^5;q^7)_\infty}{(q^2,q^2)_\infty} \\
&+ c_3(m,7,2)\frac{(q^7,q^3,q^4;q^7)_\infty}{(q^2,q^2)_\infty}.
\end{aligned}$$

where $g_m = c_i(m+2,7,2)$ satisfies the four term recurrence relation

$$g_{m+2} + q^{-1}g_{m+1} - (1+q^{-2-2m})g_m - q^{-1}g_{m-1} = 0.$$

The explicit positive forms are ($s = q^{-1}$)

$$\begin{aligned}
c_1(n+2,7,2) &= (-1)^{n-1} \sum_{2m+2j+k+1=n} \begin{bmatrix} m+j \\ j \end{bmatrix}_{s^4} \begin{bmatrix} m+k \\ k \end{bmatrix}_{s^2} s^{2m(m+1)+k} \\
c_2(n+2,7,2) &= (-1)^{n-1} \sum_{2m+2j+k+1=n} \begin{bmatrix} m+j \\ j \end{bmatrix}_{s^4} \begin{bmatrix} m+k-1 \\ k \end{bmatrix}_{s^2} s^{2m(m+1)+k}, \\
c_3(n+2,7,2) &= (-1)^{n} \sum_{2m+2j+k=n} \begin{bmatrix} m+j \\ j \end{bmatrix}_{s^4} \begin{bmatrix} m+k-1 \\ k \end{bmatrix}_{s^2} s^{2m^2+k}.
\end{aligned}$$

One may also find these explicit positive forms for $n < 0$.

Of the identities on Slater's list [5], 39 involve 3-term recurrences and 59 involve 4-term recurrences.

References

[1] L. Ahlfors, *Complex Analysis*, McGraw-Hill, New York, 1976.
[2] G. Andrews, *The Theory of Partitions*, Cambridge University Press, New York, 1976.
[3] G. Andrews, R. Baxter, D. Bressoud, W. Burge, P. Forrester, and G. Viennot, "Partitions with prescribed hook differences", Eur. J. Comb. **8** (1987), 341-350.
[4] K. Garrett, M.E.H. Ismail, and D. Stanton, "Variants of the Rogers-Ramanujan identities", Adv. Appl. Math. **23**, (1999), 274–299.
[5] L. Slater, "Further identities of the Rogers-Ramanujan type", Proc. Lon. Math. Soc. (2), **54**, (1952), 147–167.
[6] D. Stanton, "Gaussian Integrals and the Rogers-Ramanujan identities", in Symbolic computation, number theory, special functions, physics and combinatorics, eds. F. Garvan, M. Ismail, Kluwer, 2001, p. 255-265.

Department of Mathematics, University of South Florida, Tampa, Florida 33620-5700
E-mail address: ismail@hahn.math.usf.edu

School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455
E-mail address: stanton@math.umn.edu

PARTIALLY ORDERED GENERALIZED PATTERNS

SERGEY KITAEV

Abstract. We introduce partially ordered generalized patterns (POGPs), which further generalize the generalized permutation patterns (GPs) introduced by Babson and Steingrímsson [BabStein]. A POGP p is a GP some of whose letters are incomparable. Thus, in an occurrence of p in a permutation π, two letters that are incomparable in p pose no restrictions on the corresponding letters in π. We describe many relations between POGPs and GPs and give general theorems about the number of permutations avoiding certain classes of POGPs. These theorems have several known results as corollaries but also give many new results. We also give the generating function for the entire distribution of the maximum number of non-overlapping occurrences of a pattern p with no dashes, provided we know the e.g.f. for the number of permutations that avoid p.

Résumé. On étudie les motifs partiellement ordonnés généralisés (POGP), une généralisation des motifs de PERmutation généralisés (GP) définis par Babson et Steingrímsson [BabStein]. Un POGP p est un GP dont certaines lettres sont incomparables. Ainsi, dans une apparition de p dans une permutation π, deux lettres incomparables de p ne posent aucune contrainte sur les lettres correspondantes de π.

On décrit de nombreuses relations entre les PGOP et les GP et on donne des théorèmes généraux sur le nombre de permutations évitant certaines classes de POGP. Ces théorèmes ont pour corollaires plusieurs résultats connus mais mènent également à des résultats nouveaux.

Nous dérivons la fonction génératrice pour toute la distribution du nombre maximal d'apparitions sans superposition d'un motif p sans tirets, conaissant la fonction génératrice exponentielle du nombre de permutations évitant p.

1. Introduction and Background

All permutations in this paper are written as words $\pi = a_1a_2\cdots a_n$, where the a_i consist of all the integers $1, 2, \ldots, n$.

We will be concerned with *patterns* in permutations. A pattern is a word on some alphabet of letters, where some of the letters may be separated by dashes. In our notation, the classical permutation patterns, first studied systematically by Simion and Schmidt [SchSim], are of the form $p = 1{-}3{-}2$, the dashes indicating that the letters in a permutation corresponding to an occurrence of p don't have to be adjacent. In the classical case, an occurrence of a pattern p in a permutation π is a subsequence in π (of the same length as the length of p) whose letters are in the same relative order as those in p. For example, the permutation 41352 has only one occurrence of the pattern $1-2-3$, namely the subword 135.

Note that a classical pattern should, in our notation, have dashes at the beginning and end. Since all patterns considered in this paper satisfy this, we suppress these dashes from the notation. Thus, a pattern with no dashes corresponds to a contiguous subword anywhere in a permutation.

In [BabStein] Babson and Steingrímsson introduced *generalized permutation patterns (GPs)* where two adjacent letters in a pattern may be required to be adjacent in the permutation. Such an adjacency requirement is indicated by the absence of a dash between the corresponding letters in the pattern. For example, the permutation $\pi = 516423$ has only one occurrence of the pattern 2-31, namely the subword 564, but the pattern 2-3-1 occurs also in the subwords 562 and 563. The motivation for introducing these patterns in [BabStein] was the study of Mahonian statistics.

A number of interesting results on GPs were obtained by Claesson in [Claes]. Relations to several well studied combinatorial structures, such as set partitions, Dyck paths, Motzkin paths and involutions, were shown there. In [Kit] the present author investigated simultaneous avoidance of two or more 3-letter GPs with no dashes. This work is of particular interest here since avoidance of the patterns considered in this paper has a close connection to simultaneous avoidance of two or more GPs with no dashes. Also important here is the work of Elizalde and Noy [ElizNoy] where they find the distribution of several patterns with no dashes.

In this paper we introduce a further generalization of GPs — namely *partially ordered generalized patterns (POGP)*. A POGP is a GP some of whose letters are incomparable. For instance, if we write $p = 1 - 1'2'$ then we mean that in an occurrence of p in a permutation π the letter corresponding to the 1 in p can be either larger or smaller than the letters corresponding to $1'2'$. Thus, the permutation 13425 has four occurrences of p, namely 134, 125, 325 and 425.

We consider two particular classes of POGPs — *shuffle patterns* and *multi-patterns*. A multi-pattern is of the form $p = \sigma_1 - \sigma_2 - \cdots - \sigma_k$ and a shuffle pattern is of the form $p = \sigma_0 - a_1 - \sigma_1 - a_2 - \cdots - a_k - \sigma_k$, where for any i and j, the letter a_i is greater than any letter of σ_j and for any $i \neq j$ each letter of σ_i is incomparable with any letter of σ_j These patterns are investigated in Sections 4 and 5. A corollary to one of our theorems (Theorem 13) about the shuffle patterns is the result of Claesson [Claes, Proposition 2] that the number of n-permutations that avoid the pattern $12 - 3$ is the n-th Bell number.

Let $p = \sigma_1 - \sigma_2 - \cdots - \sigma_k$ be an arbitrary multi-pattern and let $A_i(x)$ be the exponential generating function (e.g.f.) for the number of permutations that avoid σ_i for each i. In Theorem 28 we find the e.g.f., in terms of the $A_i(x)$, for the number of permutations that avoid p. In particular, this allows us to find the e.g.f. for the entire *distribution* of the maximum number of non-overlapping occurrences of a pattern p with no dashes, if we only know the e.g.f. for the number of permutations that *avoid* p. In many cases, this gives nice generating functions.

We also give alternative proofs, using inclusion-exclusion, of some of the results of Elizalde and Noy [ElizNoy]. Our proofs result in explicit formulas for the e.g.f. in terms of infinite series whereas Elizalde and Noy obtained differential equations for the same e.g.f..

2. Definitions and Preliminaries

A *partially ordered generalized pattern (POGP)* is a GP where some of the letters can be incomparable.

Example 1. The simplest non-trivial example of a POGP that differs from the ordinary GPs is $p = 1' - 2 - 1''$, where the second letter is the greatest one and the first and the last letters are

incomparable to each other. The permutation 3142 has two occurrences of p, namely, the subwords 342 and 142.

It is easy to see that the number of permutations that avoid p in Example 1 is equal to 2^{n-1}. Indeed, if $\pi = a_1 \dots a_n$ and a_i is the leftmost letter in π that is smaller than its successor, then all letters to the right of a_i must be in increasing order. So any permutation π avoiding p can be written as $\pi_1 1 \pi_2$, where π_1 is decreasing and π_2 is increasing and there are 2^{n-1} ways to pick the permutation π_1, which determines π.

Definition 2. If the number of permutations in S_n, for each n, that avoid a POGP p is equal to the number of permutations that avoid a POGP q, then p and q are said to be *equivalent* and we write $p \equiv q$ in this case.

If A_n is the number of n-permutations that avoid a pattern p, then the *exponential generating function*, or *e.g.f.*, of the class of such permutations is

$$A(x) = \sum_{n \geq 0} A_n \frac{x^n}{n!}.$$

We will talk about *bivariate generating functions*, or *b.g.f.*, exclusively as generating functions of the form

$$A(u,x) = \sum_{\pi} u^{p(\pi)} \frac{x^{|\pi|}}{|\pi|!} = \sum_{n,k \geq 0} A_{n,k} u^k \frac{x^n}{n!},$$

were $A_{n,k}$ is the number of n-permutations with k occurrences of the pattern p.

The *reverse* $R(\pi)$ of a permutation $\pi = a_1 a_2 \ldots a_n$ is the permutation $a_n a_{n-1} \ldots a_1$. The *complement* $C(\pi)$ is the permutation $b_1 b_2 \ldots b_n$ where $b_i = n + 1 - a_i$. Also, $R \circ C$ is the composition of R and C. For example, $R(13254) = 45231$, $C(13254) = 53412$ and $R \circ C(13254) = 21435$. We call these bijections of $\mathcal{S}_n$ to itself *trivial*, and it is easy to see that any pattern p is equivalent to the patterns $R(p)$, $C(p)$ and $R \circ C(p)$. For example, the number of permutations that avoid the pattern 132 is the same as the number of permutations that avoid the patterns 231, 312 and 213, respectively.

It is convenient to introduce the following definition.

Definition 3. Let p be a GP without internal dashes. A permutation π *quasi-avoids* p if π has exactly one occurrence of p and this occurrence consists of the $|p|$ rightmost letters of π.

For example, the permutation 51342 quasi-avoids the pattern $p = 231$, whereas the permutations 54312 and 45231 do not. Indeed, 54312 ends with 312, which is not an occurrence of the pattern p, and 45231 has an occurrence of p, namely 452, in a forbidden place.

Proposition 4. *Let p be a non-empty GP with no dashes. Let $A(x)$ (resp. $A^\star(x)$) be the e.g.f. for the number of permutations that avoid (resp. quasi-avoid) p. Then*

$$A^\star(x) = (x-1)A(x) + 1.$$

Proof. We first show that

$$A^\star_n = nA_{n-1} - A_n. \tag{1}$$

If we consider all $(n-1)$-permutations that avoid p and all possible extending of these permutations to the n-permutations by writing one more letter to the right, then the number of obtained permutations will be nA_{n-1}. Obviously, the set of these permutations is a disjoint union of the set of all n-permutations that avoid p and the set of all n-permutations that quasi-avoid p. Thus we get (1). Multiplying both sides of (1) with $x^n/n!$ and summing over all natural numbers n, observing that $A^\star_0 = 0$, we get the desired result. □

Definition 5. Suppose $\{\sigma_0, \sigma_1, \ldots, \sigma_k\}$ is a set of GPs with no dashes and $p = \sigma_1 - \sigma_2 - \cdots - \sigma_k$ where each letter of σ_i is incomparable with any letter of σ_j whenever $i \neq j$. We call such POGPs *multi-patterns*.

Definition 6. Suppose $\{\sigma_0, \sigma_1, \dots, \sigma_k\}$ is a set of GPs with no dashes and $a_1a_2\dots a_k$ is a permutation of k letters. We define a *shuffle* pattern to be a pattern of the form

$$\sigma_0 - a_1 - \sigma_1 - a_2 - \cdots - \sigma_{k-1} - a_k - \sigma_k,$$

where for any i and j, the letter a_i is greater than any letter of σ_j and for any $i \neq j$ each letter of σ_i is incomparable with any letter of σ_j. We also allow σ_0 and σ_k, but not the other σ_i, to be empty patterns.

The pattern from Example 1 is an example of a shuffle pattern. It follows from the definitions that we can get a multi-pattern from a shuffle pattern by removing all the a_i.

Let $\mathcal{S}_\infty$ denote the disjoint union of the $\mathcal{S}_n$ for all $n \in \mathbb{N}$. The POGPs (which include the GPs, as well as the classical patterns), can be considered as functions from $\mathcal{S}_\infty$ to $\mathbb{N}$ that count the number of occurrences of the pattern in a permutation in $\mathcal{S}_\infty$. This allows us to write a POGP (as a function) as a linear combination of GPs. For example,

$$1' - 2 - 1'' \;=\; (1-3-2) \;+\; (2-3-1),$$

from which, in particular, we see that to avoid $1' - 2 - 1''$ is the same as to avoid simultaneously the patterns $1-3-2$ and $2-3-1$. A straightforward argument leads to the following proposition.

Proposition 7. *For any POGP p there exists a set S of GPs such that a permutation π avoids p if and only if π avoids all the patterns in S.*

The following theorem can be easily proved by induction on k:

Theorem 8. *Let $p_1 = \sigma_0 - a_1 - \sigma_1 - a_2 - \cdots - \sigma_{k-1} - a_k - \sigma_k$ (resp. $p_2 = \sigma_0 - \sigma_1 - \cdots - \sigma_k$) be an arbitrary shuffle pattern (resp. multi-pattern) with $|\sigma_i| = \ell_i$ for all $i = 0, \dots, k$. Then to avoid the pattern p_1 (resp. p_2) is the same as to avoid*

$$\prod_{i=1}^{k} \binom{\ell_0 + \ell_1 + \cdots + \ell_i}{\ell_i} = \binom{\ell_0 + \ell_1}{\ell_1}\binom{\ell_0 + \ell_1 + \ell_2}{\ell_2}\cdots\binom{\ell_0 + \ell_1 + \cdots + \ell_k}{\ell_k}$$

ordinary GPs.

Example 9. Let $p = 1'2' - 3 - 1''$. That is $\sigma = 12$ and $\tau = 1$. By Theorem 8, to avoid p is the same as to avoid $\binom{3}{2} = 3$ GPs simultaneously, namely $12-4-3$, $13-4-2$ and $23-4-1$.

There is a number of results on the distribution of several classes of patterns with no dashes. These results can be used as building blocks for some of the results in the present paper. The most important of these is the following result by Elizalde and Noy [ElizNoy]:

Theorem 10. [ElizNoy, Theorem 3.4] *Let m and a be positive integers with $a \leq m$, let $\sigma = 12\dots a\tau(a+1) \in \mathcal{S}_{m+2}$, where τ is any permutation of $\{a+2, a+3, \dots, m+2\}$, and let $P(u,z)$ be the b.g.f. for permutations where u marks the number of occurrences of σ. Then $P(u,z) = 1/w(u,z)$, where w is the solution of*

$$w^{a+1} + (1-u)\frac{z^{m-a+1}}{(m-a+1)!}w' = 0$$

with $w(0) = 1$, $w'(0) = -1$ and $w^{(k)}(0) = 0$ for $2 \leq k \leq a$. In particular, the distribution does not depend on τ.

3. GPs with no dashes

In order to apply our results in what follows we need to know how many patterns avoid a given ordinary GP with no dashes. We are also interested in different approaches to studying these patterns. The theorems in this section can be proved using an inclusion-exclusion argument similar to the one given in the proof of Theorem 30 and we omit these proofs. This allows us to get explicit formulas for the e.g.f. in terms of infinite series instead of having to solve differential equations as done by Elizalde and Noy [ElizNoy] for the same e.g.f.. However, in particular cases, we use certain differential equations to simplify our series.

Theorem 11. [GoulJack] *Let $A_k(x)$ be the e.g.f. for the number of permutations avoiding the pattern $p = 123\dots k$. Then*

$$A_k(x) = 1/F_k(x),$$

where $F_k(x) = \displaystyle\sum_{i\geq 0} \frac{x^{ki}}{(ki)!} - \sum_{i\geq 0} \frac{x^{ki+1}}{(ki+1)!}$.

For some k it is possible to simplify the function $F_k(x)$ in the theorems above. Indeed, $F_k(x)$ satisfies the differential equation $F_k^{(k)}(x) = F_k(x)$ with the k initial conditions $F_k(0) = 1$, $F_k'(0) = -1$ and $F_k^{(i)}(0) = 0$ for all $i = 2, 3, \dots, k-1$. For instance, if $k = 4$ then

$$F_4(x) = \frac{1}{2}(\cos x - \sin x + e^{-x}).$$

Theorem 12. *Let k and a be positive integers with $a < k$, let $p = 12\dots a\tau(a+1) \in \mathcal{S}_{k+1}$, where τ is any permutation of the elements $\{a+2, a+3, \dots, k+1\}$, and let $A_{k,a}(x)$ be the e.g.f. for the number of permutations that avoid p. Let*

$$F_{k,a}(x) = \sum_{i\geq 1} \frac{(-1)^{i+1} x^{ki+1}}{(ki+1)!} \prod_{j=2}^{i} \binom{jk-a}{k-a}.$$

Then

$$A_{k,a}(x) = 1/(1 - x + F_{k,a}(x)).$$

If $k = 2$ and $a = 1$ in the previous theorem, corresponding to the pattern $p = 132$, then from Theorem 12 the function $F_{2,1}(x)$, which is the same for the patterns p, 231, 312 and 213 because of the trivial bijections, can be written as:

$$F_{2,1}(x) = \sum_{i\geq 1} \frac{(-1)^{i+1} x^{ki+1}}{i!(k!)^i(ki+1)} = x - \int_0^x e^{-t^2/2}\, dt.$$

That is

$$A_{2,1} = \frac{1}{1 - \int_0^x e^{-t^2/2}\, dt},$$

which is a special case of Theorem 4.1 in [ElizNoy].

4. The Shuffle Patterns

We recall that according to Definition 6, a shuffle pattern is a pattern of the form $\sigma_0 - a_1 - \sigma_1 - a_2 - \dots - \sigma_{k-1} - a_k - \sigma_k$, where $\{\sigma_0, \sigma_1, \dots, \sigma_k\}$ is a set of GPs with no dashes, $a_1a_2\dots a_k$ is a permutation of k letters, for any i and j the letter a_i is greater than any letter of σ_j and for any $i \neq j$ each letter of σ_i is incomparable with any letter of σ_j.

Let us consider a shuffle pattern that in fact is an ordinary generalized pattern. This pattern is $p = \sigma - k$, where σ is an arbitrary pattern with no dashes that is built on elements $1, 2, \ldots, k-1$. So the last element of p is greater than any other element.

Theorem 13. *Let $p = \sigma - k$ and let $A(x)$ (resp. $B(x)$) be the e.g.f. for the number of permutations that avoid σ (resp. p). Then $B(x) = e^{F(x,A(y))}$, where*

$$F(x, A(y)) = \int_0^x A(y)\ dy.$$

Proof. Suppose that $\pi \in \mathcal{S}_{n+1}$ and that π avoids p. Suppose the letter $(n+1)$ is in the i-th position and $\pi = \pi_1(n+1)\pi_2$, where π_1 and π_2 might be empty.

Since π is p-avoiding, π_1 must be σ-avoiding, because otherwise an occurrence of σ in π_1 together with the letter $(n+1)$ gives an occurrence of p in π. But if π_1 is σ-avoiding then there is no interaction between π_1 and π_2, that is, if π_2 is p-avoiding and π_1 is σ-avoiding then π is p-avoiding. To see this it is enough to see that if an occurrence of σ in π contains the letter $(n+1)$, then this occurrence of σ can not lead to an occurrence of $p = \sigma - k$ containing the letter $(n+1)$.

From the above, considering all possible positions of $(n+1)$, we get the recurrence relation

$$B_{n+1} = \sum_i \binom{n}{i} A_i B_{n-i},$$

where B_j (resp. A_j) is the number of j-permutations that avoid p (resp. σ), because we can choose the elements of π_1 in $\binom{n}{i}$ ways.

Multiplying both sides of the equality by $x^n/n!$ we get

$$\frac{B_{n+1}}{n!}x^n = \sum_i \frac{A_i}{i!}x^i \frac{B_{n-i}}{(n-i)!}x^{n-i}.$$

Taking the sum over all natural numbers n leads us to

$$B'(x) = A(x)B(x)$$

where the derivative of B is with respect to x. Since $B(0) = 1$, the solution of the differential equation is $B(x) = e^{F(x,A(y))}$. $\square$

Example 14. Let $p = 1 - 2$. Here $\sigma = 1$, so $A(x) = 1$ since $A_n = 0$ for all $n \geq 1$ and $A_0 = 1$. So

$$B(x) = e^{F(x,1)} = e^x.$$

This corresponds to the fact that for each $n \geq 1$ there is exactly one permutation that avoids the pattern p, namely $\pi = n(n-1)\ldots 1$.

Example 15. Suppose $p = 12 - 3$. Here $\sigma = 12$, so $A(x) = e^x$, since there is exactly one permutation that avoids the pattern σ. So

$$B(x) = \sum_{n\geq 0} \frac{B_n}{n!}x^n = e^{F(x,e^y)} = e^{e^x-1}.$$

According to [Claes, Proposition 2], for all $n \geq 1$, B_n is the n-th Bell number and the e.g.f. for the Bell numbers is e^{e^x-1}.

The table below gives the initial values of B_n for several patterns $p = \sigma - k$. These numbers were obtained by expanding the corresponding $B(x)$. The functions $A(x)$ are taken from the previous section.

pattern	initial values for B_n
132-4	1, 2, 6, 23, 107, 585, 3671, 25986, 204738
123-4	1, 2, 6, 23, 108, 598, 3815, 27532, 221708
1234-5	1, 2, 6, 24, 119, 705, 4853, 38142, 336291
12345-6	1, 2, 6, 24, 120, 719, 5022, 40064, 359400

Theorem 16. *Let p be the shuffle pattern $\sigma - k - \tau$. So k is the greatest letter of the pattern, and each letter of σ is incomparable with any letter of τ. Let $A(x)$, $B(x)$ and $C(x)$ be the e.g.f. for the number of permutations that avoid σ, τ and p respectively. Then $C(x)$ is the solution of the differential equation*

$$C'(x) = (A(x) + B(x))C(x) - A(x)B(x),$$

with $C(0) = 1$.

Proof. As before, we consider the symmetric group $\mathcal{S}_{n+1}$ and a permutation $\pi \in \mathcal{S}_{n+1}$ that avoids p. Suppose the letter $(n+1)$ is in the i-th position and $\pi = \pi_1(n+1)\pi_2$, where π_1 and π_2 might be empty.

There are exactly four mutually exclusive possibilities:

1) π_1 does not avoid σ, π_2 does not avoid τ.
2) π_1 avoids σ, π_2 does not avoid τ;
3) π_1 does not avoid σ, π_2 avoids τ;
4) π_1 avoids σ, π_2 avoids τ;

Obviously, the situation 1) is impossible, since an occurrence of σ in π_1 with $(n+1)$ and with an occurrence of τ in π_2 gives us an occurrence of p in π. On the other hand, if p occurs in π

then it is easy to see that the letter $(n+1)$ cannot be one of the letters in the occurrences of σ or τ, so all p-avoiding permutations are described by the possibilities 2)–4). We count these permutations in the following way.

In $\binom{n}{i}$ ways we choose first i elements from the letters $1, 2 \dots n$, that is, the elements of π_1. Let A_i, B_i and C_i be the number of i-permutations that avoid σ, τ and p respectively.

If π_1 is σ-avoiding, we let π_2 be any p-avoiding permutation of the remaining $(n-i+1)$ letters. This accounts for all "good" permutations from the possibilities 2) and 4). There are $\binom{n}{i}A_iC_{n-i}$ such permutations.

If π_2 is τ-avoiding, we let π_1 be any p-avoiding permutation of chosen i letters. This covers all "good" permutations from 3) and 4). There are $\binom{n}{i}B_iC_{n-i}$ such permutations.

But we have counted p-avoiding permutations that correspond to 4) twice, so we must subtract $\binom{n}{i}A_iB_{n-i}$, which is the number of such permutations.

So we have

$$C_{n+1} = \sum_i \binom{n}{i}(A_iC_{n-i} + B_iC_{n-i} - A_iB_{n-i}).$$

Multiplying both sides of the equality by $x^n/n!$ we get

$$\frac{C_{n+1}}{n!}x^n = \sum_i \left(\frac{A_i + B_i}{i!}x^i \frac{C_{n-i}}{(n-i)!}x^{n-i} - \frac{A_i}{i!}x^i \frac{B_{n-i}}{(n-i)!}x^{n-i} \right),$$

so

$$C'(x) = (A(x) + B(x))C(x) - A(x)B(x).$$

□

Example 17. Let $p = 1' - 2 - 1''$. That is, $\sigma = 1$ and $\tau = 1$. So $A(x) = B(x) = 1$ and we need to solve the equation

$$C'(x) = 2C(x) - 1$$

with $C(0) = 1$. The solution of this equation is $C(x) = \frac{1}{2}(e^{2x} + 1)$, so for all $n \geq 1$ we have $C_n = 2^{n-1}$, as in Example 1.

In the table below we record the initial values of C_n for several patterns $p = \sigma - k - \tau$.

σ	τ	**initial values for C_n**
1	12	1, 2, 6, 21, 82, 354, 1671, 8536, 46814
1	132	1, 2, 6, 24, 116, 652, 4178, 30070, 240164
1	123	1, 2, 6, 24, 116, 657, 4260, 31144, 253400
1	1234	1, 2, 6, 24, 120, 715, 4946, 38963, 344350
12	12	1, 2, 6, 24, 114, 608, 3554, 22480, 152546
12	132	1, 2, 6, 24, 120, 710, 4800, 36298, 302780
12	123	1, 2, 6, 24, 120, 710, 4815, 36650, 308778
12	1234	1, 2, 6, 24, 120, 720, 5025, 39926, 355538
123	123	1, 2, 6, 24, 120, 720, 5020, 39790, 352470
123	132	1, 2, 6, 24, 120, 720, 5020, 39755, 351518
132	132	1, 2, 6, 24, 120, 720, 5020, 39720, 350496

Remark 18. The pattern $p = \sigma - k$ from Theorem 13 is a particular case of the pattern $p = \sigma - k - \tau$ from Theorem 16 when τ is the empty word. The e.g.f. for the number of permutations that avoid the empty word is zero, because no permutation avoids the empty word. So if τ is empty, we can use Theorem 16 to get Theorem 13. Indeed, $B(x) = 0$, and after renaming C with B we get in Theorem 16 exactly the same differential equation as we have in Theorem 13.

We now give two corollaries to Theorem 16.

Corollary 19. *Suppose we have the shuffle pattern $p = \sigma - k - \tau$. We consider the pattern $\varphi(p) = \varphi_1(\sigma) - k - \varphi_2(\tau)$, where φ_1 and φ_2 are any trivial bijections. Then $p \equiv \varphi(p)$.*

Proof. We just observe that if $A(x)$ (resp. $B(x)$) is the e.g.f. for the number of permutations that avoid σ (resp. τ) then $A(x)$ (resp. $B(x)$) is the e.g.f. for the number of permutations that avoid $\varphi_1(\sigma)$ (resp. $\varphi_2(\tau)$). □

Corollary 20. *We have $\sigma - k - \tau \equiv \tau - k - \sigma$.*

Proof. This follows directly from the differential equation of Theorem 16 ($A(x)$ and $B(x)$ are symmetric in that equation), but we can also obtain this as a corollary to Corollary 19. By Corollary 19, the pattern $\sigma - k - \tau$ is equivalent to the pattern $\sigma - k - R(\tau)$. Reversing the pattern $\sigma - k - R(\tau)$, we obtain the pattern

$$R(\sigma - k - R(\tau)) = R(R(\tau)) - k - R(\sigma) = \tau - k - R(\sigma),$$

which thus is equivalent to $\sigma - k - \tau$. Finally, we use Corollary 19 one more time to get

$$\tau - k - R(\sigma) \equiv \tau - k - R(R(\sigma)) = \tau - k - \sigma.$$

□

5. The Multi-Patterns

We recall that according to Definition 5, a multi-pattern is a pattern $p = \sigma_1 - \sigma_2 - \cdots - \sigma_k$, where $\{\sigma_0, \sigma_1, \ldots, \sigma_k\}$ is a set of GPs with no dashes and each letter of σ_i is incomparable with any letter of σ_j whenever $i \neq j$.

We first discuss patterns of the type $p = \sigma - \tau$ which are a particular case of the multi-patterns to be treated in this section.

If σ or τ is the empty word then we are dealing with ordinary GPs with no dashes, some of which were investigated in [ElizNoy] and Section 3. The analysis of the case when σ or τ is equal to 1 can also be reduced to the analysis of ordinary GPs. For example, suppose that $\sigma = 1$, that is, $p = 1 - \tau$, and we want to count the number of permutations in $\mathcal{S}_n$ that avoid p. We can choose the leftmost letter of a permutation avoiding p in n ways, then the remainder of the permutation must avoid τ, so we multiply n by the number of permutations in $\mathcal{S}_{n-1}$ that avoid τ. For instance, if $p = 1 - 1'2'$ then the number of permutations in $\mathcal{S}_n$ avoiding p is exactly n.

Theorem 21. *Let $p = \sigma - \tau$ and $q = \varphi_1(\sigma) - \varphi_2(\tau)$, where φ_1 and φ_2 are any of the trivial bijections. Then p and q are equivalent.*

Proof. The theorem is equivalent to the following statement:

Let $p = \sigma - \tau$ and $q = \sigma - \varphi(\tau)$, where φ is a trivial bijection. Then p and q are equivalent.

It is obvious that the statement follows from Theorem 21. Conversely, suppose we have $p = \sigma - \tau$. We observe that any two trivial bijections commute, that is for any trivial bijection ψ, we have $\psi(R(x)) = R(\psi(x))$. This observation, the statement and the fact that $x \equiv R(x)$ give

$$p = \sigma - \tau \equiv \sigma - \varphi_2(\tau) \equiv R(\varphi_2(\tau)) - R(\sigma) \equiv R(\varphi_2(\tau)) - \varphi_1(R(\sigma)) \equiv$$

$$R(\varphi_2(\tau)) - R(\varphi_1(\sigma)) \equiv \varphi_1(\sigma) - \varphi_2(\tau).$$

So to prove the theorem we now prove the statement.

Let $p = \sigma - \tau$ and $q = \sigma - \varphi(\tau)$, where φ is a trivial bijection. Let A_n (resp. B_n) be the number of n-permutations that avoid p (resp. q). We are going to prove that $A_n = B_n$.

Suppose π avoids p and $\pi = \pi_1\sigma'\pi_2$, where $\pi_1\sigma'$ has exactly one occurrence of the pattern σ, namely σ'. Then π_2 must avoid τ, $\varphi(\pi_2)$ must avoid $\varphi(\tau)$ and $\pi_\varphi = \pi_1\sigma'\varphi(\pi_2)$ avoids q. The converse is also true, that is, if π_φ has no occurrences of q then π has no occurrences of p. If π has no occurrences of σ then π has no occurrences of p as well as no occurrences of q. Since any permutation either avoids σ or can be factored as above, we have a bijection between the class of permutations that avoid p and the class of permutations that avoid q. Thus $A_n = B_n$. □

We get the following corollary to Theorem 21:

Corollary 22. *The pattern $\sigma - \tau$ is equivalent to the pattern $\tau - \sigma$.*

Proof. We proceed as in the proof of Corollary 20. From Theorem 21 we have:

$$\sigma - \tau \equiv \sigma - R(\tau) \equiv R(R(\tau)) - R(\sigma) \equiv \tau - R(R(\sigma)) \equiv \tau - \sigma.$$

□

We observe that the presence of the dash in the patterns in Theorem 21 is essential. That is, generally speaking, the pattern $\sigma\tau$ is not equivalent to the pattern $\varphi_1(\sigma)\varphi_2(\tau)$ for any trivial bijections φ_1 and φ_2. For example, there are 66 permutations in $\mathcal{S}_5$ that avoid the pattern $122'1'$ but only 61 that avoid $121'2'$. In Section 6 we investigate the pattern $122'1'$.

Theorem 23 and Corollary 24 generalise Theorem 21 and Corollary 22:

Theorem 23. *Suppose we have multi-patterns $p = \sigma_1 - \sigma_2 - \cdots - \sigma_k$ and $q = \tau_1 - \tau_2 - \cdots - \tau_k$, where $\tau_1\tau_2\ldots\tau_k$ is a permutation of $\sigma_1\sigma_2\ldots\sigma_k$. Then p and q are equivalent.*

Proof. We proceed by induction on k. If $k = 2$ then the statement is true by Corollary 22. Suppose the statement is true for all $k' < k$. Suppose p has exactly k blocks. If a permutation π avoiding p has no occurrences of σ_1 then it obviously avoids both p and q. Otherwise we factor π as $\pi = \pi_1\sigma_1'\pi_2$ where $\pi_1{\sigma_1}'$ has exactly one occurrence of the pattern σ_1, namely ${\sigma_1}'$. Then π_2 must avoid $\sigma_2 - \cdots - \sigma_k$. Moreover it is irrelevant from which letters $\pi_1{\sigma_1}'$ is built and therefore we can apply the inductive hypothesis. We can rearrange $\sigma_2' \ldots \sigma_k'$ of $\sigma_2 \ldots \sigma_k$ in such a way that the blocks in $\tau_1\tau_2\ldots\tau_k$ corresponding to $\sigma_2, \ldots, \sigma_k$ are arranged in the same order as the τ's. Now we consider separately two cases: $\tau_k \neq \sigma_1$ and $\tau_k = \sigma_1$. In the first case we use the following equivalences:

$$p = \sigma_1 - \sigma_2 - \cdots - \sigma_k \equiv \sigma_1 - {\sigma_2}' - \cdots - \sigma_k' \equiv R(\sigma_k') - \cdots - R(\sigma_2') - R(\sigma_1).$$

For the pattern $R(\sigma_k') - \cdots - R(\sigma_2') - R(\sigma_1)$ we use the factorisation of a permutation π avoiding this pattern, where the role of σ_1 is played by $R(\sigma_k')$. So by the inductive hypothesis we put the pattern $R(\sigma_1)$ in the right place somewhere to the left of $R(\sigma_2')$ and apply R to get that $p \equiv q$.

In the second case we have:

$$p \equiv R(\sigma_k') - \cdots - R(\sigma_2') - R(\sigma_1) \equiv R(\sigma_k') - \cdots - R(\sigma_1) - R(\sigma_2') \equiv$$

$$\sigma_2' - \sigma_1 - \cdots - \sigma_k' \equiv \sigma_2' - \cdots - \sigma_k' - \sigma_1 = q$$

The first equivalence here is taken from the considerations above; the second one uses the inductive hypothesis; then we use the fact that $R(R(x)) = x$ and apply the inductive hypothesis again. $\square$

Corollary 24. *Suppose we have multi-patterns $p = \sigma_1 - \sigma_2 - \cdots - \sigma_k$ and $q = \varphi_1(\sigma_1) - \varphi_2(\sigma_2) - \cdots - \varphi_k(\sigma_k)$, where each φ_i is an arbitrary trivial bijection. Then p and q are equivalent.*

Proof. We use induction on k, Theorem 23 and the factorisation of permutations, which is discussed in the proof of Theorem 23. If $k = 2$ then the statement is true by Theorem 21. Suppose the statement is true for all $k' < k$. Then

$$p = \sigma_1 - \sigma_2 - \cdots - \sigma_k \equiv \sigma_1 - \varphi_2(\sigma_2) - \cdots - \varphi_k(\sigma_k) \equiv$$

$$\varphi_2(\sigma_2) - \sigma_1 - \cdots - \varphi_k(\sigma_k) \equiv \varphi_2(\sigma_2) - \varphi_1(\sigma_1) - \cdots - \varphi_k(\sigma_k) \equiv$$

$$\varphi_1(\sigma_1) - \varphi_2(\sigma_2) - \cdots - \varphi_k(\sigma_k) = q,$$

where first we apply the inductive hypothesis then Theorem 23 then the inductive hypothesis and finally Theorem 23 again. $\square$

Theorem 25. *Suppose $p = \sigma - p'$, where p' is an arbitrary POGP, and the letters of σ are incomparable to the letters of p'. Let $C(x)$ (resp. $A(x)$, $B(x)$) be the e.g.f. for the number of permutations that avoid p (resp. σ, p'). Moreover let $A^\star(x)$ be the e.g.f. for the number of permutations that quasi-avoid σ. Then*

$$C(x) = A(x) + B(x)A^\star(x).$$

Proof. Let A_n, B_n, C_n be the number of n-permutations that avoid the patterns σ, p' and p respectively. Also $A_n^\star$ is the number of n-permutations that quasi-avoid σ. If a permutation π avoids σ then it avoids p. Otherwise we find the leftmost occurrence of σ in π. We assume that this occurrence consists of the $|\sigma|$ rightmost letters among the i leftmost letters of π. So the subword of π beginning at the $(i+1)$st letter must avoid p'. From this we conclude

$$C_n = A_n + \sum_{i=|\sigma|}^{n} \binom{n}{i} A_i^\star B_{n-i}.$$

We observe that we can change the lower bound in the sum above to 0, because $A_i^\star = 0$ for $i = 0, 1, \ldots, |\sigma| - 1$. Multiplying both sides by $x^n/n!$ and taking the sum over all n we get the desired result. □

Corollary 26. *Suppose $p = \sigma_1 - \sigma_2 - \cdots - \sigma_k$ is a multi-pattern where $|\sigma_i| = 2$ for all i, so each σ_i is equal to either 12 or 21. If $B(x)$ is the e.g.f. for the number of permutations that avoid p then*

$$B(x) = \frac{1 - (1 + (x-1)e^x)^k}{1 - x}.$$

Proof. We use Theorem 25, induction on k and the fact that $A(x) = e^x$ and $A^\star(x) = 1 + (x-1)e^x$. □

The following corollary to Corollary 26 can be proved combinatorially.

Theorem 27. *There are $(n-2)2^{n-1} + 2$ permutations in $\mathcal{S}_n$ that avoid the pattern $p = 12 - 1'2'$ or, according to Theorem 21, the pattern $p = 12 - 2'1'$.*

One more corollary to Theorem 25 is the following theorem that is the basis for calculating the number of permutations that avoid a multi-pattern, and therefore is the main result for multi-patterns in this paper.

Theorem 28. *Let $p = \sigma_1 - \sigma_2 - \cdots - \sigma_k$ be a multi-pattern and let $A_i(x)$ be the number of permutations that avoid σ_i. Then the e.g.f. $B(x)$ for the number of permutations that avoid p is*

$$B(x) = \sum_{i=1}^{k} A_i(x) \prod_{j=1}^{i-1} ((x-1)A_j(x) + 1).$$

Proof. We use Theorem 25 and prove by induction on k that

$$B(x) = \sum_{i=1}^{k} A_i(x) \prod_{j=1}^{i-1} A_j^\star(x).$$

Then we use Proposition 4 to get the desired result. □

Remark 29. One can consider the function $B(x)$ from Theorem 28 as a function in k variables $B(x) = B(A_1(x), A_2(x), \ldots, A_k(x))$. Then, by Theorem 23, this function is symmetric in the variables $A_1(x), A_2(x), \ldots, A_k(x)$. That means that we can rename the variables, which may simplify the calculation of $B(x)$.

6. Patterns of the Form $\sigma\tau$

Theorem 30. *Let $B(x)$ be the e.g.f. for the number of permutations that avoid the pattern $p = 122'1'$. Then*

$$B(x) = \frac{1}{2} + \frac{1}{4}\tan x(1 + e^{2x} + 2e^x \sin x) + \frac{1}{2}e^x \cos x.$$

Proof. Let B_n be the number of n-permutations that avoid p and A_n be the number of n-permutations that avoid p and begin with the pattern 12. Let also $A(x)$ be the e.g.f. for the numbers A_n. We set $B_0 = A_0 = A_1 = 1$. Suppose π is a $(n+1)$-permutation that avoids p. There are three mutually exclusive possibilities:

1) $\pi = (n+1)\pi_2$;
2) $\pi = \pi_1(n+1)$;
3) $\pi = \pi_1(n+1)\pi_2$ and $\pi_1, \pi_2 \neq \varepsilon$.

Obviously, in 1) and 2) the letter $(n+1)$ does not affect the rest of the permutation π, and therefore in each of these cases we have B_n permutations that avoid p. In 3), it is easy to see that if π_1 has more than one letter then π_1 must end with a 21 pattern whereas if π_2 has more than one letter then π_2 must begin with a 12 pattern. The key observation is that the number of n-permutations that avoid p and end with a 21 pattern is the same as the number of n-permutations that avoid p and begin with a 12 pattern. To see this it is enough to apply the reverse function to any n-permutation π that begins with 12-pattern and avoids p and observe that $R(p) = p$, that is, $R(\pi)$ avoids p and ends with a 21 pattern. Obviously this is a bijection. So if $|\pi_1| = i$ then we can choose the letters of π_1 in $\binom{n}{i}$ ways and then choose a permutation π_1 in A_i ways and a permutation π_2 in A_{n-i} ways, since the letters of π_1 and π_2 do not affect each other. From all this we get

$$B_{n+1} = 2B_n + \sum_{i=1}^{n-1} \binom{n}{i} A_i A_{n-i} = 2B_n + \sum_{i=0}^{n} \binom{n}{i} A_i A_{n-i} - 2A_n.$$

We multiply both sides of the last equality by $x^n/n!$ to get

$$B_{n+1}\frac{x^n}{n!} = 2B_n\frac{x^n}{n!} + \sum_{i=0}^{n} \frac{A_i}{i!}x^i \frac{A_{n-i}}{(n-i)!}x^{n-i} - 2A_n\frac{x^n}{n!}.$$

Summing both sides over all natural numbers n we get:

$$B'(x) = 2B(x) + A^2(x) - 2A(x). \tag{2}$$

To solve this differential equation with the initial condition $B(0) = 1$, we need to determine $A(x)$. One can observe that if a permutation π avoids p and begins with the pattern 12 then π has the structure $\pi = a_1b_1a_2b_2a_3b_3\cdots$, where $a_i < b_i$ for all i. Moreover, if $b_1 < a_2$ then we must have $a_1 < b_1 < a_2 < b_2 < a_3 < \cdots$ since otherwise we obviously have an occurrence of the pattern p. A first approximation is that $A_n = \binom{n}{2}A_{n-2}$, because we can choose a_1b_1 in π in $\binom{n}{2}$ ways and then pick an arbitrary $(n-2)$-permutation that avoids p and begins with the pattern 12, to be $a_2b_2a_3b_3\ldots$, in A_{n-2} ways. But it is possible that $b_1 < a_2$ in which case $b_1a_2b_2a_3$ can be an occurrence of p in π, and it is an occurrence of p unless $a_2 < b_2 < a_3 < \cdots$. So in order to avoid this we must subtract the number of permutations of the form $abcd\pi'$, where $a < b < c < d$ and π' is any $(n-4)$-permutation that avoids p, from the first approximation of A_n. Thus the second approximation is that $A_n = \binom{n}{2}A_{n-2} - \binom{n}{4}A_{n-4}$. We observe that in the second approximation we do not count the increasing permutation $123\ldots n$. Moreover, among the permutations counted by $\binom{n}{4}A_{n-4}$, there are the permutations that begin with 6 increasing letters. Except for the increasing permutation, such permutations are not counted by $\binom{n}{2}A_{n-2}$. We must therefore add the number of such permutations. So the third approximation is that $A_n = \binom{n}{2}A_{n-2} - \binom{n}{4}A_{n-4} + \binom{n}{6}A_{n-6}$ and so on. That is,

$$A_n = \binom{n}{2}A_{n-2} - \binom{n}{4}A_{n-4} + \binom{n}{6}A_{n-6} - \binom{n}{8}A_{n-8} + \cdots = \sum_{i\geq 1}(-1)^{i+1}\binom{n}{2i}A_{n-2i}. \tag{3}$$

We observe that if $n = 4k$ or $n = 4k + 1$ then we do not count the increasing permutation in our sum. This, together with Equation 3, gives us

$$\sum_{i \geq 0} (-1)^i \binom{n}{2i} A_{n-2i} = \begin{cases} 1, & \text{if } n = 4k \text{ or } n = 4k+1, \\ 0, & \text{if } n = 4k+2 \text{ or } n = 4k+3. \end{cases}$$

Multiplying both sides of the equality with $x^n/n!$ and summing over all natural numbers n we get

$$(A_0 + A_1 x + \frac{A_2}{2!}x^2 + \cdots)(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots) = \sum_{k=0}^{\infty} \left(\frac{x^{4k}}{(4k)!} + \frac{x^{4k+1}}{(4k+1)!} \right).$$

The left hand side of this equality is equal to $A(x)\cos x$. Let $F(x)$ be the function in the right hand side of the equality. Then it is easy to see that $F(x)$ is the solution to the differential equation $F^{(4)}(x) = F(x)$ with the initial conditions $F(0) = F'(0) = 1$, $F^{(2)}(0) = F^{(3)}(0) = 0$. So $F(x) = \frac{1}{2}(\cos x + \sin x + e^x)$ and

$$A(x) = \frac{1}{2}\left(1 + \tan x + \frac{e^x}{\cos x}\right).$$

Now we solve the differential equation (2) and get

$$B(x) = \frac{1}{2} + \frac{1}{4}\tan x(1 + e^{2x} + 2e^x \sin x) + \frac{1}{2}e^x \cos x.$$

□

Remark 31. The series expansion of $B(x)$ in Theorem 30 begins with

$$B(x) = 1 + x + x^2 + x^3 + \frac{3}{4}x^4 + \frac{11}{20}x^5 + \frac{7}{20}x^6 + \frac{7}{30}x^7 + \frac{103}{720}x^8 + \cdots.$$

That is, the initial values for B_n are 1, 2, 6, 18, 66, 252, 1176, 5768.

7. The Distribution of Non-Overlapping GPs

A descent in a permutation $\pi = a_1 a_2 \ldots a_n$ is an i such that $a_i > a_{i+1}$. The number of descents in a permutation π is denoted $\operatorname{des}\pi$ (and is equivalent to the generalized pattern 21). Any statistic with the same distribution as des is said to be *Eulerian.* The *Eulerian numbers* $A(n,k)$ count permutations in the symmetric group $\mathcal{S}_n$ with k descents and they are the coefficients of the *Eulerian polynomials* $A_n(t)$ defined by $A_n(t) = \sum_{\pi \in \mathcal{S}_n} t^{1+\operatorname{des}\pi}$. The Eulerian polynomials satisfy the identity

$$\sum_{k \geq 0} k^n t^k = \frac{A_n(t)}{(1-t)^{n+1}}.$$

Two descents i and j *overlap* if $j = i + 1$. We define a new statistic, namely the *maximum number of non-overlapping descents*, or MND, in a permutation. For instance, MND(321) $= 1$ whereas MND(41532) $= 2$. One can find the distribution of this new statistic by using Corollary 26. This distribution is given in Example 33. However, we prove a more general theorem:

Theorem 32. *Let p be a GP with no dashes. Let $A(x)$ be the e.g.f. for the number of permutations that avoid p. Let $D(x,y) = \sum_{\pi} y^{N(\pi)} \frac{x^{|\pi|}}{|\pi|!}$ where $N(\pi)$ is the maximum number of non-overlapping occurrences of p in π. Then*

$$D(x,y) = \frac{A(x)}{1 - y((x-1)A(x) + 1)}.$$

Proof. We fix the natural number k and consider an auxiliary multi-pattern $P_k = p - p - \cdots - p$ with k copies of p. If a permutation avoids P_k then it has at most $k-1$ non-overlapping occurrences of p. From Theorem 28, the e.g.f. $B_k(x)$ for the number of permutations avoiding P_k is equal to $\sum_{i=1}^{k} A(x) \prod_{j=1}^{i-1}((x-1)A(x)+1)$. If we subtract $B_k(x)$ from the e.g.f. $B_{k+1}(x) = \sum_{i=1}^{k+1} A(x) \prod_{j=1}^{i-1}((x-1)A(x)+1)$ for the number of permutations avoiding P_{k+1}, which is obtained by applying Theorem 28 to the pattern P_{k+1}, then we get the e.g.f. $D_k(x)$ for the number of permutations that have exactly k non-overlapping occurrences of the pattern p. So

$$D_k(x) = \sum_n D_{n,k} \frac{x^n}{n!} = B_{k+1}(x) - B_k(x) = A(x)((x-1)A(x)+1)^k.$$

Now

$$D(x,y) = \sum_{n,k \geq 0} D_{n,k} y^k \frac{x^n}{n!} = \sum_k D_k(x) y^k = \frac{A(x)}{1 - y((x-1)A(x)+1)}.$$

□

All of the following examples are corollaries to Theorem 32.

Example 33. If we consider descents then $A(x) = e^x$, hence the distribution of MND is given by the formula:

$$D(x,y) = \frac{e^x}{1 - y(1 + (x-1)e^x)}.$$

Example 34. Theorems 11 and 32 give the distribution of the maximum number of non-overlapping occurrences of the increasing subword of length k (the pattern $123 \ldots k$), which is equal to

$$D(x,y) = \frac{1}{(1-x)y + (1-y)F_k(x)},$$

were $F_k(x) = \sum_{i \geq 0} \frac{x^{ki}}{(ki)!} - \sum_{i \geq 0} \frac{x^{ki+1}}{(ki+1)!}$.

Example 35. If we consider the maximum number of non-overlapping occurrences of the pattern 132 then the distribution of these numbers is given by the formula

$$D(x,y) = \frac{1}{1 - yx + (y-1) \displaystyle\int_0^x e^{-t^2/2}\, dt}.$$

Example 36. The distribution of the maximum number of non-overlapping occurrences of the pattern from Theorem 12 is given by the formula:

$$D(x,y) = \frac{1}{1 - x + (1-y)F_{k,a}(x)},$$

where $F_{k,a}(x) = \sum_{i \geq 1} \frac{(-1)^{i+1} x^{ki+1}}{(ki+1)!} \prod_{j=2}^{i} \binom{jk-a}{k-a}$.

References

[BabStein] E. Babson, E. Steingrímsson: Generalised permutation patterns and a classification of the mahonian statistics, Séminaire Lotharingien de Combinatoire, B44b:18pp, 2000.

[Bon1] M. Bóna: Exact enumeration of 1342-avoiding permutations; A close link with labeled trees and planar maps, Journal of Combinatorial Theory, Series A, **80** (1997) 257-272.

[Bon2] M. Bóna: Permutations avoiding certain patterns; The case of length 4 and generalisations, Discrete Mathematics **175** (1997), 55-67.

[Bon3] M. Bóna: Permutations with one or two 132-subsequences, Discrete Mathematics **181** (1998), 267-274.

[Claes] A. Claesson: Generalised Pattern Avoidance, European Journal of Combinatorics **22** (2001), 961-971.

[ClaesMans] A. Claesson and T. Mansour: Permutations avoiding a pair of generalized patterns of length three with exactly one dash, preprint CO/0107044.

[ElizNoy] S. Elizalde and M. Noy: Enumeration of Subwords in Permutations, Proceedings of FPSAC 2001.

[GoulJack] I. P. Goulden and D. M. Jackson, *Combinatorial Enumeration*, A Wiley-Interscience Series in Discrete Mathematics, John Wiley & Sons Inc., New York, (1983).

[Kit] S. Kitaev: Multi-avoidance of generalised patterns, preprint. `http://www.math.chalmers.se/~kitaev/papers.html`

[Knuth] D. E. Knuth: *The Art of Computer Programming*, 2nd ed. Addison Wesley, Reading, MA, (1973).

[Mans] T. Mansour: Restricted 1-3-2 permutations and generalized patterns, preprint CO/0110039.

[SloPlo] N. J. A. Sloane and S. Plouffe: *The Encyclopedia of Integer Sequences*, Academic Press, (1995). http://www.research.att.com/∼njas/sequences/.

[Stan] R. Stanley: *Enumerative Combinatorics*, Volume **1**, Cambridge University Press, (1997).

[SchSim] R. Simion, F. Schmidt: Restricted permutations, European J. Combin. **6** (1985), no. 4, 383–406.

Matematik, Chalmers tekniska högskola och Göteborgs universitet, S-412 96 Göteborg, Sweden

E-mail address: `kitaev@math.chalmers.se`

GRÖBNER GEOMETRY OF SCHUBERT POLYNOMIALS (EXTENDED ABSTRACT)

ALLEN KNUTSON AND EZRA MILLER

ABSTRACT. Let $w \in S_n$ be a permutation. We provide a geometric context in which both (i) polynomial representatives for the Schubert classes $[X_w]$ in the cohomology ring $H^*(\mathcal{F}\ell_n)$ of the flag manifold are uniquely singled out, with no choices other than a Borel subgroup of GL_n; and (ii) it is geometrically obvious that these polynomials have nonnegative coefficients. These polynomials turn out to be the Schubert polynomials $\mathfrak{S}_w(x_1, \ldots, x_n)$ of Lascoux and Schützenberger [LS82a].

Our investigations lead us to replace topology on the flag manifold with multigraded commutative algebra, by generalizing the notion of degree for subschemes of projective space. Identifying the Schubert polynomials in this context then sheds light on the algebra and geometry of determinantal ideals specified by rank conditions as in [Ful92], and especially on the combinatorics of initial ideals for certain natural term orders. These initial ideals are the Stanley–Reisner ideals for simplicial complexes whose facets are in natural bijection with the rc-graphs of Fomin and Kirillov [FK96, BB93]. We give an inductive procedure on the weak Bruhat order for listing rc-graphs. Our further analysis of rc-graphs is based on the combinatorics of words and their subwords in general Coxeter groups, which give rise to shellable simplicial balls or spheres generalizing the initial ideals constructed from rc-graphs.

RÉSUMÉ. Soit $w \in S_n$ une permutation. Nous fournissons un contexte géométrique qui (i) nous permet d'exhiber de manière unique, et sans autre choix que celui d'un sous groupe de Borel de GL_n, des représentants polynômiaux des classes de Schubert $[X_w]$ dans l'anneau de cohomologie $H^*(\mathcal{F}\ell_n)$ de la variété des drapeaux; (ii) rend géométriquement évident le fait que les coefficients de ces polynômes sont tous non négatifs. Ces polynômes se trouvent être les polynômes de Schubert $\mathfrak{S}_w(x_1, \ldots, x_n)$ de Lascoux et Schützenberger [LS82a].

Nos investigations nous ont menés à remplacer la topologie de la variété des drapeaux par de l'algèbre commutative multigraduée, en généralisant la notion de degré aux sous-chémas de l'espace projectif. L'apparition des polynômes de Schubert dans ce contexte éclaire l'algèbre et la géométrie des idéaux déterminantaux spécifiés par des conditions de rang [Ful92], et spécialement la combinatoire des idéaux initiaux pour certains ordres de termes apparaissant naturellement. Ces idéaux sont les idéaux de Stanley–Reisner de certains complexes simpliciaux, dont les facettes sont en bijection naturelle avec l'ensemble des rc-graphes de Fomin et Kirillov [FK96, BB93]. Nous donnons une procédure inductive sur l'ordre de Bruhat faible pour lister les rc-graphes. Notre analyse ultérieure des rc-graphes se base sur la combinatoire des mots et sous-mots dans les groupes de Coxeter généraux, qui donne lieu à des boules ou sphères effeuillable, généralisant les idéaux construits à partir de rc-graphes.

1. GRÖBNER BASES AND MULTIDEGREES OF DETERMINANTAL IDEALS

1.1. **Matrix Schubert varieties.** Let M_n be the $n \times n$ matrices over a field $\mathbf{k}$, with coordinate ring $\mathbf{k}[\mathbf{z}]$ in indeterminates $\{z_{ij}\}_{i,j=1}^n$. Throughout the paper, q and p will be integers with $1 \leq q, p \leq n$, and Z will stand for an $n \times n$ matrix. Usually Z will be the **generic matrix** of variables (z_{ij}), although occasionally Z will be an element of M_n.

AK was partly supported by the Clay Mathematics Institute, Sloan Foundation, and NSF.
EM was supported by the Sloan Foundation and NSF.

Denote by $Z_{[q,p]}$ the northwest $q \times p$ submatrix of Z. For instance, given a permutation matrix w^T (for $w \in S_n$) with 1's in row i and column $w(i)$, we find that

$$\operatorname{rank}(w^T_{[q,p]}) \quad = \quad \#\{(i,j) \leq (q,p) \mid w(i) = j\}$$

is the number of 1's in the submatrix $w^T_{[q,p]}$.

The following definition was made by Fulton in [Ful92].

Definition 1.1.1. Let $w \in S_n$. The **matrix Schubert variety** $\overline{X}_w \subseteq M_n$ consists of the matrices $Z \in M_n$ such that $\operatorname{rank}(Z_{[q,p]}) \leq \operatorname{rank}(w^T_{[q,p]})$ for all q, p.

Let $B \subset GL_n$ denote the Borel group of *lower* triangular matrices, so that $\mathcal{F}\ell_n = B\backslash GL_n$ is the manifold of flags in $\mathbf{k}^n$. Intersecting $\overline{X}_w$ with $GL_n \subset M_n$ yields the variety $\tilde{X}_w$ of all invertible matrices mapping to the Schubert variety $X_w \subseteq \mathcal{F}\ell_n$. Here, we define X_w to be the closure in $\mathcal{F}\ell_n$ of the orbit Bw^TB^+, where B^+ denotes the upper triangular matrices in GL_n.

Heuristically, in the case $\mathbf{k} = \mathbb{C}$, the matrix Schubert variety determines a B-equivariant cohomology class $[\overline{X}_w]_B \in H^*_B(M_n)$ that maps to the corresponding class $[\tilde{X}_w]_B \in H^*_B(GL_n)$ under the inclusion $GL_n \hookrightarrow M_n$. Letting T denote the maximal torus in B, observe that $H^*_B(M_n) = H^*_T(M_n) = \mathbb{Z}[x_1, \ldots, x_n]$ because B retracts to T and M_n is T-equivariantly contractible. Therefore $[\overline{X}_w]_B \in \mathbb{Z}[\mathbf{x}]$ is well-defined as a polynomial in $\mathbf{x} = x_1, \ldots, x_n$. Since the quotient $GL_n \twoheadrightarrow \mathcal{F}\ell_n$ induces a natural isomorphism $H^*_B(GL_n) \cong H^*(\mathcal{F}\ell_n)$, we find that the polynomial $[\overline{X}_w]_B = [\overline{X}_w]_T$ represents the Schubert class $[X_w]$ on $\mathcal{F}\ell_n$.

The previous paragraph accomplishes the goal of singling out unique polynomial representatives for Schubert classes. It can be made completely precise when $\mathbf{k} = \mathbb{C}$ by arguments derived from [Kaz97] and appearing also in [FR01]. These techniques demonstrate why double Schubert polynomials (applied to the Chern roots of flagged vector bundles) are the characteristic classes for degeneracy loci [Ful92]: the mixing space construction of Borel whose cohomology is the equivariant cohomology of M_n agrees with the classifying space for maps between flagged vector bundles.

Instead of relating equivariant cohomology of M_n to ordinary cohomology of $\mathcal{F}\ell_n$, we employ the notion of multidegree—a commutative algebra approach to torus-equivariant cohomology (cf. [Tot99, Bri98, EG98]) on vector spaces. Multidegrees work over arbitrary fields $\mathbf{k}$, and are in any case essential for showing geometrically why the coefficients of $[\overline{X}_w]_T$ are positive.

1.2. **Schubert polynomials as multidegrees.** For the purpose of dealing with multidegrees in complete generality, let $\mathbf{k}[\mathbf{z}]$ be the polynomial ring in m variables $\mathbf{z} = z_1, \ldots, z_m$, with a grading by $\mathbb{Z}^n$ in which each variable z_i has **exponential weight** $\operatorname{wt}(z_i) = \mathbf{x}^{\mathbf{a}_i}$ for some vector $\mathbf{a}_i \in \mathbb{Z}^n$. In the case of interest for later purposes, $m = n^2$ with $\mathbf{z} = (z_{ij})^n_{i,j=1}$ and $\operatorname{wt}(z_{ij}) = x_i$.

Every finitely generated $\mathbb{Z}^n$-graded module $\Gamma = \bigoplus_{\mathbf{b}\in\mathbb{Z}^n} \Gamma_{\mathbf{b}}$ over $k[\mathbf{z}]$ has a graded free resolution $\mathcal{E}_\bullet : 0 \leftarrow \mathcal{E}_0 \leftarrow \mathcal{E}_1 \leftarrow \cdots \leftarrow \mathcal{E}_m \leftarrow 0$. Suppose $\mathcal{E}_i = \bigoplus_{k=1}^{\beta_i} k[\mathbf{z}](-\mathbf{b}_{ik})$, so that the k^{th} summand of $\mathcal{E}_i$ is generated in $\mathbb{Z}^n$-graded degree $\mathbf{b}_{ik}$.

Definition 1.2.1. The **K-polynomial** of Γ is $\mathcal{K}(\Gamma; \mathbf{x}) = \sum_i (-1)^i \sum_k \mathbf{x}^{\mathbf{b}_{ik}}$.

Geometrically, the K-polynomial of Γ represents the class of the sheaf $\tilde{\Gamma}$ on $\mathbf{k}^m$ in equivariant K-theory for the action of the n-torus whose weight lattice is $\mathbb{Z}^n$. Combinatorially, when the $\mathbb{Z}^n$-grading is **positive**, meaning that the ordinary weights $\mathbf{a}_1, \ldots, \mathbf{a}_n$ lie in a single open half-space in $\mathbb{Z}^n$, the K-polynomial of Γ is the numerator of the $\mathbb{Z}^n$-graded

Hilbert series $H(\Gamma;\mathbf{x})$:

$$H(\Gamma;\mathbf{x}) \;:=\; \sum_{\mathbf{b}\in\mathbb{Z}^n} \dim_{\mathbf{k}}(\Gamma_{\mathbf{b}})\cdot \mathbf{x}^{\mathbf{b}} \;=\; \frac{\mathcal{K}(\Gamma;\mathbf{x})}{\prod_{i=1}^m(1-\mathrm{wt}(z_i))}.$$

Positivity occurs when $\mathbf{z}=(z_{ij})$ and $\mathrm{wt}(z_i)=x_i$, where the denominator is $\prod_{i=1}^n(1-x_i)^n$.

Given any Laurent monomial $\mathbf{x}^{\mathbf{a}}$, the rational function $\prod_{j=1}^n(1-x_j)^{a_j}$ can be expanded as a well-defined (i.e. convergent in the $\mathbf{x}$-adic topology) formal power series $\prod_{j=1}^n(1-a_jx_j+\cdots)$ in $\mathbf{x}$. Doing the same for each monomial in an arbitrary Laurent polynomial $\mathcal{K}(\mathbf{x})$ results in a power series denoted by $\mathcal{K}(\mathbf{1}-\mathbf{x})$.

Definition 1.2.2. The **multidegree** of a $\mathbb{Z}^n$-graded $\mathbf{k}[\mathbf{z}]$-module Γ is the sum $\mathcal{C}(\Gamma;\mathbf{x})$ of the lowest degree terms in $\mathcal{K}(\Gamma;\mathbf{1}-\mathbf{x})$. If $\Gamma=\mathbf{k}[\mathbf{z}]/I$ is the coordinate ring of a subscheme $X\subseteq\mathbf{k}^m$, then write $[X]_{\mathbb{Z}^n}=\mathcal{C}(\Gamma;\mathbf{x})$.

The letters $\mathcal{C}$ and $\mathcal{K}$ stand for 'cohomology' and 'K-theory', and the relation between them ('take lowest degree terms') reflects the Grothendieck–Riemann–Roch transition from K-theory to its associated graded ring. When $\mathbf{k}=\mathbb{C}$ is the complex numbers, the (Laurent) polynomials denoted by $\mathcal{C}$ and $\mathcal{K}$ are honest torus-equivariant cohomology and K-classes on affine space.

The motivating example is the case $X=\overline{X}_w\subseteq M_n$. Recall that the i^{th} **divided difference** operator ∂_i acts on polynomials $f\in\mathbb{Z}[\mathbf{x}]$ by $\partial_i(f)=(f-s_if)/(x_i-x_{i+1})$, where the i^{th} transposition $s_i\in S_n$ switches the variables x_i and x_{i+1} in the argument of f. After calculating explicitly that $[\overline{X}_{w_0}]_{\mathbb{Z}^n}=\prod_{i=1}^n x_i^{n-i}$ for the long permutation $w_0=n\cdots 321\in S_n$, we use a direct geometric argument using multidegrees to show:

Theorem 1.2.3. *If* $\mathrm{length}(ws_i)<\mathrm{length}(w)$ *then* $[\overline{X}_{ws_i}]_{\mathbb{Z}^n}=\partial_i[\overline{X}_w]_{\mathbb{Z}^n}$. *Therefore* $[\overline{X}_w]_{\mathbb{Z}^n}$ *equals the Schubert polynomial* $\mathfrak{S}_w(\mathbf{x})$.

More generally, for the $\mathbb{Z}^{2n}$-grading in which $\mathrm{wt}(z_i)=x_i/y_j$, the proof actually shows that $[\overline{X}_w]_{\mathbb{Z}^{2n}}$ is the double Schubert polynomial $\mathfrak{S}_w(\mathbf{x},\mathbf{y})$. Although all of the results below hold in this 'double' setting, we stick to the $\mathbb{Z}^n$-grading for clarity.

1.3. **Multidegrees as positive sums.** Multidegrees, like ordinary degrees, are additive on unions of schemes with disjoint support and equal dimension. To make a precise statement, let $\mathrm{mult}_X(\Gamma)$ denote the multiplicity of Γ along the variety X.

Theorem 1.3.1. *Fix a* $\mathbb{Z}^n$*-graded module* Γ*, and let* $X_1,\ldots,X_r$ *be the maximal-dimensional irreducible components of the support variety of* Γ*. Then*

$$\mathcal{C}(\Gamma;\mathbf{x}) \;=\; \sum_{\ell=1}^{r}\mathrm{mult}_{X_\ell}(\Gamma)\cdot[X_\ell]_{\mathbb{Z}^n}.$$

Multidegrees are constant in flat families, including the Gröbner degenerations of the next lemma. For positive gradings this is easy, by the constancy of Hilbert series.

Lemma 1.3.2. *Fix a* $\mathbb{Z}^n$*-graded module* Γ*, and let* $\Gamma\cong F/K$ *be an expression of* Γ *as the quotient of a free module* F *with kernel* K*. The multidegree* $\mathcal{C}(\Gamma;\mathbf{x})$ *equals the multidegree of* $F/\mathsf{in}(K)$ *for the initial submodule* $\mathsf{in}(K)$ *of* K *under any term order.*

When $\mathbf{k}^m=M_n$ and $\mathrm{wt}(z_i)=x_i$, the multidegree $[L]_{\mathbb{Z}^n}$ of a coordinate subspace $L=$ zero set of $\langle z_{i_1j_1},\ldots,z_{i_rj_r}\rangle$ is just the monomial $\mathbf{x}^L:=x_{i_1}\cdots x_{i_r}$. Therefore multidegrees can always be expressed as positive sums of monomials.

Corollary 1.3.3. *Suppose $\mathcal{L}$ is the zero scheme of an initial ideal* $\mathsf{in}(I(\overline{X}_w))$ *of the ideal of* $\overline{X}_w$ *for some term order. The equality* $[\overline{X}_w]_{\mathbb{Z}^n} = \sum_L \mathrm{mult}_L(\mathcal{L}) \cdot \mathbf{x}^L$ *writes the multidegree* $\mathfrak{S}_w(\mathbf{x})$ *of* $\overline{X}_w$ *as a sum of monomials with positive coefficients, where the sum is over reduced subspaces of* $\mathcal{L}$ *having maximal dimension.*

1.4. **Gröbner bases, antidiagonals, and Hilbert series.** A positive sum as in Corollary 1.3.3 lends itself to combinatorial analysis only if we understand the initial ideal. Theorem 1.4.2, which determines an explicit initial ideal, will therefore be our central result. Ultimately it connects the geometry of Schubert varieties to the combinatorics of permutations, using the methods of Section 2 (which are themselves based on techniques in the proof of Theorem 1.4.2).

The matrix Schubert variety $\overline{X}_w$ is cut out set-theoretically by determinants in the generic matrix Z. These equations in fact define $\overline{X}_w$ scheme-theoretically; see [Ful92] or Corollary 1.5.1, below.

Definition 1.4.1. Let $w \in S_n$ be a permutation and $r_{qp} = \mathrm{rank}(w^T_{[q,p]})$ for each q, p.

1. The **Schubert determinantal ideal** $I_w \subset \mathbf{k}[\mathbf{z}]$ is generated by all minors in $Z_{[q,p]}$ of size $1 + r_{qp}$ for all q, p, where $Z = (z_{ij})$ is the matrix of variables.
2. The **antidiagonal ideal** J_w is generated by the antidiagonals of the minors of $Z = (z_{ij})$ generating I_w.

Here, the **antidiagonal** of a square matrix or a minor is the product of the entries on the main antidiagonal.

This broad class of determinantal ideals includes all ideals "cogenerated by a fixed minor", as in [HT92]. More generally, the generators of every ladder determinantal ideal coincide with the generators of the Schubert determinantal ideal I_w for some vexillary (also known as 2143-avoiding, or single-shaped) permutation w. Special cases of these vexillary determinantal ideals are the objects of study in [Con95, MS96a, MS96b, CH97, GM00, GL00], for example. Theorem 1.2.3 yields determinantal formulae for $\mathbb{Z}^{2n}$ and $\mathbb{Z}^n$-graded multidegrees in these cases, because vexillary double Schubert polynomials $\mathfrak{S}_w(\mathbf{x}, \mathbf{y})$ are multiSchur polynomials; see [Mac91, Ful92].

An **antidiagonal term order** is any term order in which the leading monomial of every minor is its antidiagonal; such orders are easy to construct.

Theorem 1.4.2. *For any antidiagonal term order,* $\mathsf{in}(I_w) = J_w$*; in other words, the union over q, p of the $(1 + r_{qp})$-minors in $Z_{[q,p]}$ constitute a Gröbner basis.*

The hardest step in the proof of Theorem 1.4.2 shows that the Hilbert series of $\{\mathbf{k}[\mathbf{z}]/J_w\}_{w \in S_n}$ satisfy the recursion defining the **Grothendieck polynomials** $\mathcal{G}_w(\mathbf{x})$ [LS82b]. To be precise, $\mathcal{G}_w(\mathbf{x})$ can be defined by downward induction on the weak order of S_n, in analogy with Schubert polynomials. This time, however, start with $\mathcal{G}_{w_0}(\mathbf{x}) = \prod_{i=1}^n (1 - x_i)^{n-i}$ and use the **Demazure** or **isobaric divided difference** operators $\overline{\partial}_i$, which act by $\overline{\partial}_i(f) = (x_{i+1} f - x_i s_i f)/(x_{i+1} - x_i)$. Demazure operators work just as well on power series, such as $H_w := H(\mathbf{k}[\mathbf{z}]/J_w; \mathbf{x})$ for $w \in S_n$, as they do on polynomials.

Theorem 1.4.3. *If* $\mathrm{length}(ws_i) < \mathrm{length}(w)$ *then* $H_{ws_i} = \overline{\partial}_i H_w$*. Thus*

$$H(\mathbf{k}[\mathbf{z}]/J_w; \mathbf{x}) \quad = \quad \frac{\mathcal{G}_w(\mathbf{x})}{\prod_{i=1}^n (1 - x_i)^n}.$$

This Hilbert series calculation requires substantial tailor-made combinatorics, giving rise to Section 2. Assume Theorem 1.4.3 henceforth in this exposition.

Proof sketch for Theorem 1.4.2. Directly from the definitions of ∂_i and $\overline{\partial}_i$, taking the lowest degree terms in $\mathcal{G}_w(\mathbf{1}-\mathbf{x})$ yields $\mathfrak{S}_w(\mathbf{x})$. Therefore the multidegree of $\mathbf{k}[\mathbf{z}]/J_w$ equals $\mathfrak{S}_w(\mathbf{x})$ and agrees with $[\overline{X}_w]_{\mathbb{Z}^n}$ by Theorem 1.2.3.

After (nontrivially) showing that J_w is pure of dimension $\dim \overline{X}_w$, an easy lemma concerning squarefree monomial ideals reveals why the equality $[\overline{X}_w]_{\mathbb{Z}^n} = [\mathcal{L}_w]_{\mathbb{Z}^n}$ and the containment $\mathsf{in}(I(\overline{X}_w)) \supseteq J_w$ together imply $\mathsf{in}(I(\overline{X}_w)) = J_w$. Since the containments $\mathsf{in}(I(\overline{X}_w)) \supseteq \mathsf{in}(I_w) \supseteq J_w$ are obvious, we conclude that $\mathsf{in}(I(\overline{X}_w)) = \mathsf{in}(I_w) = J_w$. □

1.5. **Applications.** The theorems in the previous section provide formulae for the K-polynomials (i.e. Hilbert series) of Schubert determinantal varieties.

Corollary 1.5.1. *The ideal of $\overline{X}_w$ is I_w, and $\mathcal{K}(\mathbf{k}[\mathbf{z}]/I_w;\mathbf{x}) = \mathcal{G}_w(\mathbf{x})$.*

Proof. That I_w is reduced follows from Theorem 1.4.2 and the fact that J_w is reduced. The advertised K-polynomial comes from Theorem 1.4.2 and Theorem 1.4.3. □

Being in K-theory rather than cohomology, Corollary 1.5.1 is substantially stronger than Theorem 1.2.3. Although the former also follows directly from known results [LS82b] with only a little work (the class of the structure sheaf of $\overline{X}_w$ in $K^\circ_T(M_n)$ maps to that of X_w in $K^\circ(\mathcal{F}\ell_n)$), the Hilbert series in Theorem 1.4.2 must in any case be computed during the course of proving Theorem 1.4.2. Therefore we recover the appropriate results from [LS82b] as consequences.

The antidiagonal ideal J_w is the Stanley–Reisner ideal for a simplicial complex $\mathcal{L}_w$ (thought of as a set of subspaces in M_n). The combinatorial implications of Theorem 1.4.3 and the following corollary will be clarified after our analysis of the simplicial complexes $\mathcal{L}_w$ for $w \in S_n$ and their generalizations to arbitrary Coxeter groups, which occupies all of Section 2.

Corollary 1.5.2. $\mathfrak{S}_w(\mathbf{x}) = [\overline{X}_w]_{\mathbb{Z}^n} = [\mathcal{L}_w]_{\mathbb{Z}^n} = \sum_{\text{facets } L \in \mathcal{L}_w} \mathbf{x}^L$.

The set of minors in I_w forming a minimal (but not reduced) Gröbner basis can be characterized in terms of essential sets [Ful92]. As a consequence, we crystallize the relation between determinantal ideals and open subsets ("opposite cells") in Schubert varieties of $\mathcal{F}\ell_n$. In particular, let $\mathfrak{C}$ be a **local condition**, meaning that $\mathfrak{C}$ holds for a variety whenever it holds on each subvariety in some open cover. Such local conditions include normality, Cohen–Macaulayness, and rational singularities.

Theorem 1.5.3. *Assume that the local condition $\mathfrak{C}$ holds for a variety X whenever it holds for the product of X with any vector space. Then $\mathfrak{C}$ holds for every Schubert variety in every flag variety if and only if $\mathfrak{C}$ holds for all matrix Schubert varieties.*

Theorem 1.5.3 serves to systematize and even reverse the flow of results from the algebraic geometry of flag manifolds to the literature on determinantal ideals. However, we know of no new consequences that can be derived from it.

1.6. **An example.** The following example illustrates the results thus far.

Example 1.6.1. Let $w = 2143 \in S_4$. The matrix Schubert variety $\overline{X}_w$ is then the set of 4×4 matrices $Z = (z_{ij})$ whose upper left entry is zero and whose upper left 3×3 block has rank at most one. The equations defining $\overline{X}_{2143}$ are the determinants

$$I_{2143} \;=\; \left\langle z_{11}, \quad \det \begin{vmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ z_{31} & z_{32} & z_{33} \end{vmatrix} = -z_{13}z_{22}z_{31} + \dots \right\rangle.$$

Note that this is *not* a Gröbner basis with respect to term orders that pick out the diagonal term $z_{11}z_{22}z_{33}$ of the second generator, since z_{11} divides that. The term orders that interest us pick out the *anti*diagonal term $-z_{13}z_{22}z_{31}$.

When we Gröbner-degenerate the matrix Schubert variety to the scheme defined by the initial ideal $J_{2143} = \langle z_{11}, -z_{13}z_{22}z_{31}\rangle$, we get a union of three coordinate subspaces

$$L_{11,13}, L_{11,22}, \text{ and } L_{11,31}, \quad \text{with ideals} \quad \langle z_{11}, z_{13}\rangle, \langle z_{11}, z_{22}\rangle, \text{ and } \langle z_{11}, z_{31}\rangle.$$

The resulting equation in T-equivariant cohomology, or on multidegrees, reads:

$$\begin{array}{rcccccc} [\overline{X}_{2143}]_{\mathbb{Z}^n} & = & [L_{11,13}]_{\mathbb{Z}^n} & + & [L_{11,22}]_{\mathbb{Z}^n} & + & [L_{11,31}]_{\mathbb{Z}^n} \\ & = & x_1^2 & + & x_1x_2 & + & x_1x_3 \end{array}$$

in $\mathbb{Z}[x_1, x_2, x_3, x_4] \cong H_T^*(M_4)$. The $\mathbb{Z}^n$-graded K-polynomial $\mathcal{K}(\mathbf{k}[\mathbf{z}]/I_{2143}; \mathbf{x})$ is equal to $\mathcal{G}_{2143}(\mathbf{x}) = (1 - x_1)(1 - x_1x_2x_3)$. Thus

$$\mathcal{K}(\mathbf{k}[\mathbf{z}]/I_{2143}; \mathbf{1} - \mathbf{x}) \;\; = \;\; x_1(x_1 + x_2 + x_3 - x_1x_2 - x_1x_3 - x_2x_3 + x_1x_2x_3),$$

whose lowest degree terms agree with $\mathfrak{S}_{2143}(\mathbf{x}) = [\overline{X}_{2143}]_{\mathbb{Z}^n} = x_1(x_1 + x_2 + x_3)$. □

2. Mitosis, rc-graphs, and subword complexes

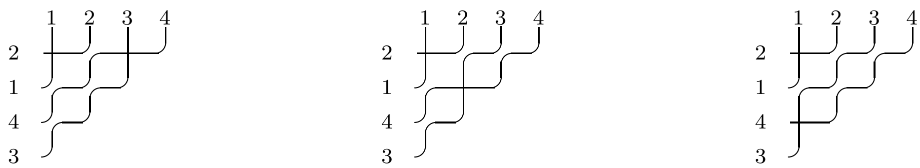
2.1. **Pipe dreams and rc-graphs.** The facets of the initial complex $\mathcal{L}_w$ correspond to certain subsets of the grid $[n] \times [n]$. The obvious question becomes: which subsets? The answer rests on drawing subsets of the grid in the "right" way.

Definition 2.1.1. A **pipe dream** is a finite subset of $\mathbb{Z}_{>0} \times \mathbb{Z}_{>0}$, identified as the set of crosses in a tiling by **crosses** $+$ and **elbow joints** ⌟⌜.

Example 2.1.2. Pictorially, each subspace $L_{11,13}$, $L_{11,22}$, and $L_{11,31}$ from Example 1.6.1 represents a subset of the 4×4 grid: place a '+' in each box containing a generator for its ideal. In other words, given a coordinate subspace L, form the diagram D_L by placing a '+' at (i, j) if every matrix in L has (i, j) entry zero:

$$\langle z_{11}, z_{13}\rangle \; = \; \begin{array}{|c|c|c|c|}\hline + & & + & \\\hline & & & \\\hline & & & \\\hline & & & \\\hline\end{array}, \quad \langle z_{11}, z_{22}\rangle \; = \; \begin{array}{|c|c|c|c|}\hline + & & & \\\hline & + & & \\\hline & & & \\\hline & & & \\\hline\end{array}, \quad \langle z_{11}, z_{31}\rangle \; = \; \begin{array}{|c|c|c|c|}\hline + & & & \\\hline & & & \\\hline + & & & \\\hline & & & \\\hline\end{array}$$

In the figures below, we draw the zero entries '+' by crossing pipes, and the nonzero entries by elbow joints (imagine that the lower-right triangle is filled with elbows).



These are the three "rc-graphs", or "planar histories", for the permutation 2143. □

The term 'rc-graph' was coined in [BB93], although [FK96] made the definition.

Definition 2.1.3. An **rc-graph** is a pipe dream in which each pair of pipes crosses at most once. If D is an rc-graph and $w \in S_n$ is the permutation such that the pipe entering row i exits from column $w(i)$, then D is said to be an **rc-graph for** w. The set of rc-graphs for w is denoted by $\mathcal{RC}(w)$.

The next theorem completes the transition from the geometry of $\mathcal{F}\ell_n$ to the combinatorics of S_n through the algebra of determinants. Recall that the itinerary has been: Schubert variety $X_w \rightsquigarrow$ matrix Schubert variety $\overline{X}_w \rightsquigarrow$ determinantal ideal $I_w \rightsquigarrow$ antidiagonal ideal $J_w \rightsquigarrow$ initial complex $\mathcal{L}_w \rightsquigarrow$ rc-graphs $\mathcal{RC}(w)$.

Theorem 2.1.4. *$\mathcal{RC}(w) = \{D_L \mid L$ is a facet of $\mathcal{L}_w\}$. In other words, rc-graphs for w are complements of maximal supports of monomials outside J_w.*

The proof uses results of [BB93]. Theorem 2.1.4 has the famous 'BJS' formula as a consequence, by Corollary 1.5.2. Previous proofs [BJS93, FS94] were combinatorial.

Corollary 2.1.5. *$\mathfrak{S}_w(\mathbf{x}) = \sum_{D\in\mathcal{RC}(w)} \mathbf{x}^D$, where $\mathbf{x}^D = \prod_{(i,j)\in D} x_i$.*

2.2. **Mitosis algorithm.** The proof in Section 1.4 that the K-polynomial of $\mathbf{k}[\mathbf{z}]/J_w$ equals $\mathcal{G}_w(\mathbf{x})$ works by constructing an operator ε_i^w that takes each monomial outside J_w to a positive sum of monomials outside J_{ws_i}. Interpreted in terms of rc-graphs, this procedure lists the coefficients of Schubert polynomials by downward induction on the weak order of S_n. Our algorithm serves as a geometrically motivated improvement on the famous conjecture of Kohnert [Mac91, Appendix to Chapter IV, by N. Bergeron], which is similarly inductive but employs Rothe diagrams instead of rc-graphs.

Given a pipe dream in $[n]\times[n]$, define the column index

$$\mathrm{start}_i(D) \quad = \quad \min(\{j \mid (i,j)\notin D\}\cup\{n+1\})$$

of the leftmost empty box in row i. Thus in the region to the left of $\mathrm{start}_i(D)$, the i^{th} row of D is filled solidly with crosses. Let

$$\mathcal{J}_i(D) \quad = \quad \{\text{columns } j \text{ strictly to the left of } \mathrm{start}_i(D) \mid (i+1,j) \text{ has no cross in } D\}$$

For $p\in\mathcal{J}_i(D)$, construct the **offspring** D_p as follows. First delete the cross at (i,p) from D. Then take all of the crosses in row i of $\mathcal{J}_i(D)$ that are to the left of column p, and move each one down to the empty box below it in row $i+1$.

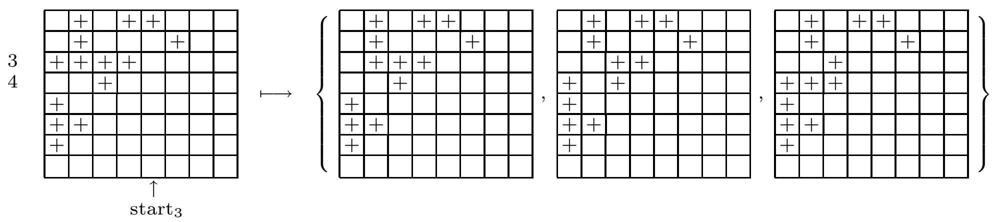
Definition 2.2.1. The i^{th} **mitosis** operator sends a pipe dream D to

$$\mathrm{mitosis}_i(D) \quad = \quad \{D_p \mid p\in\mathcal{J}_i(D)\}.$$

Write $\mathrm{mitosis}_i(\mathcal{P}) = \bigcup_{D\in\mathcal{P}} \mathrm{mitosis}_i(\mathcal{P})$ whenever $\mathcal{P}$ is a set of pipe dreams.

Observe that all of the action takes place in rows i and $i+1$, and $\mathrm{mitosis}_i(D)$ is an empty set whenever $\mathcal{J}_i(D)$ is empty.

Example 2.2.2. The left pipe dream D below is an rc-graph for $w = 13865742$.



The set of three pipe dreams on the right is obtained by applying $\mathrm{mitosis}_3$, since $\mathcal{J}_3(D)$ consists of columns 1, 2, and 4. □

Theorem 2.2.3. *$\mathcal{RC}(ws_i)$ is the disjoint union $\dot\bigcup_{D\in\mathcal{RC}(w)} \mathrm{mitosis}_i(D)$. Therefore, if $s_{i_1}\cdots s_{i_k}$ is a reduced expression for w_0w and D_0 is the unique rc-graph for w_0, then*

$$\mathcal{RC}(w) \quad = \quad \mathrm{mitosis}_{i_k}\cdots\mathrm{mitosis}_{i_1}(D_0).$$

The unique rc-graph D_0 in Theorem 2.2.3 has crosses strictly above the main antidiagonal, and no other crosses. That is, $(i,j)\in D_0$ if and only if $i+j\le n$.

Although combinatorial considerations in the proof of Theorem 1.4.2 were instrumental in figuring out how to define mitosis in the first place, it is possible to give a complete proof of Theorem 2.2.3 based entirely on the BJS formula in Corollary 2.1.5 and the characterization of Schubert polynomials by divided differences, along with elementary combinatorial properties of rc-graphs [Mil02]. This argument exploits an involution on $\mathcal{RC}(w)$ which, for grassmannian (i.e. unique descent) permutations w, reduces to a well-known involution on semistandard Young tableaux.

2.3. **Subword complexes.** Given a pipe dream D, say that a '+' at $(q, p) \in D$ sits on the i^{th} **antidiagonal** if $q + p - 1 = i$. Let Q_D be the ordered list of simple reflections s_i corresponding to the antidiagonals on which the crosses sit, starting from the northeast corner of D and reading *right to left* across each row starting from the top and snaking downwards. Induction on the number of crosses in D proves:

Lemma 2.3.1. *The list Q_D constitutes a reduced expression for w if and only if the pipe dream D is an rc-graph for w.*

For example, the rc-graph D_0 in Theorem 2.2.3 corresponds to the reduced exression $w_0 = s_4s_3s_2s_1s_4s_3s_2s_4s_3s_4$ when $n = 5$, while the full 3×3 grid yields the list $s_3s_2s_1s_4s_3s_2s_5s_4s_3$ of adjacent transpositions in S_6.

Here is the generalization of $\mathcal{L}_w$ to arbitrary words in Coxeter systems (Π, Σ).

Definition 2.3.2. A **word** of size m is an ordered list $Q = (\sigma_1, \ldots, \sigma_m)$ of elements of Σ. An ordered sublist P of Q is called a **subword** of Q.

1. Q **represents** $\pi \in \Pi$ if the ordered product of the simple reflections in Q is a reduced decomposition for π.
2. Q **contains** $\pi \in \Pi$ if some sublist of Q represents π.

The **subword complex** $\Delta(Q, \pi)$ is the set of subwords $P \subseteq Q$ whose complements $Q \smallsetminus P$ contain π.

Example 2.3.3. Theorem 2.1.4 and Lemma 2.3.1 say that if $\Pi = S_{2n}$ and $Q_{n\times n}$ is the word represented by all of $[n] \times [n]$, then $\mathcal{L}_w = \Delta(Q_{n\times n}, w)$ when $w \in S_n \subset S_{2n}$. Thus the combinatorics and Stanley–Reisner theory of general subword complexes yields information about Schubert and Grothendieck polynomials; see Corollary 2.4.5. □

Theorem 2.3.4. *Every subword complex $\Delta(Q, \pi)$ is vertex-decomposable and so Cohen–Macaulay, and even shellable. $\Delta(Q, \pi)$ is homeomorphic to a ball or sphere.*

The vertex decomposition rests on the fact that the link and deletion of the vertex $\sigma_1 \in \Delta(Q, \pi)$ are both subword complexes. The ball or sphere condition is equivalent to $\Delta(Q, \pi)$ being a manifold, given shellability; it reduces to showing that every codimension 1 face lies in at most two facets, which in turn relies on the exchange condition in Coxeter groups [Hum90, Theorem 5.8]. In view of [BW82], Theorem 2.3.4 suggests that the Bruhat and weak orders "feel" somewhat similar.

Fulton proved that $\overline{X}_w$ is Cohen–Macaulay [Ful92], but he used Cohen–Macaulayness of Schubert varieties [Ram85] to do it. Here we provide new proofs of both.

Corollary 2.3.5. *Every matrix Schubert variety $\overline{X}_w$, and hence every Schubert variety $X_w \subseteq B\backslash GL_n$, is Cohen–Macaulay.*

Proof. The Cohen–Macaulayness of $\mathcal{L}_w$ in Theorem 2.3.4 implies that of $\overline{X}_w$ by Theorem 1.4.2 and the flatness of Gröbner degeneration. Now use Theorem 1.5.3. □

2.4. **Combinatorics of Grothendieck polynomials.** Calculating the $\mathbb{Z}^m$-graded Hilbert series (that is, the K-polynomial) of the Stanley–Reisner ring for a simplicial ball or sphere Δ amounts to identifying the boundary faces of Δ. For subword complexes $\Delta(Q, \pi)$, this identification uses a standard tool from Coxeter group theory.

Definition 2.4.1. Let R be a commutative ring, and $\mathcal{D}$ a free R-module with basis $\{e_\pi \mid \pi \in \Pi\}$. Defining a multiplication on $\mathcal{D}$ by

$$(1)\qquad e_\pi e_\sigma \;=\; \begin{cases} e_{\pi\sigma} & \text{if length}(\pi\sigma) > \text{length}(\pi) \\ e_\pi & \text{if length}(\pi\sigma) < \text{length}(\pi) \end{cases}$$

for $\sigma \in \Sigma$ yields the **Demazure algebra** of (Π, Σ) over R. Define the **Demazure product** $\delta(Q)$ of the word $Q = (\sigma_1, \ldots, \sigma_m)$ by $e_{\sigma_1} \cdots e_{\sigma_m} = e_{\delta(Q)}$.

When $\Pi = S_n$ and Σ is the set of simple reflections $s_1, \ldots, s_{n-1}$, the algebra $\mathcal{D}$ is generated over R by the usual Demazure operators $\overline{\partial}_i$ (hence the name "Demazure algebra"). In general, the fact that the equations in (1) define an associative algebra is a special case of [Hum90, Theorem 7.1].

Proposition 2.4.2. *A face $Q \smallsetminus P$ is in the boundary of $\Delta(Q,\pi)$ if and only if P has Demazure product $\delta(P) \neq \pi$.*

In our final theorem, the variables $\mathbf{z} = z_1, \ldots, z_m$ are identified with the locations in the list $Q = (\sigma_1, \ldots, \sigma_m)$—that is, with the vertices of the subword complex $\Delta(Q,\pi)$. The $\mathbb{Z}^n$-grading is the finest possible, with $n = m$. So as not to confuse notation in applications with $\mathbf{z} = (z_{ij})_{i,j=1}^n$, where the Hilbert series are $\mathbb{Z}^n$-graded and expressed in variables $\mathbf{x}$, we write $\mathbb{Z}^m$-graded K-polynomials in the variables $\mathbf{z}$, with each z_i having tautological exponential weight z_i.

Theorem 2.4.3. *If* $\text{length}(\pi) = \ell$ *and J is the Stanley–Reisner ideal of $\Delta(Q,\pi)$, then*

$$\mathcal{K}(\mathbf{k}[\mathbf{z}]/J;\mathbf{z}) \;=\; \sum_{\delta(P)=\pi} (-1)^{|P|-\ell}(\mathbf{1}-\mathbf{z})^P,$$

where $(\mathbf{1}-\mathbf{z})^P = \prod_{i\in P}(1-z_i)$.

The proof uses Hochster's Betti number formula for the **Alexander dual ideal**

$$J^\star \;=\; \langle z_{i_1} \cdots z_{i_k} \mid \langle z_{i_1}, \ldots, z_{i_k}\rangle \text{ contains } J\rangle,$$

which appears in [ER98] and [BCP99] (for instance). Theorem 2.4.3 then follows from the Alexander inversion formula:

Proposition 2.4.4. *If $J \subseteq \mathbf{k}[\mathbf{z}]$ is a squarefree monomial ideal and $J^\star$ is its Alexander dual ideal, then $\mathcal{K}(\mathbf{k}[\mathbf{z}]/J;\mathbf{z}) = \mathcal{K}(J^\star;\mathbf{1}-\mathbf{z})$.*

The Alexander inversion formula can be interpreted as another way to define $J^\star$.

As a special case of Theorem 2.4.3 we recover a formula for double Grothendieck polynomials [FK94]. Although the natural "double" specialization for the exponential weight of z_{ij} is x_i/y_j, we substitute $z_{ij} \mapsto x_i y_j$ to agree with the notation in [FK94].

Corollary 2.4.5. $\mathcal{G}_w(\mathbf{1}-\mathbf{x}, \mathbf{1}-\mathbf{y}) \;=\; \displaystyle\sum_{\delta(D)=w} \prod_{(i,j)\in D} (-1)^{|D|-\ell}(x_i + y_j - x_i y_j).$

The single version, in which $z_{ij} \mapsto x_i$ and $\mathbf{x}^D$ equals $\prod_{(i,j)\in D} x_i$, reads:

$$\mathcal{G}_w(\mathbf{1}-\mathbf{x}) \;=\; \sum_{\delta(D)=w} (-1)^{|D|-\ell}\mathbf{x}^D.$$

Note that Corollary 2.1.5 is the sum of lowest degree terms in the latter formula.

Acknowledgements. The bulk of this exposition is an extended abstract of [KM02a], although the material from Section 2.4 and the second half Theorem 2.3.4 appears in [KM02b]. The authors are grateful to Bernd Sturmfels, who took part in the genesis of this project, and to Misha Kogan, as well as to Sara Billey, Francesco Brenti, Anders Buch, Cristian Lenart, Vic Reiner, Richárd Rimányi, Anne Schilling, Frank Sottile, and Richard Stanley for inspiring conversations and references. Nantel Bergeron kindly provided LaTeX macros for drawing pipe dreams.

References

[BB93] Nantel Bergeron and Sara Billey, *RC-graphs and Schubert polynomials*, Experimental Math. **2** (1993), no. 4, 257–269.

[BCP99] Dave Bayer, Hara Charalambous, and Sorin Popescu, *Extremal Betti numbers and applications to monomial ideals*, J. Algebra **221** (1999), no. 2, 497–512.

[BJS93] Sara C. Billey, William Jockusch, and Richard P. Stanley, *Some combinatorial properties of Schubert polynomials*, J. Algebraic Combin. **2** (1993), no. 4, 345–374.

[Bri98] Michel Brion, *Equivariant cohomology and equivariant intersection theory*, Representation theories and algebraic geometry (Montreal, PQ, 1997), Kluwer Acad. Publ., Dordrecht, 1998, Notes by Alvaro Rittatore, pp. 1–37.

[BW82] Anders Björner and Michelle Wachs, *Bruhat order of Coxeter groups and shellability*, Adv. in Math. **43** (1982), no. 1, 87–100.

[CH97] Aldo Conca and Jürgen Herzog, *Ladder determinantal rings have rational singularities*, Adv. Math. **132** (1997), no. 1, 120–147.

[Con95] Aldo Conca, *Ladder determinantal rings*, J. Pure Appl. Algebra **98** (1995), no. 2, 119–134.

[EG98] Dan Edidin and William Graham, *Equivariant intersection theory*, Invent. Math. **131** (1998), no. 3, 595–634.

[ER98] John A. Eagon and Victor Reiner, *Resolutions of Stanley-Reisner rings and Alexander duality*, J. Pure Appl. Algebra **130** (1998), no. 3, 265–275.

[FK94] Sergey Fomin and Anatol N. Kirillov, *Grothendieck polynomials and the Yang–Baxter equation*, 1994, Proceedings of the Sixth Conference in Formal Power Series and Algebraic Combinatorics, DIMACS, pp. 183–190.

[FK96] Sergey Fomin and Anatol N. Kirillov, *The Yang-Baxter equation, symmetric functions, and Schubert polynomials*, Discrete Math. **153** (1996), no. 1-3, 123–143, Proceedings of the 5th Conference on Formal Power Series and Algebraic Combinatorics (Florence, 1993).

[FR01] László Fehér and Richárd Rimányi, *Classes of degeneracy loci for quivers—the Thom polynomial point of view*, preprint available at http://www.cs.elte.hu/analysis/rimanyi/cikkek/, 2001.

[FS94] Sergey Fomin and Richard P. Stanley, *Schubert polynomials and the nil-Coxeter algebra*, Adv. Math. **103** (1994), no. 2, 196–207.

[Ful92] William Fulton, *Flags, Schubert polynomials, degeneracy loci, and determinantal formulas*, Duke Math. J. **65** (1992), no. 3, 381–420.

[GL00] N. Gonciulea and V. Lakshmibai, *Singular loci of ladder determinantal varieties and Schubert varieties*, J. Algebra **229** (2000), no. 2, 463–497.

[GM00] Nicolae Gonciulea and Claudia Miller, *Mixed ladder determinantal varieties*, J. Algebra **231** (2000), no. 1, 104–137.

[HT92] Jürgen Herzog and Ngô Viêt Trung, *Gröbner bases and multiplicity of determinantal and Pfaffian ideals*, Adv. Math. **96** (1992), no. 1, 1–37.

[Hum90] James E. Humphreys, *Reflection groups and Coxeter groups*, Cambridge University Press, Cambridge, 1990.

[Kaz97] M. É. Kazarian, *Characteristic classes of singularity theory*, The Arnold-Gelfand mathematical seminars, Birkhäuser Boston, Boston, MA, 1997, pp. 325–340.

[KM02a] Allen Knutson and Ezra Miller, *Gröbner geometry of Schubert polynomials*, Preprint, available at http://www-math.mit.edu/~ezra, 2002.

[KM02b] Allen Knutson and Ezra Miller, *Subword complexes in Coxeter groups*, Preprint, available at http://www-math.mit.edu/~ezra, 2002.

[LS82a] Alain Lascoux and Marcel-Paul Schützenberger, *Polynômes de Schubert*, C. R. Acad. Sci. Paris Sér. I Math. **294** (1982), no. 13, 447–450.

[LS82b] Alain Lascoux and Marcel-Paul Schützenberger, *Structure de Hopf de l'anneau de cohomologie et de l'anneau de Grothendieck d'une variété de drapeaux*, C. R. Acad. Sci. Paris Sér. I Math. **295** (1982), no. 11, 629–633.

[Mac91] I. G. Macdonald, *Notes on Schubert polynomials*, Publications du LACIM, Universitè du Québec à Montréal, 1991.

[Mil02] Ezra Miller, *Mitosis recursion for coefficients of Schubert polynomials*, Preprint, available at http://www-math.mit.edu/~ezra, 2002.

[MS96a] J. V. Motwani and M. A. Sohoni, *Divisor class groups of ladder determinantal varieties*, J. Algebra **186** (1996), no. 2, 338–367.

[MS96b] J. V. Motwani and M. A. Sohoni, *Normality of ladder determinantal rings*, J. Algebra **186** (1996), no. 2, 323–337.

[Ram85] A. Ramanathan, *Schubert varieties are arithmetically Cohen-Macaulay*, Invent. Math. **80** (1985), no. 2, 283–294.

[Tot99] Burt Totaro, *The Chow ring of a classifying space*, Algebraic K-theory (Seattle, WA, 1997), Amer. Math. Soc., Providence, RI, 1999, pp. 249–281.

Mathematics Department, UC Berkeley, Berkeley, California
E-mail address: allenk@math.berkeley.edu

Mathematics Department, MIT, Cambridge, Massachusetts
E-mail address: ezra@math.mit.edu

ALTERNATE TRANSITION MATRICES FOR BRENTI'S q-SYMMETRIC FUNCTIONS AND A CLASS OF (q,t)-SYMMETRIC FUNCTIONS ARISING FROM PLETHYSM (EXTENDED ABSTRACT)

THOMAS LANGLEY

Abstract. Brenti [4] introduces a class of q-symmetric functions based on a simple plethysm with the power-sum symmetric functions. Brenti shows that with certain specializations of q, these functions appear in connection with Jack symmetric functions, parking functions, and lattices of non-crossing partitions. Brenti develops combinatorial interpretations for the transition matrices between these new symmetric functions and the standard symmetric function bases. We provide simplified versions of many of these that are sums over significantly smaller classes of combinatorial objects. We then extend Brenti's definitions to symmetric functions on the hyperoctahedral group and give combinatorial interpretations of the analogous transition matrices. We also discuss new generating functions on permutation statistics that arise from Brenti's symmetric functions and our extensions.

Résumé. Brenti [4] présente une classe des fonctions symétriques avec un paramètre simple basé sur un plethysm simple avec les fonctions symétriques de puissance-somme. Brenti prouve qu'avec certaines spécialisations du paramètre, ces fonctions apparaissent en liaison avec des fonctions symétriques de Jack, des fonctions se garantes, et des trellis des cloisons de non-croisement. Brenti développe des traductions combinatoires pour les matrices de transition entre ces nouvelles fonctions symétriques et les bases symétriques standard de fonction. Nous fournissons des versions simplifiées de beaucoup de ces derniers qui sont des sommes au-dessus des classes sensiblement plus petites des objets combinatoires. Nous alors étendons des définitions de Brenti aux fonctions symétriques sur le groupe hyperoctahedral et donnons des traductions combinatoires des matrices analogues de transition. Nous discutons également de nouvelles fonctions se produisantes sur les statistiques de permutation qui résultent des fonctions symétriques de Brenti et de nos extensions.

1. Introduction

This work is based on a class of symmetric functions with a parameter q introduced by Brenti in [4] which arise from a simple plethysm with the power-sum symmetric functions. Brenti develops combinatorial interpretations for the transition matrices between these new symmetric functions and the standard symmetric function bases. We provide simplified versions of many of these that have three advantages. First, most involve summing over significantly smaller classes of objects and are therefore much easier to compute. Second, our expressions are given in terms of objects that appear in the standard transition matrices summarized by Beck, Remmel, and Whitehead in [2] and are therefore more recognizable as q-analogues of those matrices. Finally, part of our work involves extending Brenti's symmetric functions to the hyperoctahedral group and our expressions have natural analogues in that setting. In Section 2 below we give examples of two of our results, one of which has a nice corollary involving the expansion of a product of binomial coefficients as a sum of binomial coefficients. In Section 3 we will discuss our generalization of Brenti's results to the hyperoctahedral group B_n. Specifically, we have defined a class of symmetric functions on B_n in two parameters q and t and have developed the combinatorics of the transition matrices that expand the resulting (q,t)-analogues of the standard bases in terms of the standard bases themselves. We give the basic definitions and a sampling of the results here. Finally, in Section 4 we discuss several new generating functions on permutation statistics that arise from Brenti's symmetric functions and our B_n analogues.

2. Transition matrices arising from Brenti's symmetric functions.

We start with some notation. Let Λ_n be the space of homogeneous symmetric functions of degree n and set $\Lambda = \bigoplus_{n\geq 0} \Lambda_n$. Given two bases of Λ_n, $\{a_\lambda\}_{\lambda\vdash n}$ and $\{b_\lambda\}_{\lambda\vdash n}$, we fix some standard order of the partitions of n and then think of the bases as row vectors $\langle a_\lambda\rangle_{\lambda\vdash n}$ and $\langle b_\lambda\rangle_{\lambda\vdash n}$. We define the transition matrix $M(a,b)$ by the equation

$$\langle b_\lambda\rangle_{\lambda\vdash n} = \langle a_\lambda\rangle_{\lambda\vdash n} M(a,b)$$

So $M(a,b)$ is the matrix that transforms the basis $\langle a_\lambda\rangle_{\lambda\vdash n}$ into the basis $\langle b_\lambda\rangle_{\lambda\vdash n}$ and the (λ,μ) entry of $M(a,b)$ is given by

$$b_\mu = \sum_{\lambda\vdash n} a_\lambda M(a,b)_{\lambda\mu}$$

For λ a partition of n, let h_λ, e_λ, p_λ, s_λ, m_λ, and f_λ denote the complete homogenous, elementary, power-sum, Schur, monomial, and forgotten symmetric functions associated with λ, respectively. Brenti's symmetric functions [4] are defined on $\Lambda\otimes\mathbf{Q}(q)$ as follows. For a nonnegative integer r and an alphabet $X = x_1 + x_2 + \cdots$, define

$$p_r^q[X] = p_r[qX] = qp_r[X]$$

Then for a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)$, set

$$p_\lambda^q[X] = p_{\lambda_1}^q[X] p_{\lambda_2}^q[X]\cdots p_{\lambda_l}^q[X] = q^{l(\lambda)} p_\lambda[X]$$

For any symmetric function g, g^q is defined by expanding g in terms of the power basis so that if $g = \sum_\nu a_\nu p_\nu$, then

$$g^q[X] = \sum_\nu a_\nu p_\nu^q[X]$$

Brenti shows that h_λ^q, e_n^{n+1}, and e_n^n have appeared before in connection with Jack symmetric functions, parking functions, and lattices of non-crossing partitions, respectively.

Brenti derives transition matrices that give the expansions of m_λ^q, e_λ^q, h_λ^q and $s_{\lambda/\mu}^q$ in terms of the m's, e's, h's and s's. As mentioned above, we have simplified many of these by summing over significantly smaller classes of objects. We give two examples of these here.

We start with Brenti's expression for $M(m, e^q)$. First, for a 3-dimensional $m\times n\times p$ matrix $A = (A_{i,j,k})_{1\leq i\leq m, 1\leq j\leq n, 1\leq k\leq p}$, define

$$S_{1,2}(A) = \left(\sum_{i=1}^m\sum_{j=1}^n A_{i,j,1}, \ldots, \sum_{i=1}^m\sum_{j=1}^n A_{i,j,p}\right)$$

and analogously for $S_{1,3}(A)$ and $S_{2,3}(A)$.

Also, for a nonnegative integer i, define

$$\binom{q}{i} = \frac{q(q-1)\cdots(q-i+1)}{i!}$$

Then Brenti shows that

$$M(m,e^q)_{\lambda\mu} = \sum_{i=1}^{|\mu|} \mathcal{M}_{\mu,\lambda}(i)\binom{q}{i}$$

where $\mathcal{M}_{\mu,\lambda}(i)$ is the number of $l(\mu)\times l(\lambda)\times i$ (0,1)-matrices A such that $S_{2,3}(A) = \mu, S_{1,3}(A) = \lambda$ and $S_{1,2}(A) > 0$.

Our expression for $M(m,e^q)$ is a sum over two-dimensional matrices.

Theorem 1. *For λ and μ partitions of n,*

$$M(m,e^q)_{\lambda\mu} = \sum_{A\in M(\mu,\lambda)} \prod_{i,j} \binom{q}{A_{i,j}}$$

where $M(\mu,\lambda)$ is the set of $l(\mu)\times l(\lambda)$ matrices with nonnegative integer entries with row sums equal to μ and column sums equal to λ.

We need the following definitions for our next result.

- *μ-brick tabloids of shape λ.* For $\mu\vdash n$, create a set of bricks that have lengths equal to the parts of μ. Then place these bricks in the Ferrers diagram of λ in such a way that each brick lies in a single row and no two bricks overlap. We call each such filling a μ-brick tabloid of shape λ. For example, Figure 1 shows the four $(1,2,3)$-brick tabloids of shape $(3,3)$.
- *Strict n-brick tabloids of shape λ and type μ.* These are μ-brick tabloids of shape λ that have positive integer labels on the bricks. The labels are from the set $\{1,2,\ldots,n\}$ with repetitions allowed, and the labels must strictly increase in rows from left to right. For example, the first tabloid in Figure 1 gives rise to two 2-brick tabloids of shape $(3,3)$ and type $(1,2,3)$ as shown in Figure 2.

Also, for a nonnegative integer i, define

$$\overline{D}_{\lambda,\mu}(q) = \sum_{i=1}^{l(\mu)} \overline{d}_{\lambda,\mu}(i)\binom{q}{i} \tag{1}$$

where $\overline{d}_{\lambda,\mu}(i)$ is the number of strict i-brick tabloids T of shape λ and type μ such that all of the integers 1,2,...,i appear in T.

Brenti shows that

$$M(m,m^q)_{\lambda,\mu} = \overline{D}_{\lambda,\mu}(q)$$

Our expression is:

Theorem 2. *For λ and μ partitions of n,*

$$M(m,m^q)_{\lambda,\mu} = \sum_{T\in\mathcal{B}_{\mu,\lambda}} \prod_{i=1}^{l(\lambda)} \binom{q}{n_i(T)}$$

where $\mathcal{B}_{\mu,\lambda}$ is the set of μ-brick tabloids of shape λ and for $T\in\mathcal{B}_{\mu,\lambda}, n_i(T)$ is the number of bricks in the i^{th} row of T.

We consider a simple example to demonstrate the number of objects involved in the two expressions for $M(m,m^q)$. Let $\lambda=(3,3)$ and $\mu=(1,2,3)$. Then there are four elements of $\mathcal{B}_{\mu,\lambda}$ as shown in Figure 1. Now consider the corresponding $\overline{d}_{\lambda,\mu}(i)$ in Brenti's expression

$$M(m,m^q)_{\lambda,\mu} = \overline{D}_{(\lambda,\mu)}(q) = \sum_{i=1}^{l(\mu)} \overline{d}_{\lambda,\mu}(i)\binom{q}{i}$$

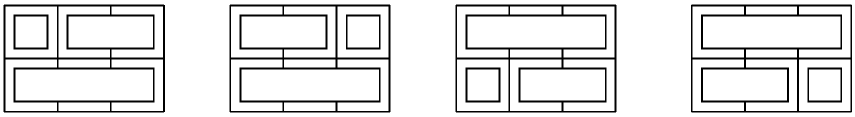


Figure 1. The four $(1,2,3)$-brick tabloids of shape $(3,3)$.

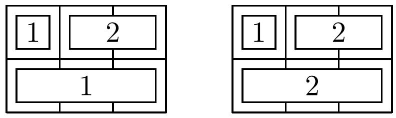


Figure 2. The two strict 2-brick tabloids of shape $(3,3)$ and type $(1,2,3)$ corresponding to the first tabloid in Figure 1.

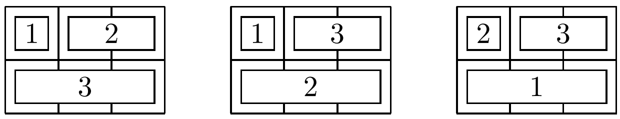


Figure 3. The three strict 3-brick tabloids of shape $(3,3)$ and type $(1,2,3)$ that contain all three of the labels 1, 2, and 3, corresponding to the first tabloid in Figure 1.

For each $i = 1$ to 3, to calculate $\overline{d}_{\lambda,\mu}(i)$ we need to label the bricks in each tabloid in Figure 1 with the integers 1 through i such that each label is used at least once and the labels strictly increase in each row. Since each tabloid has at least one row with two bricks, $\overline{d}_{\lambda,\mu}(1) = 0$. Each tabloid in Figure 1 gives rise to two 2-brick tabloids, shown in Figure 2. So $\overline{d}_{\lambda,\mu}(2) = 8$. Finally, there are three 3-brick tabloids labeled with 1, 2, and 3 corresponding to each tabloid in Figure 1, as shown in Figure 3. So $\overline{d}_{\lambda,\mu}(3) = 12$. Hence Brenti's expression involves counting twenty tabloids instead of four.

We note that our expressions for $M(m, e^q)$ and $M(m, m^q)$ follow from Brenti's by letting q be a positive integer and giving a short combinatorial argument. We can also prove our results directly, and it is these proofs that generalize nicely when we extend Brenti's results to the hyperoctahedral group.

It is interesting to note that as a corollary to Theorem 2 we obtain a combinatorial interpretation of the coefficients used to expand an arbitrary product of binomial coefficients as a sum of binomial coefficients:

Corollary 3. *For $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l) \vdash n$,*

$$\prod_{i=1}^{l} \binom{q}{\lambda_i} = \sum_{i=1}^{n} \overline{d}_{\lambda,1^n}(i) \binom{q}{i}$$

where $\overline{d}_{\lambda,1^n}(i)$ is the number of ways to fill the Ferrers diagram of λ with the integers $1, 2, \ldots, i$ so that the entries in rows strictly increase and every integer is used at least once.

3. A Class of (q,t)-Symmetric Functions on B_n

In this section we define a class of symmetric functions with two parameters on the hyperoctahedral group B_n analogous to Brenti's symmetric functions on S_n. We have a complete list of transition matrices that expand these new symmetric functions in terms of the standard bases for the symmetric functions on B_n. We will present those analogous to $M(m,e^q)$ and $M(m,m^q)$ here.

We start by reviewing the necessary definitions. Let B_n be the hyperoctahedral group on n elements, that is, the wreath product of Z_2 with S_n. We will think of B_n as the group of *signed permutations*. That is, each element $\sigma \in B_n$ can be written as a permutation $\sigma = \sigma_1\sigma_2\cdots\sigma_n$ where each σ_i can be positive or negative. We can write elements of B_n in cycle notation with cycles of the form

$$(\epsilon_1 i_1, \epsilon_2 i_2, \epsilon_3 i_3, \ldots, \epsilon_m i_m)$$

where $\epsilon_i \in \{-1,1\}$ and it is understood that i_1 is mapped to $\epsilon_2 i_2$, i_2 is mapped to $\epsilon_3 i_3$ and so on.

As an example, the following is an element of B_8:

$$\sigma = -4, 8, -2, -7, -3, -6, 1, 5$$

In cycle notation we have

$$\sigma = (1,-4,-7)(-2,8,5,-3)(-6)$$

We say that a cycle of $\sigma \in B_n$ is a *positive cycle* if the product of the signs in the cycle is positive and a *negative cycle* if the product of the signs is negative. In the example above, the first two cycles are positive cycles and the third cycle is a negative cycle. Let $(\lambda,\mu) \vdash n$ denote an ordered pair of partitions λ and μ such that $|\lambda|+|\mu| = n$. The conjugacy classes of B_n are indexed by such pairs of partitions in the following way. If $B_n(\lambda,\mu)$ denotes the conjugacy class corresponding to (λ,μ), then $B_n(\lambda,\mu)$ is the set of all $\sigma \in B_n$ such that the positive cycles of σ have lengths $\lambda_1, \lambda_2, \ldots, \lambda_{l(\lambda)}$ and the negative cycles of σ have lengths $\mu_1, \mu_2, \ldots, \mu_{l(\mu)}$.

We associate symmetric functions to B_n following Stembridge [5]. Let $X = x_1 + x_2 + \cdots$ and $Y = y_1 + y_2 + \cdots$ be commuting alphabets. The appropriate vector space is defined by

$$\Lambda_{B_n}[X,Y] = \bigoplus_{k=0}^{n} \Lambda_k[X] \otimes \Lambda_{n-k}[Y]$$

The ten standard bases of Λ_{B_n} are products of the standard S_n bases:

$$\begin{array}{ll}
\{p_\lambda[X]p_\mu[Y]\}_{(\lambda,\mu)\vdash n} & \{m_\lambda[X+Y]m_\mu[X-Y]\}_{(\lambda,\mu)\vdash n} \\
\{f_\lambda[X+Y]f_\mu[X-Y]\}_{(\lambda,\mu)\vdash n} & \{m_\lambda[X+Y]f_\mu[X-Y]\}_{(\lambda,\mu)\vdash n} \\
\{f_\lambda[X+Y]m_\mu[X-Y]\}_{(\lambda,\mu)\vdash n} & \{h_\lambda[X+Y]h_\mu[X-Y]\}_{(\lambda,\mu)\vdash n} \\
\{e_\lambda[X+Y]e_\mu[X-Y]\}_{(\lambda,\mu)\vdash n} & \{h_\lambda[X+Y]e_\mu[X-Y]\}_{(\lambda,\mu)\vdash n} \\
\{e_\lambda[X+Y]h_\mu[X-Y]\}_{(\lambda,\mu)\vdash n} & \{s_\lambda[X+Y]s_\mu[X-Y]\}_{(\lambda,\mu)\vdash n}
\end{array}$$

We are now ready to define an analogue of Brenti's symmetric functions on Λ_{B_n}. We define

$$(p_n[X]p_m[Y])^{q,t} = p_n^q[X]p_m^t[Y] = q \cdot p_n[X] \cdot t \cdot p_m[Y]$$

With this definition we have

$$(p_\lambda[X]p_\mu[Y])^{q,t} = p_\lambda^q[X]p_\mu^t[Y] = q^{l(\lambda)}t^{l(\mu)}p_\lambda[X]p_\mu[Y]$$

So in general for $P \in \Lambda_{B_n}$, we expand P in the power basis so that $P = \sum_{(\lambda,\mu)} a_{\mu,\lambda} p_\lambda[X] p_\mu[Y]$ and define $P^{q,t}$ by

$$P^{q,t} = \sum_{(\lambda,\mu)} a_{\mu,\lambda} p^q_\lambda[X] p^t_\mu[Y] = \sum_{(\lambda,\mu)} a_{\mu,\lambda} q^{l(\lambda)} t^{l(\mu)} p_\lambda[X] p_\mu[Y]$$

We will use notation similar to the S_n case to describe transition matrices. For simplicity, we will set

$$\begin{aligned} p_\lambda \tilde{p}_\mu &= p_\lambda[X] p_\mu[Y] \\ h_\lambda \tilde{e}_\mu &= h_\lambda[X+Y] e_\mu[X-Y] \end{aligned}$$

and so on. We define $M(a\tilde{b}, c\tilde{d})$ by the equation

$$\langle c_\lambda \tilde{d}_\mu \rangle_{(\lambda,\mu)\vdash n} = \langle a_\lambda \tilde{b}_\mu \rangle_{(\lambda,\mu)\vdash n} M(a\tilde{b}, c\tilde{d})$$

So $M(a\tilde{b}, c\tilde{d})$ is the matrix that transforms the basis $\langle a_\lambda \tilde{b}_\mu \rangle_{(\lambda,\mu)\vdash n}$ into the basis $\langle c_\lambda \tilde{d}_\mu \rangle_{(\lambda,\mu)\vdash n}$ and the $((\alpha,\beta),(\lambda,\mu))$ entry of $M(a\tilde{b}, c\tilde{d})$ is given by

$$c_\lambda \tilde{d}_\mu = \sum_{(\alpha,\beta)\vdash n} a_\alpha \tilde{b}_\beta M(a\tilde{b}, c\tilde{d})_{(\alpha,\beta),(\lambda,\mu)}$$

We now present results analogous to Theorems 1 and 2.

Theorem 4. *For $(\lambda,\mu) \vdash n$ and $(\rho,\epsilon) \vdash n$,*

$$\begin{aligned} &M(m\tilde{m}, (e\tilde{e})^{q,t})_{(\rho,\epsilon)(\lambda,\mu)} \\ &= \sum_{M \in M((\lambda,\mu)(\rho,\epsilon))} \prod_{\substack{1 \le i \le l(\lambda) \\ 1 \le j \le l(\rho)}} \binom{q/2 + t/2}{M_{i,j}} \prod_{\substack{1 \le i \le l(\lambda) \\ l(\rho)+1 \le j \le l(\rho)+l(\epsilon)}} \binom{q/2 - t/2}{M_{i,j}} \\ &\qquad \prod_{\substack{l(\lambda)+1 \le i \le l(\lambda)+l(\mu) \\ 1 \le j \le l(\rho)}} \binom{q/2 - t/2}{M_{i,j}} \prod_{\substack{l(\lambda)+1 \le i \le l(\lambda)+l(\mu) \\ l(\rho)+1 \le j \le l(\rho)+l(\epsilon)}} \binom{q/2 + t/2}{M_{i,j}} \end{aligned}$$

where $M((\lambda,\mu)(\rho,\epsilon))$ is the set of $(l(\lambda)+l(\mu)) \times (l(\rho)+l(\epsilon))$ matrices with nonnegative integer entries such that the row sums form the sequence $\lambda_1,\ldots,\lambda_{l(\lambda)}, \mu_1,\ldots,\mu_{l(\mu)}$ and the column sums form the sequence $\rho_1,\ldots,\rho_{l(\rho)}, \epsilon_1,\ldots,\epsilon_{l(\epsilon)}$ and $M_{i,j}$ denotes the (i,j) entry of the matrix M.

For the next result we need to define B_n versions of brick tabloids. For partitions ρ and ϵ, let $\rho * \epsilon$ be the diagram obtained by attaching the lower right corner of the Ferrers diagram of ρ to the upper left corner of the Ferrers diagram of ϵ. Then for partitions λ and μ, we distinguish λ-bricks from μ-bricks and define the set of (λ,μ)-brick tabloids of shape $\rho * \epsilon$ to be the set of fillings of the diagram $\rho * \epsilon$ with λ-bricks and μ-bricks such that

- each brick lies in a single row
- no two bricks overlap
- λ bricks come before μ bricks in each row.

We will denote this set $\mathcal{B}^{\lambda,\mu}_{\rho*\epsilon}$. For example, Figure 4 shows an element of $\mathcal{B}^{\lambda,\mu}_{\rho*\epsilon}$ for $\lambda = (1,1,2,3)$, $\mu = (1,1,2,3,3)$, $\rho = (2,7)$ and $\epsilon = (2,2,3)$.

We also need an expression analogous to $\overline{D}_{\lambda,\mu}(q)$. For $T \in \mathcal{B}^{\lambda,\mu}_{\rho*\epsilon}$, let $n^\lambda_{i,\rho}(T)$ be the number of λ-bricks in the i-th row of ρ in T and define $n^\mu_{i,\rho}(T)$, $n^\lambda_{i,\epsilon}(T)$, and $n^\mu_{i,\epsilon}(T)$ analogously.

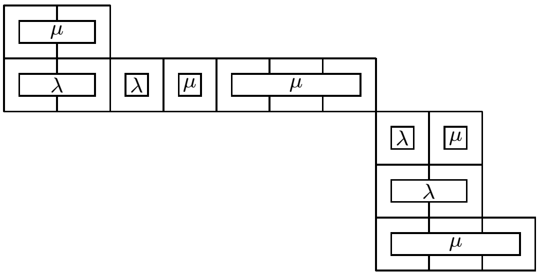


FIGURE 4. An element of $\mathcal{B}^{(1,1,2,2),(1,1,2,3,3)}_{(2,7)*(2,2,3)}$.

Then define

$$\overline{D}_{(\rho,\epsilon),(\lambda,\mu)}(q,t) = \sum_{T\in\mathcal{B}^{\lambda,\mu}_{\rho*\epsilon}} \prod_{i=1}^{l(\rho)} \binom{q/2+t/2}{n^{\lambda}_{i,\rho}(T)}\binom{q/2-t/2}{n^{\mu}_{i,\rho}(T)} \cdot \prod_{i=1}^{l(\epsilon)} \binom{q/2-t/2}{n^{\lambda}_{i,\epsilon}(T)}\binom{q/2+t/2}{n^{\mu}_{i,\epsilon}(T)}$$

The B_n version of Theorem 2 is then

Theorem 5. *For $(\lambda,\mu)\vdash n$ and $(\rho,\epsilon)\vdash n$,*

$$M(m\tilde{m},(m\tilde{m})^{q,t})_{(\rho,\epsilon)(\lambda,\mu)} = \overline{D}_{(\rho,\epsilon)(\lambda,\mu)}(q,t)$$

We note that we can also give expressions analogous to Brenti's version of $\overline{D}_{\lambda,\mu}(q)$.

Theorem 6. *For $(\lambda,\mu)\vdash n$ and $(\rho,\epsilon)\vdash n$,*

$$\overline{D}_{(\rho,\epsilon),(\lambda,\mu)}(q,t) = \sum_{\substack{1\le i+k\le l(\lambda)\\ 1\le j+l\le l(\mu)}} \overline{d}_{\lambda,\mu,\rho,\epsilon}(i,j,k,l)\binom{q/2+t/2}{i}\binom{q/2-t/2}{j}\binom{q/2-t/2}{k}\binom{q/2+t/2}{l}$$

where $\overline{d}_{\lambda,\mu,\rho,\epsilon}(i,j,k,l)$ is the number of (λ,μ)-brick tabloids of shape $\rho*\epsilon$ such that:

- *λ-bricks come before μ bricks in each row*
- *λ bricks in ρ are labelled with $1,2,\dots,i$ with every value appearing at least once and the labels strictly increasing in rows*
- *μ-bricks in ρ are labelled with $1,2,\dots,j$, with every value appearing at least once and the labels strictly increasing in rows*
- *λ-bricks in ϵ are labelled with $1,2,\dots,k$ with every value appearing at least once and the labels strictly increasing in rows*
- *μ-bricks in ϵ are labelled with $1,2,\dots,l$ with every value appearing at least once and the labels strictly increasing in rows.*

4. GENERATING FUNCTIONS

In this section we present some generating functions on permutation statistics that arise from the symmetric functions introduced in the preceding sections. First, Brenti introduces a homomorphism from Λ_n to polynomials in one variable over the rationals which, when applied to the standard bases of Λ_n, gives generating functions on permutation statistics

[3]. Brenti extends this homomorphism to $\Lambda \otimes \mathbf{Q}(q)$ in [4] and derives similar generating functions for Brenti's q-symmetric functions. We will discuss these below and then give extensions to the B_n versions. Next, we present generating functions that arise by substituting Brenti's symmetric functions and extensions into the well-known identity

$$\sum_{n\geq 0} u^n h_n = \frac{1}{\sum_{n\geq 0}(-u)^n e_n}$$

Brenti's homomorphism $\xi : \Lambda_{\mathbf{Q}} \to \mathbf{Q}[x]$ is defined by setting

$$\xi(e_n) = \frac{(1-x)^{n-1}}{n!} \tag{2}$$

Since the e_μ's are a basis of $\Lambda_{\mathbf{Q}}$, this defines ξ on $\Lambda_{\mathbf{Q}}$. The homomorphism is then extended to $\Lambda \otimes \mathbf{Q}(q)$ by letting ξ act on $\mathbf{Q}(q)$ as it does on scalars.

Recall that for $\sigma = \sigma_1\sigma_2\cdots\sigma_n \in S_n$, i is a called an *excedance* of σ if $\sigma_i > i$. We denote the number of excedances in σ by $\mathrm{exc}(\sigma)$. Also let $\mathrm{cyc}(\sigma)$ denote the number of cycles in σ.

Brenti shows the following [4]:

Theorem 7. *For ξ defined in (2) and a positive integer n,*

(1) $n!\xi\left(e_n^q\right) = (-1)^n \sum_{\sigma\in S_n} x^{\mathrm{exc}(\sigma)}(-q)^{\mathrm{cyc}(\sigma)}$
(2) $n!\xi\left(h_n^q\right) = \sum_{\sigma\in S_n} x^{\mathrm{exc}(\sigma)}q^{\mathrm{cyc}(\sigma)}$
(3) $\frac{n!}{z_\mu}\xi\left(p_\mu^q\right) = \sum_{\sigma\in S_n(\mu)} x^{\mathrm{exc}(\sigma)}q^{\mathrm{cyc}(\sigma)}$
(4) $n!\xi\left(s_\mu^q\right) = \sum_{\sigma\in S_n} \chi^\mu(\sigma)x^{\mathrm{exc}(\sigma)}q^{\mathrm{cyc}(\sigma)}$

where $S_n(\mu)$ is the conjugacy class of S_n associated with μ, χ^μ is the irreducible S_n character associated with μ, and if $m_i(\mu)$ denotes the number of parts of μ of size i, then $z_\mu = 1^{m_1(\mu)}\cdots n^{m_n(\mu)}m_1(\mu)!\cdots m_n(\mu)!$.

Beck [1] defines an analogue of Brenti's homomorphism, $\zeta : \Lambda_B \to Q[x]$, by setting

$$\zeta(e_k[X+Y]) = \frac{(1-x)^{k-1} + x(x-1)^{k-1}}{2^k k!} \tag{3}$$

$$\zeta(e_k[X-Y]) = \frac{(1-x)^{k-1} - x(1-x)^{k-1}}{2^k k!} = \frac{(1-x)^k}{2^k k!} \tag{4}$$

As with ξ, we extend ζ to our (q,t)-symmetric functions by treating q and t as scalars.

The statistic that arises here is a B_n version of *decedances.* We first define a linear ordering Θ by setting

$$1 <_\Theta 2 <_\Theta \cdots <_\Theta n <_\Theta \cdots <_\Theta -n <_\Theta \cdots <_\Theta -2 <_\Theta -1$$

Then for $\sigma \in B_n$, write σ in cycle notation as

$$\sigma = (\sigma_{1_1}\sigma_{1_2}\cdots\sigma_{1_{l(1)}})(\sigma_{2_1}\sigma_{2_2}\cdots\sigma_{2_{l(2)}})\cdots(\sigma_{k_1}\sigma_{k_2}\cdots\sigma_{k_{l(k)}})$$

A B_n decedance occurs at the j^{th} position of the i^{th} cycle if either $1 \leq j < l(i)$ and $\sigma_{i_j} >_\Theta \sigma_{i_{j+1}}$, or $j = l(i)$ and $\sigma_{i_{l(i)}} >_\Theta \sigma_{i_1}$. The number of B_n decedances of σ is denoted $\mathrm{de}_B(\sigma)$.

Beck [1] shows that:

Theorem 8. *For ζ defined in equations (3) and (4) and a positive integer n,*

$$\frac{2^n n!}{z_\lambda z_\mu}\zeta(p_\lambda[X]p_\mu[Y]) = \sum_{\sigma\in B_n(\lambda,\mu)} x^{\mathrm{de}_B(\sigma)}$$

where $B_n(\lambda,\mu)$ is the set of B_n permutations with positive cycles of type λ and negative cycles of type μ.

From this we can derive results analagous to Brenti's S_n results:

Theorem 9. *For ζ defined in equations (3) and (4) and a positive integer n,*

(1) $\frac{2^n n!}{z_\lambda z_\mu}\zeta(p_\lambda^q[X]p_\mu^t[Y]) = \sum_{\sigma\in B_n(\lambda,\mu)} x^{\mathrm{de}_B(\sigma)}q^{\mathrm{poscyc}(\sigma)}t^{\mathrm{negcyc}(\sigma)}$

(2) $2^n n!\zeta(h_n^{q,t}[X+Y]) = \sum_{\sigma\in B_n} x^{\mathrm{de}_B(\sigma)}q^{\mathrm{poscyc}(\sigma)}t^{\mathrm{negcyc}(\sigma)}$

(3) $2^n n!\zeta(h_n^{q,t}[X-Y]) = \sum_{\sigma\in B_n} x^{\mathrm{de}_B(\sigma)}q^{\mathrm{poscyc}(\sigma)}(-t)^{\mathrm{negcyc}(\sigma)}$

(4) $2^n n!\zeta(e_n^{q,t}[X+Y]) = (-1)^n\sum_{\sigma\in B_n} x^{\mathrm{de}_B(\sigma)}(-q)^{\mathrm{poscyc}(\sigma)}(-t)^{\mathrm{negcyc}(\sigma)}$

(5) $2^n n!\zeta(e_n^{q,t}[X-Y]) = (-1)^n\sum_{\sigma\in B_n} x^{\mathrm{de}_B(\sigma)}(-q)^{\mathrm{poscyc}(\sigma)}t^{\mathrm{negcyc}(\sigma)}$

(6) $2^n n!\zeta((s_\lambda[X+Y]s_\mu[X-Y])^{q,t}) = \sum_{\sigma\in B_n} \chi^{\lambda,\mu}(\sigma)x^{\mathrm{de}_B(\sigma)}q^{\mathrm{poscyc}(\sigma)}t^{\mathrm{negcyc}(\sigma)}$

where for $\sigma\in B_n$, $\mathrm{poscyc}(\sigma)$ is the number of positive cycles of σ and $\mathrm{negcyc}(\sigma)$ is the number of negative cycles of σ and $\chi^{\lambda,\mu}$ is the irreducible character of B_n associated with the conjugacy class $B_n(\lambda,\mu)$.

Next we present two results arising from the well-known identity

$$\sum_{n\geq 0} u^n h_n = \frac{1}{\sum_{n\geq 0}(-u)^n e_n} \tag{5}$$

Substituting Brenti's symmetric functions on S_n into (5), we obtain:

Theorem 10. *For a positive integer n,*

$$\sum_{n\geq 0}\frac{u^n}{n!}\sum_{\sigma\in S_n} x^{\mathrm{exc}(\sigma)}q^{\mathrm{cyc}(\sigma)} = \frac{1}{\sum_{n\geq 0}\frac{(u(x-1))^n}{n!}\sum_{k=1}^n S_{n,k}\frac{(q)\downarrow_k}{(1-x)^k}}$$

where $S_{n,k}$ is the Stirling number of the second kind, that is, the number of set partitions of $\{1,2,\ldots,n\}$ into k parts.

We conclude with the analogous result for the B_n symmetric functions.

Theorem 11. *For a positive integer n,*

$$\sum_{n\geq 0}\frac{u^n}{2^n n!}\sum_{\sigma\in B_n} x^{\mathrm{de}_B(\sigma)}q^{\mathrm{poscyc}(\sigma)}t^{\mathrm{negcyc}(\sigma)}$$
$$= \frac{1}{\exp\left\{\frac{u(x-1)(q-t)}{4}\right\}\sum_{n\geq 0}\frac{(u(x-1))^n}{2^n n!}\sum_{k=1}^n S_{n,k}^{\mathrm{odd}}\left(\frac{1+x}{1-x}\right)(q/2+t/2)\downarrow_k}$$

where $S_{n,k}^{\mathrm{odd}}(y) = \sum_{\pi\in P_{n,k}} y^{\#\text{ odd parts}(\pi)}$ and $P_{n,k}$ is the set of partitions of $\{1,2,\ldots,n\}$ into k parts.

References

[1] D. Beck. *Permutation Enumeration of the Symmetric Group and the Hyperoctahedral Group and the Combinatorics of Symmetric Functions.* PhD dissertation, University of California, San Diego, 1993.

[2] D. Beck, J. Remmel, and T. Whitehead. The combinatorics of transition matrices between the bases of the symmetric functions and the B_n analogues. *Discrete Mathematics*, 153:3–27, 1996.

[3] F. Brenti. Permutation enumeration, symmetric functions, and unimodality. *Pacific J. Math.*, 157:1–28, 1993.

[4] F. Brenti. A class of q-symmetric functions arising from plethysm. *J. Combin. Theory Ser. A*, 91(1-2):137–170, 2000.

[5] J. Stembridge. The projective representations of the hyperoctahedral group. *Journal of Algebra*, 145:396–453, 1992.

Department of Mathematics, Rose-Hulman Institute of Technology, Terre Haute, IN.
E-mail address: Thomas.Langley@rose-hulman.edu

REFINED POSITIVITY CONJECTURES AND THE MACDONALD POLYNOMIALS

L. LAPOINTE AND J. MORSE*

Abstract. We consider a filtration of the symmetric function space given by the linear span of Hall-Littlewood polynomials indexed by partitions whose first part is not larger than k. We introduce symmetric functions called the k-Schur functions, providing an analog for the Schur functions in these k-subspaces. We give several properties for the k-Schur functions including that they form a basis that reduces to the Schur basis when k is large. We also show that the connection coefficients for the k-Schur function basis with Macdonald polynomials belonging to the k-subspaces are polynomials in q and t with integral coefficients. In fact, we conjecture that these integral coefficients are actually positive, and give several other conjectures generalizing Schur function theory.

Résumé. Nous étudions la filtration de l'espace des fonctions symétriques que l'on obtient en considérant l'engendré linéaire des polynômes de Hall-Littlewood indicés par des partitions dont la première entrée n'est pas plus grande que k, pour $k = 1, 2, 3, ...$ Nous introduisons des fonctions symétriques, les k-fonctions de Schur, qui sont en quelque sorte les analogues des fonctions de Schur dans le sous-espace correspondant à k de la filtration. Nous obtenons plusieurs propriétés de ces k-fonctions de Schur, parmi lesquelles le fait qu'elles forment une base se réduisant à la base des fonctions de Schur lorsque k est grand. Nous démontrons aussi que les entrées de la matrice de changement de base entre les polynômes de Macdonald appartenant au k-ième sous-espace de la filtration et les k-fonctions de Schur sont des polynômes en q et t à coefficients entiers. Nous émettons la conjecture que ces coefficients entiers sont en fait positifs et formulons plusieurs autres conjectures se voulant des k-généralisations de propriétés des fonctions de Schur.

1. Introduction

Let Λ be the ring of symmetric functions in the variables $x_1, x_2, \ldots$, with coefficients in $\mathbb{Q}(q,t)$, for parameters q and t. The Schur functions, $s_\lambda[X]$, form a fundamental basis of Λ, with central roles in fields such as representation theory and algebraic geometry. For example, the Schur functions can be identified with the characters of irreducible representations of the symmetric group, and their products are equivalent to the Pieri formulas for multiplying Schubert varieties in the intersection ring of a Grassmannian. Furthermore, the connection coefficients of the Schur function basis with various bases such as the homogeneous symmetric functions, the Hall-Littlewood polynomials, and the Macdonald polynomials, are positive and have representation theoretic interpretations. In the case of the Macdonald polynomials, $H_\lambda[X;q,t]$, this expansion takes the form

$$H_\lambda[X;q,t] = \sum_\mu K_{\mu\lambda}(q,t)\, s_\mu[X]\,, \qquad K_{\mu\lambda}(q,t) \in \mathbb{N}[q,t]\,, \tag{1}$$

where $K_{\mu\lambda}(q,t)$ are known as the q,t-Kostka polynomials. The representation theoretic interpretation for these polynomials is given in [1, 2, 3].

* Supported by a grant from the NSF.

We study the filtration $\Lambda_t^{(1)} \subseteq \Lambda_t^{(2)} \subseteq \cdots \subseteq \Lambda_t^{(\infty)} = \Lambda$, given by the linear span of Hall-Littlewood polynomials indexed by k-bounded partitions. That is,

$$\Lambda_t^{(k)} = \mathcal{L}\{H_\lambda[X;t]\}_{\lambda;\lambda_1\leq k}\,, \qquad k=1,2,3,\ldots\,. \tag{2}$$

We introduce a new family of symmetric functions that are indexed by k-bounded partitions, denoted $s_\lambda^{(k)}[X;t]$, and give a number of properties for these functions. In particular, we show that they form bases for the subspaces, $\Lambda_t^{(k)}$. Our functions will be called the k-Schur functions since they appear to play a role for $\Lambda_t^{(k)}$ that is analogous to the role of the Schur functions for Λ. That is, the k-Schur functions give rise to the generalization of many Schur positivity properties. Details are given following a brief outline of our construction for $s_\lambda^{(k)}[X;t]$.

The characterization of $s_\lambda^{(k)}[X;t]$ relies on a t-generalization for Schur function products. More precisely, for any partition sequence $S=(\lambda^{(1)},\ldots,\lambda^{(\ell)})$, the t-analog for the product $s_{\lambda^{(1)}}[X]\cdots s_{\lambda^{(\ell)}}[X]$ has been studied in a number of papers including [9, 11, 12, 13, 14]. We find that a very particular subset of these generalized products forms a basis for $\Lambda_t^{(k)}$. The elements of our basis, denoted $G_\lambda^{(k)}[X;t]$, are the generalized Schur products with sequence S obtained by splitting λ into pieces that depend on k. $G_\lambda^{(k)}[X;t]$ are thus called k-split polynomials. These polynomials are essential in our definition for the k-Schur functions as we use a linear operator on $\Lambda_t^{(k)}$ defined by

$$T_i^{(k)}\, G_\lambda^{(k)}[X;t] = \begin{cases} G_\lambda^{(k)}[X;t] & \text{if } \lambda_1 = i \\ 0 & \text{otherwise} \end{cases}\,. \tag{3}$$

The final ingredient needed to define the k-Schur functions is the vertex operator, B_i, introduced in [4] to recursively build the Hall-Littlewood polynomials. More precisely,

$$H_{\lambda_1,\ldots,\lambda_\ell}[X;t] = B_{\lambda_1}\cdots B_{\lambda_\ell}\cdot 1\,. \tag{4}$$

Analogously to this relation, we now define the k-Schur function for k-bounded $\lambda = (\lambda_1,\ldots,\lambda_\ell)$, by

$$s_\lambda^{(k)}[X;t] = T_{\lambda_1}^{(k)} B_{\lambda_1}\cdots T_{\lambda_\ell}^{(k)} B_{\lambda_\ell}\cdot 1\,. \tag{5}$$

Our work to characterize this basis was originally motivated by two conjectures suggesting that the k-Schur functions play a central role in the understanding of the q,t-Kostka polynomials. Together, these conjectures refine relation (1). That is, for any k-bounded partition λ,

$$i) \qquad s_\lambda^{(k)}[X;t] = \textstyle\sum_{\mu\geq\lambda} v_{\mu\lambda}^{(k)}(t)\, s_\mu[X]\,, \qquad v_{\mu\lambda}^{(k)}(t)\in\mathbb{N}[t]\,, \tag{6}$$

$$ii) \quad H_\lambda[X;q,t] = \textstyle\sum_{\mu;\mu_1\leq k} K_{\mu\lambda}^{(k)}(q,t)\, s_\mu^{(k)}[X;t]\,, \qquad K_{\mu\lambda}^{(k)}(q,t)\in\mathbb{N}[q,t]\,. \tag{7}$$

Both conjectures hold when $k=2$ and we prove that $v_{\mu\lambda}^{(k)}(t)\in\mathbb{Z}[t]$ and that $K_{\mu\lambda}^{(k)}(q,t)\in\mathbb{Z}[q,t]$ for all k. Tables of coefficients $v_{\mu\lambda}^{(k)}(t)$ and $K_{\mu\lambda}^{(k)}(q,t)$ are included in section 5 to illustrate these conjectures. Our examples suggest an even stronger property,

$$0 \subseteq K_{\mu\lambda}^{(k)}(q,t) \subseteq K_{\mu\lambda}(q,t) \tag{8}$$

where for two polynomials $P,Q\in\mathbb{Z}[t]$, $P\subseteq Q$ means $Q-P\in\mathbb{N}[q,t]$.

More generally, it develops that properties of the k-Schur functions, with a number of conjectures, provide a k-generalization for the properties that make the Schur functions important to the theory of symmetric functions. In particular, we prove that the $s_\lambda^{(k)}[X;t]$ form a basis for $\Lambda_t^{(k)}$ and that the k-Schur functions of $\Lambda_t^{(\infty)} = \Lambda$ are indeed the Schur functions themselves. Conjectural evidence for the significance of the k-Schur functions includes a k-analog of partition conjugation and generalizations of Pieri and Littlewood-Richardson rules. Consequently, in the case of the multiplicative action of $h_1[X]$, a k-analog of the Young Lattice is induced. Further, we have observed that the k-Schur functions, expanded in terms of k-Schur functions in two sets of variables, have coefficients in $\mathbb{N}[t]$. This is a special property of Schur functions that is not shared by the Hall-Littlewood or Macdonald functions. Finally, the k-Schur functions of $\Lambda_t^{(k)}$, when embedded in $\Lambda_t^{(k')}$ for $k' > k$, seem to decompose positively in terms of k'-Schur functions:

$$s_\lambda^{(k)}[X;t] = s_\lambda^{(k')}[X;t] + \sum_{\mu > \lambda} v_{\mu\lambda}^{(k\to k')}(t)\, s_\mu^{(k')}[X;t]\,, \quad \text{where } v_{\mu\lambda}^{(k\to k')}(t) \in \mathbb{N}[t]\,. \tag{9}$$

Remarkably, not all of the k-Schur functions need to be constructed using (5). For each k, there is a subset of $s_\lambda^{(k)}[X;t]$, called the irreducible k-Schur functions, from which all other k-Schur functions may be constructed [7] by simply applying a succession of certain operators. The elements of this set are the k-Schur functions indexed by partitions with no more than i parts equal to $k-i$, and the operators are vertex operators [14] associated to rectangularly shaped partitions $(\ell^{k+1-\ell})$ for $\ell = 1, \ldots, k$. That is,

$$s_\lambda^{(k)}[X;t] = t^c\, B_{R_1} B_{R_2} \ldots B_{R_\ell}\, s_\mu^{(k)}[X;t] \quad \text{with} \quad c \in \mathbb{N}\,, \tag{10}$$

for an irreducible $s_\mu^{(k)}$, and vertex operators B_R, where R is a partition of rectangular shape.

Since the Hall-Littlewood polynomials at $t = 1$ are the complete symmetric functions

$$H_\lambda[X;1] = h_{\lambda_1}[X]\, h_{\lambda_2}[X] \cdots h_{\lambda_\ell}[X]\,, \tag{11}$$

we see that $\Lambda_t^{(k)}$ reduces to the polynomial ring $\Lambda^{(k)} = \mathbb{Q}[h_1, \ldots, h_k]$. Each of the properties held by the k-Schur functions has a specialization in this subring. In particular, since B_R is simply multiplication by the Schur function s_R when $t = 1$, relation (10) reduces to

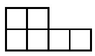
$$s_\lambda^{(k)}[X] = s_{R_1}[X] s_{R_2}[X] \ldots s_{R_\ell}[X] s_\mu^{(k)}[X]\,. \tag{12}$$

The irreducible k-Schur functions thus constitute a natural basis for the quotient ring $\Lambda^{(k)}/\mathcal{I}_k$, where $\mathcal{I}_k$ is the ideal generated by Schur functions indexed by partitions of the form $(\ell^{k+1-\ell})$.

2. Definitions

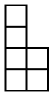
2.1. **Partitions.** Symmetric polynomials are indexed by partitions, sequences of non-negative integers $\lambda = (\lambda_1, \lambda_2, \ldots)$ with $\lambda_1 \geq \lambda_2 \geq \ldots$. The number of non-zero parts in λ is denoted $\ell(\lambda)$ and the degree of λ is $|\lambda| = \lambda_1 + \cdots + \lambda_{\ell(\lambda)}$. We use the dominance order on partitions with $|\lambda| = |\mu|$, where $\lambda \leq \mu$ when $\lambda_1 + \cdots + \lambda_i \leq \mu_1 + \cdots + \mu_i$ for all i.

Any partition λ has an associated Ferrers diagram with λ_i lattice squares in the i^{th} row, from the bottom to top. For example,

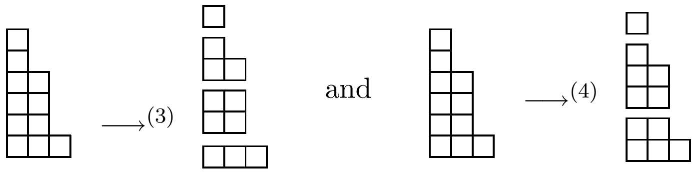
(13) $\lambda = (4,2) =$ .

For each cell $s = (i,j)$ in the diagram of λ, let $\ell'(s), \ell(s), a(s)$ and $a'(s)$ be respectively the number of cells in the diagram of λ to the south, north, east and west of the cell s. The hook-length of any cell in λ, is $h_s(\lambda) = \ell(s) + a(s) + 1$. In the example, $h_{(1,2)}(4,2) = 2+1+1$. The *main hook-length* of λ, $h_M(\lambda)$, is the hook-length of the cell $s = (1,1)$ in the diagram of

λ. Therefore, $h_M\big((4,2)\big) = 5$. The conjugate λ' of a partition λ is defined by the reflection of the Ferrers diagram about the main diagonal. For example, the conjugate of (4,2) is

(14) $\lambda' =$  $= (2,2,1,1)\,.$

A partition λ is said to be *k-bounded* if its first part is not larger than k, i.e, if $\lambda_1 \le k$. We associate to any k-bounded partition λ a sequence of partitions, $\lambda^{\to k}$, called the *k-split* of λ. $\lambda^{\to k} = (\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(r)})$ is obtained by partitioning λ (without rearranging the entries) into partitions $\lambda^{(i)}$ where $h_M(\lambda^{(i)}) = k$, for all $i < r$. For example, $(3,2,2,2,1,1)^{\to 3} = \big((3),(2,2),(2,1),(1)\big)$ and $(3,2,2,2,1,1)^{\to 4} = \big((3,2),(2,2,1),(1)\big)$. Equivalently, the diagram of λ is cut horizontally into partitions with main hook-length k.

(15)


The last partition in the sequence $\lambda^{\to k}$ may have main hook- length less than k.

2.2. **Symmetric functions.** The power sum $p_i(x_1, x_2, \ldots)$ is

$$p_i(x_1, x_2, \ldots) = x_1^i + x_2^i + \cdots\,, \tag{16}$$

and for a partition $\lambda = (\lambda_1, \lambda_2, \ldots)$,

$$p_\lambda(x_1, x_2, \ldots) = p_{\lambda_1}(x_1, x_2, \ldots)\, p_{\lambda_2}(x_1, x_2, \ldots) \cdots\,. \tag{17}$$

We employ the notation of λ-rings, needing only the formal ring of symmetric functions Λ to act on the ring of rational functions in $x_1, \ldots, x_N, q, t$, with coefficients in $\mathbb{R}$. The action of a power sum p_i on a rational function is, by definition,

$$p_i\left[\frac{\sum_\alpha c_\alpha u_\alpha}{\sum_\beta d_\beta v_\beta}\right] = \frac{\sum_\alpha c_\alpha u_\alpha^i}{\sum_\beta d_\beta v_\beta^i}, \tag{18}$$

with $c_\alpha, d_\beta \in \mathbb{R}$ and u_α, v_β monomials in $x_1, \ldots, x_N, q, t$. Since the power sums form a basis of the ring Λ, any symmetric function has a unique expression in terms of power sums, and (18) extends to an action of Λ on rational functions. In particular $f[X]$, the action of a symmetric function f on the monomial $X = x_1 + \cdots + x_N$, is simply $f(x_1, \ldots, x_N)$. In the remainder of the article, we will always consider the number of variables N to be infinite, unless otherwise specified.

The monomial symmetric function $m_\lambda[X_n]$ is

$$m_\lambda[X_n] = \sum_{\sigma \in S_n;\, \sigma(\lambda) \text{ distinct}} x^{\sigma(\lambda)}\,. \tag{19}$$

The complete symmetric function $h_r[X]$ is

$$h_r[X] = \sum_{1 \le i_1 \le i_2 \le \cdots \le i_r} x_{i_1} x_{i_2} \cdots x_{i_r}\,, \tag{20}$$

and $h_\lambda[X]$ stands for the homogeneous symmetric function

$$h_\lambda[X] = h_{\lambda_1}[X]\, h_{\lambda_2}[X] \cdots\,. \tag{21}$$

Although the Schur functions may be characterized in many ways, here it will be convenient to use the Jacobi-Trudi determinantal expression:

$$s_\lambda[X] = \det\Big|h_{\lambda_i+j-1}[X]\Big|_{1\le i,j\le \ell(\lambda)} \tag{22}$$

where $h_r[X]=0$ if $r<0$. Note, in particular, $s_r[X]=h_r[X]$.

We recall that the Macdonald scalar product, $\langle\ ,\ \rangle_{q,t}$, on $\Lambda\otimes\mathbb{Q}(q,t)$ is defined by setting

$$\langle p_\lambda[X], p_\mu[X]\rangle_{q,t} = \delta_{\lambda\mu}\, z_\lambda \prod_{i=1}^{\ell(\lambda)} \frac{1-q^{\lambda_i}}{1-t^{\lambda_i}}\,, \tag{23}$$

where for a partition λ with $m_i(\lambda)$ parts equal to i, we associate the number

$$z_\lambda = 1^{m_1} m_1!\, 2^{m_2} m_2! \cdots \tag{24}$$

If $q=t$, this expression no longer depends on a parameter and is then denoted $\langle , \rangle$, satisfying

$$\langle s_\lambda[X], s_\mu[X]\rangle = \delta_{\lambda\mu}\,. \tag{25}$$

The Macdonald integral forms $J_\lambda[X;q,t]$ are uniquely characterized [10] by

$$\text{(i) } \langle J_\lambda, J_\mu\rangle_{q,t} = 0, \qquad \text{if } \lambda\neq\mu, \tag{26}$$

$$\text{(ii) } J_\lambda[X;q,t] = \sum_{\mu\le\lambda} v_{\lambda\ mu}(q,t) s_\mu[X] \quad \text{with} \quad v_{\lambda\mu}(q,t)\in\mathbb{Q}(q,t)\,, \tag{27}$$

$$\text{(iii) } v_{\lambda\lambda}(q,t) = \prod_{s\in\lambda}(1-q^{a(s)}t^{\ell(s)+1}), \tag{28}$$

Here, we use a modification of the Macdonald integral forms that is obtained by setting

$$H_\lambda[X;q,t] = J_\lambda[X/(1-t);q,t] = \sum_\mu K_{\mu\lambda}(q,t)\, s_\mu[X]\,, \tag{29}$$

with the coefficients $K_{\mu\lambda}(q,t)\in\mathbb{N}[q,t]$ known as the q,t-Kostka polynomials. When $q=0$, $J_\lambda[X;q,t]$ reduces to the Hall-Littlewood polynomial, $J_\lambda[X;0,t]=Q_\lambda[X;t]$. Again, we use a modification;

$$H_\lambda[X;t] = H_\lambda[X;0,t] = Q_\lambda[X/(1-t);t] = s_\lambda[X] + \sum_{\mu>\lambda} K_{\mu\lambda}(t)\, s_\mu[X]\,, \tag{30}$$

with the coefficients $K_{\mu\lambda}(t)\in\mathbb{N}[t]$ known as the Kostka-Foulkes polynomials.

3. Results

The ring of symmetric polynomials over rational functions in an extra parameter t has proven to be of interest in many fields of mathematics and physics. One natural basis of this space is given by the Hall-Littlewood polynomials, $H_\lambda[X;t]$, which provide t-analogs of the homogeneous symmetric functions $h_\lambda[X]$. Our approach employs vertex operators that arise in the recursive construction for the Hall-Littlewood polynomials [4]. These operators can be defined [14] for $\ell\in\mathbb{Z}$, by

$$B_\ell = \sum_{i=0}^{\infty} s_{i+\ell}[X]\, s_i[X(t-1)]^\perp\,, \tag{31}$$

where for f,g and h arbitrary symmetric functions, $f^\perp$ is such that on the scalar product (25),

$$\langle f^\perp g, h\rangle = \langle g, fh\rangle\,. \tag{32}$$

The operators add an entry to the Hall-Littlewood polynomials, that is,

$$H_\lambda[X;t] = B_{\lambda_1} H_{\lambda_2,\ldots,\lambda_\ell}[X;t]\,, \qquad \text{for} \qquad \lambda_1 \geq \lambda_2\,. \tag{33}$$

We study the subspaces given by

$$\Lambda_t^{(k)} = \mathcal{L}\left\{H_\lambda[X;t]\right\}_{\lambda;\lambda_1\leq k}\,. \tag{34}$$

It is clear that $\Lambda_t^{(1)} \subseteq \Lambda_t^{(2)} \subseteq \cdots \subseteq \Lambda_t^{(\infty)} = \Lambda$ and thus that these subspaces provide a filtration for Λ. Note that $\Lambda_t^{(k)}$ can be equivalently defined as

$$\Lambda_t^{(k)} = \mathcal{L}\left\{H_\lambda[X;q,t]\right\}_{\lambda;\lambda_1\leq k}\,. \tag{35}$$

We seek elements that play the role in $\Lambda_t^{(k)}$ that the Schur functions play in Λ. In particular, since $\Lambda_t^{(\infty)} = \Lambda$, we want a basis for $\Lambda_t^{(k)}$ that reduces to the Schur functions when k is large.

3.1. **k-split polynomials.** Important in our work to find such functions is a family of polynomials, studied in many recent papers such as [9, 11, 12, 13, 14], that give a t-analog of the product of Schur functions. These functions, indexed by a sequence of partitions, can be built recursively using vertex operators [14]. For a partition λ of length m, define

$$B_\lambda \equiv \prod_{1\leq i<j\leq m} (1 - t e_{ij}) B_{\lambda_1}\cdots B_{\lambda_m}\,, \tag{36}$$

where e_{ij} acts by

$$e_{ij}\left(B_{\lambda_1}\cdots B_{\lambda_m}\right) = B_{\lambda_1}\cdots B_{\lambda_i+1}\cdots B_{\lambda_j-1}\cdots B_{\lambda_m}\,. \tag{37}$$

For any sequence of partitions $(\lambda^{(1)},\lambda^{(2)},\ldots)$, the generalized Schur function product can then be defined recursively by

$$\mathcal{H}_{(\lambda^{(1)},\lambda^{(2)},\lambda^{(3)},\ldots)}[X;t] = B_{\lambda^{(1)}}\mathcal{H}_{(\lambda^{(2)},\lambda^{(3)},\ldots)}[X;t]\,, \tag{38}$$

starting with $\mathcal{H}_{()} = 1$. Note that since $B_\lambda \cdot 1 = s_\lambda[X]$, we have that

$$\mathcal{H}_{(\lambda)}[X;t] = s_\lambda[X]\,. \tag{39}$$

For our purposes, we consider only the $\mathcal{H}_S[X;t]$ indexed by a dominant sequence S. That is, sequences of partitions $S = (\lambda^{(1)},\lambda^{(2)},\ldots)$ such that the concatenation of $\lambda^{(1)},\lambda^{(2)},\ldots$, denoted $\bar{S}$, forms a partition. In this case, we prove that $\mathcal{H}_S[X;t]$ obeys important unitriangular relations.

Property 1. [6] *If S is dominant, with $\bar{S} = \lambda$, then*

$$\mathcal{H}_S[X;t] = s_\lambda[X] + \sum_{\mu>\lambda} K_{\mu;S}(t) s_\mu[X]\,, \quad \textit{where} \quad K_{\mu;S}(t) \in \mathbb{Z}[t]\,, \tag{40}$$

$$\mathcal{H}_S[X;t] = H_\lambda[X] + \sum_{\mu>\lambda} C_{\mu;S}(t) H_\mu[X]\,, \quad \textit{where} \quad C_{\mu;S}(t) \in \mathbb{Z}[t]\,. \tag{41}$$

Furthermore, we discovered that a particular subset of the $\mathcal{H}_S$ not only form a basis for $\Lambda_t^{(k)}$, but are essential in the construction of the k-Schur functions. This subset consists only of the elements that are indexed by the *k-split* of a k-bounded partition. More precisely,

Definition 1. *The k-split polynomials are defined, for a k-bounded partition λ, by*

$$G_\lambda^{(k)}[X;t] = \mathcal{H}_S[X;t] \quad \textit{where} \quad S = \lambda^{\to k} \quad \textit{is the } k\textit{-split of } \lambda\,. \tag{42}$$

We give a number of properties for the k-split polynomials, starting with the fact that the $G_\lambda^{(k)}[X;t]$ actually lie in the space $\Lambda_t^{(k)}$.

Property 2. [6] *For any k-bounded partition λ, we have that*

$$G_\lambda^{(k)}[X;t] \in \Lambda_t^{(k)}\,. \tag{43}$$

It also happens that the k-split polynomials are triangularly related to the Hall-Littlewood polynomials.

Property 3. [6] *We have*

$$G_\lambda^{(k)}[X;t] = H_\lambda[X;t] + \sum_{\mu>\lambda;\mu_1\leq k} g_{\lambda\mu}^{(k)}(t)\, H_\mu[X;t]\,, \quad \textit{where} \quad g_{\lambda\mu}^{(k)}(t) \in \mathbb{Z}[t]\,. \tag{44}$$

In fact, k-split polynomials reduce to usual Schur functions when k is large enough.

Property 4. [6] *Let λ be such that $h_M(\lambda) \leq k$. Then,*

$$G_\lambda^{(k)}[X;t] = s_\lambda[X]\,. \tag{45}$$

Given these properties, we are able to prove that the k-split polynomials form a basis for the k-subspaces.

Theorem 1. [6] *The k-split polynomials form a basis of $\Lambda_t^{(k)}$.*

3.2. **k-Schur functions.** Although the k-split polynomials form a basis for $\Lambda_t^{(k)}$, they do not play the fundamental role that the Schur functions do for Λ. That is, the positivity properties stated in Section 4 do not hold for the k-split polynomials. However, these polynomials are needed in the construction of our Schur analog, $s_\lambda^{(k)}[X;t]$. Using a projection operator, $T_j^{(k)}$, for $j \leq k$, that acts linearly on $\Lambda_t^{(k)}$ by

$$T_j^{(k)}\, G_\lambda^{(k)}[X;t] = \begin{cases} G_\lambda^{(k)}[X;t] & \text{if } \lambda_1 = j \\ 0 & \text{otherwise} \end{cases}\,, \tag{46}$$

we can define our functions that play an important role in the refinement of symmetric function theory.

Definition 2. *For k-bounded partition λ, the k-Schur functions are recursively defined*

$$s_\lambda^{(k)}[X;t] = T_{\lambda_1}^{(k)} B_{\lambda_1} s_{(\lambda_1,\lambda_2,\dots)}^{(k)}[X;t]\,, \quad \textit{where} \quad s_{()}^{(k)}[X;t] = 1\,. \tag{47}$$

Tables of k-Schur functions in terms of Schur functions can be found in Section 5.1.

We prove several properties satisfied by the k-Schur functions, including that they are triangularly related to the k-split polynomials and that they form a basis for the k-subspaces.

Property 5. [6] *For λ a k-bounded partition, we have*

$$s_\lambda^{(k)}[X;t] = G_\lambda^{(k)}[X;t] + \sum_{\substack{\mu>\lambda\\ \mu_1=\lambda_1}} u_{\mu\lambda}^{(k)}(t)\, G_\mu^{(k)}[X;t]\,, \quad \textit{where} \quad u_{\mu\lambda}^{(k)}(t) \in \mathbb{Z}[t]\,. \tag{48}$$

Theorem 2. [6] *The k-Schur functions form a basis of $\Lambda_t^{(k)}$.*

We can thus refine the expansion of Hall-Littlewood polynomials in terms of Schur functions (30). That is, for any k-bounded partition λ,

$$H_\lambda[X;t] = s_\lambda^{(k)}[X;t] + \sum_{\mu>\lambda:\mu_1\le k} K_{\mu\lambda}^{(k)}(t)\, s_\mu^{(k)}[X;t]\,, \quad \text{where} \quad K_{\mu\lambda}^{(k)}(t) \in \mathbb{Z}[t]\,. \tag{49}$$

The integrality of $K_{\mu\lambda}^{(k)}(t)$ follows from the unitriangularity and integrality in Properties 3 and 5. Moreover, by our triangularity and integrality properties and (30), we also have integrality of the coefficients in Conjecture 6:

Property 6. [6] *For any k-bounded partition λ,*

$$s_\lambda^{(k)}[X;t] = s_\lambda[X] + \sum_{\mu>\lambda} v_{\mu\lambda}^{(k)}(t)\, s_\mu[X]\,, \quad \textit{where} \quad v_{\mu\lambda}^{(k)}(t) \in \mathbb{Z}[t]\,. \tag{50}$$

This unitriangularity property, given that the coefficients in the Schur function expansion of the Macdonald polynomials are polynomials in q and t with integral coefficients, implies

Property 7. [6] *For any k-bounded partition λ,*

$$H_\lambda[X;q,t] = \sum_{\mu:\mu_1\le k} K_{\mu\lambda}^{(k)}(q,t)\, s_\mu^{(k)}[X;t]\,, \quad \textit{where} \quad K_{\mu\lambda}^{(k)}(q,t) \in \mathbb{Z}[q,t]\,. \tag{51}$$

Now, to further support the idea that the $s_\lambda^{(k)}[X;t]$ provide a refinement for Schur function theory, we show that they reduce to the usual $s_\lambda[X]$ when $k\to\infty$.

Property 8. [6] *For any k-bounded partition λ,*

$$s_\lambda^{(k)}[X;t] = s_\lambda[X] \quad \textit{if} \quad h_M(\lambda) \le k\,. \tag{52}$$

3.3. **Irreducibility.** Further results with the k-Schur functions naturally impose an irreducible structure on the space $\Lambda_t^{(k)}$. We prove that the action of such an operator on a k-Schur function produces only one k-Schur function. Namely, for $\ell = 1,2,\ldots,k$,

$$B_{\ell^{k+1-\ell}}\, s_\lambda^{(k)}[X;t] = t^d s_\mu^{(k)}[X;t]\,, \tag{53}$$

where μ is the partition rearrangement of the entries in $(\ell^{k+1-\ell})$ and λ, and t^d is a positive power of t given explicitly in [7]. This result has the important consequence of simplifying the construction of the k-Schur functions. In effect, for each k, there is a subset of $k!$ k-Schur functions called the irreducible k-Schur functions, from which all other $s_\lambda^{(k)}[X;t]$ may be constructed by successive application of operators indexed by rectangular partitions. That is,

Property 9. [7] *For any k-bounded partition λ,*

$$s_\lambda^{(k)}[X;t] = t^c B_{R_1}\cdots B_{R_j}\, s_\mu^{(k)}[X;t] \qquad c\in\mathbb{N}\,, \tag{54}$$

where $s_\mu^{(k)}[X;t]$ is an irreducible k-Schur function and $R_1,\ldots,R_j$ are rectangular partitions.

Since the Hall-Littlewood polynomials at $t=1$ are the homogeneous symmetric functions, $h_\lambda[X]$, $\Lambda_t^{(k)}$ reduces to the polynomial ring $\Lambda^{(k)} = \mathbb{Q}[h_1,\ldots,h_k]$. Since B_R is simply multiplication by the Schur function s_R when $t=1$, relation (54) reduces to

$$s_\lambda^{(k)}[X] = s_{R_1}[X]s_{R_2}[X]\ldots s_{R_\ell}[X]s_\mu^{(k)}[X]\,. \tag{55}$$

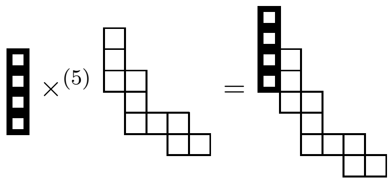
It follows that the irreducible k-Schur functions thus constitute a natural basis for the quotient ring $\Lambda_t^{(k)}/\mathcal{I}_k$, where $\mathcal{I}_k$ is the ideal generated by Schur functions indexed by rectangular shapes of the type $(\ell^{k+1-\ell})$.

4. Positivity conjectures

Computer experimentation reveals that many of the properties making the Schur function basis so important are generalized by the k-Schur functions. We now state several of these properties.

4.1. **The k-conjugation of a partition.** We give a generalization of partition conjugation that is an involution on k-bounded partitions, and reduces to usual conjugation of partitions for large k.

A skew diagram D has hook-lengths bounded by k if the hook-length of any cell in D is not larger than k. For a positive integer $m \leq k$, the k-multiplication $m \times^{(k)} D$ is the skew diagram $\overline{D}$ obtained by prepending a column of length m to D such that the number of rows of $\overline{D}$ is as small as possible while ensuring that its hook-lengths are bounded by k. For example,

(56)  .

Definition 3. *The k-conjugate of a k-bounded partition $\lambda = (\lambda_1, \ldots, \lambda_n)$, denoted λ^{ω_k}, is the vector obtained by reading the number of boxes in each row of the skew diagram,*

$$D = \lambda_1 \times^{(k)} \cdots \times^{(k)} \lambda_n \,, \tag{57}$$

arising by k-multiplying the entries of λ from right to left.

When $k \to \infty$, $\lambda^{\omega_k} = \lambda' = D$ since each k-multiplication step reduces to concatenating a column of height λ_i. Further, the k-conjugate is an involution on k-bounded partitions:

Theorem 3. [5] *ω_k is an involution on partitions bounded by k. That is, for λ with $\lambda_1 \leq k$,*

$$(\lambda^{\omega_k})^{\omega_k} = \lambda \,. \tag{58}$$

We have observed that the k-conjugation of a partition plays a natural role in the generalization of classical Schur function properties. We now give two examples.

4.2. **The involution ω.** It is known that the involution in (25) preserves the space Λ and acts on a Schur function by

$$\omega s_\lambda[X] = s_{\lambda'}[X] \,. \tag{59}$$

We prove that a simple generalization for this involution,

$$\omega_t f = (\omega f)\big|_{t \to 1/t} \quad \text{for} \quad f \in \Lambda \tag{60}$$

preserves the space $\Lambda_t^{(k)}$. Note also that ω_t is an involution since ω is an involution. This given, many examples support the following natural generalization of (59):

Conjecture 1. [6] *For any k-bounded partition λ,*

$$\omega_t \, s_\lambda^{(k)}[X;t] = t^{-c(\lambda)} s_{\lambda^{\omega_k}}^{(k)}[X;t] \,, \tag{61}$$

where $c(\lambda)$ is some nonnegative integer.

The conjecture holds when $h_M(\lambda) \leq k$ since then, $\lambda^{\omega_k} = \lambda'$ and by Property 8, $s_\lambda^{(k)}[X;t] = s_\lambda[X]$.

4.3. **Pieri Rules.** Beautiful combinatorial algorithms are known for the Littlewood-Richardson coefficients that appear in a product of Schur functions;

$$s_\lambda[X]\, s_\mu[X] = \sum_\nu c_{\lambda\mu}^{\nu}\, s_\nu[X] \quad \text{where} \quad c_{\lambda\mu}^{\nu} \in \mathbb{N}\,. \tag{62}$$

Since $\Lambda^{(k)} \equiv \Lambda_{t=1}^{(k)}$ is a ring, and $s_\lambda^{(k)}$ forms a basis for this space, a similar expression holds for the product of two k-Schur functions. That is, for k-bounded partitions λ and μ,

$$s_\lambda^{(k)}[X]\, s_\mu^{(k)}[X] = \sum_\nu c_{\lambda\mu}^{\nu\,(k)} s_\nu^{(k)}[X] \quad \text{where} \quad c_{\lambda\mu}^{\nu\,(k)} \in \mathbb{Z}\,, \tag{63}$$

Further, Property 8 says that the k-Schur functions are simply the Schur functions when k is large, and therefore $c_{\lambda\mu}^{\nu\,(k)} = c_{\lambda\mu}^{\nu}$ for $k \geq |\nu|$. In fact, we believe the coefficients are nonnegative for all k.

Conjecture 2. [6] *For all k-bounded partitions λ, μ, ν, we have $0 \leq c_{\lambda\mu}^{\nu\,(k)} \leq c_{\lambda\mu}^{\nu}$.*

In particular, (63) reduces to a k-generalization of the Pieri rule when λ is a row (resp. column) of length $\ell \leq k$ since $s_\lambda^{(k)}[X]$ reduces to $h_\ell[X]$ (resp. $e_\ell[X]$). That is, for $\ell \leq k$,

$$h_\ell[X]\, s_\lambda^{(k)}[X] = \sum_{\mu \in E_{\lambda,\ell}^{(k)}} s_\mu^{(k)}[X] \quad \text{and} \quad e_\ell[X]\, s_\lambda^{(k)}[X] = \sum_{\mu \in \bar{E}_{\lambda,\ell}^{(k)}} s_\mu^{(k)}[X]\,, \tag{64}$$

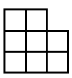
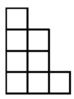
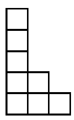
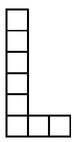
for some sets of partitions $E_{\lambda,\ell}^{(k)}$ and $\bar{E}_{\lambda,\ell}^{(k)}$, which we believe naturally extend the Pieri rules by:

Conjecture 3. [6] *For any positive integer $\ell \leq k$,*

$$\begin{aligned} E_{\lambda,\ell}^{(k)} &= \{\mu \,|\, \mu/\lambda \textit{ is a horizontal } \ell\textit{-strip and } \mu^{\omega_k}/\lambda^{\omega_k} \textit{ is a vertical } \ell\textit{-strip}\}\,, \\ \bar{E}_{\lambda,\ell}^{(k)} &= \{\mu \,|\, \mu/\lambda \textit{ is a vertical } \ell\textit{-strip and } \mu^{\omega_k}/\lambda^{\omega_k} \textit{ is a horizontal } \ell\textit{-strip}\}\,. \end{aligned} \tag{65}$$

For example, to obtain the indices of elements occurring in $e_2\, s_{3,2,1}^{(4)}$, we find $(3,2,1)^{\omega_4} = (2,2,1,1)$ by definition. Adding a horizontal 2-strip to (2,2,1,1) in all ways, we obtain (2,2,2,1,1),(3,2,1,1,1),(3,2,2,1) and (4,2,1,1) of which all are 4-bounded. Our set then consists of all the 4-conjugates of these partitions that leave a vertical 2-strip when $(3,2,1)$ is extracted from them. The 4-conjugates are

(66)

$(2,2,2,1,1)^{\omega_4} =$ , $(3,2,1,1,1)^{\omega_4} =$ , $(3,2,2,1)^{\omega_4} =$ , $(4,2,1,1)^{\omega_4} =$ ,

and of these, the first three are such that a vertical 2-strip remains when $(3,2,1)$ is extracted. Thus

$$e_2[X]\, s_{3,2,1}^{(4)}[X] \;=\; s_{3,3,2}^{(4)}[X] + s_{3,2,2,1}^{(4)}[X] + s_{3,2,1,1,1}^{(4)}[X]\,. \tag{67}$$

4.4. **Coproduct expansion.** Another unique property of the Schur functions is the expansion,

$$s_\lambda[X+Y] = \sum_{|\mu|+|\rho|=|\lambda|} c^\lambda_{\mu\rho}\, s_\mu[X]\, s_\rho[Y] \quad \text{where} \quad c^\lambda_{\mu\rho} \in \mathbb{N}\,. \tag{68}$$

We have found by experimentation that the k-Schur functions also satisfy a similar relation,

Conjecture 4. [6] *For any k-bounded partition,*

$$s^{(k)}_\lambda[X+Y;t] = \sum_{|\mu|+|\rho|=|\lambda|} g^\lambda_{\mu\rho}(t)\, s^{(k)}_\mu[X;t]\, s^{(k)}_\rho[Y;t] \quad \textit{where} \quad g^\lambda_{\mu\rho} \in \mathbb{N}[t]\,. \tag{69}$$

5. Tables

In the tables below, we have not included the cases when $k \geq |\lambda|$, which, from Property 8, simply correspond to the trivial cases $s_\lambda[X;t] = s_\lambda[X]$.

5.1. k-Schur functions in terms of Schur functions.

$k=2$	1^3	21	3
1^3	1	t	
21		1	t

$k=2$	1^4	21^2	2^2	31	4
1^4	1	t	t^2		
21^2		1		t	
2^2			1	t	t^2

$k=2$	1^5	21^3	2^21	31^2	32	41	5
1^5	1	$t+t^2$	t^2+t^3	t^3	t^4		
21^3		1	t	$t+t^2$	t^2	t^3	
2^21			1	t	$t+t^2$	t^2+t^3	t^4

$k=3$	1^4	21^2	2^2	31	4
1^4	1	t			
21^2		1		t	
2^2			1		
31				1	t

$k=3$	1^5	21^3	2^21	31^2	32	41	5
1^5	1	t	t^2				
21^3		1		t			
2^21			1		t		
31^2				1		t	
32					1	t	t^2

$k=3$	1^6	21^4	2^21^2	2^3	31^3	321	3^2	41^2	42	51	6
1^6	1	t	t^2	t^3							
21^4		1	t		t	t^2					
2^21^2			1			t	t^2				
2^3				1		t			t^2		
31^3					1			t			
321						1		t	t	t^2	
3^3							1		t	t^2	t^3

5.2. Macdonald polynomials in terms of k-Schur functions.

$k=2$	1^3	21
1^3	1	t^2
21	q	1

$k=2$	1^4	21^2	2^2
1^4	1	t^2+t^3	t^4
21^2	q	$1+qt^2$	t
2^2	q^2	$q+qt$	1

$k=2$	1^5	21^3	2^21
1^5	1	t^3+t^4	t^6
21^3	q	$1+qt^3$	t^2
2^21	q^2	$q+qt$	1

$k=2$	1^6	21^4	2^21^2	2^3
1^6	1	$t^3+t^4+t^5$	$t^6+t^7+t^8$	t^9
21^4	q	$1+qt^3+qt^4$	$t^2+t^3+qt^6$	t^4
2^21^2	q^2	$q+qt+q^2t^3$	$1+qt^2+qt^3$	t
2^3	q^3	$q^2+q^2t+q^2t^2$	$q+qt+qt^2$	1

$k=3$	1^4	21^2	2^2	31
1^4	1	t^2+t^3	t^2+t^4	t^5
21^2	q	$1+qt^2$	$t+qt^2$	t^2
2^2	q^2	$q+qt$	$1+q^2t^2$	t
31	q^3	$q+q^2$	$q+q^2t$	1

$k=3$	1^5	21^3	2^21	31^2	32
1^5	1	$t^2+t^3+t^4$	$t^3+t^4+t^5+t^6$	$t^5+t^6+t^7$	t^8
21^3	q	$1+qt^2+qt^3$	$t+t^2+qt^3+qt^4$	$t^2+t^3+qt^5$	t^4
2^21	q^2	$q+qt+q^2t^2$	$1+qt+qt^2+q^2t^3$	$t+qt^2+qt^3$	t^2
31^2	q^3	$q+q^2+q^3t^2$	$q+qt+q^2t+q^2t^2$	$1+qt^2+q^2t^2$	t
32	q^4	$q^2+q^3+q^3t$	$q+q^2+q^2t+q^3t$	$q+qt+q^2t$	1

$k=4$	1^5	21^3	2^21	31^2	32	41
1^5	1	$t^2+t^3+t^4$	$t^2+t^3+t^4+t^5+t^6$	$t^5+t^6+t^7$	$t^4+t^5+t^6+t^7+t^8$	t^9
21^3	q	$1+qt^2+qt^3$	$t+t^2+qt^2+qt^3+qt^4$	$t^2+t^3+qt^5$	$t^2+t^3+t^4+qt^4+qt^5$	t^5
2^21	q^2	$q+qt+q^2t^2$	$1+qt+qt^2+q^2t^2+q^2t^3$	$t+qt^2+qt^3$	$t+t^2+qt^2+qt^3+q^2t^4$	t^3
31^2	q^3	$q+q^2+q^3t^2$	$q+qt+q^2t+q^2t^2+q^3t^2$	$1+qt^2+q^2t^2$	$t+qt+qt^2+q^2t^2+q^2t^3$	t^2
32	q^4	$q^2+q^3+q^3t$	$q+q^2+q^2t+q^3t+q^4t^2$	$q+qt+q^2t$	$1+qt+q^2t+q^2t^2+q^3t^2$	t
41	q^6	$q^3+q^4+q^5$	$q^2+q^3+q^4+q^4t+q^5t$	$q+q^2+q^3$	$q+q^2+q^2t+q^3t+q^4t$	1

References

[1] A. M. Garsia and M. Haiman, *A graded representation module for Macdonald's polynomials*, Proc. Natl. Acad. Sci. USA **90** (1993) 3607-3610.

[2] A. M. Garsia and C. Procesi, *On certain graded S_n-modules and the q-Kostka polynomials*, Adv. Math. **87** (1992), 82–138.

[3] M. Haiman, *Hilbert schemes, polygraphs, and the Macdonald positivity conjecture*, J Am Math Soc **14** (2001), 941–1006.

[4] N. Jing, *Vertex operators and Hall-Littlewood symmetric functions*, Adv. Math. **87** (1991), 226–248.

[5] L. Lapointe, A. Lascoux and J. Morse, *A filtration of the symmetric function space and a refinement of the Macdonald positivity conjecture*, math.CO/9812001.

[6] L. Lapointe and J. Morse, *Schur function analogs for a filtration of the symmetric function space*, http://www.math.upenn.edu/˜morsej/new_research.html.

[7] L. Lapointe and J. Morse, *Schur function identities, their t-analogs, and k-Schur irreducibility*, http://www.math.upenn.edu/˜morsej/new_research.html.

[8] L. Lapointe and J. Morse, *Tableaux statistics for two part Macdonald polynomials*, math.CO/9812001.

[9] A. Lascoux, B. Leclerc and J.-Y. Thibon, *Ribbon tableaux, Hall-Littlewood functions, quantum affine algebras and unipotent varieties*, J. Math. Phys. **38** (1997), 1041–1068.

[10] I. G. Macdonald, Symmetric Functions and Hall Polynomials, 2nd edition, Clarendon Press, Oxford, 1995.

[11] A. Schilling and S. Warnaar, *Inhomogeneous lattice paths, generalized Kostka-Foulkes polynomials, and A_{n-1}-supernomials*, Comm. Math. Phys. **202** (1999), 359–401.

[12] M. Shimozono, *A cyclage poset structure for Littlewood-Richardson tableaux*, European J Combin **22** (2001), 365–393.

[13] M. Shimozono and J. Weyman, *Graded characters of modules supported in the closure of a nilpotent conjugacy class*, European J. Combin. **21** (2000), 257–288.

[14] M. Shimozono and M. Zabrocki, *Hall-Littlewood vertex operators and generalized Kostka polynomials*, Adv. Math. **158** (2001), 66–85.

[15] S. Veigneau, *ACE Version 3.0* (1998) http://phalanstere.univ-mlv.fr/˜ace/.

[16] M.A. Zabrocki, *A Macdonald vertex operator and standard tableaux statistics for the two-column (q,t)-Kostka coefficients*, Electron. J. Combinat. **5**, R45 (1998).

Department of Mathematics, McGill University, Montreal, Canada
E-mail address: lapointe@math.mcgill.ca

Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104, USA
E-mail address: morsej@math.upenn.edu

GENERALIZED PSEUDO-PERMUTATIONS

LE MINH HA AND PHAN HA DUONG

Abstract. We introduce and study combinatorial properties of pseudo-permutations with multiple occurences. We prove that the set of all such generalized pseudo-permutations of a given type have a lattice structure and derive a recursive formula to compute its cardinality, which turns out to be related to Stirling numbers of the second kind. Results about the longest and shortest chains are also obtained.

Résumé. Nous introduisons et étudions des propiétés combinatoires des pseudo-permutations avec occurrences multiples. Nous montrons que l'ensemble de toutes ces pseudo-permutations généralisées d'un type donné possède une structure de treillis et dérivons une formule récursive pour calculer sa cardinalité, qui est reliée aux nombres de Stirling du second type. Résultats sur les chaînes les plus longues et les chaînes les plus courtes sont également obtenus.

1. Introduction

One of the most active and important research area in Computer Science and in particular, Artificial Intelligence is representation of temporal knowledge (see, *e.g.*, [All81, Yu83], where one needs to consider a set of events which happen at certain dates and wants to use this information to solve a problem, take a decision. In this context, it is natural to represent the temporal relations between n events by an ordered sequence of nonempty parts, each part corresponds to events which happen at the same time.

For example, the sequence $(2,4)(1,2)(3)$ means that the events 2 and 4 occur first at the same time, then event 1 and event 2 (for the second time) occur, and event 3 occurs last. If we add the constraint that each event occurred exactly once, then we have a so-called *pseudo-permutation* of order n. This is a new combinatorial object which was introduced in the five author paper [KLNPS] and studied futher in [BHKN01]. They showed that there is a natural partial order on the set of all pseudo-permutation of a given order, and surprisingly, turns out to be a lattice. This new object is not only combinatorially interesting in its own right, but turns out to be closely related to *S-arrangement* in the field of formal languages [DS00, Sch97].

In this paper, we deal with the more general (and more natural) case where each event can occur several times; in other words, pseudo-permutations with multiple occurrences. We show that most of the properties of pseudo-permutations can be generalized to this situation. Our main results are proofs of the lattice property of the set of all generalized (pseudo-)permutation of a given type K, and a recursive formula to compute its size by using Stirling numbers of the second kind. Results about the longest and shortest chains are also obtained. Geometric applications, and applications in representation theory of the symmetric groups will appear in a subsequence paper.

Let us now recall some basic definitions and introduce our cast of characters. Let n be a positive integer number. Let $\mathfrak{S}(n)$ be the group of all permutations of n numbers $1 \le i \le n$. There is a natural partial order on $\mathfrak{S}(n)$ by requiring that σ' is *covered* by σ, and write $\sigma \succ \sigma'$, if σ' is obtained from σ by switching two *consecutive* numbers i and j in σ where $i < j$. The resulting poset is what usually called *permutohedron* of order n, and is denoted by $\mathcal{P}(n)$. A *pseudo-permutation* of order n is a sequence of non-empty parentheses of numbers from 1 to n. Let $\mathcal{PP}(n)$ denote the set of all such pseudo-permutations of order n. Note that the order inside each parentheses is irrelevant (events occur at the same time!), and we make the convention that the numbers in each parentheses will always be written in increasing order.

As in the case of permutations, there is also a partial order on $\mathcal{PP}(n)$. Say that a pseudo-permutation $\mathfrak{s}'$ is *covered* by $\mathfrak{s}$, and write $\mathfrak{s} \succ \mathfrak{s}'$, if and only if it can be obtained from the latter by applying one of the following two operators:

- The merging operators M_i: If each element of the parentheses $(i)^{th}$ is smaller than all elements of the $(i+1)^{th}$, then M can combine these two parentheses into a single one.
- The splitting operators $S_{i,j}$: the $(i)^{th}$ parentheses is split into two, the *second* one is composed of the first j smallest elements.

For example, M_2 ((3567) (1) (2) (4)) = (3567) (12) (4); and $S_{1,3}$ ((3567) (12) (4)) = (7) (356) (12) (4). We also use the notation $\mathcal{PP}(n)$ for the resulting poset of all pseudo-permutations of order n and call it *pseudo-permutohedron of order* n. Figure 1 is an example of such a pseudo-permutohedron for $n = 3$. Of course, the complexity of these posets increase exponentially with n.

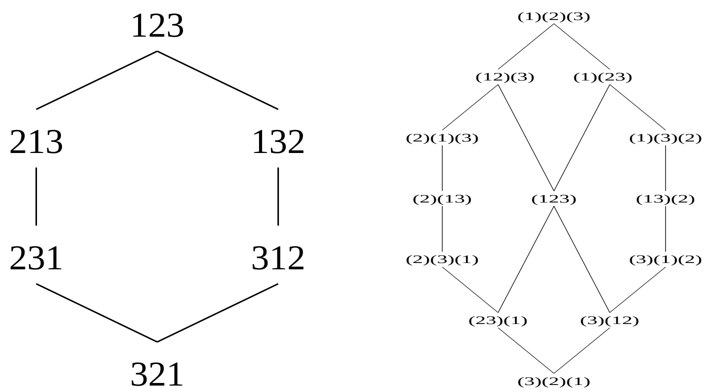


FIGURE 1. The permutohedron $\mathcal{P}(3)$ (left) and the pseudo-permutohedron $\mathcal{PP}(3)$ (right). The order orientation is top-down.

We will now define generalized (pseudo-)permutation. Let m be a positive integer, and $K = (k_1, \ldots, k_n)$ be a composition of m, *i.e.* k_i are positive integers and $\sum k_i = m$. Let I be the sequence of integers $(1, \ldots, 1, \ldots, n, \ldots, n)$ where each integer i appears k_i times. Sometimes, in order to distinguish between numbers of the same symbols, we write $\mathcal{I} = (1_1, \ldots 1_{k_1}, \ldots, n_1 \ldots n_{k_n})$. A *generalized permutation of type* K is, by definition, a permutation of I.

Let $\mathcal{GP}(K)$ be the set of all generalized permutations of type K. The *generalized permutohedron* of type K, also denoted by $\mathcal{GP}(K)$, is the poset over $\mathcal{GP}(K)$ where σ' is *covered* by σ if σ' is obtained from σ by switching two *consecutive* numbers i and j in σ where $i < j$.

In the special case when each letter i appear exactly once, then $\mathcal{GP}(K)$ is just the usual group $\mathfrak{S}(m)$ of all permutations of order m. Given a generalized-permutation, again one can try to put their numbers into non-empty parts by parentheses, such that in every parentheses, each number appears at most once. Such a partition is called a *generalized pseudo-permutation.* Let $\mathcal{GPP}(K)$ be the set of all generalized pseudo-permutations of type K. As in the case of pseudo-permutation, one also has the merging operator M, and the splitting operator S on $\mathcal{GPP}(K)$. The *generalized pseudo-permutohedron* of type K, also denoted by $\mathcal{GPP}(K)$, is the poset over $\mathcal{GPP}(K)$ where $\mathfrak{s}'$ is *covered* by $\mathfrak{s}$ if it can be obtained from $\mathfrak{s}$ by applying one of these two operators.

Figures 2 and 3 are pictures of the generalized-permutohedron $\mathcal{GP}(K)$ and the generalized pseudo-permutohedron $\mathcal{GPP}(K)$ where $K = (2, 1, 1)$.

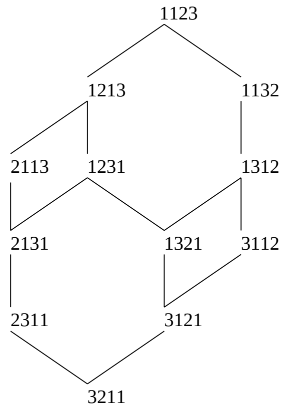


FIGURE 2. The generalized permutohedron of (211).

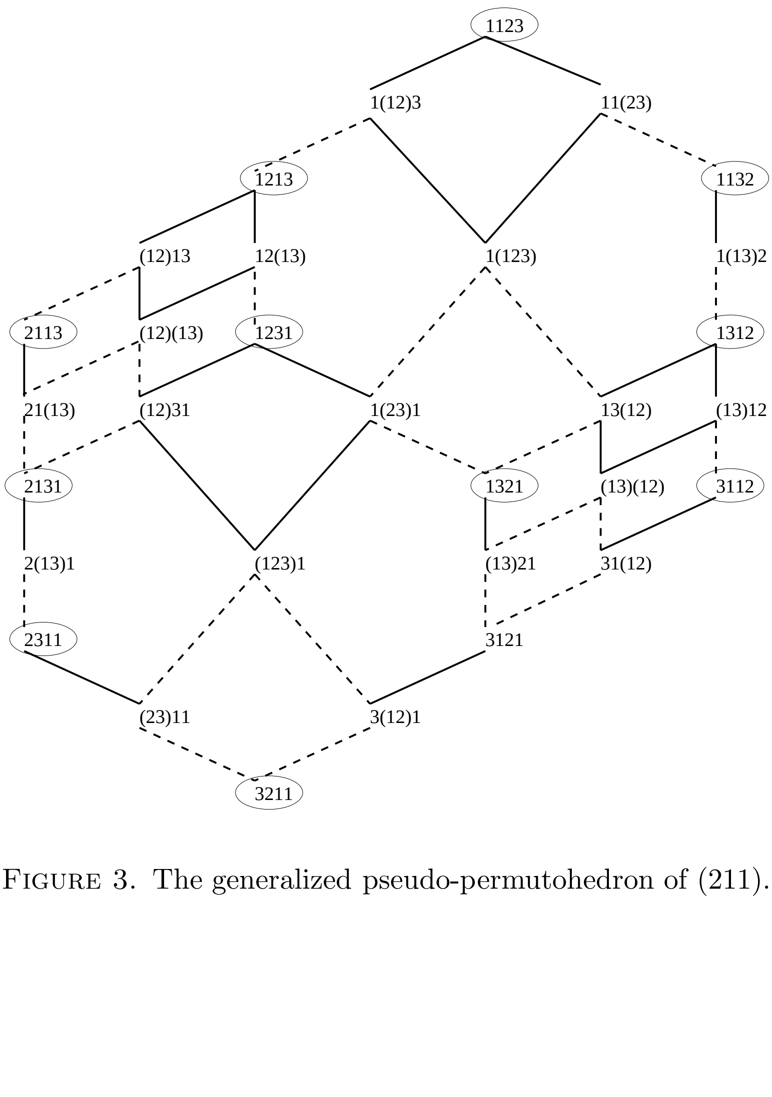


FIGURE 3. The generalized pseudo-permutohedron of (211).

2. Lattice properties of $\mathcal{GP}(K)$ and $\mathcal{GPP}(K)$

Let m be a positive integer and let $K = (k_1, \dots, k_n)$ be a composition of m. In [KLNPS], it is showed that the poset $\mathcal{P}(m)$ is in fact a lattice. Their proof make use of the inversion table associated with each pseudo-permutation, and complicated construction of the infimum and supremum. We are going to show that the newly introduced posets $\mathcal{GP}(K)$ as well as $\mathcal{GPP}(K)$ enjoy similar property. But in order to prove this, we just have to show that there is a canonical embedding ϕ of $\mathcal{GP}(K)$ into $\mathcal{P}(m)$ whose image is a sub-lattice of $\mathcal{P}(m)$; similar statement holds for $\mathcal{GPP}(K)$. We refer to [DP90] for basic reference about lattice theory.

Let $\mathcal{I}$ be the sequence $(1_1, \dots, 1_{k_1}, \dots, n_1, \dots, n_{k_n})$. We say that the elements i_ts are of the same symbol. We define a natural order between elements of $\mathcal{I}$ as follows: $1_1 < \dots < 1_{k_1} < \dots < n_1 < \dots, n_{k_n}$. The permutohedron $\mathcal{P}(\mathcal{I})$ and the pseudo-permutohedron $\mathcal{PP}(\mathcal{I})$ are defined, similar to $\mathcal{P}(m)$ and $\mathcal{PP}(m)$. In fact, it it evident that $\mathcal{P}(\mathcal{I})$ (resp. $\mathcal{PP}(\mathcal{I})$) is isomorphic as a lattice to $\mathcal{P}(m)$ (resp. $\mathcal{PP}(m)$).

The embedding ϕ from $\mathcal{GP}(K)$ (resp. $\mathcal{GPP}(K)$) to $\mathcal{P}(\mathcal{I})$ (resp. $\mathcal{PP}(\mathcal{I})$) is done by just replacing the sequence $I = (\underbrace{1, \dots, 1}_{k_1 \text{ times}}, \dots, \underbrace{n, \dots, n}_{k_n \text{ times}})$ with $\mathcal{I} = (1_1, \dots, 1_{k_1}, \dots, n_1, \dots, n_{k_n})$. It is easy to see that the image of ϕ of $\mathcal{GP}(K)$ (resp. $\mathcal{GPP}(K)$) is the set of all elements such that for any sub-indices $r < s$, the parentheses containing i_r is on the left of that for i_s. In other words, the image of $\mathcal{GP}(K)$ (resp. $\mathcal{GPP}(K)$) is the interval from the element $(1_1 \dots 1_{k_1} \dots n_1 \dots n_{k_n})$ to the element $(n_1 \dots n_{k_n} \dots 1_1 \dots 1_{k_1})$. It is a standard fact in lattice theory [DP90] that an interval in a lattice is a sub-lattice, hence we have:

Theorem 1. *The generalized permutohedron $\mathcal{GP}(K)$ is isomorphic to a sub-lattice of $\mathcal{P}(m)$, and the generalized pseudo-permutohedron $\mathcal{GPP}(K)$ is isomorphic to a sub-lattice of $\mathcal{PP}(m)$.*

3. Combinatorial properties of $\mathcal{GP}(K)$ and $\mathcal{GPP}(K)$

Our purpose in this section is to compute the cardinalities of $\mathcal{GP}(K)$ and $\mathcal{GPP}(K)$. We also obtain a decomposition of $\mathcal{P}(m)$ (resp. $\mathcal{PP}(m)$) as a disjoint union of $\mathcal{GP}(L)$ (resp. $\mathcal{GPP}(L)$) where L are subsequences of K. As a corollary, we obtain a recursive formula for the cardinality of $\mathcal{GPP}(K)$ in terms of $\mathcal{GPP}(L)$.

Recall that the Young subgroup $\mathfrak{S}(K)$ of $\mathfrak{S}(m)$ is defined as the product:

$$\mathfrak{S}(K) = \mathfrak{S}(k_1) \times \dots \times \mathfrak{S}(k_n),$$

where the i^{th} component $\mathfrak{S}(k_i)$ acts on the set $\{i_1, \dots, i_{k_i}\}$. It is clear that $\mathcal{GP}(K)$ is isomorphic to the coset $\mathcal{P}(\mathcal{I})/\mathfrak{S}(K)$, and its cardinality can be computed by the following well-known formula:

$$|\mathcal{GP}(K)| = \binom{m}{k_1, \dots, k_n} = \frac{m!}{k_1! \dots k_n!}.$$

For each element σ of $\mathfrak{S}(K)$, we define a map $\phi_\sigma : \mathcal{GP}(K) \to \mathcal{P}(\mathcal{I})$ by letting $\phi_\sigma(\mathfrak{s}) = \sigma(\mathfrak{s})$. For example, if $\mathcal{I} = (1_1, 1_2, 1_3, 2, 3_1, 3_2)$ and if $\sigma = (321) \times id \times (21)$ and $\mathfrak{s} = (23)1(13)1$, then $\phi_\sigma(\mathfrak{s}) = (23_2)1_3(1_2 3_1)1_1$.

It is easy to see that ϕ_σ is an injection which preserves the lattice structure. In other words, its image, denoted by $\mathcal{GP}(\sigma, K)$, is a sub-lattice of $\mathcal{P}(\mathcal{I})$. Note that all $\mathcal{GP}(\sigma, K)$ are isomorphic to $\mathcal{GP}(K)$ (by convention, $\mathcal{GP}(K)$ is $\mathcal{GP}(id, K)$). See Figure 4 for an example of lattice decomposition of $\mathcal{P}(1_1, 1_2, 2)$. We obtain then a decomposition of $\mathcal{P}(m)$:

Proposition 2. *$\mathcal{P}(m)$ is a disjoint union of lattices, each one is isomorphic to $\mathcal{GP}(K)$:*

$$\mathcal{P}(m) \cong \mathcal{P}(\mathcal{I}) = \bigsqcup_{\sigma \in \mathfrak{S}(K)} \mathcal{GP}(\sigma, K).$$

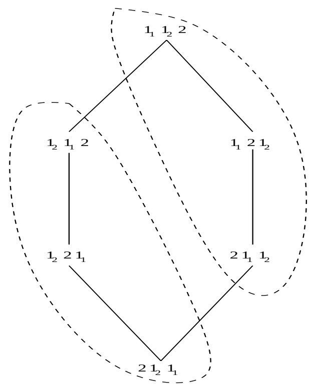


FIGURE 4. The decomposition of $\mathcal{P}(1_1, 1_2, 2)$ for $K = (2, 1)$.

Let us now consider the generalized pseudo-permutohedron $\mathcal{GPP}(K)$. For each σ in $\mathfrak{S}(K)$, we can also construct a lattice-monomorphism

$$\phi_\sigma : \mathcal{GPP}(K) \to \mathcal{GPP}(\sigma, K) \subset \mathcal{PP}(m),$$

which takes $\mathfrak{s}$ to $\sigma(\mathfrak{s})$. For every $\sigma \in \mathfrak{S}(K)$, the image $\mathcal{GPP}(\sigma, K)$ of ϕ_σ is isomorphic to $\mathcal{GP}P(K)$. But unlike the previous situation for $\mathcal{P}(\mathcal{I})$, $\mathcal{PP}(\mathcal{I})$ is not the disjoint union of all $\mathcal{GP}P(\sigma, K)$. Indeed, consider the case when $K = (2, 1)$ (see Figure 5). The three elements

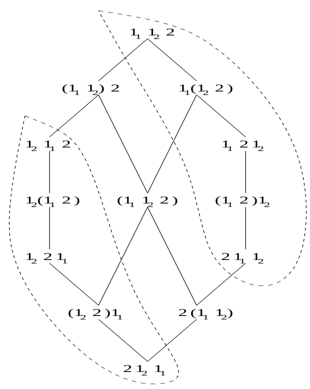


FIGURE 5. The decomposition of $\mathcal{PP}((2, 1))$.

$(1_1 1_2)(2), (1_1 1_2 2)$ and $(2)(1_1 1_2)$ are not in any of $\mathcal{GP}P(\sigma, K)$. The reason for this is that in $\mathcal{PP}(\mathcal{I})$, an element may have a parentheses containing a pair of numbers i_r, i_s of the same symbol, while this is forbidden in $\mathcal{GPP}(\sigma, K)$.

It remains to compute the set of all those extra elements in $\mathcal{PP}(\mathcal{I})$. Let us introduce the set $Par(\mathcal{I})$ of all sequences $\mathcal{J}$ obtained from $\mathcal{I}$ by parenthesizing several elements of the same symbol. For example, if $\mathcal{I} = (1_1, 1_2, 1_3, 2, 3)$, then $\mathcal{J}$ can be $1_3(1_1 1_2)23$ or $(1_1 1_3)1_2 23$ or $1_2 1_1 1_3 23$, etc. We define $\mathcal{GPP}(\mathcal{J})$ as the lattice of all elements $\mathfrak{s}$ obtained from $\mathcal{J}$ by using the two operators M and S, but numbers of the same symbol are not allowed to be merged or split. In other words, for any pair i_r, i_s, the relative position between parentheses containing i_r and that containing i_s does not change.

It is clear that the disjoint union of all lattices $\mathcal{GPP}(\mathcal{J})$ is $\mathcal{PP}(\mathcal{I})$.

$$\mathcal{PP}(\mathcal{I}) = \bigsqcup_{J \in Par(\mathcal{I})} \mathcal{GPP}(J).$$

On the other hand, let $L_J = (l_1, \dots l_n)$ where l_i is the number of parentheses of the symbol i in $\mathcal{J}$, then we see that the lattice $\mathcal{GPP}(\mathcal{J})$ is isomorphic to the lattice $\mathcal{GPP}(L_J)$ (in fact, if we identify a parentheses of the same symbol i in $\mathcal{J}$ to the number i, then $\mathcal{J}$ corresponds to the sequence $(\underbrace{1, \dots, 1}_{l_1 \text{ times}}, \dots, \underbrace{n, \dots, n}_{l_n \text{ times}})$). Write $L \leq K$ if $1 \leq l_i \leq k_i$ for any i, moreover $L < K$ if $L \neq K$. Let $N(K, L)$ be the number of sequences $\mathcal{J}$ such that $L_J = L$, we obtain immediately:

Theorem 3. *The lattice $\mathcal{PP}(m)$ is isomorphic to a disjoint union of lattices $\mathcal{GPP}(L)$ with multiplicity $N(K, L)$.*

$$\mathcal{PP}(m) \cong \mathcal{PP}(\mathcal{I}) = \bigsqcup_{L \leq K} N(K, L)\mathcal{GPP}(L). \tag{1}$$

From the above theorem, we need only to compute the multiplicity $N(K, L)$. It is easy to see that $N(K, L)$ is equal to the product $N(k_1, l_1) \times \dots \times N(k_n, l_n)$, where $N(k_i, l_i)$ is the number of partitions of k_i elements into l_i (ordered) parentheses. Hence, $N(k_i, l_i) = l_i! S(k_i, l_i)$, where $S(k_i, l_i)$ is the Stirling number of the second kind [Sta98]:

$$S(p, q) = \frac{1}{q!} \sum_{j=0}^{j=q} (-1)^{(q-j)} \binom{j}{q} j^p,$$

We deduce that:

$$N(k_i, l_i) = \sum_{j=0}^{j=l_i} (-1)^{(l_i - j)} \binom{j}{l_i} j^{k_i}. \tag{2}$$

In particular, $N(K, K) = k_1! \times \dots \times k_n! = |\mathfrak{S}(K)|$, the above relation implies a recursive formula to compute $|\mathcal{GPP}(K)|$:

$$|\mathcal{GPP}(K)| = \frac{|\mathcal{PP}(m)| - \sum_{L<K} N(K, L)|\mathcal{GPP}(L)|}{k_1! \times \dots \times k_n!}. \tag{3}$$

The cardinality of the pseudo-permutohedron $\mathcal{PP}(m)$ has been computed in [KLNPS], which turns out to be related to Eulerian numbers $a_{m,i}$ (See [FS70], [Com70]).

$$|\mathcal{PP}(m)| = \sum_{i=0}^{m-1} 2^i a_{m,i}.$$

Recall that $a_{m,i}$ is the number of permutations in $\mathfrak{S}(m)$ whose descent number is exactly $(i - 1)$. By convention, $a_{m,0} = 1$.

Note that from the recursive formula for $|\mathcal{GPP}(K)|$ above, one can show by induction that the cardinality of $\mathcal{GPP}(K)$ is a symmetric function on $(k_1, \dots, k_n)$. That is, if K' is a permutation of K, then $|\mathcal{GPP}(K)| = |\mathcal{GPP}(K')|$. Of course, the same result holds for $|\mathcal{GP}(K)|$. However, $\mathcal{GP}(K')$ (resp. $\mathcal{GPP}(K''))$ is, in general, not isomorphic as a lattice to $\mathcal{GP}(K)$ (resp. $\mathcal{GPP}(K)$). Figure 6 is a picture of $\mathcal{GP}(1, 2, 1)$ which is not isomorphic to $\mathcal{GP}(2, 1, 1)$ (in Figure 6).

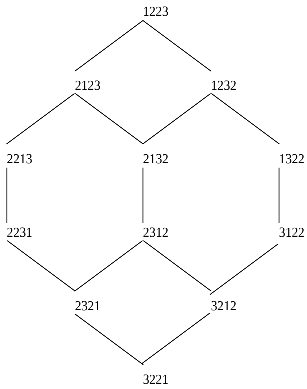


FIGURE 6. The generalized permutohedron of (121)

4. Longest and shortest chains

In this section, we will compute the length of the longest and shortest chains in a pseudo-permutohedron as well as in a generalized pseudo-permutohedron. Here we mean a *chain* in a lattice is a path from the maximal element to the minimal element. In order to prove our results, we need to make use of the notion of inversion of a pair (i, j) and the table of inversions of a pseudo-permutation. This is a crucial ingredient in the proof of the lattice property of the set of pseudo-permutations of order n [KLNPS].

Let $\mathfrak{s}$ be a pseudo-permutation of order n. For every pair of numbers $1 \leq i < j \leq n$, we associate a rational number called inversion, and denoted by *inv(i,j)* as follows:

$$inv(i,j) = \begin{cases} 1/2 & \text{if } i \text{ and } j \text{ are in the same parentheses,} \\ 0 & \text{if the parentheses of } i \text{ is on the left of that of } j, \\ 1 & \text{if the parentheses of } i \text{ is on the right of that of } j. \end{cases}$$

The *table of inversions* of $\mathfrak{s}$ is just the list of all non-zero inv(i,j) in $\mathfrak{s}$. The summation of all such inversion numbers is called the *inversion number* of $\mathfrak{s}$. For example, if $\mathfrak{s} = (4)(13)(2)$, then its table of inversions is:

$$\{\frac{1}{2}(1,3), 1(1,4), 1(2,3), 1(2,4), 1(3,4)\}$$

and its inversion number is $4\frac{1}{2}$. The number on the left of each pair (i, j) is its inversion.

It is easy to see that in the lattice $\mathcal{P}(m)$, each operator increase inversion number by 1. As a result, all chains in $\mathcal{P}(m)$ are of the same length. However, this is different in $\mathcal{PP}(m)$, M and S still increase inversion numbers, the least increase is $\frac{1}{2}$, which correspond to either merging two parentheses, each of a single element, into one; or splitting a parentheses which contains two elements. It follows from this remark that the longest chains should be those which contain only operations described above.

Proposition 4. *There is a natural bijection between the set of chains in $\mathcal{P}(n)$ and the set of longest chains in $\mathcal{PP}(n)$. Moreover, the length of the longest chains in $\mathcal{PP}(n)$ is $n(n-1)$.*

Proof. The bijection is given as follows. Given a chain of $\mathcal{P}(n)$, consider it as a subchain in $\mathcal{PP}(n)$ via the obvious inclusion (parenthesizing every single number); then between any two consecutive pseudo-permutations

$$\ldots(i)(j)\ldots, \ldots(j)(i)\ldots,$$

where $i < j$, insert a pseudo-permutation $\dots (ij) \dots$. The result is a chain of pseudo-permutations of the longest length possible, since it is easy to check that the inversion number increase by $\frac{1}{2}$ in the entire chain. For example, $123, 132, 312, 321$ is a chain in $\mathcal{P}(3)$, the corresponding chain in $\mathcal{PP}(3)$ is:

$$(1)(2)(3), (1)(23), (1)(3)(2), (13)(2), (3)(1)(2), (3)(12), (3)(2)(1).$$

The converse is also clear, using the observation in the previous paragraph. Note that the length of the longest chain in $\mathcal{PP}(n)$ is twice the difference of inversion number between the maximum element $(1)(2)\dots(n)$ and the minimum element $(n)(n-1)\dots(1)$ of $\mathcal{PP}(n)$. Since the inversion number of of $(1)(2)\dots(n)$ is 0, and the inversion number of $(n)(n-1)\dots(1)$ is $\frac{n(n-1)}{2}$, we conclude that the longest length is $n(n-1)$. $\square$

The discussion above also applies to the "generalized" version, *i.e.* all chains in $\mathcal{GP}(K)$ are of the same length and there is a bijection between the set of chains in $\mathcal{GP}(K)$ and the set of longest chains in $\mathcal{GPP}(K)$. The inversion number of the minimal element $(n)\dots(n)\dots(1)\dots(1)$ is $\sum_{i<j} k_i k_j$, so every chain in $\mathcal{GP}(K)$ is of length $\sum_{i<j} k_i k_j$ and the longest length in $\mathcal{GPP}(K)$ is $2\sum_{i<j} k_i k_j$.

We next consider the shortest chains in $\mathcal{PP}(n)$.

Proposition 5. *There are* $[(n-1)!]^2$ *chains of shortest length in* $\mathcal{PP}(n)$*, and they each have length* $(2n-2)$*.*

Proof. We first show that any chain in $\mathcal{PP}(n)$ have length at least $(2n-2)$. Our key observation is that the operations M or S can either eliminate or increase at most one parentheses between any two numbers i and j at a time. Consider a chain from the maximum element $(1)(2)\dots(n)$ to the minimum one $(n)\dots(1)$. At the beginning, there are (n-1) parentheses between 1 and n, so one needs at least (n-1) steps to put 1 and n in the same parentheses. Again, one needs at least $(n-1)$ more moves to move 1 to the position on the far right of n, so that there are $(n-1)$ parentheses between them. Therefore, any chain in $\mathcal{PP}(n)$ has length at least $(2n-2)$. It remains now to provide an explicit chain which has this length $(2n-2)$. Here is an example,

$$\begin{aligned}&(1)(2)\dots(n), (12)(3)\dots(n), \dots, (12\dots n),\\ &\qquad (n)(12\dots n-1), (n)(n-1)(12\dots n-2), \dots, (n)(n-1)\dots(1).\end{aligned}$$

Since in a shortest chain, the first $(n-1)$ steps decrease the number of parentheses between 1 and n, the n^{th} elements in the chain must be $(12\dots n)$. It is easy to see that there are $(n-1)!$ different paths from $(1)(2)\dots(n)$ to $(12\dots n)$, all have length $(n-1)$. By symmetry, there are also $(n-1)!$ paths from $(12\dots n)$ to $(n)(n-1)\dots(1)$. A chain of minimum length is obtained by combining these two paths. Conversely, every chain containing $(12\dots n)$ has length $2n-2$. The proposition follows immediately. $\square$

Our method of computing the shortest chains does not work for generalized pseudo-permutations, however. Since there are more constraints over the merging operation M. For example, it is not allowed to merge the two parentheses (12) and (13) because 1 appears in both.

References

[All81] J. F. Allen, *An Interval-Based Representation of Temporal Knowledge*, Proc. 7th International Joint Conference on Artificial Intelligence, (1981) 221-226.

[BHKN01] F. Boulier, F. Hivert, D. Krob, and J.C. Novelli, *Pseudo-permutations 2: Geometry and representation theory*, in M. Morvan, Ed., DM-CCG (2001) 123–132.

[Com70] L. Comtet, *Analyse combinatoire T. 2*, Presses Univeritaires de France, 1970.

[DP90] B. A. Davey and H. A. Priestley, *Introduction to Lattices and Order*, Cambridge University Press, 1990.

[DS00] F. Dubois and S. R. Schwer, *Classification topologique des ensembles convexes de Allen*, Proceedings of the RFIA, (2000).

[FS70] D. Foata and M. P. Schützenberger, *Théorie géométrique des polynômes eulériens*, Lecture notes, Springer, 1970.

[KLNPS] D. Krob, M. Latapy, J.-C. Novelli, H. D. Phan and S. Schwer, *Pseudo-Permutations I: First Combinatorial and Lattice Properties*, Proceddings of the FPSAC'01, H. Barcelo, ed. 2000.

[Sch97] S. R. Schwer, *Dependances temporelles: les mots pour le dire*, rapport interne LIPN, 1997.

[Sta98] R. P. Stanley, *Enumerative combinatorics, Vol. 1*, Cambridge University Press, 1997.

[Yu83] K. E. Yu, *Representation of Temporal Knowledge*, Proc. 8th International Joint Conference on Artificial Intelligence, (1983).

Institut des Hautes Etudes Scientifiques - Le Bois-Marie, 35, route de Chartres - F 91440 Bures sur Yvette - France.

E-mail address: `lha@ihes.fr`

LIAFA Université Denis Diderot, Paris 7 - Case 7014 - 2, Place Jussieu - 75256 Paris Cedex 05 - France

E-mail address: `phan@liafa.jussieu.fr`

ENUMERATION OF UNICURSAL PLANAR MAPS

VALERY A. LISKOVETS* AND TIMOTHY R. S. WALSH†

Abstract. Sum-free enumerative formulae are derived for several classes of rooted eulerian planar maps and maps with two vertices of odd valency. As corollaries we obtain simple formulae for the numbers of unrooted eulerian and unicursal planar maps.

Résumé. Nous présentation des formules énumératives sans sommation pour plusieurs classes de cartes planaires eulériennes pointées et cartes unicursalles (avec deux sommets de degré impair). En tant que corollaires nous obtenons des formules simples pour les nombres de cartes planaires eulériennes et unicursalles non-pointées.

1. Introduction

1.1. Eulerian maps have played a crucial role in enumerative map theory since its beginning in the early sixties. In particular, Tutte's sum-free formula [Tut62] for the number eulerian planar maps, all of whose vertices are labelled and contain a distinguished edge-end, with a given sequence of (even) vertex valencies was an essential step in obtaining his groundbreaking formula for counting rooted planar maps by number of edges [Tut63]. Several new results on the subject have been published since then (see, e.g., [Wal75, Lis85, BouS00]) but a number of natural enumerative problems for eulerian maps have remained unsolved.

Here we consider a slight generalization of eulerian maps: unicursal maps. We continue our previous investigation and count planar maps of several classes. Our main results concern rooted unicursal maps (see Section 2) and unrooted eulerian as well as unicursal maps (see Section 3) specified by the number of edges. We present several new 1-parametric sum-free enumerative formulae.

In the last section, rooted eulerian bipartite and non-separable maps are enumerated. Although these results may be considered as implicitly known, they, apparently, had never been published.

There are some open questions related to the settled ones; they (together with asymptotic corollaries) are to be discussed in the full version of this article.

1.2. Basic definitions. A *map* means a planar map: a 2-cell imbedding of a planar connected graph (loops and multiple edges allowed) in an oriented sphere. A map is *rooted* if one of its edge-ends (variously known as edge-vertex incidence pairs, darts, semi-edges, or "brins" in French) is distinguished as the *root.* The corresponding edge and vertex are called the *root-edge* and *root-vertex*, respectively; the *root-face* is the face incident to the root-edge and on its left as one face away from the root-vertex.

A map (and a graph in general) is *eulerian* if it has an eulerian circuit, that is, a circuit containing all the edges exactly once. It is well-known that a map is eulerian if and only if all its vertices are of even valency.

We call a graph *unicursal* if it possesses an eulerian tour, not necessarily a circuit. A connected graph is unicursal if and only if it contains no more than two vertices of odd valency.

* Supported by the INTAS (Grant INTAS-BELARUS 97-0093).

† Supported by National Science and Engineering Research Council (NSERC) under grant RGPIN-44703.

However, throughout this paper, *par abus de langage*, we will use the term "unicursal maps" in the restricted sense to mean maps with *exactly two* vertices of odd valency.

2. Rooted unicursal maps

2.1. Unicursal maps together with eulerian maps are the very maps considered by Tutte in his seminal paper [Tut62]. But all the enumerative results obtained so far for unicursal maps concerned maps with specified vertex valencies. Accordingly, no formula has been known for the number of n-edged rooted unicursal maps. Evidently, the question cannot be reduced straightforwardly to eulerian maps (by adding an edge connecting the two odd-valent vertices, etc.) since the vertices of odd valency may not be incident to a common face. So, this natural class requires independent consideration.

Let $U'(n)$ denote the number of rooted unicursal maps with n edges. Let also $U_i'(n)$, $i = 0, 1, 2$, denote the number of rooted unicursal maps having i endpoints, i.e. vertices of valency 1. The next theorem is the main result of this paper.

2.2. Theorem.

$$U'(n) = 2^{n-2}\binom{2n}{n}, \qquad n \geq 1, \tag{2.1}$$

and for $n \geq 2$,

$$U_0'(n) = 2^{n-2}\,\frac{n-2}{n}\binom{2n-2}{n-1}, \tag{2.2}$$

$$U_1'(n) = 2^{n-1}\binom{2n-2}{n-1} \tag{2.3}$$

and

$$U_2'(n) = 2^{n-2}\binom{2n-2}{n-1}. \tag{2.4}$$

Proof. The number of unicursal planar maps with n edges and v labelled vertices of valencies $2d_1+1, 2d_2+1, 2d_3, \ldots, 2d_v$, each vertex rooted by distinguishing one of its edge-ends, is given in [Tut62, p. 772] as

$$C(2d_1+1, 2d_2+1, 2d_3, \ldots, 2d_v) = \frac{(n-1)!}{(n-v+2)!}\,\frac{(2d_1+1)!(2d_2+1)!}{d_1!^2 d_2!^2}\prod_{i=3}^{v}\frac{(2d_i)!}{d_i!(d_i-1)!}.$$

The number of rooted planar maps with n edges and v vertices, exactly two of which are of odd valency, is found from the previous equation by multiplying by the number of ways to root a map with n edges and dividing by the number of ways to label and root all the vertices of the same map so that the two vertices of odd valency get labels 1 and 2 (we multiply by $2n$ and divide by the product of the valencies and by $n!$ and then multiply by $n(n-1)/2$ to account for the fact that the two vertices of odd valency get labels 1 and 2) and then summing over the sequences of valencies that add to $2n$:

$$\frac{n!}{(v-2)!(n-v+2)!}\sum_{d_1+\cdots+d_v=n-1}\left\{\frac{(2d_1)!(2d_2)!}{d_1!^2 d_2!^2}\prod_{i=3}^{v}\frac{(2d_i-1)!}{d_i!(d_i-1)!}\right\}.$$

To obtain $U'(n)$ we evaluate the sum and then add over all possible values of v : from 2 to $n+1$.

Since $\sum_{j=0}^{\infty} \frac{(2j)!}{j!^2} x^j = (1-4x)^{-1/2}$ and $\sum_{j=1}^{\infty} \frac{(2j-1)!}{j!(j-1)!} x^j = \frac{(1-4x)^{-1/2}-1}{2}$, we have

$$U'(n) = [x^{n-1}] \sum_{v=2}^{n+1} \frac{n!}{(v-2)!(n-v+2)!} (1-4x)^{-1} \Big[\frac{(1-4x)^{-1/2}-1}{2}\Big]^{v-2},$$

where $[x^n]b$ means the coefficient of x^n in the power series $b = b(x)$.

We set $z := x(z+1)^2$ so that $\frac{(1-4x)^{-1/2}-1}{2} = \frac{z}{1-z}$ and $(1-4x)^{-1} = \left(\frac{1+z}{1-z}\right)^2$. Then

$$\begin{aligned} U'(n) &= [x^{n-1}] \sum_{v=2}^{n+1} \binom{n}{v-2} \left(\frac{1+z}{1-z}\right)^2 \left(\frac{z}{1-z}\right)^{v-2} \\ &= [x^{n-1}] \left(\frac{1+z}{1-z}\right)^2 \sum_{v=0}^{n-1} \binom{n}{v} \left(\frac{z}{1-z}\right)^{v} \\ &= [x^{n-1}] \left(\frac{1+z}{1-z}\right)^2 \left[\left(1+\frac{z}{1-z}\right)^n - \left(\frac{z}{1-z}\right)^n\right] \\ &= [x^{n-1}] \Big[(1+z)^2(1-z)^{-(n+2)} - (1+z)^2 z^n (1-z)^{-(n+2)}\Big]. \end{aligned}$$

By Lagrange's inversion formula (see, e.g., [Lab81]),

$$U'(n) = \frac{1}{n-1}[z^{n-2}]\Big\{(1+z)^{2n-2}\frac{d}{dz}\Big[(1+z)^2(1-z)^{-(n+2)} - (1+z)^2 z^n (1-z)^{-(n+2)}\Big]\Big\}.$$

Now a factor of z^n means that the coefficient of z^{n-2} will be zero even in the derivative. We have $\frac{d}{dz}\Big[(1+z)^2(1-z)^{-(n+2)}\Big] = 2(1+z)(1-z)^{-(n+2)} + (n+2)(1+z)^2(1-z)^{-(n+3)}$, so that

$$\begin{aligned} U'(n) &= \frac{1}{n-1}[x^{n-2}]\Big[2(1+z)^{2n-1}(1-z)^{-(n+2)} + (n+2)(1+z)^{2n}(1-z)^{-(n+3)}\Big] \\ &= \frac{1}{n-1}\Big[2\sum_{i=0}^{n-2}\binom{2n-1}{n-2-i}\binom{i+n+1}{i} + (n+2)\sum_{i=0}^{n-2}\binom{2n}{n-2-i}\binom{i+n+2}{i}\Big] \\ &= \frac{1}{n-1}\Big[2\frac{(2n-1)!}{(n+1)!(n-2)!}\sum_{i=0}^{n-2}\binom{n-2}{i} + (n+2)\frac{(2n)!}{(n+2)!(n-2)!}\sum_{i=0}^{n-2}\binom{n-2}{i}\Big] \\ &= \frac{2^{n-2}}{n-1}\Big[2\frac{(2n-1)!}{(n+1)!(n-2)!} + \frac{(2n)!}{(n+1)!(n-2)!}\Big], \end{aligned}$$

which simplifies to (2.1). This derivation is valid only for $n \geq 2$ since we are taking coefficients of z^{n-2}, but (2.1) turns out to be valid for $n = 1$ as well.

To prove formula (2.4) we set d_1 and d_2 to 0, so that the first and second vertices have valency 1. Proceeding as above, we find that

$$U_2'(n) = [x^{n-1}] \sum_{v=2}^{n+1} \frac{n!}{(v-2)!(n-v+2)!} \Big[\frac{(1-4x)^{-1/2}-1}{2}\Big]^{v-2},$$

which simplifies to (2.4).

If instead we just set d_1 to 0, then the first vertex has valency 1 and the second vertex can have any odd valency $2d_2+1$, including 1. If $d_2 = 0$ then, as before, we multiply by $n(n-1)/2$ to account for the fact that the two vertices of valency 1 get labels 1 and 2, but if $d_2 > 0$, then we instead multiply by $n(n-1)$ to account for the fact the vertex of valency 1 gets label 1 and the other odd-valent vertex gets label 2; so

$$\begin{aligned} 2U_2'(n) + U_1'(n) &= 2[x^{n-1}]\sum_{v=2}^{n+1} \frac{n!}{(v-2)!(n-v+2)!}(1-4x)^{-1/2}\left[\frac{(1-4x)^{-1/2}-1}{2}\right]^{v-2} \\ &= 2^n\binom{2n-2}{n-1}, \end{aligned}$$

from which (2.3) follows.

Finally, formula (2.2) follows from the other formulae since

$$U_0'(n) + U_1'(n) + U_2'(n) = U'(n).$$

Formulae (2.2), (2.3) and (2.4) are valid only for $n \geq 2$. □

More generally, a simple formula can be deduced in the same way for the number of rooted unicursal maps with n edges and specified valencies of the odd-valent vertices. Moreover, a sum-free formula is valid for the number of n-edged unicursal maps rooted at an odd-valent vertex. In all these cases, in our opinion, the most significant novelty is the existence, per se, of such formulae.

2.3. Rooted eulerian maps. The number $E'(n)$ of rooted eulerian planar maps with n edges is expressed by the following well-known formula [Wal75] (see also the same formula in [Tut63, p. 269] for the number of rooted trivalent planar maps and the bijection between these two classes of maps shown in [Mul66]):

$$E'(n) = \frac{3\cdot 2^{n-1}}{(n+1)(n+2)}\binom{2n}{n}. \tag{2.5}$$

Denoting $e(z) := \sum_n E'(n)z^n$, it can be easily verified that

$$e(z) = \frac{8z^2+12z-1+(1-8z)^{3/2}}{32z^2}. \tag{2.6}$$

2.4. Comparing formula (2.5) with (2.1) we obtain the following identity:

$$U'(n) = \tfrac{1}{6}(n+1)(n+2)E'(n).$$

We do not know whether this identity (or anything similar to it) can be proved directly. The same question concerns another curious identity: $U_1'(n) = 2U_2'(n)$.

3. Unrooted eulerian and unicursal maps

Formulae (2.1) and (2.3) – (2.5) enable us to complete the solution of the long-standing problem of the enumeration of unrooted eulerian planar maps. Namely, the formulae obtained in [Lis85] can be transformed into an explicit formula with single sums over the divisors of n.

3.1. Theorem. *The number $E^+(n)$ of non-isomorphic eulerian planar maps with n edges, $n \geq 2$, is expressed as follows:*

$$
\begin{aligned}
E^+(n) \;=\; & \frac{1}{2n}\left[\frac{3\cdot 2^{n-1}}{(n+1)(n+2)}\binom{2n}{n} + 3\sum_{k<n,\,k|n}\phi\Big(\frac{n}{k}\Big)2^{k-2}\binom{2k}{k}\right.\\
& \left. + \begin{cases} \dfrac{n\cdot 2^{(n+1)/2}}{n+1}\dbinom{n-1}{\frac{n-1}{2}}, & n \text{ odd},\\[2ex] \displaystyle\sum_{k|\frac{n}{2}}\phi\Big(\frac{n}{k}\Big)2^{k-2}\binom{2k}{k} + \frac{n\cdot 2^{(n-2)/2}}{n+2}\binom{n}{\frac{n}{2}}, & n \text{ even},\end{cases}\right]
\end{aligned}
$$

where $\phi(n)$ is the Euler totient function.

Proof. This is an easy consequence of the following result.

3.2. Theorem [Lis85]**.**

$$
\begin{aligned}
E^+(n) \;=\; & \frac{1}{2n}\left[E'(n) + \frac{1}{2}\sum_{k<n,\,k|n}\phi\Big(\frac{n}{k}\Big)(k+2)(k+1)E'(k)\right.\\
& \left. + \begin{cases} U'_*\Big(\dfrac{n+1}{2}\Big), & n \text{ odd},\\[2ex] \displaystyle\sum_{k|\frac{n}{2}}\phi\Big(\frac{n}{k}\Big)U'_1(k) + U'_{**}\Big(\frac{n+2}{2}\Big), & n \text{ even},\end{cases}\right]
\end{aligned}
$$

where U'_* and U'_{**} denote the numbers of rooted unicursal maps with one and two singular vertices respectively; a singular *vertex means a vertex of valency 1 in which the map is not allowed to be rooted.*

It is clear that $U'_{**}(n) = \frac{n-1}{n}U'_2(n)$ since any map with n edges and two singular vertices contains $2n$ edge-ends, of which exactly two are ineligible to be the root.

Likewise $U'_*(n) = \frac{2n-1}{2n}U'_1(n) + \frac{2n-1}{n}U'_2(n)$. Indeed, the first summand reflects the fact that we may take any unrooted unicursal map with a unique 1-valent vertex, declare this vertex to be singular and choose a root in one of $2n-1$ ways. The second summand is obtained by considering the contribution to the set of rooted maps with one singular vertex made by a map Γ with two 1-valent vertices. If Γ has no non-trivial symmetries, then we must declare one of its endpoints to be singular and then choose a root in one of $2n-1$ ways; so Γ contributes $2(2n-1)$ to the set of rooted maps with one singular vertex instead of the usual $2n$ rootings. Now suppose that Γ has a rotational symmetry of order 2 (the only possible non-trivial orientation-preserving automorphism). Both endpoints are equivalent, and after we declare one of them to be singular (which destroys the symmetry), there are $2n-1$ (instead of n) possible rootings. Therefore, in both cases the proportion $(2n-1):n$ is the same.

Finally, taking into account formulae (2.3) and (2.4) we obtain

$$U'_*(n) = \frac{2n-1}{n}U'_1(n). \tag{3.1}$$

□

Similarly we obtain the following.

3.3. Theorem. *Let $U^+(n)$ denote the number of non-isomorphic unicursal planar maps with n edges, $n \geq 2$, then*

$$U^+(n) = \frac{1}{2n} \sum_{k|n,\, n/k \text{ odd}} \phi\Big(\frac{n}{k}\Big) 2^{k-2} \binom{2k}{k} + \begin{cases} 2^{(n-3)/2} \binom{n-1}{\frac{n-1}{2}}, & n \text{ odd,} \\ 2^{(n-6)/2} \binom{n}{\frac{n}{2}}, & n \text{ even.} \end{cases}$$

Proof. We exploit the method developed in [Lis85]. Unicursal maps are similar to but simpler than eulerian maps with respect to possible rotational symmetries. Namely, only three types of rotations exist:

(I_k) rotations of an odd order k around the two odd-valent vertices (that is, around an axis that intersects the map in the two odd-valent vertices);

(I_2) rotations of order 2 around two even-valent vertices or an even-valent vertex and (the center of) a face;

(T) rotations of order 2 around the middle of an edge and a vertex or a face.

In every case, the quotient map is a unicursal map; it contains one singular vertex in the last case. Now consider the possible liftings. In the first case, the axial cells (the vertices, edges or faces in which the axis of rotation intersects the map) are determined uniquely. For I_2 we choose one odd-valent vertex of the quotient map as axial; the other axial cell is an arbitrary vertex or face except for the second odd-valent vertex. These are possible choices of the second axial vertex for rotations of the type T as well, while the first axial cell is necessarily the singular vertex. Now, by the main theorem of [Lis85] we obtain immediately the formula

$$U^+(n) = \frac{1}{2n} \left[\sum_{k|n,\, n/k \text{ odd}} \phi\Big(\frac{n}{k}\Big) U'(k) + \begin{cases} nU'\Big(\frac{n}{2}\Big), & n \text{ odd,} \\ \frac{n+1}{2} U'_*\Big(\frac{n+1}{2}\Big), & n \text{ even} \end{cases} \right].$$

This, together with formulae (2.1), (2.3) and (3.1), gives rise to Theorem 3.3. □

Specializing this proof to unrooted unicursal maps with two endpoints we obtain the following expression for their number $U_2^+(n)$.

3.4. Proposition. $U_2^+(1) = U_2^+(2) = 1$ *and for* $n \geq 3$,

$$U_2^+(n) = \frac{1}{n} 2^{n-3} \binom{2n-2}{n-1} + 2^{m-3} \binom{2m-2}{m-1}$$

where $m = \lfloor (n+1)/2 \rfloor$.

The first term in this formula can be written as $2^{n-3} C_{n-1}$, where C_n is the n-th Catalan number. Notice also that unicursal maps with one endpoint do not have non-trivial symmetries; therefore by (2.3),

$$U_1^+(n) = 2^{n-2} C_{n-1}.$$

4. Rooted bipartite and non-separable eulerian maps

4.1. Bi-eulerian maps. We call a map *bi-eulerian* if all its vertices and faces are of even valency; thus, both the map and its dual possess eulerian circuits. It is well known (see, e.g., [Wal75]) that the dual of an eulerian planar map is bipartite, and vice versa (so $E'(n)$ is also the number of rooted bipartite maps). Therefore,

4.2. Lemma. *A planar map is bi-eulerian if and only if it is bipartite and eulerian.*

Thus, this is a purely graph-theoretical property, not depending on the embedding.
Since the edges of an eulerian circuit switch alternate between the two parts, from this lemma we obtain the following.

4.3. Corollary. *Any bi-eulerian map contains an even number of edges.*

4.4. According to [KazSW96] (see also [SzaW97]), the cubic equation

$$3z^2y^3 - y + 1 = 0 \tag{4.1}$$

and

$$b(z) = (1 + 3y - y^2)/3$$

determine the generating function $b(z) = 1 + \sum_{n>0} B'(2n)z^{2n}$ of the number of rooted bipartite eulerian planar maps. This remarkable result has been obtained by a strong physical method known as the method of matrix integrals (see [Zvo96]) with the help of character expansion techniques.

From formula (4.1) one can easily obtain the following explicit sum-free formula:

4.5. Proposition.

$$B'(2n) = \frac{3^{n-1}}{n(2n+1)}\binom{3n}{n+1}. \tag{4.2}$$

Proof. Represent (4.1) in the form $w = 3z(w+1)^3$ where $w = y - 1$. Now $b = b(z) = (3 + w - w^2)/3$. Applying Lagrange's inversion formula $[z^n]\, b = [w^{n-1}]\, b' f^n/n$, where $f = f(w) = 3(w+1)^3$, we obtain

$$[z^n]\, b = \frac{1}{n}[w^{n-1}]\frac{1-2w}{3}3^n(1+w)^{3n} = \frac{3^{n-1}}{n}\left\{\binom{3n}{n-1} - 2\binom{3n}{n-2}\right\},$$

which gives rise to (4.2). □

This is, apparently, a new result (announced in [Lis00]); although as we learned not long ago [Sch00], D. Poulalhon and G. Schaeffer deduced formula (4.2) directly, based on the combinatorial technique developed in [BouS00].

4.6. Corollary. *The following identity is valid:*

$$B'(2n) = 3^{n-1}S'(n+1),$$

where $S'(n)$ denotes the number of rooted non-separable *maps with n edges.*

This follows immediately from the well-known formula of Tutte [Tut63] for $S'(n)$. It would be nice to find a direct bijective proof.

4.7. Remarks. 1. There is a simple 1:2:3 correspondence between, resp., rooted bi-eulerian planar maps with $2n$ edges, tetravalent bi-eulerian maps with $4n$ edges and trivalent maps with all face sizes multiple to 3 and with $6n$ edges. This has been established by Szabo and Wheater [SzaW97].

2. It is an easy matter to prove that bi-eulerian maps form a degenerate class of maps in the sense that they cannot be 3-connected (that is, polyhedral).

4.8. Non-separable eulerian and bi-eulerian maps. It is a useful but not difficult matter to count rooted non-separable maps of the classes under consideration. So, we mention the results only sketchily (cf. also [Sch98, p. 45]). Suppose again that $e = e(z) := \sum_n E'(n)z^n$. Then

$$e(z) = d(ze(z)^2)$$

where $d = d(z) := 1 + \sum_{n>0} E'_{\mathrm{NS}}(n)z^n$ is the generating function for non-separable eulerian maps. This functional equation together with expression (2.6) uniquely determines $d(z)$.

Likewise, for $b = b(z) := \sum_n B'(2n)z^{2n}$ we have

$$b(z) = f(zb(z)^2)$$

where $f = f(z) := 1 + \sum_{n>0} B'_{\mathrm{NS}}(2n)z^{2n}$ is the generating function for non-separable bi-eulerian maps.

Here we used the general functional equation for non-separable maps [Tut63, WalL75]: $g(z) = f(zg(z)^2)$. It is applicable in our cases: all 2-connected components of a (bi-)eulerian map are (bi-)eulerian (consider end components in the component tree, and so on).

Unfortunately (and somewhat unexpectedly), the explicit expressions that can be extracted from these equations look very tedious (double sums); so we do not provide them here.

Tables 1 and 2 contain numerical data for maps of the classes under consideration. The values for $n \le 6$ may be verified by the Atlas of maps [JacV00] (for some quantities, in fact, we first guessed the formulae from data extracted from the Atlas).

Acknowledgement. We are thankful to Gilles Schaeffer and Cedric Chauve for fruitful discussion of various results concerning the enumeration of eulerian maps.

References

[BouS00] Bousquet-Mélou, M., and Schaeffer, G., "Enumeration of planar constellations." *Adv. Appl. Math.* **24** (2000), No 4, 337–368.

[JacV00] Jackson, D. M., and Visentin, T. I., *An Atlas of the Smaller Maps in Orientable and Nonorientable Surfaces.* CRC Press, 2000.

[KazSW96] Kazakov, V. A., Staudacher, M., and Wynter, Th., "Character expansion methods for matrix models of dually weighted graphs." *Commun. Math. Phys.* **177** (1996), No 1–2, 451–468 (also preprint arXiv hep-th/9502132).

[Lab81] Labelle, G., "Une nouvelle démonstration combinatoire des formules d'inversion de Lagrange." *Advances in Math.* **42** (1981), 217–247.

[Lis85] Liskovets, V. A., "Enumeration of nonisomorphic planar maps." *Selecta Math. Soviet.* **4** (1985), No 4, 304–323.

[Lis00] Liskovets, V. A., "Enumeration of planar rooted bipartite eulerian maps." *Abstracts of VIII Belarus. Mathem. Conf., June 2000*, Minsk, Institute of Mathematics of NAS of Belarus (2000), Part 4, 41 (in Russian).

[Mul66] Mullin, R. C., "On the average number of trees in certain maps." *Canad. J. Math.* **18** (1966), No 1, 33–41.

[Sch98] Schaeffer, G., *Conjugaison d'arbes et cartes combinatoires aléatoires.* Thèse, L'Université Bordeaux, 1998, 231 p. (in French).

[Sch00] Schaeffer, G., Private communication (May 2000).

[SzaW97] Szabo, R. J., and Wheater, J. F., "Curvature matrix models for dynamical triangulations and the Itzykson-Di Francesco formula." *Nucl. Phys.* **B 491** (1997), No 3, 689–723 (also preprint arXiv hep-th/9609237).

[Tut62] Tutte, W. T., "A census of slicings." *Canad. J. Math.* **14** (1962), No 4, 708–722.

[Tut63] Tutte, W. T., "A census of planar maps." *Canad. J. Math.* **15** (1963), No 2, 249–271.

[Wal75] Walsh, T. R. S., "Hypermaps versus bipartite maps." *J. Comb. Theory*, Ser. B **18** (1975), No 2, 155–163.

[WalL75] Walsh, T. R. S., and Lehman, A. B., "Counting rooted maps by genus. III: Nonseparable maps." *J. Comb. Theory*, Ser. B **18** (1975), No 3, 222–259.

[Zvo96] Zvonkin, A., "Matrix integrals and map enumeration: An accessible introduction." *Mathem. Comput. Modeling* **26** (1997), 281–304.

Table 1: Unicursal maps

n	$U^+(n)$	$U_2^+(n)$	$U'(n)$	$U_2'(n) = U_1'(n)/2$
2	2	1	6	2
3	9	3	40	12
4	38	11	280	80
5	214	62	2016	560
6	1253	342	14784	4032
7	7925	2152	109824	29568
8	51620	13768	823680	219648
9	346307	91800	6223360	1647360
10	2365886	622616	47297536	12446720
11	16421359	4301792	361181184	94595072
12	115384738	30100448	2769055744	722362368
13	819276830	213019072	21300428800	5538111488
14	5868540399	1521473984	164317593600	42600857600
15	42357643916	10954616064	1270722723840	328635187200
16	307753571520	79420280064	9848101109760	2541445447680
17	2249048959624	579300888960	76467608616960	19696202219520
18	16520782751969	4248201302400	594748067020800	152935217233920
19	121915128678131	31302536066560	4632774416793600	1189496134041600
20	903391034923548	231638727063040	36135640450990080	9265548833587200

Table 2: Eulerian and bi-eulerian maps

n	$E^+(n)$	$E'(n)$	$E'_{\mathrm{NS}}(n)$	$B'(n)$	$B'_{\mathrm{NS}}(n)$
1	1	1	1		
2	2	3	1	1	1
3	4	12	1		
4	12	56	2	6	2
5	34	288	6		
6	154	1584	19	54	8
7	675	9152	64		
8	3534	54912	230	594	54
9	18985	339456	865		
10	108070	2149888	3364	7371	442
11	632109	13891584	13443		
12	3807254	91287552	54938	99144	4032
13	23411290	608583680	228749		
14	146734695	4107939840	967628	1412802	39706
15	934382820	28030648320	4149024		
16	6034524474	193100021760	18000758	21025818	413358
17	39457153432	1341536993280	78905518		
18	260855420489	9390758952960	349037335	323686935	4487693
19	1741645762265	66182491668480	1556494270		
20	11732357675908	469294031831040	6991433386	5120138790	50348500

Institute of Mathematics, National Academy of Sciences, 220072, Minsk, Belarus
E-mail address: liskov@im.bas-net.by

Département D'Informatique, Université du Québec à Montréal, Montréal (Québec), Canada, H3C 3P8
E-mail address: walsh.timothy@uqam.ca

LATTICES INDUCED BY CHIP FIRING GAMES AND RELATED MODELS

CLÉMENCE MAGNIEN

Abstract. In this paper we study three classes of models widely used in physics, computer science and social science: the Chip Firing Game, the Abelian Sandpile Model and the Chip Firing Game on a mutating graph. We study the sets of configurations reachable from a given initial configuration, called the *configuration space* of a model, and try to determine their main properties. In order to acheive this, we study the order induced over the configurations by the evolution rule. This makes it possible to compare the power of expression of these models, *i.e.* determine which orders can be obtained by each model. It is known that these orders all are lattices, a special kind of partially ordered set. Although the Chip Firing Game on a mutating graph is a generalisation of the usual Chip Firing Game, we prove that both models exactly generate the same configuration spaces. We prove also that the class of lattices induced by the Abelian Sandpile Model is strictly included in the class of lattices induced by Chip Firing Game, but contains the class of distributive lattices, a very well known class. This leads to interesting questions concerning orders and lattices theory, since the class of lattices induced by Chip Firing Game is itself an uncharacterised class of lattices, therefore two new classes of lattices naturally appeared in this context.

Résumé. Dans cet article nous étudions trois classes de modèles étudiés en physiques, en informatique et en sciences sociales: le Chip Firing Game, l'Abelian Sandpile Model et le Chip Firing Game sur des graphes changeants. Nous étudions les ensembles de configurations atteignables à partir d'une configuration initiale donnée. Ces ensembles sont appelés espaces des configurations du modèle, et nous essayons de déterminer leurs propriétés principales. Pour cela, nous étudions l'ordre induit sur les configurations par la règle d'évolution.

Cela permet de comparer le pouvoir d'expression de ces modèles, c'est à dire de déterminer quels ordres peuventêtre obtenus par chaque modèle. Il est connu que ces ordres sont tous des treillis, un genre d'ordre particulier. Bien que le Chip Firing Game sur des graphes changeants soit une généralisation du Chip Firing Game habituel, nous prouvons que ces deux modèles génèrent exactement les mêmes espaces de configurations.

Nous prouvons également que la classe de treillis induits par l'Abelian Sandpile Model est incluse strictement dans la classe des treillis induits par Chip Firing Game, mais qu'elle contient la classe des treillis distributifs, une classe très connue. Cela mène à des questions intéressantes concernant la théorie des ordres et des treillis, car la classe des treillis induits par Chip Firing Game est elle-mme une classe de treillis non caractérisée, donc deux nouvelles classes de treillis apparaissent naturellement dans ce contexte.

1. Introduction

The Chip Firing Game (CFG) and the Abelian Sandpile Model (ASM) are closely related models studied in physics [BTW87], computer science [GMP98] and social science [Big97, Big99, Heu99]. These three models are variations on the following game: given a graph and a distribution of chips on its vertices, one may select a vertex that contains at least as many chips as its outdegree, and move one chip from this vertex to each of its neighbours along each outgoing edge. Many questions have naturally arisen and given matter for reseach about this game: given a graph and an initial distribution of chips, does the game stop after some time, or can it be played forever [BLS91, BL92, Eri96] ? For a given graph, what are the properties of the distribution of chips such that no move is possible ? This has led to the algebraic study of some of these distributions called *recurrent configurations*, [DRSV95, CR00, Big99]. Given a graph, what can be said about all the distributions that can be reached from a given initial distribution [LP01, MPV01] ?

This paper is concerned with this last question. Our purpose is to study the set of reachable distribution of chips, or *configurations* of a game, which we will call the *configuration space* of the game. This set is naturally ordered by the relation of reachability induced by the evolution rule. We will study the structure of such sets, and try to determine the differences between the models with respect to this aspect, *i.e.* given such an order, which is the configuration space of a game, we will try to decide if it is isomorphic to the configuration space of another game. For instance, given the configuration space of a CFG, does there exist a MCFG or an ASM such that its configuration space is isomorphic to it ? We will do so by trying to transform the CFG into an ASM or a MCFG without changing its configuration space.

In Section 2, we give the definitions and known results used in this paper, as well as the exact definition of each model. In Section 3, we will compare the configuration spaces of CFGs and ASMs. Since an ASM is a special CFG, we will study under which condition a CFG can be transformed into an ASM. We will give a sufficient (but not necessary) condition of such an transformation, and we will give an example of a CFG that cannot be transformed into an ASM.

2. Definitions and known results

We will need some definitions from order and lattice theory to describe the configuration spaces of the models. We five them first, after what we give the precise definitions for each models, as well as the previously known results about them.

2.1. **Posets and lattices.** A *partially ordered set* (or *poset*) is a set equipped with an order relation $\leq$ (*i.e.* a transitive, reflexive and antisymmetric relation). If x and y are two elements of a poset, we say that x is *covered* by y (or y *covers* x), and write $x \prec y$ (or $y \succ x$) if $x < y$ and $x \leq z < y$ implies $z = x$. The *interval* $[x, y]$ is the set $\{z, z \geq x, z \leq y\}$. To represent a poset P we will use its Hasse diagram, defined as follows :

- each element x of P is represented by a point p_x of the plane,
- if $x < y$, then p_x is lower than p_y, and
- p_x and p_y are joined by a line if and only if $x \prec y$.

Two posets P and P' are *isomorphic* if there exists a bijection $\varphi : P \longrightarrow P'$ satisfying: for all $x, y \in P$, $x \leq y \iff \varphi(x) \leq \varphi(y)$.

A poset L is a *lattice* if any two elements x, y of L have a least upper bound (called *join* and denoted by $x \vee y$) and a greatest lower bound (called *meet* and denoted by $x \wedge y$). $x \vee y$ is the (unique) smallest element greater than both x and y. $x \wedge y$ is defined dually. All the lattices considered here are finite, therefore they have a least and a greatest element, respectively denoted by 0_L and 1_L.

A lattice is a *hypercube of dimension* n if it is isomorphic to the set of all subsets of a set of n elements, ordered by inclusion. A lattice is *upper locally distributive* (denoted by *ULD*) [Mon90] if the interval between any element and the join of all its upper covers is a hypercube. All ULD lattices are ranked, *i.e.* all the paths in the covering relation from the minimal to the maximal element have the same length. A lattice L is *distributive* if it satisfies one of two following laws of distributivity (which are equivalent and imply each other):

$$\text{for all } x, y, z \in L, x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$$

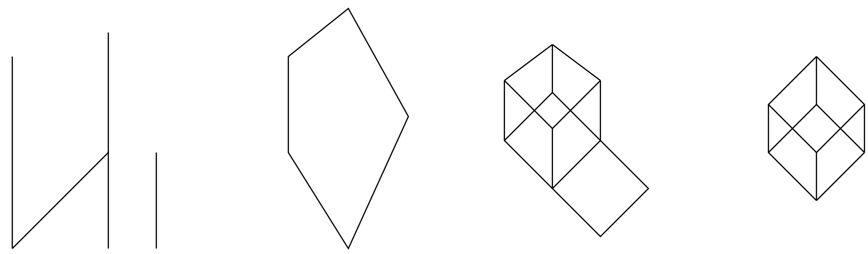


FIGURE 1. From left to right: a poset, a lattice, a distributive lattice and a hypercube

$$\text{for all } x, y, z \in L, x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z).$$

A distributive lattice is a lattice that is at the same time upper *and* lower locally distributive, *i.e.* if it is an ULD lattice, and if the interval between any element and the meet of all its lower covers is also a hypercube. In the sequel we will only be concerned with distributive and ULD lattices. Figure 1 shows examples of posets and different lattices.

For a more complete introduction to psoets and lattices, see for instance [DP90].

2.2. **The different kinds of chip firing games.** In this section we give the definition of the Abelian Sandpile Model, the Chip Firing Game and the Mutating Chip Firing Game. We begin by presenting the features shared by the three models, then we detail what is specific about each of them. We will see that some models are generalisations of others. Finally, we give some results about the CFG which will be useful in the paper.

2.2.1. *Definitions.* Each model is defined over a graph $G = (V, E)$, called the *support graph* of the game (undirected graphs will be regarded as directed by replacing each undirected edge $\{i, j\}$ by the two directed edges (i, j) and (j, i)). All graphs are supposed to be multigraphs, *i.e.* there might be more than one edge between two vertices, therefore all edge sets are supposed to be multisets. A *configuration* of the game is a mapping $\sigma : V \mapsto \mathbb{N}$ which associates a weight to each vertex; this weight can be considered as a number of *chips* stored in the vertex. The game is played with respect to the following evolution rule, also called the *firing* rule: if a vertex v contains at least as many chips as its outdegree, we can transfer a chip from v along each of its outgoing edges to the corresponding vertex. We call this process *firing* v. If σ is the configuration we start from, and σ' is obtained from σ by firing v, we write $\sigma \xrightarrow{v} \sigma'$, and we call σ a *predecessor* of σ'.

Note 2.1. *We consider that the firing rule cannot be applied to a sink (a vertex with no outgoing edges) because firing a sink does not change the configuration of the game and is therefore of no interest to us.*

The Chip Firing Game (CFG) [BL92] is defined over a directed graph. does not change throughout the game. We give an example of a CFG together with its configuration space in Figure 2 (in this example we have label led each edge (σ, σ') with the name of the vertex fired to reach σ' from σ). The Abelian Sandpile Model (ASM) [BTW87] is defined over an undirected graph, with a distinguished vertex called the *sink*. The sink can never be fired.

These three games are strongly convergent games [Eri93], which implies that, given an initial configuration, either a given game can be played forever, or it reaches a unique fixed point (where no firing is possible), called the *final configuration*, that does not depend on the order in which the vertices were fired. We will only consider games that reach a fixed point. We call these *convergent* games. We call *execution* of a game any sequence of firing which, starting from the initial configuration, reaches the final configuration.

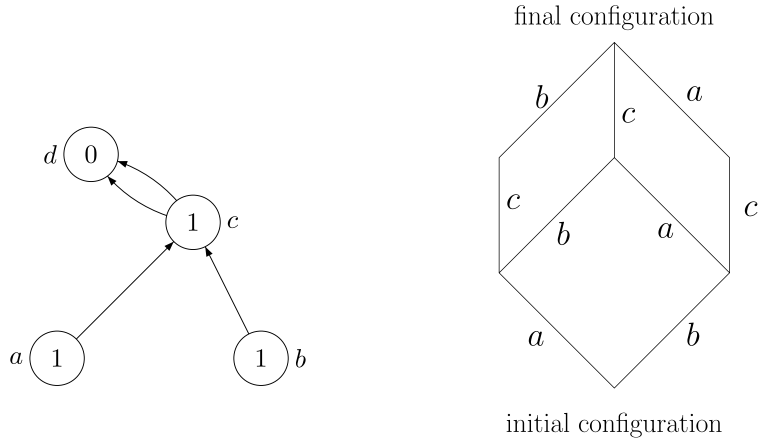


Figure 2. A CFG in its initial configuration and its configuration space

We call *configuration space* of a game the set of all configurations reachable from the initial configuration, ordered by the reflexive and transitive closure of the predecessor relation. It is known [BLS91, LP01, Eri96] that when a game is convergent, its configuration space is an ULD lattice. Given a game C, we denote its configuration space by $L(C)$. We will say that two games C and C' are *equivalent* if $L(C)$ is isomorphic to $L(C')$.

We denote by $L(CFG)$, $L(ASM)$ and $L(MCFG)$ the classes of lattices that are the configurations spaces of convergent CFGs, ASMs and MCFGs. It is possible to guarantee that a CFG is convergent by the presence in the support graph of a sink reachable from all vertices [LP01]. In the ASM, the presence of the sink which can never be fired (although it is not a sink in the undirected graph) guarantees that the game is always convergent.

These three models are very close to one another, and it is easy to see that some of them are generalisations of others (the class of lattices they induce are included in one another): since undirected graphs are particular directed graphs, one can consider an ASM as a particular CFG, where each undirected edge is replaced by two opposite directed edges, and where each undirected edge $\{v, \perp\}$ is replaced only by the directed edge $(v, \perp)$. Therefore, we obtain that $L(ASM) \subseteq L(CFG)$. Likewise, a CFG can be regarded as a MCFG where the graph remains the same after each firing, therefore $L(CFG) \subseteq L(MCFG)$.

2.2.2. *Previous results.* Now let us give some definitions and known results about CFGs, useful for simplifying the notations and proofs in the sequel.

Definition 2.2. *A convergent game is* simple *if, during an execution, each vertex is fired at most once.*

Theorem 2.3. [MPV01] *Any convergent CFG is equivalent to a simple CFG.*

All CFGs considered in the sequel will be such that such that their support graph has exactly one sink (denoted by $\perp$) and such that all vertices except $\perp$ are fired during an execution. This is always possible because, if this is not the case:

- either there is no sink and all vertices are fired during an execution; then we can add an isolated vertex to the graph, which becomes the sink ,
- or there exists a vertex v that is never fired during any execution but is not a sink; then we can remove its outgoing edges without changing the configuration space of the CFG, and v becomes the sink,
- or there is more than one sink; then we can merge them into a single vertex without changing the configuration space.

In the sequel, we will restrict ourselves to CFGs for which the support graph has no loops (*i.e.* no (v, v) edges). This is always possible because, for a simple CFG C, if there is a loop on a vertex v (v cannot be the sink), we can replace it by an edge $(v, \perp)$, and the resulting CFG is equivalent to C, since C is *simple.*

Given a vertex v, we denote by $d^-(v)$ its indegree, by $d^+(v)$ its outdegree, by $d_\perp(v)$ the number of edges from v to $\perp$, and we define $d(v) = d^+(v) - d_\perp(v)$. Given two vertices u and v, we denote by $d(u, v)$ the number of edges from u to v. Given a CFG and a vertex v of its support graph, the initial number of chips in v is denoted by $\sigma_0(v)$, and the number of chips in v in the final configuration by $\sigma_f(v)$ (since all the CFGs considered are simple, we have $\sigma_f(v) = \sigma_0(v) + d^-(v) - d^+(v)$). The number of chips that are *needed to fire* v is the difference between $\sigma_0(v)$ and $d^+(v)$, *i.e.* 0 if $\sigma_0(v) \geq d^+(v)$, and $d^+(v) - \sigma_0(v)$ otherwise.

The configuration spaces of simple CFGs can be described more easily than in the general case. Indeed we have the following results [MPV01, LP01]:

Lemma 2.4. *In a simple CFG, if, starting from the same configuration, two sequences of firing lead to the same configuration, then the vertices fired in each sequence are the same.*

This allows us to define the *shot-set* $s(\sigma)$ of a configuration σ as the set of vertices fired to reach σ from the initial configuration. Given a CFG with support graph (V, E), we say that a subset $X \subseteq V$ is a *valid* shot-set if there exists a configuration σ reachable from the initial configuration such that $s(\sigma) = X$. A list $(v_1, \ldots, v_n)$ of vertices is a *valid firing sequence* if, for each i, $\{v_1, \ldots, v_i\}$ is a valid shot-set. Moreover, the configuration space of a CFG is isomorphic to the lattice of the shot-sets of its configurations ordered by inclusion. The join of any two elements a and b is given by the following formula [LP01]:

$$s(a \vee b) = s(a) \cup s(b).$$

We give here another way of characterising the configuration space of a CFG with respect to the shot-sets: given a CFG and its vertex set, we can associate to each vertex v the configurations in which v can be fired. Amongst these, we distinguish the smallest configurations (the ones such that their shot-set is minimal with respect to the inclusion), and we say that their shot-set represent the *first moments* at which v can be fired. For instance, in Figure 2, the minimal configuration in which a and b can be fired is the minimal point of the lattice, and its shot-set is the empty set. We say that a and b can be fired from the beginning of any execution. The shot-sets of the minimal configurations in which c can be fired are $\{a\}$ and $\{b\}$, and we say that c can be fired after a *or* after b. The knowledge, for all vertices, of the first moments at which they can be fired, is a characterisation of the configuration space of a CFG, as it is stated in the following proposition:

Proposition 2.5. *Let C be a simple CFG and let $L = L(C)$. Let $\{v_1, \ldots, v_n, \perp\}$ be the set of vertices of C. For each $i, i = 1, \ldots, n$, let X_i be the set representing the first moments at which v_i can be fired,* i.e. *X_i contains the shot-sets of the minimal configurations at which v_i can be fired. Then L is completely determined by $\{v_1, \ldots, v_n\}$ and $\{X_1, \ldots, X_n\}$.*

3. Comparison of CFGs and ASMs

In this section, we compare the classes of configuration spaces induced by CFGs and ASMs. Since any ASM is equivalent to a CFG, we try to determine at which conditions a CFG is equivalent to an ASM. We will show that this is always the case when the support graph of the CFG has no cycle. Since we know that any distributive lattice is the configuration space of a CFG without cycles, we obtain as a corollary that the class of lattices induced by ASM contains the distributive lattices. However, not all CFGs are equivalent to some ASM. Since one can easily find examples of lattices in $L(ASM)$ that are not distributive, we obtain the result that $L(ASM)$is strictly between the distributive lattices and the lattices induced by CFG, which is surprising because these classes are very close.

We will try to transform CFGs into ASMs by using local transformations on the support graph of a game that do not change its configuration space. By using a combination of these transformations, it is possible in some cases to obtain a CFG such that its support graph contains one edge (u, v) for each edge (v, u), *i.e.* the graph is undirected therefore we have an ASM. We begin by giving the two basic transformations we will use on CFGs, which preserve the configuration space of a CFG. These transformations do not change the vertex set of a CFG, only its edge set and its initial configuration. Since we know from Proposition 2.5 that the configuration space is preserved if the first moments at which each vertex can be fired are not changed, we will ensure that these modifications do not change the first moments at which each vertex can be fired. Morover, a modification may not change this, but still allow a vertex to be fired more than once during an execution. We will also ensure that this does not happen.

Modification 1: Grounding
The grounding modification applied on a vertex v of a CFG consists in adding one chip to the initial configuration of v and adding one edge from v to the sink.

The purpose of the next modification is to multiply the indegree and the initial configuration of a given vertex by a given integer n.

Modification 2: Multipliying
The multiplying modification consists in multiplying by an integer factor the indegree and initial configuration of a given vertex v, without modifying the rest of the CFG. It consists in:

- multiplying by n the initial configuration of v,
- add $(n-1)d^+(v)$ edges from v to the sink,
- for each immediate predecessor u of v, multiplying by n the number of edges (u, v), and
- for each immediate predecessor u of v, adding $(n-1)d(u, v)$ chips to the initial configuration of u.

The next two lemmas prove that these modifications do not change the configuration space of the CFG to which they are applied. The first result being quite obvious, the proof is ommitted.

Lemma 3.1. *The CFG obtained by aplying the Grounding Modification to a CFG C is equivalent to C.*

Lemma 3.2. *The CFG obtained by applying the Multiplying Modification to a CFG C is equivalent to C.*

Now we give the main theorem of this Section:

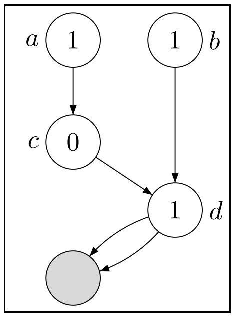
Theorem 3.3. *Algorithm 1 transforms a simple CFG with no cycle into an ASM in linear time.*

We will now give an example of the execution of Algorithm 1.

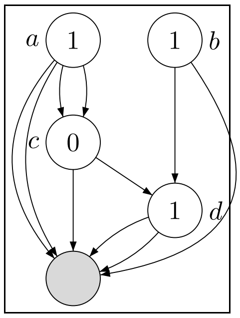
Algorithm 1 Transformation of a simple CFG graph with no cycle into an equivalent ASM

Input: A simple CFG C with support graph $G = (V, E)$ and initial configuration σ_0, such that G has no cycle
Output: An ASM equivalent to C
Compute $L(C)$.
For each $v \in V \setminus \{\perp\}$, $D[v] \leftarrow \max_{\sigma \in L(C)}(d^+(v) - \sigma(v))$.
Mark $\perp$.
$L = [\,]$.
while *there are unmarked vertices* **do**
  Choose an unmarked vertex v with all successors marked.
  Mark v.
  $L \leftarrow L\&[v]$.
Step 1: while L *is not empty* **do**
  $v \leftarrow \text{first}(L)$
  if $D[v] \leq d(v)$ **then**
    Apply the Multiplying Modification to v with a factor $\lceil (d(v) + 1)/D[v] \rceil$.
    For each edge (u, v) added by the Multiplying Modification, add 1 to $D[u]$.
  $L \leftarrow \text{tail}(L)$.
Step 2: for each $v \in V \setminus \{\perp\}$ **do**
  if $\sigma_0(v) + d^-(v) + d(v) \geq 2d^+(v)$ **then**
    Apply the Grounding Modification to v with a factor
    $\lceil (d(v) + d^-(v) + 1)/(2 \cdot d^+(v) - \sigma_0(v)) \rceil$.
Step 3: for each $v \in V \setminus \{\perp\}$ **do**
  for each *edge* (u, v) *in* E *ending on* v **do**
    Add one chip to the initial configuration of v.
    Add one edge (v, u) to E.

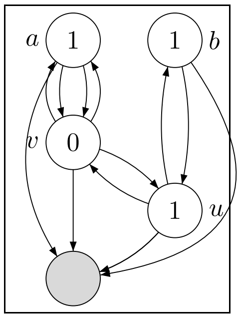
Original CFG. The first loops gives the list of vertices: $L = [d, c, b, a]$.



Step 1. Notice that in the obtained CFG, we have, for any vertex v and any configuration σ: $d^+(v) - \sigma(v) > d(v)$. Step 2. The CFG is not modified because we have, for any vertex v: $\sigma_0(v) + d^-(v) + d(v), 2 \cdot d^+(v)$.



Step 3. The adjuction of the edges (u, v) do not change the configuration space of the CFG because: the modifications applied to u are similar to applying the Grounding Modification to u, with v playing the role of the sink. After Step 2 we had $\sigma_0(v) + d^-(v) + d(v) < 2d^+(v)$. Therefore, after we have added $d(v, u)$ edges from u to v, we still have $\sigma_0(v) + d^-(v) < 2d^+(v)$, which means that v can still be fired only once. After Step 1 we had, for all configuration σ: $d^+(v) - \sigma(v) > d(v)$. Therefore, there is no configuration σ in which v cannot be fired, and in which adding edges from u to v allows v to be fired. The moments at which v can be fired are not modified.



Corollary 3.4. *Let C be a simple CFG with support graph $G = (V, E)$ such that G has no cycle. Then C is equivalent to an ASM.*

Theorem 3.5. [MPV01] *Given any distributive lattice L, there exists a CFG C with no cycle, such that $L(C) = L$.*

Corollary 3.6. *Any distributive lattice is the configuration space of an ASM.*

The distributive lattices are not the only ones that can be obtained by ASM. Indeed there are many CFGs without cycle that do not induce distributive lattices. Moreover, some CFGs with a cycle in their support graph can be transformed in ASM. An exemple of these is given in Figure 3.

Finally, it can be proved that the lattice of Figure 4, which is in $L(CFG)$, in not the configuration space of any ASM. Therefore we have the following theorem.

Theorem 3.7. $L(ASM) \subsetneq L(CFG)$.

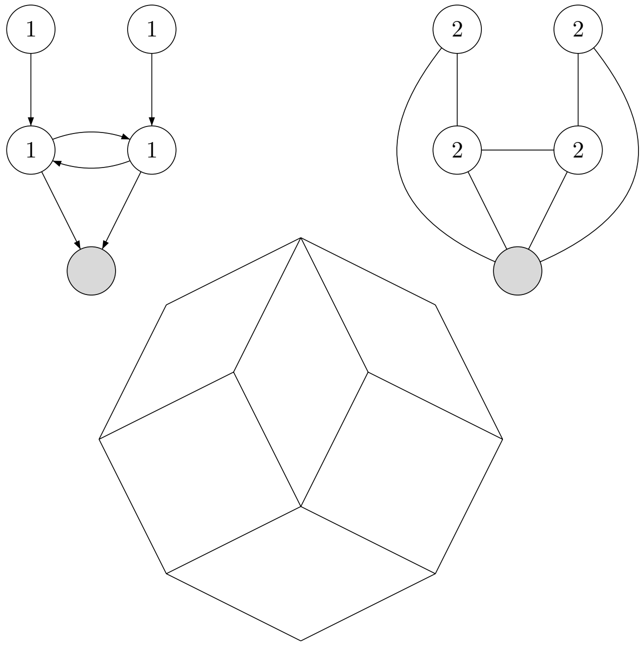


FIGURE 3. A CFG with a cycle in its support graph that is equivalent to an ASM. We give here the CFG, an equivalent ASM, and their configuration space

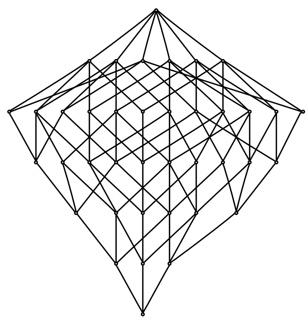


FIGURE 4. A lattice that can be obtained by CFG but not by ASM

By combining Theorem 3.7 and Corollary 3.6, we obtain that $L(ASM)$ is situated strictly between the distributive lattices and $L(CFG)$. This shows the complexity of the problems raised by the Chip Firing Game and the Abelian Sandpile Model in lattice theory: $L(CFG)$ and $L(ASM)$ are both between the distributive and ULD lattices, while there is no previously known lattices class satisfying this condition.

Conclusion and perspectives

In this paper, we have studied from the lattice point of view the Chip Firing Game and a closely related model: the Abelian Sandpile Model. It was already known that these models generate lattices, and that these lattices are in the class of Upper Locally Distributive (ULD) lattices. Our goal was to characterise the classes of lattices $L(CFG)$, $L(ASM)$ induced by each model (*i.e.* determine, given a ULD lattice, by which model(s) it can be obtained).

We have given an example of a CFG which is equivalent to no ASM. Since any ASM is equivalent (in lattice terms) to a CFG, this implies that $L(ASM) \subsetneq L(CFG)$. Then we have given a sufficient but not necessary condition at which a CFG can be transformed into an ASM, from which it can be shown that the class D of distributive lattices is included in $L(ASM)$. The class of lattices induced by CFG being somewhere between the distributive and the ULD lattices, we obtain that the class of lattices induced by ASM is a new class between the distributive lattices and $L(CFG)$. In other words, we have proved the following relation:

$$D \subsetneq L(ASM) \subsetneq L(CFG) \subsetneq \text{ULD}$$

The CFG and the ASM share the same definition, except that the first is defined on a directed graph, and the last on an undirected graph. This might seem like a strong difference, and the first idea that comes to mind is that $L(ASM)$ is much smaller than $L(CFG)$. The sufficient but not necessary condition we have given at which a CFG is equivalent to an ASM is that the support graph must contain no directed cycle. This is a strong condition, but in nonetheless shows that $L(ASM)$ is a very significant part of $L(CFG)$ (most of the CFGs we have studied are in fact equivalent to CFGs with no cycle). It also shows that the difference between these models does not reside in the fact that the graph is oriented or not, but on the existence or not of oriented cycles in the graph.

The fact that two important models, used in various domains like physics, computer science and social science, both induce strongly structured sets precisely situated between two classical types of lattices, shows the importance of the use of order theory in the context of dynamical models studies. In our case it is even more interesting to notice that the models introduce new classes of lattices which one may study from the order theoretical point of view.

References

[Big97] N. Biggs. Algebraic potential theory on graphs. *Bull. London Math. Soc.*, 29:641–682, 1997.

[Big99] N. Biggs. Chip firing and the critical group of a graph. *Journal of Algebraic Combinatorics*, 9:25–45, 1999.

[BL92] A. Björner and L. Lovász. Chip-firing games on directed graphs. *Journal of Algebraic Combinatorics*, 1:304–328, 1992.

[BLS91] A. Björner, L. Lovász, and W. Shor. Chip-firing games on graphs. *European Journal of Combinatorics*, 12:283–291, 1991.

[BTW87] P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality: an explanation of the $1/f$ noise. *Physics Review Letters*, 59(4):381–384, 1987.

[CR00] Robert Cori and Dominique Rossin. On the sandpile group of a graph. *European Journal of Combinatorics*, 21(4):447–459, 2000.

[DP90] B.A. Davey and H.A. Priestley. *Introduction to Lattices and Orders*. Cambridge university press, 1990.

[DRSV95] Deepak Dhar, P. Ruelle, S. Sen, and D. Verma. Algebraic aspects of sandpile models. *Journal of Physics*, A 28:805–831, 1995.

[Eri93] Kimmo Eriksson. *Strongly Convergent Games and Coxeter Groups*. PhD thesis, Kungl Tekniska Hogskolan, Sweden, 1993.

[Eri96] Kimmo Eriksson. Chip firing games on mutating graphs. *SIAM J. Discrete Math.*, 9:118–128, 1996.

[GMP98] E. Goles, M. Morvan, and H.D. Phan. Lattice structure and convergence of a game of cards. 1998. To appear in Annals of Combinatorics.

[Heu99] Jan van den Heuvel. Algorithmic aspects of a chip firing game. *London School of Economics, CDAM Research Reports.*, 1999. Preprint available at `http://www.cdam.lse.ac.uk/Reports/reports99.html`.

[LP01] Matthieu Latapy and Ha Duong Phan. The lattice structure of chip firing games. *Physica D*, 115:69–82, 2001. Preprint available at `http://www.liafa.jussieu.fr/~latapy/`.

[Mon90] Bernard Monjardet. The consequences of dilworth's work on lattices with unique irreductible decompositions. In K. P. Bogart, R. Freese, and J. Kung, editors, *The Dilworth theorems Selected papers of Robert P. Dilworth*, pages 192–201. Birkhauser, Boston, 1990.

[MPV01] Clémence Magnien, Ha Duong Phan, and Laurent Vuillon. Characterization of lattices induced by (extended) chip firing games. In *Discrete Mathematics and Theoretical Computer Science, Proceedings of the 1-st international conference Discrete Models: Combinatorics, Computation, and Geometry (DM-CCG'01)*, pages 229–244. MIMD, July 2001. Preprint available at `http://www.liafa.jussieu.fr/~magnien/`.

LIAFA, Univ. Paris 7, 2, place Jusseiu, 75251 Paris Cedex 05, Franc
E-mail address: `magnien@liafa.jussieu.fr`

COUNTING OCCURRENCES OF 132 IN A PERMUTATION

TOUFIK MANSOUR AND ALEK VAINSHTEIN

Abstract. We study the generating function for the number of permutations on n letters containing exactly $r \geq 0$ occurrences of 132. It is shown that finding this function for a given r amounts to a routine check of all permutations in S_{2r}.

Résumé. On è tudie la fonction gè nè ratrice pour les nombres de permutations de n lettres contenant exactement $r \geq 0$ occurrences de 132. Il est montrè que trouver cette fonction pour un r donnè se rè duit á un passage et examination de routine de toutes les permutations dans S_{2r}.

1. Introduction

Let $\pi \in S_n$ and $\tau \in S_m$ be two permutations. An *occurrence* of τ in π is a subsequence $1 \leq i_1 < i_2 < \cdots < i_m \leq n$ such that $(\pi(i_1), \ldots, \pi(i_m))$ is order-isomorphic to τ; in such a context, τ is usually called a *pattern*.

Recently, much attention has been paid to the problem of counting the number $\psi_r^\tau(n)$ of permutations of length n containing a given number $r \geq 0$ of occurrences of a certain pattern τ. Most of the authors consider only the case $r = 0$, thus studying permutations *avoiding* a given pattern. Only a few papers consider the case $r > 0$, usually restricting themselves to the patterns of length 3. In fact, simple algebraic considerations show that there are only two essentially different cases for $\tau \in S_3$, namely, $\tau = 123$ and $\tau = 132$. Noonan [No] has proved that $\psi_1^{123}(n) = \frac{3}{n}\binom{2n}{n-3}$. A general approach to the problem was suggested by Noonan and Zeilberger [NZ]; they gave another proof of Noonan's result, and conjectured that

$$\psi_2^{123}(n) = \frac{59n^2 + 117n + 100}{2n(2n-1)(n+5)}\binom{2n}{n-4}$$

and $\psi_1^{132}(n) = \binom{2n-3}{n-3}$. The latter conjecture was proved by Bóna in [B2]. A general conjecture of Noonan and Zeilberger states that $\psi_r^\tau(n)$ is P-recursive in n for any r and τ. It was proved by Bóna [B1] for $\tau = 132$. However, as stated in [B1], a challenging question is to describe $\psi_r^\tau(n)$, $\tau \in S_3$, explicitly for any given r.

In this note we suggest a new approach to this problem in the case $\tau = 132$, which allows to get an explicit expression for $\psi_r(n) = \psi_r^{132}(n)$ for any given r. More precisely, we present an algorithm that computes the generating function $\Psi_r(x) = \sum_{n \geq 0} \psi_r(n)x^n$ for any $r \geq 0$. To get the result for a given r, the algorithm performs certain routine checks for each element of the symmetric group S_{2r}. The algorithm has been implemented in C, and yielded explicit results for $1 \leq r \leq 6$.

The authors are sincerely grateful to M. Fulmek and A. Robertson for inspiring discussions.

2. Preliminary results

To any $\pi \in S_n$ we assign a bipartite graph G_π in the following way. The vertices in one part of G_π, denoted V_1, are the entries of π, and the vertices of the second part, denoted V_3, are the occurrences of 132 in π. Entry $i \in V_1$ is connected by an edge to occurrence

$j \in V_3$ if i enters j. For example, let $\pi = 57614283$, then π contains 5 occurrences of 132, and the graph G_π is presented on Figure 1.

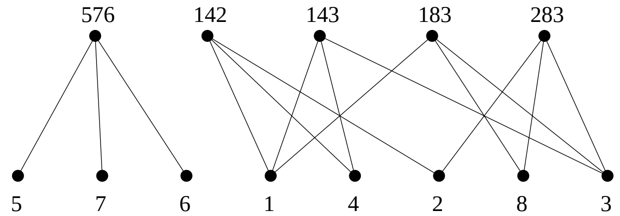


FIGURE 1. Graph G_π for $\pi = 57614283$

Let $\widetilde{G}$ be an arbitrary connected component of G_π, and let $\widetilde{V}$ be its vertex set. We denote $\widetilde{V}_1 = \widetilde{V} \cap V_1$, $\widetilde{V}_3 = \widetilde{V} \cap V_3$, $t_1 = |\widetilde{V}_1|$, $t_3 = |\widetilde{V}_3|$.

Lemma 2.1. *For any connected component $\widetilde{G}$ of G_π one has $t_1 \le 2t_3 + 1$.*

Proof. Assume to the contrary that the above statement is not true. Consider the smallest n for which there exists $\pi \in S_n$ such that for some connected component $\widetilde{G}$ of G_π one has

$$t_1 > 2t_3 + 1. \tag{$*$}$$

Evidently, $\widetilde{G}$ contains more than one vertex, since otherwise $t_1 = 1$, $t_3 = 0$, which contradicts $(*)$. Let l be the number of leaves in $\widetilde{G}$ (recall that a leaf is a vertex of degree 1). Clearly, all the leaves belong to $\widetilde{V}_1$; the degree of any other vertex in $\widetilde{V}_1$ is at least 2, while the degree of any vertex in $\widetilde{V}_3$ equals 3. Calculating the number of edges in $\widetilde{G}$ by two different ways, we get $l + 2(t_1 - l) \le 3t_3$, which together with $(*)$ gives $l > t_3 + 2$. This means that there exist two leaves $u, v \in \widetilde{V}_1$ incident to the same vertex $a \in \widetilde{V}_3$.

Let $w \in \widetilde{V}_1$ be the third vertex incident to a. If w is a leaf, then $\widetilde{G}$ contains only four vertices a, u, v, w, and hence $t_1 = 3$, $t_3 = 1$, which contradicts $(*)$. Hence, the degree of w is at least 2. Delete the entries u, v from π and consider the corresponding permutation $\pi' \in S_{n-2}$. Denote by $\widetilde{G}'$ the connected component of $G_{\pi'}$ containing w. Since the degree of w in $\widetilde{G}$ was at least 2, we see that $\widetilde{G}'$ is obtained from $\widetilde{G}$ by deleting vertices u, v, and a. Therefore, $t_1' = t_1 - 2$, $t_3' = t_3 - 1$, and hence $t_1' > 2t_3' + 1$, a contradiction to the minimality of n. □ □

Denote by G_π^n the connected component of G_π containing entry n. Let $\pi(i_1), \dots, \pi(i_s)$ be the entries of π belonging to G_π^n, and let $\sigma = \sigma_\pi \in S_s$ be the corresponding permutation. We say that $\pi(i_1), \dots, \pi(i_s)$ is the *kernel* of π and denote it $\ker \pi$; σ is called the *shape* of the kernel, or the *kernel shape*, s is called the *size* of the kernel, and the number of occurrences of 132 in $\ker \pi$ is called the *capacity* of the kernel. For example, for $\pi = 57614283$ as above, the kernel equals 14283, its shape is 14253, the size equals 5, and the capacity equals 4.

The following statement is implied immediately by Lemma 2.1.

Theorem 2.2. *Let $\pi \in S_n$ contain exactly r occurrences of* 132*, then the size of the kernel of π is at most $2r + 1$.*

We say that ρ is a *kernel permutation* if it is the kernel shape for some permutation π. Evidently ρ is a kernel permutation if and only if $\sigma_\rho = \rho$.

Let $\rho \in S_s$ be an arbitrary kernel permutation. We denote by $S(\rho)$ the set of all the permutations of all possible sizes whose kernel shape equals ρ. For any $\pi \in S(\rho)$ we define the *kernel cell decomposition* as follows. The number of cells in the decomposition equals $s(s+1)$. Let $\ker\pi = \pi(i_1),\dots,\pi(i_s)$; the *cell* $C_{ml} = C_{ml}(\pi)$ for $1 \le l \le s+1$ and $1 \le m \le s$ is defined by

$$C_{ml}(\pi) = \{\pi(j)\; i_{l-1} < j < i_l,\ \pi(i_{\rho^{-1}(m-1)}) < \pi(j) < \pi(i_{\rho^{-1}(m)})\},$$

where $i_0 = 0$, $i_{s+1} = n+1$, and $\alpha(0) = 0$ for any α. If π coincides with ρ itself, then all the cells in the decomposition are empty. An arbitrary permutation in $S(\rho)$ is obtained by filling in some of the cells in the cell decomposition. A cell C is called *infeasible* if the existence of an entry $a \in C$ would imply an occurrence of 132 that contains a and two other entries $x, y \in \ker\pi$. Clearly, all infeasible cells are empty for any $\pi \in S(\rho)$. All the remaining cells are called *feasible*; a feasible cell may, or may not, be empty. Consider the permutation $\pi = 67382451$. The kernel of π equals 3845, its shape is 1423. The cell decomposition of π contains four feasible cells: $C_{13} = \{2\}$, $C_{14} = \emptyset$, $C_{15} = \{1\}$, and $C_{41} = \{6,7\}$, see Figure 2. All the other cells are infeasible; for example, C_{32} is infeasible, since if $a \in C_{32}$, then $a\pi'(i_2)\pi'(i_4)$ is an occurrence of 132 for any π' whose kernel is of shape 1423.

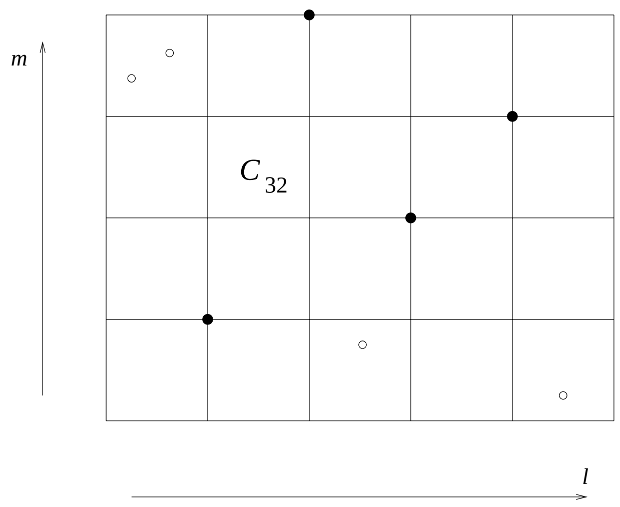


FIGURE 2. Kernel cell decomposition for $\pi \in S(1423)$

As another example, permutation $\widetilde{\pi} = 10\,11\,7\,12\,4\,6\,5\,8\,3\,9\,2\,1$ belongs to the same class $S(1423)$. Its kernel is $7\,12\,8\,9$, and the feasible cells are $C_{13} = \{4,6,5\}$, $C_{14} = \{3\}$, $C_{15} = \{2,1\}$, $C_{41} = \{10,11\}$.

Given a cell C_{ij} in the kernel cell decomposition, all the kernel entries can be positioned with respect to C_{ij}. We say that $x = \pi(i_k) \in \ker\pi$ lies *below* C_{ij} if $\rho(k) < i$, and *above* C_{ij} if $\rho(k) \ge i$. Similarly, x lies to the *left* of C_{ij} if $k < j$, and to the *right* of C_{ij} if $k \ge j$. As usual, we say that x lies to the *southwest* of C_{ij} if it lies below C_{ij} and to the left of it; the other three directions, northwest, southeast, and northeast, are defined similarly.

The following statement plays a crucial role in our considerations.

Lemma 2.3. *Let $\pi \in S(\rho)$ and $\pi(i_k) \in \ker\pi$, then any cell C_{ml} such that $l > k$ and $m > \rho(k)$ is infeasible.*

Proof. Assume to the contrary that there exist l and m as above such that C_{ml} is feasible, and consider the partition of the entries in $\ker\pi$ into two subsets: κ_1 containing the entries of $\ker\pi$ that lie to the southwest of C_{ml} and κ_2 containing the rest of the entries. Observe that $\pi(i_k) \neq n$, since $m > \rho(k)$ implies $n \geq \pi(i_{\rho^{-1}(m)}) > \pi(i_k)$. Moreover, $\pi(i_k) \in \kappa_1$ and $n \in \kappa_2$. Since $\pi(i_k)$ and n belong to the same connected component of G_π, there exists at least one occurrence of 132 whose elements are distributed between κ_1 and κ_2. Let a denote the minimal entry in this occurrence, let c denote its maximal entry, and let b denote the remaining entry.

Evidently, $a \in \kappa_1$. Assume first that $c \in \kappa_2$.

If c lies to the left of C_{ml}, then the existence of $z \in C_{ml}$ would imply that acz is an occurrence of 132, and hence C_{ml} is infeasible.

If c lies to the northeast of C_{ml} and b lies above C_{ml}, then the existence of $z \in C_{ml}$ would imply that zcb is an occurrence of 132, and hence C_{ml} is infeasible. If b lies below C_{ml}, then the existence of $z \in C_{ml}$ would imply that azb is an occurrence of 132, and hence C_{ml} is infeasible.

If c lies to the southeast of C_{ml}, then the existence of $z \in C_{ml}$ would imply that azc is an occurrence of 132, and hence C_{ml} is infeasible.

It remains to consider the case $c \in \kappa_1$, which means that b belongs to κ_2 and lies to the southeast of C_{ml}. Hence, the existence of $z \in C_{ml}$ would imply that azb is an occurrence of 132, and hence C_{ml} is infeasible. □ □

As an easy corollary of Lemma 2.3, we get the following proposition. Let us define a partial order $\prec$ on the set of all feasible cells by saying that $C_{ml} \prec C_{m'l'} \neq C_{ml}$ if $m \geq m'$ and $l \leq l'$.

Lemma 2.4. *$\prec$ is a linear order.*

Proof. Assume to the contrary that there exist two feasible cells C_{ml} and $C_{m'l'}$ such that $l < l'$ and $m < m'$, and consider the entry $x = \pi(i_l) \in \ker \pi$. By Lemma 2.3, $x > \pi(i_{\rho^{-1}(m'-1)})$, that is, x lies above the cell $C_{m'l'}$, since otherwise $C_{m'l'}$ would be infeasible. For the same reason, $y = \pi(i_{\rho^{-1}(m'-1)})$ lies to the right of $C_{m'l'}$, and hence to the right of x. Therefore, the existence of $z \in C_{ml}$ would imply that zxy is an occurrence of 132, and hence C_{ml} is infeasible, a contradiction. □ □

Consider now the dependence between two nonempty feasible cells lying on the same horizontal or vertical level.

Lemma 2.5. *Let C_{ml} and $C_{ml'}$ be two nonempty feasible cells such that $l < l'$. Then for any pair of entries $a \in C_{ml}$, $b \in C_{ml'}$, one has $a > b$.*

Proof. Assume to the contrary that there exists a pair $a \in C_{ml}$, $b \in C_{ml'}$ such that $a < b$. Consider the entry $x = \pi(i_l) \in \ker \pi$. By Lemma 2.3, $x > b$, since otherwise $C_{ml'}$ would be infeasible. Hence axb is an occurrence of 132, which means that both a and b belong to $\ker \pi$, a contradiction. □ □

Lemma 2.6. *Let C_{ml} and $C_{m'l}$ be two nonempty feasible cells such that $m < m'$. Then any entry $a \in C_{ml}$ lies to the right of any entry $b \in C_{m'l}$.*

Proof. Assume to the contrary that there exists a pair $a \in C_{ml}$, $b \in C_{m'l}$ such that a lies to the left of b. Consider the entry $y = \pi(i_{\rho^{-1}(m'-1)}) \in \ker \pi$. By Lemma 2.3, y lies to the right of b, since otherwise $C_{m'l}$ would be infeasible. Hence aby is an occurrence of 132, which means that both a and b belong to $\ker \pi$, a contradiction. □ □

Lemmas 2.4–2.6 yield immediately the following two results.

Theorem 2.7. *Let $\widetilde{G}$ be a connected component of G_π distinct from G_π^n. Then all the vertices in $\widetilde{V}_1$ belong to the same feasible cell in the kernel cell decomposition of π.*

Let $F(\rho)$ be the set of all feasible cells in the kernel cell decomposition corresponding to permutations in $S(\rho)$, and let $f(\rho) = |F(\rho)|$. We denote the cells in $F(\rho)$ by $C^1, \dots, C^{f(\rho)}$ in such a way that $C^i \prec C^j$ whenever $i < j$.

Theorem 2.8. *For any given sequence $\alpha_1, \dots, \alpha_{f(\rho)}$ of arbitrary permutations there exists a unique $\pi \in S(\rho)$ such that the content of C^i is order-isomorphic to α_i.*

3. Main Theorem and explicit results

Let ρ be a kernel permutation, and let $s(\rho)$, $c(\rho)$, and $f(\rho)$ be the size of ρ, the capacity of ρ, and the number of feasible cells in the cell decomposition associated with ρ, respectively. Denote by K the set of all kernel permutations, and by K_t the set of all kernel shapes for permutations in S_t. The main result of this note can be formulated as follows.

Theorem 3.1. *For any $r \geq 1$,*

$$\Psi_r(x) = \sum_{\rho \in K_{2r+1}} \left(x^{s(\rho)} \sum_{r_1 + \dots + r_{f(\rho)} = r - c(\rho)} \prod_{j=1}^{f(\rho)} \Psi_{r_j}(x) \right), \qquad (**)$$

where $r_j \geq 0$ for $1 \leq j \leq f(\rho)$.

Proof. For any $\rho \in K$, denote by $\Psi_r^\rho(x)$ the generating function for the number of permutations in $\pi \in S_n \cap S(\rho)$ containing exactly r occurrences of 132. Evidently, $\Psi_r(x) = \sum_{\rho \in K} \Psi_r^\rho(x)$. To find $\Psi_r^\rho(x)$, recall that the kernel of any π as above contains exactly $c(\rho)$ occurrences of 132. The remaining $r - c(\rho)$ occurrences of 132 are distributed between the feasible cells of the kernel cell decomposition of π. By Theorem 2.7, each occurrence of 132 belongs entirely to one feasible cell. Besides, it follows from Theorem 2.8, that occurrences of 132 in different cells do not influence one another. Therefore,

$$\Psi_r^\rho(x) = x^{s(\rho)} \sum_{r_1 + \dots + r_{f(\rho)} = r - c(\rho)} \prod_{j=1}^{f(\rho)} \Psi_{r_j}(x),$$

and we get the expression similar to $(**)$ with the only difference that the outer sum is over all $\rho \in K$. However, if $\rho \in K_t$ for $t > 2r + 1$, then by Theorem 2.2, $c(\rho) > r$, and hence $\Psi_r^\rho(x) = 0$. □ □

Theorem 3.1 provides a finite algorithm for finding $\Psi_r(x)$ for any given $r > 0$, since we have to consider all permutations in S_{2r+1}, and to perform certain routine operations with all shapes found so far. Moreover, the amount of searching can be decreased substantially due to the following proposition.

Proposition 3.2. *The only kernel permutation of capacity $r \geq 1$ and size $2r + 1$ is $2r - 1, 2r + 1, 2r - 3, 2r, \dots, 2r - 2j - 3, 2r - 2j, \dots, 1, 4, 2$. Its contribution to $\Psi_r(x)$ equals $x^{2r+1}\Psi_0^{r+2}(x)$.*

This proposition is proved easily by induction, similarly to Lemma 2.1. The feasible cells in the corresponding cell decomposition are $C_{2r-2j+1,2j+1}$, $j = 0, \dots, r$, and $C_{1,2r+2}$, hence the contribution to $\Psi_r(x)$ is as described.

By the above proposition, it suffices to search only permutations in S_{2r}. Below we present several explicit calculations.

Let us start from the case $r = 0$. Observe that $(**)$ remains valid for $r = 0$, provided the left hand side is replaced by $\Psi_r(x) - 1$; subtracting 1 here accounts for the empty permutation. So, we begin with finding kernel shapes for all permutations in S_1. The only shape obtained is $\rho_1 = 1$, and it is easy to see that $s(\rho_1) = 1$, $c(\rho_1) = 0$, and $f(\rho_1) = 2$

(since both cells C_{11} and C_{12} are feasible). Therefore, we get $\Psi_0(x) - 1 = x\Psi_0^2(x)$, which means that

$$\Psi_0(x) = \frac{1-\sqrt{1-4x}}{2x},$$

the generating function of Catalan numbers.

Let now $r = 1$. Since permutations in S_2 do not exhibit kernel shapes distinct from ρ_1, the only possible new shape is the exceptional one, $\rho_2 = 132$, whose contribution equals $x^3\Psi_0^3(x)$. Therefore, $(**)$ amounts to

$$\Psi_1(x) = 2x\Psi_0(x)\Psi_1(x) + x^3\Psi_0^3(x),$$

and we get the following result.

Corollary 3.3. (Bóna [B2, Theorem 5])

$$\Psi_1(x) = \frac{1}{2}\left(x - 1 + (1-3x)(1-4x)^{-1/2}\right);$$

equivalently,

$$\psi_1(n) = \binom{2n-3}{n-3}$$

for $n \geq 3$.

Let $r = 2$. We have to check the kernel shapes of permutations in S_4. An exhaustive search adds four new shapes to the previous list; these are 1243, 1342, 1423, and 2143; besides, there is the exceptional $35142 \in S_5$. Calculation of the parameters s, c, f is straightforward, and we get

Corollary 3.4.

$$\Psi_2(x) = \frac{1}{2}\left(x^2 + 3x - 2 + (2x^4 - 4x^3 + 29x^2 - 15x + 2)(1-4x)^{-3/2}\right);$$

equivalently,

$$\psi_2(n) = \frac{n^3 + 17n^2 - 80n + 80}{2n(n-1)}\binom{2n-6}{n-2}$$

for $n \geq 4$.

Let $r = 3, 4, 5, 6$; exhaustive search in S_6, S_8, S_{10}, and S_{12} reveals 20, 104, 503, and 2576 new nonexceptional kernel shapes, respectively, and we get

Corollary 3.5. *Let* $3 \leq r \leq 6$, *then*

$$\Psi_r(x) = \frac{1}{2}\left(P_r(x) + Q_r(x)(1-4x)^{-r+1/2}\right),$$

where

$$\begin{aligned} P_3(x) &= 2x^3 - 5x^2 + 7x - 2, \\ P_4(x) &= 5x^4 - 7x^3 + 2x^2 + 8x - 3, \\ P_5(x) &= 14x^5 - 17x^4 + x^3 - 16x^2 + 14x - 2 \end{aligned}$$

and

$$\begin{aligned} Q_3(x) &= -22x^6 - 106x^5 + 292x^4 - 302x^3 + 135x^2 - 27x + 2, \\ Q_4(x) &= 2x^9 + 218x^8 + 1074x^7 - 1754x^6 + 388x^5 + 1087x^4 \\ &\qquad - 945x^3 + 320x^2 - 50x + 3, \\ Q_5(x) &= -50x^{11} - 2568x^{10} - 10826x^9 + 16252x^8 - 12466x^7 + 16184x^6 - 16480x^5 \\ &\qquad + 9191x^4 - 2893x^3 + 520x^2 - 50x + 2. \end{aligned}$$

Equivalently,

$$\psi_r(n) = R_r(n)\frac{(2n-3r)!}{n!r!(n-r-2)!},$$

for $n \geq r+2$, where

$$\begin{aligned}
R_3(n) &= n^6 + 51n^5 - 407n^4 - 99n^3 + 7750n^2 - 22416n + 20160,\\
R_4(n) &= n^9 + 102n^8 - 282n^7 - 12264n^6 + 32589n^5 + 891978n^4\\
&\qquad - 7589428n^3 + 25452024n^2 - 39821760n + 23950080,\\
R_5(n) &= n^{12} + 170n^{11} + 1861n^{10} - 88090n^9 - 307617n^8 + 27882510n^7\\
&\qquad - 348117457n^6 + 2119611370n^5 - 6970280884n^4\\
&\qquad + 10530947320n^3 + 2614396896n^2 - 30327454080n + 29059430400.
\end{aligned}$$

The expressions for $P_6(x)$, $Q_6(x)$, and $R_6(n)$ are too long to be presented here.

4. Further results and open questions

As an easy consequence of Theorem 3.1 we get the following results due to Bóna [B1].

Corollary 4.1. *Let $r \geq 0$, then $\Psi_r(x)$ is a rational function in the variables x and $\sqrt{1-4x}$.*

In fact, Bóna has proved a stronger result, claiming that

$$\Psi_r(x) = P_r(x) + Q_r(x)(1-4x)^{-r+1/2}, \qquad (***)$$

where $P_r(x)$ and $Q_r(x)$ are polynomials and $1-4x$ does not divide $Q_r(x)$. We were unable to prove this result; however, it stems almost immediately from the following conjecture.

Conjecture 4.2. *For any kernel permutation $\rho \neq 1$,*

$$s(\rho) \geq f(\rho).$$

Indeed, it is easy to see that $\Psi_r(x)$ enters the right hand side of $(**)$ with the coefficient $2x\Psi_0(x)$, which is a partial contribution of the kernel shape $\rho_1 = 1$. Since $1 - 2x\Psi_0(x) = \sqrt{1-4x}$, we get by induction from $(**)$ that $\sqrt{1-4x}\Psi_r(x)$ equals the sum of fractions whose denominators are of the form $x^d(1-4x)^{r-c(\rho)-f(\rho)/2}$, where $d \leq f(\rho)$. On the other hand, each fraction is multiplied by $x^{s(\rho)}$, hence if $s(\rho) \geq f(\rho)$ as conjectured, then x^d in the denominator is cancelled. The maximal degree of $(1-4x)$ is attained for $\rho = \rho_1$, and is equal to $r-1$, and we thus arrive at $(***)$.

In view of our explicit results, we have even a stronger conjecture.

Conjecture 4.3. *The polynomials $P_r(x)$ and $Q_r(x)$ in $(***)$ have halfinteger coefficients.*

Another direction would be to match the approach of this note with the previous results on restricted 132-avoiding permutations. Let $\Phi_r(x;k)$ be the generating function for the number of permutations in S_n containing r occurrences of 132 and avoiding the pattern $12\dots k \in S_k$. It was shown previously that $\Phi_r(x;k)$ can be expressed via Chebyshev polynomials of the second kind for $r=0$ ([CW]) and $r=1$ ([MV]). Our new approach allows to get a recursion for $\Phi_r(x;k)$ for any given $r \geq 0$.

Let ρ be a kernel permutation, and assume that the feasible cells of the kernel cell decomposition associated with ρ are ordered linearly according to $\prec$. We denote by $l_j(\rho)$ the length of the longest increasing subsequence of ρ that lies to the north-east from C^j. For example, let $\rho = 1423$, as on Figure 2. Then $l_1(\rho)=1$, $l_2(\rho)=2$, $l_3(\rho)=1$, $l_4(\rho)=0$.

Theorem 4.4. *For any $r \geq 1$ and $k \geq 3$,*

$$\Phi_r(x;k) = \sum_{\rho \in K_{2r+1}} \left(x^{s(\rho)} \sum_{r_1+\cdots+r_{f(\rho)}=r-c(\rho)} \prod_{j=1}^{f(\rho)} \Phi_{r_j}(x;k-l_j(\rho)) \right),$$

where $r_j \geq 0$ for $1 \leq j \leq f(\rho)$ and $\Phi_r(x;m) \equiv 0$ for $m \leq 0$.

As in the case of $\Psi_r(x)$, the statement of the theorem remains valid for $r = 0$, provided the left hand side is replaced by $\Phi_r(x;k) - 1$. This allows to recover known explicit expressions for $\Phi_r(x;k)$ for $r = 0, 1$, and to get an expression for $r = 2$, which is too long to be presented here.

This approach can be extended even further, to cover also permutations containing r occurrences of 132 and avoiding other permutations in S_k, for example, $23\dots k1$.

References

[B1] M. Bona, The number of permutations with exactly r 132-subsequences is P-recursive in the size! *Adv. Appl. Math.* **18** (1997), 510–522.

[B2] M. Bona, Permutations with one or two 132-subsequences, *Discr. Math.* 1998 **181** (1998), 267–274.

[CW] T. Chow and J. West, Forbidden subsequences and Chebyshev polynomials, *Discr. Math.* **204** (1999), 119–128.

[MV] T. Mansour and A. Vainshtein, Restricted permutations, continued fractions, and Chebyshev polynomials, *Electron. J. Combin.* **7** (2000), R17.

[No] J. Noonan, The number of permutations containing exactly one increasing subsequence of length three, *Discr. Math.* **152** (1996), 307–313.

[NZ] J. Noonan and D. Zeilberger, The enumeration of permutations with a prescribed number of "forbidden" patterns, *Adv. Appl. Math.* **17** (1996), 381–407.

LaBRI, Université Bordeaux 1, 351 cours de Libération, 33405 Talence Cedex, France
E-mail address: toufik@labri.fr

Department of Mathematics and Department of Computer Science, University of Haifa, Mount Carmel, 31905 Haifa, Israel
E-mail address: alek@cs.haifa.ac.il

PLETHYSM OF SCHUR FUNCTIONS AND THE BASIC REPRESENTATION OF $A_2^{(2)}$

HIROSHI MIZUKAWA AND HIRO-FUMI YAMADA

Abstract. Some formulas of plethysms $p_r \circ S_\lambda$ and $p_r \circ Q_\lambda$ are presented. In particular, for the case $p_2 \circ S_\lambda$ with rectangular λ, a relation with the basic representation of the affine Lie algebra of type $A_2^{(2)}$ is discussed.

Résumé. Quelques formules de plethysms $p_r \circ S_\lambda$ et $p_r \circ Q_\lambda$ sont présenties. En particulier, pour le cas $p_2 \circ S_\lambda$ avec rectangulaire λ, une relation avec la repréntation basique de l'algèbre de Lie affine de type $A_2^{(2)}$ est discuté.

1. Introduction

Plethysm is one of fundamental problems in the theory of symmetric functions. Many authors have computed the plethysm $f \circ g$ for particular choices of symmetric functions f and g. This abstract deals with the plethysms $p_r \circ S_\lambda$ and $p_r \circ Q_\lambda$ involving the power sum symmetric functions, where S_λ is the Schur function indexed by the partition λ and Q_λ is Schur's Q-function indexed by the strict partition λ. The case $p_2 \circ S_\lambda$ where λ is the rectangular Young diagram is of particular interest. A nice formula is known for this plethysm which is proved by Carré-Leclerc [1] and Carini-Remmel [2].

One of our main contributions in this abstract is to understand their formula from the viewpoint of representations of affine Lie algebras. By looking at the homogeneous realization of the basic representation of $A_2^{(2)}$ we obtain another formula for $p_2 \circ S_\lambda$ where λ is rectangular.

Our formula should be regarded as an explicit expression of a homogeneous τ-function of a hierarchy of nonlinear differential equations, similar to the case of $A_1^{(1)}$ ([5]). Yet another result in this abstract is a formula for $p_r \circ Q_\lambda$, which is derived from certain factorization theorem of Q-functions.

As for this formula the relation with affine Lie algebras is unclear at this moment. However our proof requires a formula shown by You [17]. Therefore we believe that there is a nice explanation of our formula from the affine Lie algebra point of view.

The abstract is organized as follows. In Section 2 we fix some notations concerning the Schur functions. Section 3 is devoted to a brief review of bar-cores and bar-quotients of strict partitions. In Section 4 we discuss the basic representation of the affine Lie algebra of type $A_2^{(2)}$. The first main result is proved in Section 5. In Section 6 we announce some formulas of plethysms involving Schur's Q-functions. Proofs of these formulas are not given in this abstract.

2. Schur Functions

We denote by P_n the set of all partitions of n, SP_n the set of all strict partitions of n and OP_n the set of those partitions of n whose parts are odd numbers. Let χ_ρ^λ be the irreducible character of the symmetric group S_n, indexed by $\lambda \in P_n$ and evaluated at the conjugacy class ρ, and ζ_ρ^λ be the irreducible *negative* character of the double cover $\tilde{S}_n$ (cf.

[4]), indexed by $\lambda \in SP_n$ and evaluated at the conjugacy class ρ. Here we recall symmetric functions of variables $\boldsymbol{x} = (x_1, x_2, \cdots)$ which are discussed in this abstract.

Let $p_r(\boldsymbol{x}) = \sum_{i\geq 1} x_i^r$ be the power sum symmetric function for $r \geq 1$. The Schur functions and the "big" Schur functions are defined as follows:

$$S_\lambda(\boldsymbol{x}) = \sum_{\rho\in P_n} z_\rho^{-1}\chi_\rho^\lambda p_\rho(\boldsymbol{x}),$$
$$T_\lambda(\boldsymbol{x}) = \sum_{\rho\in OP_n} z_\rho^{-1}\chi_\rho^\lambda 2^{l(\rho)} p_\rho(\boldsymbol{x}).$$

For $\lambda \in SP_n$ define Schur's Q-function and P-function by

$$Q_\lambda(\boldsymbol{x}) = \sum_{\rho\in OP_n} z_\rho^{-1}\zeta_\rho^\lambda 2^{(l(\lambda)+l(\rho)+\epsilon(\lambda))/2} p_\rho(\boldsymbol{x}),$$
$$P_\lambda(\boldsymbol{x}) = 2^{-l(\lambda)}Q_\lambda(\boldsymbol{x}),$$

where

$$\epsilon(\lambda) = \begin{cases} 0 & \text{if } n - l(\lambda) \text{ is even,} \\ 1 & \text{if } n - l(\lambda) \text{ is odd.} \end{cases}$$

Let $\boldsymbol{x} = (x_1, x_2, \cdots)$ and $\boldsymbol{y} = (y_1, y_2, \cdots)$ be variables. We write

$$\boldsymbol{x}^r = (x_1^r, x_2^r, \cdots),$$
$$\boldsymbol{xy} = (x_i y_j; i \geq 1, j \geq 1).$$

When $\boldsymbol{y}$ is specialized as $\boldsymbol{y} = (1, \omega, \omega^2, \cdots, \omega^{r-1}, 0, 0, \cdots)$ for $\omega = \exp(2\pi\sqrt{-1}/r)$, we write

$$\boldsymbol{x}\omega_r = \boldsymbol{xy}.$$

3. Bar-Cores and Bar-Quotients of Partitions

Fix a positive integer r. An $(r+1)$-tuple of partitions $(\lambda^{c(r)}, \lambda^0, \ldots, \lambda^{r-1})$ is attached to λ ; $\lambda^{c(r)}$ is the r-core of λ and the collection $\lambda^{q(r)} = (\lambda^0, \ldots, \lambda^{r-1})$ is the r-quotient of λ (cf. [15]).

Definition 3.1. Let $\lambda = (\lambda_1, \cdots, \lambda_l)$ be a strict partition. We define the double of λ by

$$D(\lambda) = (\lambda_1, \cdots, \lambda_l \mid \lambda_1 - 1, \cdots, \lambda_l - 1),$$

in the Frobenius notation.

Example 3.2. Take $\lambda = (4, 2, 1) = \begin{matrix} \times & \times & \times & \times \\ \times & \times & & \\ \times & & & \end{matrix}$. The double of λ is

$$D(\lambda) = (5, 4, 4, 1) = \begin{matrix} \circ & \times & \times & \times & \times \\ \circ & \circ & \times & \times & \\ \circ & \circ & \circ & \times & \\ \circ & & & & \end{matrix}.$$

There is a remarkable property of $D(\lambda)$ as follows.

Proposition 3.3. ([10]) *Let λ be a strict partition and r be a positive odd integer.*
(1) *There exist strict partitions $\lambda^{bc(r)}$ and $\lambda^{b(0)}$ such that*

$$D(\lambda^{bc(r)}) = D(\lambda)^{c(r)},$$
$$D(\lambda^{b(0)}) = D(\lambda)^0.$$

(2)$D(\lambda)^{r-i}$ *is the partition conjugate to* $D(\lambda)^i$ *for* $1 \le i \le (r-1)/2$.

Definition 3.4. (1)The strict partition $\lambda^{bc(r)}$ is called the *r-bar core* of λ.
(2)The collection

$$\lambda^{bq(r)} = (\lambda^{b(0)}, \lambda^{b(1)}, \cdots, \lambda^{b(t)})$$

is called the *r-bar quotient* of λ where $t = (r-1)/2$ and $\lambda^{b(i)} = D(\lambda)^i$ for $1 \le i \le t$.

Example 3.5. We compute the 5-bar quotient of $\lambda = (15, 14, 13, 7, 6, 5, 3, 1)$. Adding 0's in the tail of $D(\lambda)$, if necessary, we always suppose that the size of the vector $D(\lambda)$ is the multiple of r. We see that

$$D(\lambda) + \delta_{15} = (30, 29, 28, 22, 21, 20, 18, 16, 13, 11, 7, 6, 5, 4, 3),$$

where $\delta_{15} = (14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0)$. To compute the 5-quotient of $D(\lambda)$, we put a set of beads on the positions assigned by $D(\lambda) + \delta_{15}$ as follows.

$$\begin{array}{ccccc} 0 & 1 & 2 & \textcircled{3} & \textcircled{4} \\ \textcircled{5} & \textcircled{6} & \textcircled{7} & 8 & 9 \\ 10 & \textcircled{11} & 12 & \textcircled{13} & 14 \\ 15 & \textcircled{16} & 17 & \textcircled{18} & 19 \\ \textcircled{20} & \textcircled{21} & \textcircled{22} & 23 & 24 \\ 25 & 26 & 27 & \textcircled{28} & \textcircled{29} \\ \textcircled{30} & 31 & 32 & 33 & 34 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \to ((4,3,1), (1^4), (3,1), (2,1^2), (4)).$$

Read each runner from the bottom and count the number of vacancies above each bead. Thus we have the corresponding partition $D(\lambda)^i$ $(i = 0, \cdots, 4)$. In this case the 5-quotient of $D(\lambda)$ reads $((4,3,1), (1^4), (3,1), (2,1^2), (4))$. This is the 5-bar quotient of λ; $\lambda^{bq(5)} = ((3,1), (1^4), (3,1))$. Moving each bead upwards in the runner successively as far as possible, we get

$$\begin{array}{ccccc} \textcircled{0} & \textcircled{1} & \textcircled{2} & \textcircled{3} & \textcircled{4} \\ \textcircled{5} & \textcircled{6} & \textcircled{7} & \textcircled{8} & \textcircled{9} \\ \textcircled{10} & \textcircled{11} & 12 & \textcircled{13} & 14 \\ 15 & \textcircled{16} & 17 & \textcircled{18} & 19 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \to (4,3,1) = D(3,1).$$

Thus we see that $D(\lambda)^{c(5)} = (4,3,1)$ and $\lambda^{bc(5)} = (3,1)$. Next we explain the *r-sign* and *r-bar sign* through the example above. Number the beads in the following two ways.
(1)The natural numbering according to the increasing order.
(2)The 5-numbering according to the *layers*.

$$\begin{array}{ccccc} \multicolumn{5}{c}{\text{natural numbering}} \\ 0 & 1 & 2 & \textcircled{3}_1 & \textcircled{4}_2 \\ \textcircled{5}_3 & \textcircled{6}_4 & \textcircled{7}_5 & 8 & 9 \\ 10 & \textcircled{11}_6 & 12 & \textcircled{13}_7 & 14 \\ 15 & \textcircled{16}_8 & 17 & \textcircled{18}_9 & 19 \\ \textcircled{20}_{10} & \textcircled{21}_{11} & \textcircled{22}_{12} & 23 & 24 \\ 25 & 26 & 27 & \textcircled{28}_{13} & \textcircled{29}_{14} \\ \textcircled{30}_{15} & 31 & 32 & 33 & 34 \end{array} \longrightarrow \begin{array}{ccccc} \multicolumn{5}{c}{\text{5-numbering}} \\ 0 & 1 & 2 & \textcircled{3}_4 & \textcircled{4}_5 \\ \textcircled{5}_1 & \textcircled{6}_2 & \textcircled{7}_3 & 8 & 9 \\ 10 & \textcircled{11}_7 & 12 & \textcircled{13}_9 & 14 \\ 15 & \textcircled{16}_{12} & 17 & \textcircled{18}_{13} & 19 \\ \textcircled{20}_6 & \textcircled{21}_{14} & \textcircled{22}_8 & 23 & 24 \\ 25 & 26 & 27 & \textcircled{28}_{15} & \textcircled{29}_{10} \\ \textcircled{30}_{11} & 31 & 32 & 33 & 34 \end{array}$$

When we compare the natural numbering with the 5-numbering we get a permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \\ 4 & 5 & 1 & 2 & 3 & 7 & 9 & 12 & 13 & 6 & 14 & 8 & 15 & 10 & 11 \end{pmatrix}.$$

The 5-sign of $D(\lambda)$, which is denoted by $\delta_5(D(\lambda))$, is defined to be the sign of σ.

$$\delta_5(D(\lambda)) = \mathrm{sgn}\sigma = 1.$$

To define the *r-bar sign* $\bar{\delta}_r(\lambda)$ of a strict partition λ, we draw the r-bar abacus of λ. If we take $\lambda = (15, 14, 13, 7, 6, 5, 3, 1)$ as before, then the 5-bar abacus of λ is

$$\begin{array}{ccccc} 0 & ① & 2 & ③ & 4 \\ ⑤ & ⑥ & ⑦ & 8 & 9 \\ 10 & 11 & 12 & ⑬ & ⑭ \\ ⑮ & 16 & 17 & 18 & 19 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array}.$$

Number the beads in the following two ways.
(1)The natural numbering according to the increasing order.
(2)The 5-bar numbering according to the *layers*.

$$\begin{array}{ccccc} \multicolumn{5}{c}{\text{natural numbering}} \\ 0 & ①_1 & 2 & ③_2 & 4 \\ ⑤_3 & ⑥_4 & ⑦_5 & 8 & 9 \\ 10 & 11 & 12 & ⑬_6 & ⑭_7 \\ ⑮_8 & 16 & 17 & 18 & 19 \end{array} \longrightarrow \begin{array}{ccccc} \multicolumn{5}{c}{\text{5-bar numbering}} \\ 0 & ①_4 & 2 & ③_6 & 4 \\ ⑤_1 & ⑥_7 & ⑦_5 & 8 & 9 \\ 10 & 11 & 12 & ⑬_8 & ⑭_3 \\ ⑮_2 & 16 & 17 & 18 & 19 \end{array}$$

Since the full description of the r-bar numbering is rather complicated, we refer the reader to [13, pp. 64–65] or [15, pp. 32–33]. When we compare the natural numbering with the 5-bar numbering we get a permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 6 & 1 & 7 & 5 & 8 & 3 & 2 \end{pmatrix}.$$

The 5-bar sign of λ is defined to be the sign of

$$\bar{\delta}_5(\lambda) = \mathrm{sgn}\sigma = -1.$$

Remark 3.6. (1)If $\lambda \in P_{rn}$ $(r, n \in \mathbb{N})$ has empty r-core, i.e., $\lambda^{c(r)} = \emptyset$, then

$$\delta_r(\lambda) = \chi^\lambda_{(r^n)} / |\chi^\lambda_{(r^n)}|.$$

(2) If $\lambda \in SP_{rn}$ $(r, n \in \mathbb{N}$, r is odd $)$ has empty r-bar core, i.e., $\lambda^{bc(r)} = \emptyset$, then

$$\bar{\delta}_r(\lambda) = \zeta^\lambda_{(r^n)} / |\zeta^\lambda_{(r^n)}|.$$

4. Basic Representation of $A_2^{(2)}$

We discuss in this section the basic representation of the affine Lie algebra of type $A_2^{(2)}$ following [7, 8]. Here the Schur functions, Schur's P and Q-functions are described in terms of the so called Sato variables: $u_j = p_j/j$ $(j \geq 1)$ for S_λ, $t_j = 2p_j/j$ $(j \geq 1,\ odd)$ for P_λ and Q_λ. We will denote them by $S_\lambda(u),\ P_\lambda(t),\ Q_\lambda(t)$, etc. Put $\Gamma = \mathbb{C}[t_j;\ j \geq 1,\ odd]$, whose basis is chosen as $\{P_\lambda;\ \lambda \in SP_n,\ n \in \mathbb{N}\}$. Associated with the Cartan matrix

$$(a_{ij})_{i,j \in \{0,1\}} = \begin{pmatrix} 2 & -4 \\ -1 & 2 \end{pmatrix},$$

the Lie algebra $\mathfrak{g}$ of type $A_2^{(2)}$ is generated by $e_i, f_i, \alpha_i^\vee (i = 0, 1)$ and d subject to the relations

$$[\alpha_i^\vee, \alpha_j^\vee] = 0, \qquad [\alpha_i^\vee, e_j] = a_{ij} e_j, \qquad [\alpha_i^\vee, f_j] = -a_{ij} f_j,$$
$$[e_i, f_j] = \delta_{i,j}\alpha_i^\vee, \qquad (\mathrm{ad} e_i)^{1-a_{ij}} e_j = (\mathrm{ad} f_i)^{1-a_{ij}} f_j = 0 \quad (i \neq j),$$

and

$$[d, \alpha_i^\vee] = 0, \qquad [d, e_j] = \delta_{j,0} e_j, \qquad [d, f_j] = -\delta_{j,0} f_j.$$

The Cartan subalgebra $\mathfrak{h}$ of $\mathfrak{g}$ is spanned by $\alpha_0^\vee, \alpha_1^\vee$ and d. Choose the basis $\{\alpha_0, \alpha_1, \Lambda_0\}$ for the dual space $\mathfrak{h}^*$ of $\mathfrak{h}$ by the pairing

$$< \alpha_i^\vee, \alpha_i > = a_{ij}, \qquad < \alpha_i^\vee, \Lambda_0 >= \delta_{i,0},$$
$$< d, \alpha_j > = \delta_{0,j}, \qquad < d, \Lambda_0 >= 0.$$

The fundamental imaginary root is $\delta = 2\alpha_0 + \alpha_1$.

The basic representation of $\mathfrak{g}$ is by definition the irreducible highest weight $\mathfrak{g}$-module with highest weight Λ_0. The weight system of the basic representation is well known:

$$P(\Lambda_0) = \{\Lambda_0 - p\delta + q\alpha_1 \ ; \ p \leq q^2, p \in \mathbb{N}, \ q \in \frac{1}{2}\mathbb{Z}\}.$$

A weight Λ on the parabola $\Lambda_0 - p\delta + q\alpha_1$ is said to be maximal in the sense that $\Lambda + \delta$ is no longer a weight. For any maximal weight Λ, the multiplicity of $\Lambda - n\delta$ $(n \in \mathbb{N})$ is known to be equal to $p(n)$, the number of partitions of n. A construction of the basic representation in "principal" grading is realized on the space $\Gamma^{(3)} = \mathbb{C}[t_j; \ j \geq 1, \ \text{odd}, \ j \not\equiv 0(\text{mod}3)]$ ([8]). A P-function $P_\lambda(t)$ is not necessarily contained in $\Gamma^{(3)}$. However, if the strict partition λ is a 3-bar core, then $P_\lambda(t) \in \Gamma^{(3)}$ and in fact $P_\lambda(t)$ is a maximal weight vector. More generally we "kill" the variables t_{3j} $(j \geq 1, \ \mathit{odd})$ in the P-function $P_\lambda(t)$ and consider the reduced P-function:

$$P_\lambda^{(3)}(t) := P_\lambda(t)|_{t_3 = t_9 = \cdots = 0} \in \Gamma^{(3)}.$$

It is shown in [14] that $P_\lambda^{(3)}(t)$ is a weight vector for any strict partition λ, and that

$$\begin{aligned} &\{P_\lambda^{(3)}(t); \ \lambda \text{ is a strict partition with no part divisible by 3}\} \\ =&\{P_\lambda^{(3)}(t); \ \lambda \text{ is a strict partition with } \lambda^{bq(3)} = (\emptyset, \lambda^{b(1)})\} \end{aligned}$$

form a weight basis for $\Gamma^{(3)}$. The weight of a reduced P-function with a given strict partition λ is known as follows. Draw the Young diagram λ and fill each cell with 0 or 1 in such a way that, in each row the sequence (010) repeats from the left as long as possible. If k_0 (resp. k_1) is the number of 0's (resp. 1's) written in the Young diagram, then the weight of the corresponding reduced P-function is $\Lambda_0 - k_0\alpha_0 - k_1\alpha_1$. A removable i-node (i=0,1) is a node $\boxed{i}$ of the boundary of λ which can be removed. An indent i-node (i=0,1) is a concave corner on the rim of λ where a node $\boxed{i}$ can be added. The action of $\mathfrak{g}$ to the reduced P-function $P_\lambda^{(3)}(t)$ is described as follows:

$$e_i P_\lambda^{(3)} = \sum_{\mu \in \mathcal{E}_i^1(\lambda)} P_\mu^{(3)},$$

where $\mathcal{E}_i^1(\lambda)$ is the set of the strict partitions which can be obtained by removing a removable i-node from λ, and

$$f_i P_\lambda^{(3)} = \sum_{\mu \in \mathcal{F}_i^1(\lambda)} P_\mu^{(3)},$$

where $\mathcal{F}_i^1(\lambda)$ is the set of the strict partitions which can be obtained by adding an indent i-node to λ. For instance

$$e_0 P^{(3)}_{(4,3,1)} = P^{(3)}_{(4,2,1)} + P^{(3)}_{(4,3)},$$
$$f_1 P^{(3)}_{(4,3,1)} = P^{(3)}_{(5,2,1)} + P^{(3)}_{(4,3,2)}.$$

Another realization of the basic representation is known, one in the homogeneous grading. The isomorphism between principal and homogeneous realizations is given by Leidwanger [9]. Put

$$\mathcal{B} = \mathbb{C}[u_j,\ s_{2j-1};\ j \geq 1].$$

Define the mapping Φ by

$$\begin{aligned} \Phi\ :\ \Gamma &\xrightarrow{\sim} \mathcal{B} \otimes \mathbb{C}[q, q^{-1}], \\ P_\lambda(t) &\longmapsto 2^{p(\lambda)} \bar{\delta}_3(\lambda) P_{\lambda^{b(0)}}(s) S_{\lambda^{b(1)}}(u) \otimes q^{m(\lambda)}, \end{aligned}$$

where

$$p(\lambda) = \sum_{\lambda_i \not\equiv 0 \pmod 3} \left[\frac{\lambda_i - 1}{3} \right],$$

and $m(\lambda)$ is determined by

$$\begin{aligned} m(\lambda) = &\ (\text{number of beads on the first runner of } \lambda) \\ &- (\text{number of beads on the second runner of } \lambda). \end{aligned}$$

For example

$$\Phi(P_{(7,5,3,1)}(t)) = 8 P_{(1)}(s) S_{(2,1,1)}(u) \otimes q.$$

Leidwanger [9] shows that Φ is indeed an isomorphism and that, if we denote by V the subalgebra of $\mathcal{B}$ generated by u_{2j} and $2^{2j-1}u_{2j-1} - s_{2j-1}$ $(j \geq 1)$, then

$$\Phi(\Gamma^{(3)}) = V \otimes \mathbb{C}[q, q^{-1}].$$

The representation of $\mathfrak{g}$ on $V \otimes \mathbb{C}[q, q^{-1}]$, which is induced by Φ, is the basic representation in the homogeneous grading. In fact, if we define the degree in $V \otimes \mathbb{C}[q, q^{-1}]$ by

$$\deg f(u, s) \otimes q^m = 2 \deg f(u, s) + m^2,$$

then $\deg \Phi(P_\lambda^{(3)})$ is equal to the number of 0-nodes in λ.

5. Main Result

We first recall a formula of plethysm which is due to Carini and Remmel [2]. For positive integers $l \geq m \geq 0$, let $W(l-m, m)$ denote the set of partitions $\mu = (\mu_1, \cdots, \mu_{2m})$ of $2m(l-m)$ such that $\mu_i + \mu_{2(l-m)+1-i} = 2m$ for $1 \leq i \leq m$. It is easily seen that $|W(l-m, m)| = \binom{l}{m}$. Denote by $\square(l-m, m)$ the rectangular Young diagram (m^{l-m}).

Theorem 5.1. ([2])

$$p_2 \circ S_{\square(l-m,m)} = \sum_{\mu \in W(l-m,m)} (-1)^{\sum_{i=1}^{l-m} \mu_i} S_\mu.$$

Now fix a strict partition $\Lambda_l = (3l-2, 3l-5, \cdots, 7, 4, 1)$ of length l, which corresponds to a maximal weight vector of the basic representation of $A_2^{(2)}$ in the principal grading. As before, each cell of the Young diagram of Λ_l is supposed to be filled with 0 or 1. Note that each concave corner on the rim of Λ_l is an indent 1-node. Let $\mathcal{F}^m = \mathcal{F}_1^m(\Lambda_l)$ $(0 \le m \le l)$ be the set of the strict partitions which can be obtained by adding m indent 1-nodes to Λ_l. It is obvious that $|\mathcal{F}^m| = \binom{l}{m}$. Our main result in this abstract is the following.

Theorem 5.2.

$$p_2 \circ S_{\square(l-m,m)} = \varepsilon(l,m) \sum_{\mu \in \mathcal{F}_1^m(\Lambda_l)} \bar{\delta}_3(\mu) S_{\mu^{b(1)}},$$

where

$$\varepsilon(l,m) = \begin{cases} (-1)^{\binom{m}{2}} & (0 \le m \le \frac{l}{2}) \\ (-1)^{l(m+1)+\binom{l-m}{2}} & (\frac{l}{2} \le m \le l). \end{cases}$$

Proof. Besides the sign factor $\varepsilon(l,m)$, we only have to show that $\mu^{b(1)} \in W(l-m, m)$ for any $\mu \in \mathcal{F}^m$. Recall the relation between 3-bar quotients and 3-quotients:

$$\mu^{b(1)} = D(\mu)^1.$$

First we see that the 3-abacus of $D(\Lambda_l)$ is

$$\begin{array}{ccc|c}
0 & 1 & 2 & 1 \\
3 & 4 & 5 & \\
 & \vdots & & \\
3l-3 & 3l-2 & 3l-1 & \\
3l & 3l+1 & 3l+2 & l+1 \\
3l+3 & 3l+4 & 3l+5 & \\
 & \vdots & & \\
6l-3 & 6l-2 & 6l-1 & 2l
\end{array}$$

If we add the indent 1-node to Λ_l at the i-th row, then, in the 3-abacus of $D(\Lambda_l)$, the beads at $3i-2$ and $6l-3i+1$ move to $3i-1$ and $6l-3i+2$, respectively:

$$\begin{array}{ccc|c}
0 & 1 & 2 & 1 \\
3 & 4 & 5 & \\
 & \vdots & & \\
3i-3 & 3i-2 & 3i-1 & i \\
 & \vdots & & \\
3l-3 & 3l-2 & 3l-1 & \\
3l & 3l+1 & 3l+2 & l+1 \\
3l+3 & 3l+4 & 3l+5 & \\
 & \vdots & & \\
6l-3i & 6l-3i+1 & 6l-3i+2 & 2l-i+1 \\
 & \vdots & & \\
6l-3 & 6l-2 & 6l-1 & 2l
\end{array}$$

Adding indent 1-nodes successively, we see that, in the 3-abacus of $D(\mu)$ $(\mu \in \mathcal{F}^m)$, the beads at i_1-th, i_2-th, $\cdots$, i_m-th rows in the first runner shift to the second runner as well as the beads at i_1-th, i_2-th, $\cdots$, i_m-th rows from the bottom. Then $D(\mu)_i^1 + D(\mu)_{2(l-m)+1-i}^1$ counts the number of the vacancies of the first runner up to the $2l$-th row; that is $2m$. This proves that $\mu^{b(1)} = D(\mu)^1 \in W(l-m,m)$. By a rather tedious computation, we can fit the sign factor $\varepsilon(l,m)$. $\square$

6. Plethysms Involving Q-functions

In this section we consider the plethysm $p_r \circ Q_\lambda$. For this purpose we need the following *factorization* theorem.

Theorem 6.1. *Let r be a positive odd integer. If $\lambda \in SP_{rn}$ has empty r-bar core, then*

$$2^{-l(\lambda)/2} Q_\lambda(\boldsymbol{x}\omega_r) = \bar{\delta}_r(\lambda) 2^{-l(\lambda^{b(0)})/2} Q_{\lambda^{b(0)}}(\boldsymbol{x}^r) T_{\lambda^{b(1)}}(\boldsymbol{x}^r) \cdots T_{\lambda^{b(t)}}(\boldsymbol{x}^r).$$

Remark 6.2. Since we assume that $\lambda^{bc(r)} = \emptyset$ in Theorem 6.1, we easily verify that

$$l(\lambda) \equiv l(\lambda^{b(0)}) (\mathrm{mod}\ 2).$$

Therefore $2^{(l(\lambda)-l(\lambda^{b(0)}))/2}$ is an integer.

Our proof of Theorem 6.1 relies on a formula shown by You[17]. In the process of the proof we obtain

Corollary 6.3.

$$\delta_r(D(\lambda)) = 1.$$

A special case of Theorem 6.1 is found in [11, 12]. Comparing the spin character table of $\tilde{S}_4$ with that of $\tilde{S}_{12}$, one observes that each character of $\tilde{S}_{12}$ evaluated at the conjugacy classes 3ρ is a linear combination of the characters of $\tilde{S}_4$ with integral coefficients which are simultaneously non-negative or non-positive.

$n=4$	(1^4)	$(3,1)$
$<4>$	2	1
$<3,1>$	4	-1

$n=12$	(3^4)	$(9,3)$	$3-$ *bar core*	
$<12>$	2	1	$\emptyset$	$+1<4>+0<3,1>$
$<11,1>$	-4	-2	$\emptyset$	$-2<4>-0<3,1>$
$<10,2>$	4	2	$\emptyset$	$+2<4>+0<3,1>$
$<9,3>$	4	-1	$\emptyset$	$+0<4>+1<3,1>$
$<8,4>$	-12	0	$\emptyset$	$-2<4>-2<3,1>$
$<7,5>$	12	0	$\emptyset$	$+2<4>+2<3,1>$
$<5,4,2,1>$	8	-2	$\emptyset$	$+0<4>+2<3,1>$
$<5,4,3>$	-16	1	$\emptyset$	$-2<4>-3<3,1>$
$<7,3,2>$	-8	-1	$\emptyset$	$-2<4>-1<3,1>$
$<8,3,1>$	8	1	$\emptyset$	$+2<4>+1<3,1>$
$<6,4,2>$	12	0	$\emptyset$	$+2<4>+2<3,1>$
$<6,5,1>$	-12	0	$\emptyset$	$-2<4>-2<3,1>$
$<9,2,1>$	-8	-1	$\emptyset$	$-2<4>-1<3,1>$
$<6,3,2,1>$	-8	2	$\emptyset$	$-0<4>-2<3,1>$
$<7,4,1>$	0	0	$(7,4,1)$	$+0<4>+0<3,1>$

We give a description of these coefficients. Put

$$t_{\lambda,\mu} = [Q_{\lambda^{b(0)}} T_{\lambda^{b(1)}} \cdots T_{\lambda^{b(t)}}, P_\mu],$$

where $[,]$ is an inner product satisfying $[P_\lambda, Q_\mu] = \delta_{\lambda,\mu}$. It is known that $t_{\lambda,\mu}$ is a non-negative integer [16]. The following theorem is a spin character version of *Littlewood's multiple formula* [11, 12].

Theorem 6.4. *Let r be a positive odd integer. If $\lambda \in SP_{rn}$ has empty r-bar core, then*

$$\zeta^{\lambda}_{r\rho} = \bar{\delta}_r(\lambda) \sum_{\mu \in SP_n} 2^{(l(\mu)-l(\lambda^{b(0)})+\epsilon(\mu)-\epsilon(\lambda))/2} t_{\lambda,\mu}\zeta^{\mu}_{\rho}.$$

for any $\rho \in OP_n$.

Another application of Theorem 6.1 is a formula of plethysm involving P and Q-functions.

Theorem 6.5. *For a positive odd integer r, we have*

$$p_r \circ P_\lambda = \sum_{\mu^{bc(r)}=\emptyset} \bar{\delta}_r(\mu) 2^{(l(\mu)-l(\mu^{b(0)}))/2} t_{\lambda,\mu} P_\mu,$$

$$p_r \circ Q_\lambda = 2^{l(\lambda)} \sum_{\mu^{bc(r)}=\emptyset} \bar{\delta}_r(\mu) 2^{(-l(\mu)-l(\mu^{b(0)}))/2} t_{\lambda,\mu} Q_\mu.$$

For the special case $\lambda = (n)$, the right hand side becomes simpler.

Theorem 6.6. *For a positive odd integer r, we have*

$$(p_r \circ Q_{(n)}) = \sum_{\mu \in H_{r,n}} \bar{\delta}_r(\mu) Q_\mu.$$

Here $H_{r,n} = \{\mu = (\mu_1 \dots \mu_l) \in SP_{rn} |\ \exists i \ s.t.\ \mu_i \equiv k (\mathrm{mod}\ r) \Rightarrow \exists_1 j \ s.t.\ \mu_j \equiv r-k (\mathrm{mod}\ r)\}$.

This formula should be compared with Theorem 5.1.

Example 6.7.

$$p_3 \circ Q_{(1)} = Q_{(3)} - Q_{(2,1)},$$
$$p_3 \circ Q_{(2)} = Q_{(6)} - Q_{(5,1)} + Q_{(4,2)} - Q_{(3,2,1)},$$
$$p_3 \circ Q_{(3)} = Q_{(9)} - Q_{(8,1)} + Q_{(7,2)} - Q_{(5,4)} - Q_{(6,2,1)} + Q_{(5,3,1)} - Q_{(4,3,2)},$$
$$p_5 \circ Q_{(2)} = Q_{(10)} - Q_{(9,1)} + Q_{(8,2)} - Q_{(7,3)} + Q_{(6,4)} - Q_{(5,4,1)} + Q_{(5,3,2)} - Q_{(4,3,2,1)}.$$

References

[1] C. Carré and B. Leclerc, Splitting the square of a Schur function into its symmetric and antisymmetric parts, J. Algebraic Combin. 4 (1995), no. 3, 201-231.

[2] L. Carini and J. Remmel, Formulas for the expansion of the plethysms $s_2[s_{(a,b)}]$ and $s_2[s_{(n^k)}]$, Discrete Math. 193 (1998), no. 1-3, 225-233.

[3] Y. Chen, A. Garsia and J. Remmel, Algorithms for plethysm, in Combinatorics and Algebra, Contemporary Math. 34(1984), 109-153.

[4] P. N. Hoffman and J. F. Humphreys, *Projective Representations of the Symmetric Groups,* Oxford, 1992.

[5] T. Ikeda and H. -F. Yamada, Polynomial τ-functions of the NLS-Toda hierarchy and the Virasoro singular vectors, Preprint, 2001

[6] A. T. James and A. Kerber, *The Representation Theory of the Symmetric Group,* Addison-Wesley Reading, MA, 1981.

[7] V. G. Kac, Infinite Dimentional Lie Algebras, 3rd. ed. , Cambridge, 1990.

[8] V. G. Kac, D. A. Kazhdan, J. Lepowsky and R. L. Wilson, Realization of the basic representations of the Euclidean Lie algebras, Adv. Math. 42 (1981), no. 1, 83-112.

[9] S. Leidwanger, Basic representations of $A^{(1)}_{n-1}$ and $A^{(2)}_{2n}$ and the combinatorics of partitions, Adv. Math. 141 (1999), no. 1, 119-154.

[10] I. G. Macdonald, *Symmetric Functions and Hall Polynomials, 2nd. ed.* , Oxford, 1995.

[11] H. Mizukawa and H.-F. Yamada, Littlewood's multiple formula for spin characters of symmetric groups, J. London Math. Soc. (2) 65 (2002), 1-9; doi: 10.1112/S0024610701002800.

[12] H. Mizukawa and H.-F. Yamada, Littlewood's multiple formula for spin characters of symmetric groups, Local Proceedings of FPSAC 2001, Arizona USA, 381-388.

[13] A. O. Morris and J. B. Olsson, On p-quotient for spin characters, J. Alg. 119 (1988), 51-82.

[14] T. Nakajima and H. -F. Yamada, Schur's Q-functions and twisted affine Lie algebras, Adv. Stud. in Pure Math. 28 (2000), 241-259.
[15] J. B. Olsson, Combinatorics and Representations of Finite Groups, Essen, 1993.
[16] J. R. Stembridge, Shifted tableaux and the projective representations of symmetric groups, Adv. Math. 74 (1989), no. 1, 87-134.
[17] Y. You, Polynomial solutions of the BKP hierarchy and projective representations of symmetric groups, Infinite-dimensional Lie algebras and groups (Luminy-Marseille, 1988), 449-464, Adv. Ser. Math. Phys. 7, World Sci. Publishing, 1989.

Department of Mathematics, Hokkaido University, Sapporo 060-0810, Japan
E-mail address: `mzh@math.sci.hokudai.ac.jp, mzh@math.okayama-u.ac.jp`

Department of Mathematics, Okayama University, Okayama 700-8530, Japan
E-mail address: `yamada@math.okayama-u.ac.jp`

SATURATED SIMPLICIAL COMPLEXES (EXTENDED ABSTRACT)

VALERIY MNUKHIN

Abstract. Among shellable complexes a certain class has maximal modular homology, and these are the so-called *saturated* complexes. We give a brief survey of their properties and characterize saturated complexes via p-ranks of incidence matrices and via structure of links.

Résumé. Parmi les complexes analysables, une certaine catégorie, que l'on appelle les *complexes saturés*, a une homologie modulaire maximale; nous donnons un bref aperçu des propriétés des complexes analysables et décrivons les complexes saturés grâce aux p-rangs d'incidence des matrices et à la structure de leur liens.

1. Introduction

The standard homology for a simplicial complex Δ is concerned with the $\mathbb{Z}$-module $\mathbb{Z}\Delta$ with basis Δ and the boundary map

$$\tau \mapsto \sigma_1 - \sigma_2 + \sigma_3 - \ldots \pm \sigma_k$$

which assigns to the face τ the alternating sum of the co-dimension 1 faces of τ. This defines a homological sequence over $\mathbb{Z}$ and hence over any domain with identity.

In [12] we started to investigate the same module with respect to a different homomorphism. This is the *inclusion map* $\partial : \mathbb{Z}\Delta \to \mathbb{Z}\Delta$ given by

$$\partial : \quad \tau \mapsto \sigma_1 + \sigma_2 + \sigma_3 + \ldots + \sigma_k.$$

Clearly, $\partial^2 \neq 0$. However, when coefficients are taken modulo an integer p then a simple calculation shows that in fact $\partial^p = 0$. One may attempt therefore to build a generalized *modular homology theory* of simplicial complexes, in particular when p is a prime. This kind of homology appears to be mentioned first in W Mayer [9] in 1947, further historical remarks and references can be found in [1, 12]. More recent papers on nilpotent homomorphisms include Dubois-Violette [6] and Kapranov [8].

We showed in [12] that in general modular homology does not behave nicely: It is not homotopy invariant and there are shellable complexes with the same h-vector but with different modular homology. Nevertheless, homology of any shellable complex can be embedded into a well-understood module constructed purely from the shelling of the complex. It follows in particular that the modular Betti numbers for an arbitrary shellable complex are bounded by functions of its h-vector only.

Shellable complexes which attain these bounds are of special interest and are called *saturated*. Here we investigate conditions which guarantee saturation. Our main results are Theorems 4.1 and 5.3 which characterize saturated complexes via p-ranks of incidence matrices and via structure of links respectively. As a corollary we prove that rank-selected subcomplexes of a saturated complex are saturated.

Supported in part by the London Mathematical Society.

2. Modular Homology of Shellable Complexes

Let F be a field, Ω be a finite set and k a non-negative integer. Let then M_k denote the F-vector space with k-element subsets of Ω as basis and put $M := \oplus_{0\le k} M_k$. The *inclusion map* is the linear map $\partial : M_k \to M_{k-1}$ defined on a basis by mapping each k-element subset of Ω to the sum of all its $(k-1)$-element subsets. If $\Delta \subseteq 2^\Omega$ is a simplicial complex, denote by M^Δ the subspace of M with basis Δ and let $M_k^\Delta := M^\Delta \cap M_k$. Then ∂ restricts to maps $M_k^\Delta \to M_{k-1}^\Delta$ for all k, and so we can attach to the complex Δ the sequence

$$\mathcal{M}^\Delta : \quad 0 \xleftarrow{\partial} M_0^\Delta \xleftarrow{\partial} M_1^\Delta \xleftarrow{\partial} \dots \xleftarrow{\partial} M_{k-1}^\Delta \xleftarrow{\partial} M_k^\Delta \xleftarrow{\partial} \dots$$

of submodules of M.

Throughout we suppose that p is a fixed prime and that F is a field of characteristic p. For any j and $0 < i < p$ consider the sequence

$$\dots \xleftarrow{\partial^*} M_{j-p}^\Delta \xleftarrow{\partial^*} M_{j-i}^\Delta \xleftarrow{\partial^*} M_j^\Delta \xleftarrow{\partial^*} M_{j+p-i}^\Delta \xleftarrow{\partial^*} M_{j+p}^\Delta \xleftarrow{\partial^*} \dots$$

in which ∂^* is the appropriate power of ∂. This sequence is determined uniquely by any arrow $M_l^\Delta \leftarrow M_r^\Delta$ in it, and so is denoted by $\mathcal{M}_{(l,r)}^\Delta$. The unique arrow $M_a^\Delta \leftarrow M_b^\Delta$ in it for which $0 \le a+b < p$ is the *initial arrow*. We regard M_b^Δ as the 0-*position* of $\mathcal{M}_{(l,r)}^\Delta$ and while a may be negative b is always positive. The position of any other module in $\mathcal{M}_{(l,r)}^\Delta$ will be counted from this 0-position and (a,b) is referred to as the *type* of $\mathcal{M}_{(l,r)}^\Delta$.

As F has characteristic $p > 0$ it follows immediately that $\partial^p = 0$. In particular, in $\mathcal{M}_{(l,r)}^\Delta$ we have $(\partial^*)^2 = 0$ and so this sequence is homological. The homology at $M_{j-i}^\Delta \leftarrow M_j^\Delta \leftarrow M_{j+p-i}^\Delta$ is referred as the *p-modular homology* and is denoted by

$$H_{j,i}^\Delta := (\operatorname{Ker} \partial^i \cap M_j^\Delta)/\partial^{p-i}(M_{j+p-i}^\Delta).$$

with the corresponding Betti number $\beta_{j,i}^\Delta := \dim H_{j,i}^\Delta$.

If $\mathcal{M}_{(l,r)}^\Delta$ has at most one non-vanishing homology then it is said to be *almost exact* and the only non-trivial homology then is denoted by $H_{(l,r)}^\Delta$. In general, when referring to a particular sequence $\mathcal{M}_{(l,r)}^\Delta$, the homology at position t is denoted by H_t^Δ and $\beta_t^\Delta := \dim H_t^\Delta$ is the corresponding Betti number. It is useful to allow the possibility $t = \infty$ so that an almost exact sequence $\mathcal{M}_{(l,r)}^\Delta$ is exact iff either $\beta_t^\Delta := \beta_{(l,r)}^\Delta = 0$ or $t = \infty$. Finally, if $\mathcal{M}_{(l,r)}^\Delta$ is almost exact for every choice of l and r, then $\mathcal{M}^\Delta$ is *almost p-exact*.

To formulate further results we shall need the following functions on sequences $\mathcal{M}_{(l,r)}^\Delta$: If Δ is any complex of dimension $n-1$ suppose that $\mathcal{M}_{(l,r)}^\Delta$ has type (a,b). We put

$$(1) \qquad d_{(l,r)}^n := \begin{cases} \left\lfloor \frac{n-a-b}{p} \right\rfloor & \text{if } n-a-b \not\equiv 0 \pmod p, \\ \infty & \text{if } n-a-b \equiv 0 \pmod p \end{cases}$$

and let the *weight* of $\mathcal{M}_{(l,r)}^\Delta$ be the integer $0 < w \le p$ with $w \equiv l + r - n \pmod p$. It is useful to call the finite number $d := \min\{ d_{(l,r)}^n , d_{(l,r)}^{n+1} \}$ the *middle position* or just the *middle* of $\mathcal{M}_{(l,r)}^\Delta$.

Now we are in position to formulate a result from [10] and [1] about the p-modular homology of the $(n-1)$-dimensional simplex Σ^n on n vertices. For this we shall throughout use the notation $\mathcal{M}_{(l,r)}^n := \mathcal{M}_{(l,r)}^{\Sigma^n}$ and $H_{(l,r)}^n := H_{(l,r)}^{\Sigma^n}$.

Theorem 2.1. *The sequence $\mathcal{M}^n$ is almost p-exact. For any l, r with $0 < r - l < p$ the Betti number of $\mathcal{M}^n_{(l,r)}$ is*

$$\beta^n_{(l,r)} := \left| \sum_{t=-\infty}^{+\infty} \binom{n}{l-pt} - \binom{n}{r-pt} \right| \tag{2}$$

at position $d^n_{(l,r)}$.

For $p = 3$ the numbers $\beta^n_{(l,r)}$ could be 0 or 1 only, while for $p = 5$ these are Fibonacci numbers. The structure of $H^n_{(l,r)}$ as a $Sym(n)$-module has been determined in [1] and [2].

The structure of modular homology of shellable complexes has been determined in [12]. Note that in this paper shellable complexes are always pure, see [3, 4] for standard notions of shellability and h-vector.

Theorem 2.2. *Let Δ be an $(n-1)$-dimensional shellable complex with h-vector $(h_0,\ldots,h_n)$. For a fixed sequence $\mathcal{M}^{\Delta}_{(l,r)}$ let d be its middle position and let w be its weight. Then $H^{\Delta}_t = 0$ for $t < d$ and for all $s \geq 0$ there is an embedding*

$$H^{\Delta}_{d+s} \hookrightarrow \bigoplus_{j=w+(s-1)p+1}^{w+sp} \left[H^{n-j}_{(l-j,r-j)} \right]^{h_j} . \tag{3}$$

Note that in (3) we use the convention that $[H]^0$ is the zero module. The result of Theorem 2.2 cannot be improved in general: there are examples of 7-dimensional complexes with the same h-vector which have the same 3-modular homologies but different 5-modular homologies, see [12].

3. Saturated Complexes

The result of Theorem 2.2 motivates the following definition:

Definition 3.1. A shellable complex Δ is *(l, r)-saturated* in characteristic p if the embedding (3) is an isomorphism for all $s \geq 0$. The complex Δ is *saturated* if it is (l, r)-saturated for all (l, r).

Thus, saturation is defined with respect to a prime p and it is not clear if there are complexes which are saturated for some primes but not for others. Note that there are examples of complexes which are (l, r)-saturated for certain values of (l, r) but not for others.

It follows immediately from Theorem 2.2 that for a fixed p the saturated complexes have the maximal possible modular homology, in the following sense:

Proposition 3.2. *Let Δ' and Δ be shellable complexes of the same dimension and with the same h-vector. Suppose that Δ is (l, r)-saturated. Then the Betti numbers of $\mathcal{M}^{\Delta'}_{(l,r)}$ and $\mathcal{M}^{\Delta}_{(l,r)}$ satisfy $\beta^{\Delta'}_t \leq \beta^{\Delta}_t$ for all $t \in \mathbb{Z}$. Furthermore, Δ' is (l, r)-saturated if and only if $\beta^{\Delta'}_t = \beta^{\Delta}_t$ for each $t \in \mathbb{Z}$.*

Thus, for a saturated complex all Betti numbers are determined entirely by the h-vector. For instance, if Δ is a 5-dimensional complex with $h = (h_0, h_1, \ldots, h_6)$ which is saturated for $p = 3$ then its Betti numbers are the following:

(l,r)	w			
(1,2)	3	$\beta_{4,2} = h_1 + h_2$;		$\beta_{5,1} = h_4 + h_5$
(1,3)	1	$\beta_{3,2} = h_0$;	$\beta_{4,1} = h_2 + h_3$;	$\beta_{6,2} = h_5 + h_6$
(2,3)	2	$\beta_{3,1} = h_0 + h_1$;	$\beta_{5,2} = h_3 + h_4$;	$\beta_{6,1} = h_6$

If the same Δ is saturated for $p = 5$ then its Betti numbers are the following:

(l,r)	w		
(1,2)	2	$\beta_{2,1} = 8h_0 + 3h_1$;	$\beta_{6,4} = h_3 + h_4 + h_5 + h_6$
(1,3)	3	$\beta_{3,2} = 13h_0 + 8h_1 + 3h_2$;	$\beta_{6,3} = h_4 + h_5 + h_6$
(1,4)	4	$\beta_{4,3} = 8h_0 + 8h_1 + 5h_2 + 2h_3$;	$\beta_{6,2} = h_5 + h_6$
(1,5)	5	$\beta_{5,4} = 3h_1 + 3h_2 + 2h_3 + h_4$;	$\beta_{6,1} = h_6$
(2,3)	4	$\beta_{3,1} = 5h_0 + 5h_1 + 3h_2 + h_3$;	
(2,4)	5	$\beta_{4,2} = 5h_1 + 5h_2 + 3h_3 + h_4$;	
(2,5)	1	$\beta_{2,2} = 8h_0$;	$\beta_{5,3} = 3h_2 + 3h_3 + 2h_4 + h_5$
(3,4)	1	$\beta_{3,4} = 5h_0$;	$\beta_{4,1} = 2h_2 + 2h_3 + h_4$
(3,5)	2	$\beta_{3,3} = 13h_0 + 5h_1$;	$\beta_{5,2} = 2h_3 + 2h_4 + h_5$
(4,5)	3	$\beta_{4,4} = 8h_0 + 5h_1 + 2h_2$;	$\beta_{5,1} = h_4 + h_5$

A number of examples of saturated complexes have been found in [12] and [13]:

Example 1: Let Δ be a $(n-1)$-dimensional complex with m facets and with h-vector of the form $(1, m-1, 0, \ldots, 0)$. Every such Δ is saturated for every p. Moreover, every sequence $\mathcal{M}^{\Delta}_{(l,r)}$ is almost p-exact with homology

$$H^{\Delta}_{(l,r)} \simeq H^{n}_{(l,r)} \oplus \left[H^{n-1}_{(l-1,r-1)}\right]^{m-1}$$

in the middle. In particular, a simplex Σ^n is trivially saturated.

Example 2: The $(n-1)$-dimensional *hyperoctahedron* or *cross-polytope* is obtained by performing successive suspensions over vertex pairs α_i, β_i, or alternatively, as the dual of the $(n-1)$-dimensional cube. It is shellable and it follows from results of [13] that it is saturated for all primes.

Example 3: Finite Coxeter complexes and spherical buildings are saturated for every prime, see [13].

4. Saturated Complexes and Ranks of Incidence Matrices

Here we give an alternative definition of saturated complexes. Let $\mathrm{rk}^{\Delta}_p(s,t)$ be the p-rank of the incidence matrix of s-faces versus t-faces of a complex Δ. When Δ is a simplex Σ^n, we denote corresponding ranks by $\mathrm{rk}^{n}_p(s,t)$. It is well-known [7, 10, 15] that for $s+t<n$,

$$\mathrm{rk}^{n}_p(s,t) = \sum_{k=0} \binom{n}{s-pk} - \binom{n}{t-p-pk} \tag{4}$$

A similar relation holds for arbitrary shellable complexes:

Theorem 4.1. *Let Δ be a shellable $(n-1)$-dimensional complex and $p > 2$ be a prime. Let $s < t \leq n$ be non-negative integers such that $t - s < p$. If $s + t < n$ then*

$$\mathrm{rk}^{\Delta}_p(s,t) = \sum_{i=0}^{n} h_i \mathrm{rk}^{n-i}_p(s-i, t-i) \tag{5}$$

Moreover, a shellable complex Δ is saturated if and only if the relation (5) holds also for $s+t \geq n$.

Proof. First, let Δ be an arbitrary shellable complex with f-vector $(f_0, f_1, \dots, f_n)$. In view of the condition $0 < t - s < p$ we may look at $\mathrm{rk}_p^{\Delta}(s,t)$ as the p-rank of the map $\partial^{t-s} : M_t^{\Delta} \to M_s^{\Delta}$. According to Theorem 2.2, in the sequence $\mathcal{M}_{(s,t)}^{\Delta}$ all homologies on the left from the middle are trivial. Equivalently (see [12, Corollary 5.6]), for $s + t < n$,

$$\mathrm{rk}_p^{\Delta}(s,t) = f_s - f_{t-p} + f_{s-p} - f_{t-2p} + f_{s-2p} - f_{t-2p} + \dots .$$

The result follows now from the well-known formula

$$h_k = \sum_{i=0}^{n} (-1)^{i+k} f_i \binom{n-i}{k-i}. \tag{6}$$

after substituting it into (4).

Now let Δ be saturated, so that its Betti numbers are defined by (3). For $s + t \geq n$ we need to take these into account when evaluating rank:

$$\mathrm{rk}_p^{\Delta}(s,t) = \sum_{k=0} (f_{s-kp} - f_{t-p-kp}) - (\beta^{\Delta}_{s-kp,p-t+s} - \beta^{\Delta}_{t-p-kp,t-s}). \tag{7}$$

Also

$$\mathrm{rk}_p^{n}(s,t) = \sum_{k=0} \binom{n}{s-pk} - \binom{n}{t-p-pk} \pm \beta^{n}_{(s,t)} , \tag{8}$$

where the sign of Betti number is determined by its position in the sequence $\mathcal{M}^n_{(s,t)}$. Now put (6) and (8) into right-hand side of (5). After transforming dimensions into positions we obtain (7). Thus, for saturated Δ the relation (5) holds also for $s + t \geq n$.

Finally, since Betti numbers are completely determined by ranks, (5) implies saturation of Δ in view of Proposition 3.2. $\square$

We note that by using the *r-step modular homology* [1] it can be shown that the condition $t - s < p$ in Theorem 4.1 is redundant.

5. Combinatorial Characterization of Saturated Complexes

Two previous definitions of saturated complexes were algebraic. Now we shall state a combinatorial description of saturated complexes. We show that certain conditions on the links of the complex imply saturation.

Let Γ be an $(n-1)$-dimensional complex and let $\Delta = \Gamma \overset{k}{\cup} \Sigma^n$ be obtained by gluing Σ^n onto Γ along some k facets of Σ^n.

Definition 5.1. We say that $\Delta := \Gamma \overset{k}{\cup} \Sigma^n$ is *(l,r)-saturated over* Γ, if Δ has the same homologies as Γ in all positions but $u := d^{n+k}_{(l,r)}$, where $H^{\Delta}_u \simeq H^{\Gamma}_u \oplus H^{n-k}_{(l-k,r-k)}$. We say that Δ is *saturated over* Γ, if Δ is (l,r)-saturated over Γ for all (l,r).

Proposition 5.2. *Let Γ be (l,r)-saturated. Then $\Delta = \Gamma \overset{k}{\cup} \Sigma^n$ is (l,r)-saturated iff Δ is (l,r)-saturated over Γ.*

In particular, a shellable complex Δ is *(l,r)-saturated* if and only if Δ has a shelling $\Delta_1, \Delta_2, \dots, \Delta_m = \Delta$ in which Δ_i is (l,r)-saturated over Δ_{i-1} for every $2 < i \leq m$.

Let σ denote the vertex set of Σ^n and let $\Delta = \Gamma \overset{k}{\cup} \Sigma^n$. Then the *restriction* $\mathrm{res}(\sigma)$ is the set of all vertices $\beta \in \sigma$ such that $\sigma \setminus \{\beta\}$ is contained in Γ, see Björner [5]. So $\mathrm{res}(\sigma)$ is a $(k-1)$-face of Σ^n and one may regard it as the 'outer face' under gluing. Its complement $t(\sigma) := \sigma \setminus \mathrm{res}(\sigma)$ is the 'inner face' under gluing. If x is a face of Δ then the subcomplex

$\mathrm{star}_\Delta(x)$ is generated by all facets that contain x and $\mathrm{link}_\Delta(x)$ is the subcomplex of all faces of $\mathrm{star}_\Delta(x)$ that do not contain x. So the dimension of $\mathrm{link}_\Delta(x)$ is $n-|x|-1$.

The next result gives combinatorial characterization of saturated complexes. Note that its sufficiency has been proved in [13].

Theorem 5.3. *Let Γ be a complex and let $\Delta = \Gamma \overset{k}{\cup} \Sigma^n$. Then Δ is saturated over Γ if and only if $\mathrm{res}(\sigma)$ is a 1-cycle of Δ relative to $\mathrm{link}_\Gamma(t(\sigma))$.*

Note that when saying that $\mathrm{res}(\sigma)$ is a 1-*cycle of* Δ *relative to* $\mathrm{link}_\Gamma(t(\sigma))$ we mean, as usual, that there is some $f \in M_k^\Gamma \subset M^\Delta$ such that $\mathrm{supp}(f) \cap t(\sigma) = \emptyset$, $f \cup t(\sigma) \in M^\Gamma$ and $\partial(\mathrm{res}(\sigma) + f) = 0$.

There is a simple geometrical condition which implies saturation.

Definition 5.4. Let Δ be a pure $(n-1)$-dimensional complex with facets $\sigma_1, \ldots, \sigma_m$. Then Δ is *null over F with respect to ∂*, or just *null* for short, if there are non-zero $c_1, \ldots, c_m \in F$ such that $\partial(c_1\sigma_1 + \ldots + c_m\sigma_m) = 0$.

We say that a complex is 2-*colourable* if its facets can be 2-coloured in such a way that facets with a common co-dimension 1 face have different colours. Further, in a *pseudomanifold without boundary*, see Definition 3.15 in [14], each co-dimension 1 face is contained in exactly 2 facets. Therefore a 2-colourable pseudomanifold without boundary is null: Choose all $c_i = \pm 1$, suitably according to the 2-colouring. In particular, even cyclic graphs are null over every field, and odd cyclic graphs are null only over fields of characteristic 2.

Corollary 5.5. *Let Γ be a complex and let $\Delta = \Gamma \overset{k}{\cup} [\sigma]$ for some $k \geq 1$. Suppose that $\mathrm{link}_\Delta(t(\sigma))$ is null. (In particular, suppose that $\mathrm{link}_\Delta(t(\sigma))$ is a 2-colourable triangulation of a sphere, or a 2-colourable pseudomanifold without boundary.) Then Δ is saturated over Γ.*

The next result follows from Theorem 5.3:

Theorem 5.6. *Let Δ be a pure $(n-1)$-dimensional completely balanced complex. For every $R \subseteq \{0, \ldots, n-1\}$ let Δ_R be the type-selected subcomplex. If Δ is saturated then Δ_R is also saturated.*

In particular, let P be a poset of a finite rank with saturated order complex $\Delta(P)$. Then all rank-selected subcomplexes $\Delta(P)_R$ are saturated.

References

[1] S Bell, P Jones *and* I J Siemons, On modular homology in the Boolean Algebra II, *Journal of Algebra*, **199** (1998), 556–580.

[2] P Jones *and* I J Siemons, On modular homology in the Boolean Algebra III, *Journal of Algebra*, **243** (2001), 409–426.

[3] A Björner, Topological methods, *in: Handbook of Combinatorics*, Chapter 34, (eds RL Graham, M Grötschel and L Lovász), North-Holland, 1995.

[4] A Björner, The homology and shellability of matroids and geometric lattices, *in: Matroid Applications* (ed. by N. White), Cambridge University Press, 1991, 226–283.

[5] A Björner, Some combinatorial and algebraic properties of Coxeter complexes and Tits buildings, *Adv. Math.*, **52**(1984), 173–212.

[6] M Dubois-Violette, $d^N = 0$: Generalized homology, *K-Theory*, **14** (1998), 104-118.

[7] P Frankl, Intersection theorems and mod p rank of inclusion matrices, *J. Comb. Theory*, **A54** (1990), 85–94.

[8] M M Kapranov, On the q-analogue of homological algebra, *Cornell University Reprint*, 1991.

[9] W Mayer, A new homology theory, *Annalen der Mathematik*, **48** (1947) 370–380 *and* 594–605.

[10] V B Mnukhin *and* I J Siemons, The modular homology of inclusion maps and group actions, *J. Comb. Theory*, **A74** (1996) 287–300.

[11] V B Mnukhin *and* I J Siemons, On the modular homology in the Boolean algebra, *Journal of Algebra*, **179** (1996) 191–199.

[12] V B Mnukhin *and* I J Siemons, On modular homology of simplicial complexes: Shellability, *J. Comb. Theory*, **A93** (2001), 350–370.

[13] V B Mnukhin *and* I J Siemons, On modular homology of simplicial complexes: Saturation, *J. Comb. Theory*, in press.

[14] R P Stanley, *Combinatorics and Commutative Algebra*, (Progress in Mathematics, **41**), Birkhäuser Boston 2nd ed., 1996.

[15] R M Wilson, A diagonal form for the incidence matrix of t-subsets versus k-subsets. *European Journal of Combinatorics*, **11**(1990) 609–615.

Department of Mathematics and Computer Science, University of the South Pacific, Suva, Fiji Islands

E-mail address: `Mnukhin_V@usp.ac.fj`

A DETERMINANTAL FORMULA FOR SUPERSYMMETRIC SCHUR POLYNOMIALS

E.M. MOENS AND J. VAN DER JEUGT

Abstract. We derive a new formula for the supersymmetric Schur polynomial $s_\lambda(x/y)$. The origin of this formula goes back to representation theory of the Lie superalgebra $\mathfrak{gl}(m/n)$. In particular, we show how a character formula due to Kac and Wakimoto can be applied to covariant representations, leading to a new expression for $s_\lambda(x/y)$. This new expression gives rise to a determinantal formula for $s_\lambda(x/y)$. In particular, the denominator identity for $\mathfrak{gl}(m/n)$ corresponds to a determinantal identity combining Cauchy's double alternant with Vandermonde's determinant. We also provide an independent proof of the new determinantal formula.

Résumé Nous dérivons une nouvelle formule pour les polynômes de Schur supersymétriques $s_\lambda(x/y)$. Cette formule remonte à la théorie de représentation de la superalgèbre de Lie $\mathfrak{gl}(m/n)$. En fait, nous démontrons de quelle façon l'application de la formule de caractère de Kac et Wakimoto aux représentations covariantes mène à une nouvelle expression de $s_\lambda(x/y)$. Cette expression peut être reformulée uniquement en termes de déterminants. En particulier, dans le domaine de l'algèbre $\mathfrak{gl}(m/n)$ l'identité du dénominateur correspond à une identité de déterminants dans laquelle une combinaison d'un déterminant double alterné de Cauchy et d'un déterminant de Vandermonde peut être retrouvée. Une preuve directe de l'expression en déterminants de $s_\lambda(x/y)$ est également présentée.

1. Introduction

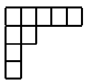
This paper deals with a new formula for supersymmetric Schur polynomials $s_\lambda(x/y)$, parametrized by a partition λ, and symmetric in two sets of variables $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_n)$. Supersymmetric Schur polynomials, or S-functions, were studied in detail in the work of Berele and Regev [3], who also showed that these polynomials are closely related to characters of certain representations of the Lie superalgebra $\mathfrak{gl}(m/n)$ or $\mathfrak{sl}(m/n)$. More precisely, the characters of irreducible covariant tensor representations of $\mathfrak{gl}(m/n)$ are supersymmetric S-functions (just as the characters of irreducible covariant tensor representations of $\mathfrak{gl}(m)$ are ordinary Schur functions). It is in the context of $\mathfrak{gl}(m/n)$ representations that our new formula for supersymmetric S-functions was discovered, as a consequence of a certain character formula of Kac and Wakimoto [6]. In their general study of characters for classical Lie superalgebras and affine superalgebras, Kac and Wakimoto [6] developed the idea of so-called tame representations. These tame representations allow for the construction of an explicit character formula. Unfortunately, it is not easy to give a simple characterization of which representations are tame. In this paper, we show that the covariant tensor representations of $\mathfrak{gl}(m/n)$ are indeed tame. As a consequence we can apply the character formula of Kac and Wakimoto, and obtain a new expression for supersymmetric S-functions. This new expression is not particularly elegant, but after certain manipulations we show that it is equivalent to a nice determinantal formula. Once

Key words and phrases. supersymmetric Schur polynomials, Lie superalgebra $\mathfrak{gl}(m/n)$, characters, covariant tensor representations, determinantal identities.

this determinantal formula for $s_\lambda(x/y)$ was found, we realized that its validity could also be proved independently. Indeed, in [10, §I.3, exercise 23] or [13] it is shown that the supersymmetric S-functions are characterized by four properties. Using our determinantal expression, it is possible to show that these four properties are indeed satisfied. Secondly, one can also prove the determinantal formula using a double Laplace expansion and a special case of the so-called Sergeev-Pragacz formula (we are due to the referee for pointing this out).

The structure of this paper is as follows. In the first section, we fix the notation and recall some known formulas for ordinary and supersymmetric Schur functions. The simplest form of our new determinantal formula for $s_\lambda(x/y)$ is already given in (1.17). Furthermore, we point out that the case $\lambda = 0$ gives rise to an interesting determinantal identity, which can be called "the denominator identity for $\mathfrak{gl}(m/n)$". Section 2 is devoted to showing that covariant tensor representations of the Lie superalgebra $\mathfrak{gl}(m/n)$ are tame, and to applying the Kac-Wakimoto character formula. This section uses a lot of representation theory, and closes with a formula for the character, i.e. a formula for $s_\lambda(x/y)$. This formula is not in an optimal form, and in Section 3 we use a number of intricate but elementary manipulations to derive from this form our main result, Theorem 3.2, giving the determinantal expression for $s_\lambda(x/y)$. Finally, we provide a straightforward proof, without using representation theory, based on the formula of Sergeev-Pragacz and Laplace's theorem. So in fact we provide two proofs : one in the context of representation theory of $\mathfrak{gl}(m/n)$, and one by means of the formula of Sergeev-Pragacz. A reader who is not so familiar with representation theory can easily skip Section 2 (apart from Definition 2.1). In fact, we could have presented our main result without any reference to representation theory. However, we did not want to hide the natural background of this new formula, and therefore we have chosen to present also its representation theoretic origin. This is also clear from the determinantal formula itself : in this expression, the so-called (m, n)-index k of λ is crucial; this definition of k has a natural interpretation in representation theory of $\mathfrak{gl}(m/n)$.

Let $\lambda = (\lambda_1, \ldots, \lambda_p, 0, 0, \ldots)$ be a partition of the non-negative integer N, with $\lambda_1 \geq \ldots \geq \lambda_p > 0$ and $\sum_i \lambda_i = |\lambda| = N$. The number $p = \ell(\lambda)$ is the length of λ. The Young diagram F^λ of shape λ is the set of left-adjusted rows of squares with λ_i squares (or boxes) in the ith row (reading from top to bottom). For example, the Young diagram of $(5, 2, 1, 1)$ is given by :

(1.1) $F^\lambda =$ 

As usual, λ' denotes the conjugate to λ; e.g. if $\lambda = (5, 2, 1, 1)$ then $\lambda' = (4, 2, 1, 1, 1)$. Denote by $\mathcal{S}(x)$ the ring of symmetric functions in m independent variables $x_1, \ldots, x_m$ [10]. The Schur functions or S-functions s_λ form a $\mathbb{Z}$-basis of $\mathcal{S}(x)$. There are various ways to define the S-functions [10]. For a partition λ with $\ell(\lambda) \leq m$, there is the determinantal formula (as a quotient of two alternants) :

$$s_\lambda(x) = \frac{\det(x_i^{\lambda_j+m-j})_{1\leq i,j\leq m}}{\det(x_i^{m-j})_{1\leq i,j\leq m}}. \tag{1.2}$$

The numerator can be rewritten as

$$\det(x_i^{\lambda_j+m-j})_{1\leq i,j\leq m} = \det(x^{\lambda+\delta_m}) = \sum_{w\in S_m} \varepsilon(w)w(x^{\lambda+\delta_m}), \tag{1.3}$$

where S_m is the symmetric group acting on $x = (x_1, \ldots, x_m)$, $\varepsilon(w)$ the signature of w, and

$$\delta_m = (m-1, m-2, \ldots, 1, 0). \tag{1.4}$$

Thus (1.2) is essentially equivalent to Weyl's character formula for the Lie algebra $\mathfrak{gl}(m)$, where an irreducible representation of $\mathfrak{gl}(m)$ is characterized by a partition λ and its character is given by $s_\lambda(x)$. The denominator on the right hand side of (1.2) is the Vandermonde determinant, equal to the product $\prod_{1\le i<j\le m}(x_i - x_j)$; this is Weyl's denominator formula for $\mathfrak{gl}(m)$.

When $\lambda = (r)$, $s_\lambda(x)$ is the complete symmetric function $h_r(x)$, and when $\lambda = (1^r)$, $s_\lambda(x)$ is the elementary symmetric function $e_r(x)$. The Jacobi-Trudi formula and the Nägelsbach-Kostka formula give s_λ in terms of these [10] :

$$s_\lambda(x) = \det\Big(h_{\lambda_i - i + j}(x)\Big)_{1\le i,j\le \ell(\lambda)} = \det\Big(e_{\lambda'_i - i + j}(x)\Big)_{1\le i,j\le \ell(\lambda')}. \tag{1.5}$$

Other determinantal formulas for s_λ are Giambelli's formula [10] and the ribbon formula [9]. Finally, recall there is a combinatorial formula for $s_\lambda(x)$ as a sum of monomials summed over all column-strict Young tableaux of shape λ [10].

All these formulas, except (1.2), have their extensions to skew Schur functions $s_{\lambda/\mu}(x)$, where λ and μ are two partitions with $\lambda_i \ge \mu_i$ for all i.

Let us now recall some notions of *supersymmetric S-functions* [3, 7, 17]. The ring of doubly symmetric polynomials in $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_n)$ is $\mathcal{S}(x, y) = \mathcal{S}(x) \otimes_{\mathbb{Z}} \mathcal{S}(y)$. An element $p \in \mathcal{S}(x, y)$ has the *cancellation property* if it satisfies the following : when the substitution $x_1 = t$, $y_1 = -t$ is made in p, the resulting polynomial is independent of t. We denote $\mathcal{S}(x/y)$ the subring of $\mathcal{S}(x, y)$ consisting of all elements satisfying the cancellation property. The elements of $\mathcal{S}(x/y)$ are the supersymmetric polynomials [17].

The complete supersymmetric functions $h_r(x/y)$ belong to $\mathcal{S}(x/y)$, and are defined by

$$h_r(x/y) = \sum_{k=0}^{r} h_{r-k}(x) e_k(y). \tag{1.6}$$

The following gives a first formula for the supersymmetric S-functions :

$$s_\lambda(x/y) = \det\Big(h_{\lambda_i - i + j}(x/y)\Big)_{1\le i,j\le \ell(\lambda)}. \tag{1.7}$$

The polynomials $s_\lambda(x/y)$ are identically zero when $\lambda_{m+1} > n$. Denote by $\mathcal{H}_{m,n}$ the set of partitions with $\lambda_{m+1} \le n$, i.e. the partitions (with their Young diagram) inside the (m, n)-hook. Stembridge [17] showed that the set of $s_\lambda(x/y)$ with $\lambda \in \mathcal{H}_{m,n}$ form a $\mathbb{Z}$-basis of $\mathcal{S}(x/y)$.

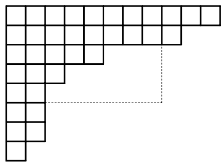
There exist a number of other formulas for the supersymmetric S-functions. One is a combinatorial formula in terms of supertableaux of shape λ, see [3]. From the combinatorial formula, one finds expansions of $s_\lambda(x/y)$ in terms of ordinary S-functions :

$$s_\lambda(x/y) = \sum_{\mu} s_\mu(x) s_{(\lambda/\mu)'}(y) = \sum_{\mu,\nu} c^\lambda_{\mu,\nu} s_\mu(x) s_{\nu'}(y), \tag{1.8}$$

where $c^\lambda_{\mu,\nu}$ are the Littlewood-Richardson coefficients [10].

Just as the functions $s_\lambda(x)$ are characters of simple modules of the Lie algebra $\mathfrak{gl}(m)$, the supersymmetric S-functions are characters of (a class of) simple modules of the Lie superalgebra $\mathfrak{gl}(m/n)$ [3]. In this context, a different formula for $s_\lambda(x/y)$ was found by

Sergeev (see [12]) and in [19]; the first proof of this formula was given by Pragacz [12]. To describe the so-called Sergeev-Pragacz formula, let λ be a partition with $\lambda_{m+1} \leq n$. Consider the Young diagram F^λ, let F^κ be the part of F^λ that falls within the $m \times n$ rectangle, and let F^τ, resp. F^η, be the remaining part to the right, resp. underneath this rectangle; i.e. $\lambda = (\kappa + \tau) \cup \eta$. This is illustrated, for $m = 5$, $n = 8$ and $\lambda = (11, 9, 5, 3, 2, 2, 2, 1)$, as follows :

(1.9) $F^\lambda =$  hence $\begin{aligned} \kappa &= (8,8,5,3,2) \\ \tau &= (3,1) \\ \eta &= (2,2,1) \end{aligned}$

Then, the Sergeev-Pragacz formula for $s_\lambda(x/y)$ is given by

$$s_\lambda(x/y) = D_0^{-1} \sum_{w \in S_m \times S_n} \varepsilon(w)\, w\Big(x^{\tau+\delta_m} y^{\eta'+\delta_n} \prod_{(i,j)\in F^\kappa} (x_i + y_j) \Big), \tag{1.10}$$

where $(i,j) \in F^\kappa$ if and only if the box with row-index i (read from left to right) and column-index j (read from top to bottom) belongs to F^κ, and

$$D_0 = \prod_{1\leq i<j\leq m} (x_i - x_j) \prod_{1\leq i<j\leq n} (y_i - y_j). \tag{1.11}$$

This formula is useful for the computation of $s_\lambda(x/y)$, and even for the computation of Littlewood-Richardson coefficients [2, 18]. Note that for the special case that $\lambda_m \geq n$, (1.10) becomes :

$$s_\lambda(x/y) = s_\tau(x) s_{\eta'}(y) \prod_{i=1}^{m}\prod_{j=1}^{n} (x_i + y_j). \tag{1.12}$$

This is the case of *factorization*. Furthermore, from (1.10) one easily deduces duality :

$$s_\lambda(x/y) = s_{\lambda'}(y/x). \tag{1.13}$$

Observe, on the other hand, that D_0 is just Weyl's denominator for $\mathfrak{gl}(m) \oplus \mathfrak{gl}(n)$. So when $\lambda = 0$, (1.10) does not yield a new denominator identity related to $\mathfrak{gl}(m/n)$; it only gives the denominator identity for $\mathfrak{gl}(m)$ and $\mathfrak{gl}(n)$.

In this paper, we shall give a new formula for $s_\lambda(x/y)$. In its simplest form, this yields a new determinantal formula for supersymmetric S-functions. Furthermore, this formula yields a genuine denominator identity related to $\mathfrak{gl}(m/n)$. Let us briefly describe one of the forms of the new formula. First, we introduce some new notations. Let

$$D(x) = \prod_{1\leq i<j\leq m} (x_i - x_j) \qquad \text{and} \qquad R(x,y) = \prod_{i=1}^{m}\prod_{j=1}^{n} (x_i - y_j) \tag{1.14}$$

and define D by

$$D = \frac{\prod_{1\leq i<j\leq m}(x_i - x_j) \prod_{1\leq i<j\leq n}(y_i - y_j)}{\prod_{i=1}^{m}\prod_{j=1}^{n}(x_i + y_j)} = \frac{D(x)D(y)}{R(x,-y)}. \tag{1.15}$$

Let λ be a partition with $\lambda_{m+1} \leq n$, i.e. $\lambda \in \mathcal{H}_{m,n}$, and put

$$k = \min\{j | \lambda_j + m + 1 - j \leq n\}; \tag{1.16}$$

since $\lambda_{m+1} \leq n$, we have that $1 \leq k \leq m+1$. Then the new formula reads

$$s_\lambda(x/y) = (-1)^{mn-m+k-1} D^{-1} \det \begin{pmatrix} R & X_\lambda \\ Y_\lambda & 0 \end{pmatrix}, \tag{1.17}$$

where the (rectangular) blocks of the determinant are given by

$$R = \left(\frac{1}{x_i + y_j} \right)_{1 \leq i \leq m,\ 1 \leq j \leq n},$$

$$X_\lambda = \left(x_i^{\lambda_j + m - n - j} \right)_{1 \leq i \leq m,\ 1 \leq j \leq k-1}, \qquad Y_\lambda = \left(y_j^{\lambda'_i + n - m - i} \right)_{1 \leq i \leq n-m+k-1,\ 1 \leq j \leq n}.$$

When $\lambda = 0$ it follows from (1.7) or (1.10) that $s_\lambda(x/y) = 1$. The new formula (1.17) gives rise to a denominator identity for $\mathfrak{gl}(m/n)$. Suppose $m \leq n$ ($m \geq n$ is similar); when $\lambda = 0$, it follows from (1.16) that $k = 1$. So the X_λ-block and 0-block disappear in (1.17). Changing the order of the R-block and Y_λ-block, implies

$$\det \begin{pmatrix} y_1^{n-m-1} & \cdots & y_n^{n-m-1} \\ \vdots & & \vdots \\ y_1^0 & \cdots & y_n^0 \\ \frac{1}{x_1+y_1} & \cdots & \frac{1}{x_1+y_n} \\ \vdots & & \vdots \\ \frac{1}{x_m+y_1} & \cdots & \frac{1}{x_m+y_n} \end{pmatrix} = \frac{\prod_{1 \leq i < j \leq m}(x_i - x_j) \prod_{1 \leq i < j \leq n}(y_i - y_j)}{\prod_{i=1}^m \prod_{j=1}^n (x_i + y_j)}. \tag{1.18}$$

Clearly, when $m = n$, this is simply Cauchy's double alternant; when $m = 0$, it is just Vandermonde's determinant. When $0 < m < n$, it is a combination of the two. These type of determinants have already been encountered in a different context [1] (we found this reference in [8]); here they are for the first time related to a denominator identity.

2. Covariant modules for the Lie superalgebra $\mathfrak{gl}(m/n)$

For general theory on classical Lie superalgebras and their representations, we refer to [4, 5, 14]; for representations of the Lie superalgebra $gl(m/n)$, see [19, 20, 21].

Let $\mathfrak{g} = \mathfrak{gl}(m/n)$, $\mathfrak{h} \subset \mathfrak{g}$ its Cartan subalgebra, and $\mathfrak{g} = \mathfrak{g}_{-1} \oplus \mathfrak{g}_0 \oplus \mathfrak{g}_{+1}$ the consistent $\mathbb{Z}$-grading. Note that $\mathfrak{g}_0 = \mathfrak{gl}(m) \oplus \mathfrak{gl}(n)$, and put $\mathfrak{g}^+ = \mathfrak{g}_0 \oplus \mathfrak{g}_{+1}$ and $\mathfrak{g}^- = \mathfrak{g}_0 \oplus \mathfrak{g}_{-1}$. The dual space $\mathfrak{h}^*$ has a natural basis $\{\epsilon_1, \ldots, \epsilon_m, \delta_1, \ldots, \delta_n\}$, and the roots of $\mathfrak{g}$ can be expressed in terms of this basis. Let Δ be the set of all roots, Δ_0 the set of even roots, and Δ_1 the set of odd roots. One can choose a set of simple roots (or, equivalently, a triangular decomposition), but note that contrary to the case of simple Lie algebras not all such choices are equivalent. The so-called *distinguished choice* [4] for a triangular decomposition $\mathfrak{g} = \mathfrak{n}^- \oplus \mathfrak{h} \oplus \mathfrak{n}^+$ is such that $\mathfrak{g}_{+1} \subset \mathfrak{n}^+$ and $\mathfrak{g}_{-1} \subset \mathfrak{n}^-$. Then $\mathfrak{h} \oplus \mathfrak{n}^+$ is the corresponding distinguished Borel subalgebra, and Δ_+ the set of positive roots. For this choice we have explicitly :

$$\begin{aligned} \Delta_{0,+} &= \{\epsilon_i - \epsilon_j | 1 \leq i < j \leq m\} \cup \{\delta_i - \delta_j | 1 \leq i < j \leq n\}, \\ \Delta_{1,+} &= \{\beta_{ij} = \epsilon_i - \delta_j | 1 \leq i \leq m,\ 1 \leq j \leq n\}, \end{aligned} \tag{2.1}$$

and the corresponding set of simple roots (the distinguished set) is given by

$$\Pi = \{\epsilon_1 - \epsilon_2, \ldots, \epsilon_{m-1} - \epsilon_m, \epsilon_m - \delta_1, \delta_1 - \delta_2, \ldots, \delta_{n-1} - \delta_n\}. \tag{2.2}$$

Thus in the distinguished basis there is only one simple root which is odd. As usual, we put

$$\rho_0 = \frac{1}{2}\Big(\sum_{\alpha\in\Delta_{0,+}} \alpha\Big), \qquad \rho_1 = \frac{1}{2}\Big(\sum_{\alpha\in\Delta_{1,+}} \alpha\Big), \qquad \rho = \rho_0 - \rho_1. \tag{2.3}$$

There is a symmetric form $(\ ,\ )$ on $\mathfrak{h}^*$ induced by the invariant symmetric form on $\mathfrak{g}$, and in the natural basis it takes the form $(\epsilon_i, \epsilon_j) = \delta_{ij}$, $(\epsilon_i, \delta_j) = 0$ and $(\delta_i, \delta_j) = -\delta_{ij}$. The odd roots are isotropic : $(\alpha, \alpha) = 0$ if $\alpha \in \Delta_1$.

The Weyl group of $\mathfrak{g}$ is the Weyl group W of $\mathfrak{g}_0$, hence it is the direct product of symmetric groups $S_m \times S_n$. For $w \in W$, we denote by $\varepsilon(w)$ its signature.

Let $\Lambda \in \mathfrak{h}^*$; the *atypicality* of Λ, denoted by $\mathrm{atyp}(\Lambda)$, is the maximal number of linearly independent roots β_i such that $(\beta_i, \beta_j) = 0$ and $(\Lambda, \beta_i) = 0$ for all i and j [6]. Such a set $\{\beta_i\}$ is called a Λ-maximal isotropic subset of Δ.

Given a set of positive roots Δ_+ of Δ, and a simple odd root α, one may construct a new set of positive roots [11, 6] by

$$\Delta'_+ = (\Delta_+ \cup \{-\alpha\}) \setminus \{\alpha\}. \tag{2.4}$$

The set Δ'_+ is called a simple reflection of Δ_+. Since we use only simple reflections with respect to simple odd roots, $\Delta_{0,+}$ remains invariant, but $\Delta_{1,+}$ will change and the new ρ is given by :

$$\rho' = \rho + \alpha. \tag{2.5}$$

Let V be a finite-dimensional irreducible $\mathfrak{g}$-module. Such modules are $\mathfrak{h}$-diagonalizable with weight decomposition $V = \oplus_\mu V(\mu)$, and the character is defined to be $\mathrm{ch}\, V = \sum_\mu \dim V(\mu)\, e^\mu$, where e^μ $(\mu \in \mathfrak{h}^*)$ is the formal exponential. Consider such a module V. If we fix a set of positive roots Δ_+, we may talk about the highest weight Λ of V and about the corresponding ρ. If Δ'_+ is obtained from Δ_+ by a simple α-reflection, where α is odd, and Λ' denotes the highest weight of V with respect to Δ'_+, then [6]

$$\Lambda' = \Lambda - \alpha \text{ if } (\Lambda, \alpha) \neq 0; \qquad \Lambda' = \Lambda \text{ if } (\Lambda, \alpha) = 0. \tag{2.6}$$

From this, one deduces that for the $\mathfrak{g}$-module V, $\mathrm{atyp}(\Lambda + \rho)$ is independent of the choice of Δ_+; then $\mathrm{atyp}(\Lambda + \rho)$ is referred to as the atypicality of V (if $\mathrm{atyp}(\Lambda + \rho) = 0$, V is typical, otherwise it is atypical). If one can choose a $(\Lambda + \rho)$-maximal isotropic subset S_Λ in Δ_+ such that $S_\Lambda \subset \Pi \subset \Delta_+$ (Π is the set of simple roots with respect to Δ_+), then the $\mathfrak{g}$-module V is called *tame*, and a character formula is known due to Kac and Wakimoto [6]. It reads :

$$\mathrm{ch}\, V = j_\Lambda^{-1} e^{-\rho} R^{-1} \sum_{w\in W} \varepsilon(w) w\Big(e^{\Lambda+\rho} \prod_{\beta\in S_\Lambda} (1 + e^{-\beta})^{-1}\Big), \tag{2.7}$$

where

$$R = \prod_{\alpha\in\Delta_{0,+}} (1 - e^{-\alpha}) \Big/ \prod_{\alpha\in\Delta_{1,+}} (1 + e^{-\alpha}) \tag{2.8}$$

and j_Λ is a normalization coefficient to make sure that the coefficient of e^Λ on the rhs of (2.7) is 1.

The rest of this section is now devoted to a particular class of finite-dimensional irreducible $\mathfrak{g}$-modules, the *covariant* modules, and to showing that these modules are tame.

Berele and Regev [3], and Sergeev [15], showed that the tensor product of N copies of the natural $(m+n)$-dimensional representation of $\mathfrak{g} = \mathfrak{gl}(m/n)$ is completely reducible, and that the irreducible components V_λ can be labeled by a partition λ of N such that λ is inside the (m,n)-hook, i.e. such that $\lambda_{m+1} \leq n$. These representations V_λ are known as covariant modules. Let us first consider V_λ in the distinguished basis fixed by (2.2). Then, the highest weight Λ_λ of V_λ in the standard ϵ-δ-basis is given by [19]

$$\Lambda_\lambda = \lambda_1\epsilon_1 + \cdots + \lambda_m\epsilon_m + \nu'_1\delta_1 + \cdots + \nu'_n\delta_n, \tag{2.9}$$

where $\nu'_j = \max\{0, \lambda'_j - m\}$ for $1 \leq j \leq n$. Let us consider the atypicality of V_λ, in the distinguished basis. For this purpose, it is sufficient to compute the numbers $(\Lambda_\lambda + \rho, \beta_{ij})$, with $\beta_{ij} = \epsilon_i - \delta_j$, for $1 \leq i \leq m$ and $1 \leq j \leq n$, and count the number of zeros. It is convenient to put the numbers $(\Lambda_\lambda + \rho, \beta_{ij})$ in a $(m \times n)$-matrix (the atypicality matrix [19, 20]), and give the matrix entries in the (m,n)-rectangle together with the Young frame of λ. This is illustrated here for example (1.9) :

18	16	13	12	11	10	9	8			
15	13	10	9	8	7	6	5			
10	8	5	4	3	2	1	0			
7	5	2	1	0	−1	−2	−3			
5	3	0	−1	−2	−3	−4	−5			

(2.10)

So for this example, $\operatorname{atyp}(\Lambda_\lambda + \rho) = 3$, since the atypicality matrix contains three zeros. In the following, we shall sometimes refer to (i,j) as the position of β_{ij}; the position is also identified by a box in the (m,n)-rectangle.

The maximal isotropic subset S_{Λ_λ} consist now of those odd roots β_{ij} with $(\Lambda_\lambda + \rho, \beta_{ij}) = 0$, i.e. it corresponds to the zeros in the atypicality matrix. From the combinatorics of atypicality matrices [20], it follows that

$$S_{\Lambda_\lambda} = \{\beta_{i,\lambda_i+m+1-i} | 1 \leq i \leq m,\ 1 \leq \lambda_i + m + 1 - i \leq n\}.$$

That is to say, one finds the zeros in the atypicality matrix as follows : on row m in column $\lambda_m + 1$; on row $m-1$ in column $\lambda_{m-1} + 2$; etc. (as long as these column indices are not exceeding n); see also in (2.10).

Clearly, S_{Λ_λ} is in general not a subset of the set of simple roots Π (since Π contains only one odd root, $\epsilon_m - \delta_1$). So formula (2.7) cannot be applied. The purpose is to show that there exists a sequence of simple odd α-reflections such that for the new Δ'_+, where the module V_λ has highest weight Λ', there exists a $(\Lambda' + \rho')$-maximal isotropic subset $S_{\Lambda'}$ with $S_{\Lambda'} \subset \Pi' \subset \Delta'_+$.

Definition 2.1. *For $\lambda \in \mathcal{H}_{m,n}$, the (m,n)-index of λ is the number*

$$k = \min\{i | \lambda_i + m + 1 - i \leq n\}, \qquad (1 \leq k \leq m+1). \tag{2.11}$$

In what follows, k will always denote this number; it will be a crucial entity for our developments. If $k = m+1$, then $S_{\Lambda_\lambda} = \emptyset$ and V_λ is typical and trivially tame. Thus in the following, we shall assume that $k \leq m$. To begin with, Δ_+ corresponds to the distinguished choice, and Π is the distinguished set of simple roots (2.2). The highest weight of V_λ is given by Λ_λ. Denote $\Lambda^{(1)} = \Lambda_\lambda$, $\rho^{(1)} = \rho$ and $\Pi^{(1)} = \Pi$. Now we perform a sequence of simple odd $\alpha^{(i)}$-reflections; each of these reflections preserve $\Delta_{0,+}$ but may change $\Lambda^{(i)}+\rho^{(i)}$ and $\Pi^{(i)}$. Denote the sequence of reflections by :

$$\Lambda^{(1)}+\rho^{(1)},\,\Pi^{(1)} \quad \xrightarrow{\alpha^{(1)}} \quad \Lambda^{(2)}+\rho^{(2)},\,\Pi^{(2)} \quad \xrightarrow{\alpha^{(2)}} \quad \cdots \quad \xrightarrow{\alpha^{(f)}} \quad \Lambda'+\rho',\,\Pi' \tag{2.12}$$

where, at each stage, $\alpha^{(i)}$ is an odd root from $\Pi^{(i)}$. For given λ, consider the following sequence of odd roots (with positions on row m, row $m-1$, ..., row k) :

$$\begin{array}{ll} \text{row } m \ : & \beta_{m,1},\beta_{m,2},\ldots,\beta_{m,\lambda_k-k+m} \\ \text{row } m-1 \ : & \beta_{m-1,1},\beta_{m-1,2},\ldots,\beta_{m-1,\lambda_k-k+m-1} \\ \vdots & \vdots \\ \text{row } k \ : & \beta_{k,1},\beta_{k,2},\ldots,\beta_{k,\lambda_k} \end{array} \tag{2.13}$$

in this particular order (i.e. starting with $\beta_{m,1}$ and ending with β_{k,λ_k}). Then we have :

Lemma 2.2. *The sequence (2.13) is a proper sequence of simple odd reflections for Λ_λ, i.e. $\alpha^{(i)}$ is a simple odd root from $\Pi^{(i)}$. At the end of the sequence, one finds :*

$$\begin{aligned} \Pi' \ = \ & \{\epsilon_1-\epsilon_2,\epsilon_2-\epsilon_3,\ldots,\epsilon_{k-2}-\epsilon_{k-1},\epsilon_{k-1}-\delta_1,\delta_1-\delta_2,\delta_2-\delta_3,\ldots,\delta_{\lambda_k-1}-\delta_{\lambda_k}, \\ & \delta_{\lambda_k}-\epsilon_k,\epsilon_k-\delta_{\lambda_k+1},\delta_{\lambda_k+1}-\epsilon_{k+1},\epsilon_{k+1}-\delta_{\lambda_k+2},\ldots,\delta_{\lambda_k+m-k}-\epsilon_m,\epsilon_m-\delta_{\lambda_k+m+1-k}, \\ & \delta_{\lambda_k+m+1-k}-\delta_{\lambda_k+m+2-k},\ldots,\delta_{n-1}-\delta_n\}. \end{aligned} \tag{2.14}$$

Furthermore,

$$\Lambda'+\rho' = \Lambda_\lambda+\rho+\sum_{i=k+1}^{m}\sum_{j=\lambda_i+1}^{\lambda_k-k+i}\beta_{i,j}. \tag{2.15}$$

Proof. Let us consider, in the first stage, the reflections with respect to the roots in row m. Clearly, $\alpha^{(1)}=\beta_{m,1}=\epsilon_m-\delta_1$ is an odd root from $\Pi^{(1)}=\Pi$. Performing the reflection with respect to $\beta_{m,1}$ implies that $\Pi^{(2)}$ contains $\epsilon_{m-1}-\delta_1,\delta_1-\epsilon_m,\epsilon_m-\delta_2$ as simple odd roots. Thus $\Pi^{(2)}$ contains $\alpha^{(2)}=\beta_{m,2}$. A reflection with respect to β_{m-2} implies that $\Pi^{(3)}$ contains $\epsilon_{m-1}-\delta_1,\delta_2-\epsilon_m,\epsilon_m-\delta_3$ as simple odd roots. So this process continues, and after λ_k-k+m such reflections (i.e. at the end of row m), we have

$$\begin{aligned} \Pi^{(\lambda_k-k+m+1)} = \ & \{\epsilon_1-\epsilon_2,\ldots,\epsilon_{m-2}-\epsilon_{m-1},\epsilon_{m-1}-\delta_1,\delta_1-\delta_2,\ldots,\delta_{\lambda_k-k+m-1}-\delta_{\lambda_k-k+m}, \\ & \delta_{\lambda_k-k+m}-\epsilon_m,\epsilon_m-\delta_{\lambda_k-k+m+1},\delta_{\lambda_k+m+1-k}-\delta_{\lambda_k+m+2-k},\ldots,\delta_{n-1}-\delta_n\}. \end{aligned} \tag{2.16}$$

Observe that this process can continue since $\lambda_k-k+m<n$ by definition of the (m,n)-index k of λ. So after the first stage (i.e. after the reflections with respect to odd roots of row m) there are three odd roots in $\Pi^{(\lambda_k-k+m+1)}$, and the set is ready to continue the reflections with respect to the elements of row $m-1$, since $\beta_{m-1,1}$ belongs to $\Pi^{(\lambda_k-k+m+1)}$. From (2.5) and (2.6) we have that

$$\begin{aligned} \Lambda^{(i+1)}+\rho^{(i+1)} &= \Lambda^{(i)}+\rho^{(i)} \ \text{ if } \ (\Lambda^{(i)}+\rho^{(i)},\alpha^{(i)})\neq 0, \\ \Lambda^{(i+1)}+\rho^{(i+1)} &= \Lambda^{(i)}+\rho^{(i)}+\alpha^{(i)} \ \text{ if } \ (\Lambda^{(i)}+\rho^{(i)},\alpha^{(i)})=0. \end{aligned}$$

Examining this explicitly for the elements of row m yields

$$\Lambda^{(\lambda_k-k+m+1)}+\rho^{(\lambda_k-k+m+1)} = \Lambda_\lambda+\rho+\sum_{j=\lambda_m+1}^{\lambda_k-k+m}\beta_{m,j}. \tag{2.17}$$

If $k=m$ the lemma follows. If $k<m$ the process continues; suppose this is the case. But now we are in a situation where the elements of row $m-1$ play completely the same role

as those of row m in the first stage. This means that at the end of the second stage, the new set of simple roots is given by

$$\Pi^{(i)} = \{\epsilon_1 - \epsilon_2, \ldots, \epsilon_{m-3} - \epsilon_{m-2}, \epsilon_{m-2} - \delta_1, \delta_1 - \delta_2, \ldots, \delta_{\lambda_k-k+m-2} - \delta_{\lambda_k-k+m-1},$$
$$\delta_{\lambda_k-k+m-1} - \epsilon_{m-1}, \epsilon_{m-1} - \delta_{\lambda_k-k+m}, \delta_{\lambda_k-k+m} - \epsilon_m, \epsilon_m - \delta_{\lambda_k-k+m+1},$$
$$\delta_{\lambda_k+m+1-k} - \delta_{\lambda_k+m+2-k}, \ldots, \delta_{n-1} - \delta_n\}, \tag{2.18}$$

and the new $\Lambda^{(i)} + \rho^{(i)}$ by

$$\Lambda^{(i)} + \rho^{(i)} = \Lambda_\lambda + \rho + \sum_{j=\lambda_m+1}^{\lambda_k-k+m} \beta_{m,j} + \sum_{j=\lambda_{m-1}+1}^{\lambda_{m-1}-k+m-1} \beta_{m-1,j} \tag{2.19}$$

(the last addition follows by inspecting the atypicality matrix). Continuing with the remaining stages (i.e. rows in (2.13)) leads to (2.14) and (2.15). □

Corollary 2.3. *The covariant module V_λ is tame.*

Proof. Having performed the simple odd reflections (2.13), one can compute the atypicality matrix for $\Lambda' + \rho'$ using (2.15). This gives :

$$(\Lambda' + \rho', \beta_{ij}) = 0 \text{ for all } (i,j) \text{ with } k \le i \le m, \lambda_k + 1 \le j \le \lambda_k + m + 1 - k. \tag{2.20}$$

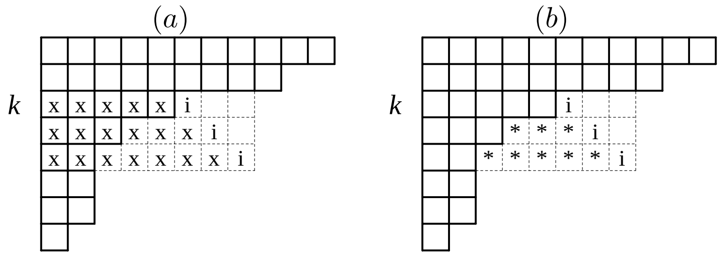
Therefore the set

$$S'_\Lambda = \{\epsilon_k - \delta_{\lambda_k+1}, \epsilon_{k+1} - \delta_{\lambda_k+2}, \ldots, \epsilon_m - \delta_{\lambda_k+m+1-k}\} \tag{2.21}$$

is a $(\Lambda' + \rho')$-maximal isotropic subset. Furthermore, $S'_\Lambda \subset \Pi'$, see (2.14). □

Let us illustrate some of these notions for example (1.9).

(2.22)



In (2.22)(a) the positions marked with "i" refer to the $(\Lambda'+\rho')$-maximal isotropic set (2.21). The first element, with row index k, is simply the position of the box just to the right of the Young diagram F^λ in row k. For the remaining positions one continues in the direction of the diagonal until one reaches row m. For convenience, let us refer to these positions as "the isotropic diagonal." The positions of the odd roots that have been used for the sequence of reflections to go from Λ_λ and Π to Λ' and Π' are marked by "x" in (2.22)(a). So, they are simply all positions to the left of the isotropic diagonal. Finally, (2.22)(b) shows the positions of those β_{ij} that appear on the rhs of (2.15); they are marked by "*". These are all positions to the left of the isotropic diagonal that are not inside F^λ. One can see from these examples that the (m,n)-index k determines all other necessary ingredients.

We are now in a position to evaluate the character formula (2.7),

$$\operatorname{ch} V_\lambda = j_{\Lambda'}^{-1} e^{-\rho'} R'^{-1} \sum_{w\in W} \varepsilon(w) w\Big(e^{\Lambda'+\rho'} \prod_{\beta\in S_{\Lambda'}} (1+e^{-\beta})^{-1}\Big), \tag{2.23}$$

where R' is given by (2.8) with $\Delta_{1,+}$ replaced by $\Delta'_{1,+}$ ($\Delta_{0,+}$ remains unchanged). The mn elements of $\Delta'_{1,+}$ are $\pm\beta_{ij}$, where one must take the minus-sign if β_{ij} appears in the list (2.13) (i.e. if its position is marked by "x" in (2.22)(a)) and the plus-sign otherwise. However, all this information is not necessary here, since by definition of ρ and R

$$e^{-\rho'}R'^{-1} = e^{-\rho}R^{-1}.$$

Putting, as usual in this context,

$$x_i = e^{\epsilon_i}, \qquad y_j = e^{\delta_j} \qquad (1 \le i \le m, 1 \le j \le n), \tag{2.24}$$

one has

$$e^{-\rho}R^{-1} = D^{-1}\prod_{i=1}^{m} x_i^{(m-n-1)/2}\prod_{j=1}^{n} y_j^{(n-m-1)/2},$$

with D given in (1.15). Using (2.15) and (2.21), one finds

$$e^{\Lambda'+\rho'}\prod_{\beta\in S_{\Lambda'}}(1+e^{-\beta})^{-1} = \prod_{i=1}^{m} x_i^{(1-m+n)/2}\prod_{j=1}^{n} y_j^{(1+m-n)/2}\prod_{i=1}^{k-1} x_i^{\lambda_i+m-i-n}\prod_{i=k}^{m} x_i^{\lambda_k+m+1-k-n}$$

$$\times\prod_{j=1}^{\lambda_k} y_j^{\lambda'_j+n-j-m}\prod_{j=\lambda_k+1}^{\lambda_k+m+1-k} y_j^{n-\lambda_k-m-1+k}\prod_{j=\lambda_k+m+2-k}^{n} y_j^{n-j}\Big/\prod_{i=k}^{m}(x_i+y_{\lambda_k+i+1-k}). \tag{2.25}$$

In order to rewrite this in a more appropriate form, let us introduce some further notation related to the partition $\lambda \in \mathcal{H}_{m,n}$. Clearly, the (m,n)-index k defined in (2.11) plays again an essential role. Related to this, let us also put

$$l = \lambda_k + 1, \qquad r = n - m + k - l. \tag{2.26}$$

Now we have

$$\operatorname{ch} V_\lambda = j_{\Lambda'}^{-1} D^{-1}\sum_{w\in W}\varepsilon(w)w(t_\lambda),$$

with

$$t_\lambda = \prod_{i=1}^{k-1} x_i^{\lambda_i+m-i-n}\prod_{j=1}^{l-1} y_j^{\lambda'_j+n-j-m}\prod_{i=k}^{m}\frac{y_{l+i-k}^r}{x_i^r(x_i+y_{l+i-k})}\prod_{j=l+m+1-k}^{n} y_j^{n-j}. \tag{2.27}$$

This form also allows us to deduce $j_{\Lambda'} = j_{\Lambda_\lambda}$. Indeed, consider in $W = S_m \times S_n$ the subgroup H of elements $w = \sigma_x \times \sigma_y$, where σ_x is a permutation of $(x_k, x_{k+1}, \ldots, x_m)$ and σ_y is the same permutation of $(y_l, y_{l+1}, \ldots, y_{l+m-k})$. Each element of H leaves t_λ invariant. Furthermore, $\varepsilon(w) = 1$ for $w \in H$. Since H is isomorphic to S_{m-k+1},

$$\sum_{w\in H}\varepsilon(w)w(t_\lambda) = (m-k+1)!\; t_\lambda.$$

So, in order to have multiplicity one for the highest weight term, one must take $j_{\Lambda'} = (m-k+1)!$.

We can now conlude this section by the following two alternative formulas for the computation of $\operatorname{ch} V_\lambda$ (i.e. of $s_\lambda(x/y)$) :

$$\operatorname{ch} V_\lambda = \frac{D^{-1}}{(m-k+1)!}\sum_{w\in S_m\times S_n}\varepsilon(w)w(t_\lambda) \tag{2.28}$$

$$= D^{-1}\sum_{w\in(S_m\times S_n)/S_{m-k+1}}\varepsilon(w)w(t_\lambda), \tag{2.29}$$

where the second sum is over the cosets of $H = S_{m-k+1}$ in $S_m \times S_n$.

3. A determinantal formula for $s_\lambda(x/y)$

Let $\lambda \in \mathcal{H}_{m,n}$; the important quantities related to λ are given by the (m,n)-index k (see (2.11)) and the related numbers l and r (see (2.26)). Since $\operatorname{ch} V_\lambda = s_\lambda(x/y)$, (2.28) or (2.29), together with the expression (2.27) for t_λ, yield a (new) expression for the supersymmetric S-function. In this section we shall rewrite (2.28) in a nicer form; in particular we shall show that it is equivalent to a determinantal form for $s_\lambda(x/y)$. Then we shall end with an alternative proof.

The first step in this process is the following :

Lemma 3.1. *Let t_λ be given by (2.27). Then*

$$\frac{1}{(m-k+1)!}\sum_{w\in S_m\times S_n}\varepsilon(w)w(t_\lambda) = (-1)^{(k-1)(l-1)}\det(C), \tag{3.1}$$

where C is the following square matrix of order $n+k-1$:

$$C = \begin{pmatrix} 0 & Y_\lambda^{(1)} \\ X_\lambda & R^{(r)} \\ 0 & Y^{(r)} \end{pmatrix} \tag{3.2}$$

with

$$\text{(3.3)}\quad X_\lambda = \left(x_i^{\lambda_j+m-n-j}\right)_{1\le i\le m,\ 1\le j\le k-1}, \qquad R^{(r)} = \left(\frac{y_j^r}{x_i^r(x_i+y_j)}\right)_{1\le i\le m,\ 1\le j\le n}$$

$$\text{(3.4)}\quad Y_\lambda^{(1)} = \left(y_j^{\lambda'_i+n-m-i}\right)_{1\le i\le l-1,\ 1\le j\le n}, \qquad Y^{(r)} = \left(y_j^{r-i}\right)_{1\le i\le r,\ 1\le j\le n}$$

Proof. The result is obtained by applying Laplace's theorem [16, Section 1.8] for the expansion of $\det(C)$ with respect to columns $1, 2, \ldots, k-1$. □

This determinantal formula can be rewritten, leading to our main result :

Theorem 3.2. *Let $\lambda \in \mathcal{H}_{m,n}$ and k be the (m,n)-index of λ. Then*

$$s_\lambda(x/y) = (-1)^{mn-m+k-1}D^{-1}\det\begin{pmatrix} \left(\frac{1}{x_i+y_j}\right)_{\substack{1\le i\le m\\ 1\le j\le n}} & \left(x_i^{\lambda_j+m-n-j}\right)_{\substack{1\le i\le m\\ 1\le j\le k-1}} \\ \left(y_j^{\lambda'_i+n-m-i}\right)_{\substack{1\le i\le n-m+k-1\\ 1\le j\le n}} & 0 \end{pmatrix}. \tag{3.5}$$

Proof. Consider C as defined in the previous lemma and let R_i denote its row i. Let C' denote the modified matrix C where every column R_{l-1+i} has been multiplied by x_i^r (for all i with $1 \le i \le m$). The result can be found by applying the following row operations on the modified matrix C' :

$$R'_{l-1+i} \longrightarrow R'_{l-1+i} - \sum_{s=0}^{r-1}(-1)^s x_i^s R'_{l+m+s} \qquad (1 \le i \le m)$$

taking into account that $\lambda_{k-1} - \lambda_k \ge r \ge 0$ and $\lambda'_{l-1+i} = k-1$ for $1 \le i \le r$ (k, l and r defined as in (2.11) and (2.26)). □

In Section 1 we have already pointed out that for $\lambda = 0$ and $m \le n$, (3.5) gives rise to the determinantal identity (1.18) combining Cauchy's double alternant with Vandermonde's determinant. It is easy to verify that for $\lambda = 0$ and $m > n$ we have that $k = n - m + 1$, and then (3.5) gives rise to an identity equivalent to (1.18).

Finally, let us consider the case with $k = m+1$ (corresponding to a typical representation V_λ in terms of the previous section). Then the blocks X_λ and Y_λ are square matrices, and one finds :

$$\begin{aligned} s_\lambda(x/y) &= D^{-1}\det(x_i^{\lambda_j+m-n-j})_{1\le i,j\le m}\det(y_j^{\lambda'_i+n-m-i})_{1\le i,j\le n} \\ &= s_\tau(x)s_{\eta'}(y)\prod_{i=1}^m\prod_{j=1}^n(x_i+y_j) = s_\tau(x)s_{\eta'}(y)R(x,-y), \end{aligned} \tag{3.6}$$

where τ and η are the parts of λ defined in (1.9). This is in agreement with (1.12).

Once the determinantal expression has been established, there are several ways to prove that the rhs of (3.5) is indeed a supersymmetric S-polynomial. A first proof is based on the fact that, in [10, §I.3, exercise 23], Macdonald showed that the supersymmetric S-polynomials satisfy four properties (see also [13]) which also characterize these polynomials. [1] We have shown that the polynomials defined by means of the rhs of (3.5) do indeed satisfy these four properties (see an earlier version of this paper, available from the authors).

Another and more direct proof was outlined by a referee and is based on the formula (1.10) of Sergeev-Pragracz (see [12] and [19]) and Laplace's theorem [16, Section 1.8].

Let us write $x = x' + x''$ for a decomposition of $x = (x_1, x_2, \dots, x_m)$ into two disjoint subsets of fixed size, say $|x'| = p$ and $|x''| = q$ with $p + q = m$. Recall the definition of $R(x', x'')$ in (1.14).

Lemma 3.3. *For $m = p + q$, let $\mu = (\mu_1, \dots, \mu_p)$, $\nu = (\nu_1, \dots, \nu_q)$ be two partitions and $\lambda = (\mu_1, \dots, \mu_p, \nu_1, \dots, \nu_q)$. Then*

$$\sum_{x'+x''}\frac{s_{\mu+(q^p)}(x')s_\nu(x'')}{R(x',x'')} = s_\lambda(x) \tag{3.7}$$

where the sum is over all possible decompositions $x = x' + x''$ with the size of x' equal to p and the size of x'' equal to q.

Proof. We can rewrite the lhs of expression (3.7) using the determinantal formula for S-functions and the equality $D(x) = (-1)^{\frac{p(p+1)}{2}+r_1+\cdots+r_p}D(x')D(x'')R(x',x'')$ with the elements of x' denoted by $x_{r_1}, \dots, x_{r_p}$ and those of x'' by $x_{s_1}, \dots, x_{s_q}$:

$$\sum_{x'+x''}\frac{s_{\mu+(q^p)}(x')s_\nu(x'')}{R(x',x'')} = \frac{(-1)^{\frac{p(p+1)}{2}}}{D(x)}\sum_{x'+x''}(-1)^{r_1+\cdots+r_p}|x_{r_i}^{\mu_j+q+m-j}||x_{s_i}^{\nu_j+m-j}|. \tag{3.8}$$

The numerator of this sum is the Laplace expansion of the following determinant with respect to columns $1, 2, \dots, p$:

$$\left|\ x^{\mu+(q^p)+\delta_p} \quad x^{\nu+\delta_q}\ \right| = \left|x^{\lambda+\delta_m}\right| \tag{3.9}$$

with $\lambda = (\mu_1, \dots, \mu_p, \nu_1, \dots, \nu_q)$ and δ_l given in (1.4). The result follows. □

[1] Our convention for the $s_\lambda(x/y)$ is slightly different from that of Macdonald's $s_\lambda^{\mathrm{Mac}}(x/y)$: $s_\lambda(x/y) = s_\lambda^{\mathrm{Mac}}(x/-y)$.

Lemma 3.4. *Let $m = p+q$ and $\lambda = (\lambda_1, \lambda_2, \ldots)$ be a partition. Denote $\lambda^{(p)} = (\lambda_1, \ldots, \lambda_p) + (q^p)$ and $\lambda_*^{(p)} = (\lambda_{p+1}, \lambda_{p+2}, \ldots)$. Then,*

$$s_\lambda(x/-y) = \sum_{x'+x''} \frac{s_{\lambda^{(p)}}(x'/-y)s_{\lambda_*^{(p)}}(x''/-y)}{R(x',x'')} \tag{3.10}$$

where the sum is over all possible decompositions $x = x' + x''$ with the size of x' equal to p and the size of x'' equal to q.

Proof. One can use, e.g., induction on n, i.e. the number of variables $y = (y_1, \ldots, y_n)$. If $n = 0$ then we are reduced to the symmetric case and the result follows from Lemma 3.3. Otherwise, one separates the variable y_n and sums over all partitions μ such that $\lambda - \mu$ is a vertical strip (v.s.) :

$$\begin{aligned}
s_\lambda(x/-y^{(n)}) &= \sum_{\mu\subseteq\lambda,(\text{v.s.})} s_\mu(x/-y^{(n-1)})s_{(r)}(-y_n) \\
&= \sum_{\mu\subseteq\lambda,(\text{v.s.})}\sum_{x'+x''} \frac{s_{\mu^{(p)}}(x'/-y^{(n-1)})s_{\mu_*^{(p)}}(x''/-y^{(n-1)})}{R(x',x'')}(-y_n)^r \\
&= \sum_{x'+x''}\left[\left(\sum_{\mu^{(p)}\subseteq\lambda^{(p)},(\text{v.s.})} s_{\mu^{(p)}}(x'/-y^{(n-1)})(-y_n)^s\right)\left(\sum_{\mu_*^{(p)}\subseteq\lambda_*^{(p)},(\text{v.s.})} s_{\mu_*^{(p)}}(x''/-y^{(n-1)})(-y_n)^t\right)\right] \\
&\quad \times\frac{1}{R(x',x'')} - \sum_{\substack{\mu\subseteq\lambda,(\text{v.s.})\\ \mu_{p+1}=\mu_p+1}}\sum_{x'+x''} \frac{s_{\mu^{(p)}}(x'/-y^{(n-1)})s_{\mu_*^{(p)}}(x''/-y^{(n-1)})}{R(x',x'')}(-y_n)^r \qquad (\text{with } r = s+t)
\end{aligned}$$

The first part in this expression is the result we want. For the second part, one can by induction use (3.10); then it is easy to see that these contributions are zero. ◻

We can now give a direct proof of the determinantal formula.

Theorem 3.5. *Let $\lambda \in \mathcal{H}_{m,n}$ and k be the (m,n)-index of λ. Then*

$$\frac{R(x,y)}{D(x)D(y)} \det\begin{pmatrix} \frac{1}{x-y} & X_\lambda \\ Y_\lambda & 0 \end{pmatrix} = \pm s_\lambda(x/-y), \tag{3.11}$$

where the (rectangular) blocks of the determinant are given by

$$\frac{1}{x-y} = \left(\frac{1}{x_i - y_j}\right)_{1\le i\le m,\, 1\le j\le n},$$
$$X_\lambda = \left(x_i^{\lambda_j+m-n-j}\right)_{1\le i\le m,\, 1\le j\le k-1}, \qquad Y_\lambda = \left(y_j^{\lambda'_i+n-m-i}\right)_{1\le i\le n-m+k-1,\, 1\le j\le n}.$$

Proof. Suppose $|x'| = k-1 \equiv p$ and $|y'| = n-m+k-1 \equiv q$. We will indicate the indices of the elements of x' by i_t ($t = 1, \ldots, p$) and those of y' by j_t ($t = 1, \ldots, q$). The determinant in (3.11) has a double Laplace expansion, with partitions $\alpha = (\lambda_1, \ldots, \lambda_{k-1}) - (q^p)$ and

$\beta = (\lambda'_1, \ldots, \lambda'_q) - (p^q)$ determined by X_λ and Y_λ, so the lhs of (3.11) equals :

$$\begin{aligned}
&\frac{R(x,y)}{D(x)D(y)} \sum_{x=x'+x''} \sum_{y=y'+y''} (-1)^P D(x')D(y')s_\alpha(x')s_\beta(y') \det\left(\frac{1}{x''-y''}\right) \\
&(\text{ with } P = \frac{p(p+1)}{2} + \frac{q(q+1)}{2} + \sum_{t=1}^{p} i_t + \sum_{t=1}^{q} j_t) \\
= \ &\frac{R(x,y)}{D(x)D(y)} \sum_{x'+x''} \sum_{y'+y''} (-1)^P D(x')D(y')s_\alpha(x')s_\beta(y')D(x'')D(y'')/R(x'',y'') \\
= \ &\sum_{x'+x''} \sum_{y'+y''} s_\alpha(x')s_\beta(y') \frac{R(x,y')R(x',y'')}{R(x',x'')R(y',y'')} \\
= \ &\sum_{y'+y''} \left(\sum_{x'+x''} \frac{s_\alpha(x')R(x',y'')}{R(x',x'')} \right) \frac{s_\beta(y')(-1)^{mq}R(y',x)}{R(y',y'')}.
\end{aligned}$$

Now we can apply the special case (1.12) of the Sergeev-Pragacz formula twice. Putting $\eta = \alpha + ((m-k+1)^p)$ and $\chi = \beta + (m^q) = (\lambda'_1, \ldots, \lambda'_q) + ((m-k+1)^q)$, there comes

$$\pm \sum_{y'+y''} \left(\sum_{x'+x''} \frac{s_\eta(x'/-y''))}{R(x',x'')} \right) \frac{s_\chi(y'/-x)}{R(y',y'')} = \pm \sum_{y'+y''} \left(\sum_{x'+x''} \frac{s_\eta(x'/-y''))}{R(x',x'')} \right) \frac{s_\chi(-y'/x)}{R(y',y'')}.$$

Finally, we use Lemma 3.4, and duality (1.13); the last expression becomes :

$$\pm \sum_{y'+y''} s_\alpha(x/-y'')s_\chi(-y'/x)\frac{1}{R(y',y'')} = \pm(-1)^{|\alpha|} \sum_{y'+y''} \frac{s_{\alpha'}(-y''/x)s_\chi(-y'/x)}{R(y',y'')}.$$

As $\alpha' = (\lambda'_{q+1}, \lambda'_{q+2}, \ldots)$ and $\chi = (\lambda'_1, \ldots, \lambda'_q) + ((m-k+1)^q)$, this is equal to $\pm s_{\lambda'}(-y/x) = \pm s_\lambda(x/-y)$, using once more Lemma 3.4 and duality. □

Thus, the proof uses essentially a double Laplace expansion, twice the application of the factorization case of the Sergeev-Pragacz formula, and twice Lemma 3.4. Finally, observe that the sign in Theorem 3.11 depends on the partition λ, in particular it is equal to

$$(-1)^{\sum_{i=1}^{n-m+k-1} \lambda'_i + \frac{m(m-1)}{2} + \frac{n(n-1)}{2} - \frac{k(k-1)}{2} - 1}$$

Acknowledgements

The authors would like to thank the referee for pointing out a second proof of the superdeterminant formula, now given here in Section 3.

References

[1] E.L. Basor and P.J. Forrester, Formulas for the evaluation of Toeplitz determinants with rational generating functions, Mathematische Nachrichten, 170 (1994), 5-18.
[2] N. Bergeron and A. Garsia, Sergeev's formula and the Littlewood-Richardson rule, Linear and Multilinear Algebra, 27 (1990), 79-100.
[3] A. Berele and A. Regev, Hook Young diagrams with applications to combinatorics and to representations of Lie superalgebras, Advances in Mathematics, 64 (1987), 118-175.
[4] V.G. Kac, Lie superalgebras, Advances in Mathematics, 26 (1977), 8-96.
[5] V.G. Kac, Representations of classical Lie superalgebras, Lecture Notes in Mathematics 676, pp. 597-626, Springer, Berlin, 1978.

[6] V.G. Kac and M. Wakimoto, Integrable highest weight modules over affine superalgebras and number theory, Progress in Mathematics, 123 (1994), 415-456.

[7] R.C. King, Supersymmetric functions and the Lie supergroup $U(m/n)$, Ars Combinatoria, 16 A (1983), 269-287.

[8] C. Krattenthaler, Advanced determinant calculus, Séminaire Lotharingien Combinatoire, 42 (1999), Article B42q, 67 pp.

[9] A. Lascoux and P. Pragacz, Ribbon Schur functions, European Journal of Combinatorics, 9 (1988), 561-574.

[10] I.G. Macdonald, *Symmetric functions and Hall polynomials*, Oxford University Press, Oxford, 1979.

[11] I. Penkov and V. Serganova, On irreducible representations of classical Lie superalgebras, Indagationes Mathematica, 3 (1992), 419-466.

[12] P. Pragacz, Algebro–geometric applications of Schur S- and Q-polynomials, Topics in invariant theory (Paris, 1989/1990), Lecture Notes in Mathematics 1478, pp. 130-191, Springer, Berlin, 1991.

[13] P. Pragacz and A. Thorup, On a Jacobi-Trudi identity for supersymmetric polynomials, Advances in Mathematics, 95 (1992), 8-17.

[14] M. Scheunert, *The theory of Lie superalgebras*, Springer, Berlin, 1979.

[15] A.N. Sergeev, Tensor algebra of the identity representation as a module over the Lie superalgebras $Gl(n,m)$ and $Q(n)$ (Russian), MatematicheskiĭSbornik, 123 (1984), 422-430.

[16] G.E. Shilov, *Linear Algebra*, Dover Publications, New York, 1977.

[17] J. Stembridge, A characterization of supersymmetric polynomials, Journal of Algebra, 95 (1985), 439-444.

[18] J. Van der Jeugt and V. Fack, The Pragacz identity and a new algorithm for Littlewood-Richardson coefficients, Computers and Mathematics with Applications, 21 (1991), 39-47.

[19] J. Van der Jeugt, J.W.B. Hughes, R.C. King and J. Thierry-Mieg, Character formulas for irreducible modules of the Lie superalgebra $sl(m/n)$, Journal of Mathematical Physics, 31 (1990), 2278-2304.

[20] J. Van der Jeugt, J.W.B. Hughes, R.C. King and J. Thierry-Mieg, A character formula for singly atypical modules of the Lie superalgebra $sl(m/n)$, Communications in Algebra, 18 (1990), 3453-3480.

[21] J. Van der Jeugt and R.B. Zhang, Characters and composition factor multiplicities for the Lie superalgebra $gl(m/n)$, Letters in Mathematical Physics, 47 (1999), 49-61.

Department of Applied Mathematics and Computer Science, University of Ghent, Krijgslaan 281-S9, B-9000 Gent, Belgium

E-mail address: {ElsM.Moens, Joris.VanderJeugt}@rug.ac.be

ASYMPTOTICS OF MULTIVARIATE SEQUENCES

ROBIN PEMANTLE AND MARK WILSON

Abstract. We discuss the problem of coefficient extraction from multivariate generating functions. We concentrate on the authors' ongoing project (see `http://www.math.auckland.ac.nz/~wilson/Research/mvGF/`). Some representative results are given without proof.

Résumé. Nous discutons du problème de l'extraction de coefficients pour les fonctions génératrices en plusieurs variables. Nous nous concentrons sur le projet continu des auteurs (voir `http://www.math.auckland.ac.nz/ wilson/Research/mvGF/`). Quelques résultats représentatifs sont donnés sans preuve.

1. Introduction

The generating function $F(z) := \sum_{r=0}^{\infty} a_r z^r$ for the sequence $a_0, a_1, a_2, \ldots$ is one of the most useful constructions in combinatorics. If the function F has a simple description, it is usually not too hard to obtain F as a formal power series once one understands a recursive or combinatorial description of the numbers $\{a_r\}$. One may then analyze the analytic properties of F in order to obtain asymptotic information about the sequence $\{a_r\}$. While still part art and part science, this latter analytic step has become quite systematized. [Sta97] in his introduction to enumerative combinatorics gives the example of the function $F(z) = \exp(z + \frac{z^2}{2})$, from which he says "it is routine (for someone sufficiently versed in complex variable theory) to obtain the asymptotic formula $a_r = 2^{-1/2} r^{r/2} e^{-r/2+\sqrt{r}-1/4}$." Routine, in this case, means a single application of the saddle point method. When F has singularities in the complex plane, the analysis is often more direct: the location of the singularities and the behavior of F near these determine almost algorithmically the asymptotic behavior of the sequence $\{a_r\}$. For those not sufficiently versed in complex variable theory, two useful sources are [Hen77] and [Odl95]. The transfer theorems of [FO90] encapsulate much of this knowledge in a very useful way; see also [Wil94] for an elementary introduction.

When the sequence $a_0, a_1, a_2, \ldots$ is replaced by a multidimensional array $\{a_{r_1,\ldots,r_d}\}$, things become much more hit and miss. Let us use boldface to denote vectors in $\mathbb{C}^d$ or $\mathbb{N}^d$, and use multi-index notation, so that $a_{\mathbf{r}}$ denotes the multi-index $a_{r_1,\ldots,r_d}$ and $\mathbf{z}^{\mathbf{r}}$ denotes the product $z_1^{r_1} \cdots z_d^{r_d}$. The generating function $F : \mathbb{C}^d \to \mathbb{C}$ is defined analogously to the one-dimensional generating function by

$$F(\mathbf{z}) = \sum_{\mathbf{r} \in \mathbb{N}^d} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}.$$

Surprisingly, techniques for extracting asymptotics of $\{a_{\mathbf{r}}\}$ from the analytic properties of F were, until recently, almost entirely missing. In a survey of asymptotic methods, [Ben74] says:

> Practically nothing is known about asymptotics for recursions in two variables even when a generating function is available. Techniques for obtaining asymptotics from bivariate generating functions would be quite useful.

In the intervening 25 years, some results have appeared, addressing chiefly the case where the array $\{a_{\mathbf{r}}\}$ obeys a central limit theorem. Common to all of these is the following method. Treat $\{a_{\mathbf{r}}\}$ as a sequence of $(d-1)$-dimensional arrays indexed by r_d; show that the n^{th} $(d-1)$-dimensional generating function is roughly the n^{th} power of a given function; use this approximation to invert the characteristic function and obtain a Central Limit Theorem. We refer to these methods as GF-sequence methods. The other body of work on multivariate sequences, which we will call the diagonal method, is based on algebraic extraction of the diagonal, as found in [HK71] (see also [Fur67] and later [Lip88] for an algebraic description of the scope of this method; variants are described in [Sta99] and [Pip]).

The most fundamental GF-sequence result is probably [BR83], with extensions appearing in later work of the same authors. [FS] presents a version of the same idea which holds in much greater generality. [GR92] go beyond the central limit case, using the transfer theorems of [FO90] to handle functions that are products of powers with powers of logs. Recent work of Bender and Richmond [BR96, BR99] extends the applicability of the central limit results to many problems of combinatorial interest; see also [Hwa95, Hwa98b], where more precise asymptotics are given, and [Hwa98a], which extends some results to the combinatorial schemes of [FS93]. This does not exhaust the recent work on the problem of multivariable coefficient extraction, but does circumscribe it. It is interesting to note that Odlyzko's survey article [Odl95] devoted only 4% of its space to multivariate problems.

The present authors' recent research on multivariate coefficient extraction has concentrated on functions belonging to a certain class that includes rational functions, although the basic ideas surely extend to a larger class. Desirable features in any theory we develop include: generality of application, complete expansions and not just leading terms, expansions that are as uniform as possible in the direction $\mathbf{r}/|\mathbf{r}|$, and formulae for expansions that lead to effective computation. An ultimate goal is to systematize the extraction of multivariate asymptotics sufficiently that it may be automated in a computer algebra system.

Our methods are analytic, based on complex contour integration, and are outlined below.

For the remainder of this paper, we will assume that the formal power series F converges in a neighborhood of the origin and may be analytically continued everywhere except a set $\mathcal{V}$ of complex dimension $d-1$ which we call the *singular variety*.

A crude preliminary step in approximating $a_{\mathbf{r}}$ is to determine its exponential rate; in other words, to estimate $\log|a_{\mathbf{r}}|$ up to a factor of $1+o(1)$. Let $\mathcal{D}$ denote the (open) domain of convergence of F and let $\log\mathcal{D}$ denote the logarithmic domain in $\mathbb{R}^{d+1}$, that is, the set of $\mathbf{x}\in\mathbb{R}^{d+1}$ such that $e^{\mathbf{x}}\in\mathcal{D}$. If $\mathbf{z}\in\mathcal{D}$ then Cauchy's integral formula

$$a_{\mathbf{r}} = \left(\frac{1}{2\pi i}\right)^{d+1}\int_{T(\mathbf{z})}\frac{F(\mathbf{w})}{\mathbf{w}^{\mathbf{r}+\mathbf{1}}}\,d\mathbf{w} \tag{1}$$

shows that $a_{\mathbf{r}} = O(|\mathbf{z}|^{-\mathbf{r}})$. Letting $\mathbf{z}\to\partial\mathcal{D}$ gives

$$\log|a_{\mathbf{r}}| \le -\mathbf{r}\cdot\log|\mathbf{z}| + o(|\mathbf{r}|),$$

and optimizing in $\mathbf{z}$ gives $\log|a_{\mathbf{r}}|\le\gamma(\mathbf{r})+o(|\mathbf{r}|)$ where

$$\gamma(\mathbf{r}) := -\sup_{\mathbf{x}\in\log\mathcal{D}}\mathbf{r}\cdot\mathbf{x}\,. \tag{2}$$

The cases in which the most is known about $a_{\mathbf{r}}$ are those in which this upper bound is correct, that is, $\log|a_{\mathbf{r}}| = \gamma(\mathbf{r})+o(|\mathbf{r}|)$. To explain this, note first that the supremum in (2) is equal to $\mathbf{r}\cdot\mathbf{x}$ for some $\mathbf{x}\in\partial\log\mathcal{D}$. The torus centred at the origin and containing $\mathbf{x}$ must contain some minimal singularity $\mathbf{z} \in \mathcal{V} \cap \partial\mathcal{D}$. Asking that $\log |a_{\mathbf{r}}| \sim -\mathbf{r} \cdot \mathbf{z}$ is then precisely the same as requiring the Cauchy integral (1) — or the residue integral mentioned above — to be of roughly the same order as its integrand. This is the situation in which it easiest to estimate the integral.

Our approach may now be summarized as follows. Associated to each minimal singularity $\mathbf{z}$ is a cone $\boldsymbol{\kappa}(\mathbf{z}) \subseteq (\mathbb{R}^+)^{d+1}$. Given $\mathbf{r}$, we find one or more $\mathbf{z} = \mathbf{z}(\mathbf{r}) \in \mathcal{V} \cap \partial\mathcal{D}$ where the upper bound is least. We then attempt to compute a residue integral there. This works only if $\mathbf{r} \in \boldsymbol{\kappa}(\mathbf{z})$ and if the residue computation is of a type we can handle.

So far we have performed this residue computation in an explicit way by representing $a_{\mathbf{r}}$ (up to an exponentially smaller term) as a $(d-1)$-dimensional integral of one-variable residues. We have not used the more symmetric Leray residue theory largely because it has not been clear to us how to obtain sufficiently explicit results. Although we have broken the symmetry between coordinates, many of our formulae can be re-symmetrized with moderate effort.

The key analytic tool in extracting asymptotics has been the theory of oscillatory integrals. We reduce the residue computation to the computation of integrals of the form $\int_D \exp(-\lambda f(\mathbf{z}))\psi(\mathbf{z})\, d\mathbf{z}$ where f, ψ are smooth complex and complex-valued with $\operatorname{Re} f \geq 0$, and D is a compact product of simplices and intervals. Several challenges arise because the existing literature apparently does not contain the exact results needed; in particular the boundary terms arising in these integrals cannot always be neglected.

To amplify on this, define a point $\mathbf{z} \in \mathcal{V}$ to be *minimal* if $\mathbf{z} \in \partial\mathcal{D}$ and each z_j is nonzero. Note that $\mathbf{z}$ is minimal if $D(\mathbf{z}) \cap \mathcal{V} \subseteq T(\mathbf{z}$, where $T(\mathbf{z})$ and $D(\mathbf{z})$ are respectively the torus and disk centred at $\mathbf{0}$ and containing $\mathbf{z}$. A minimal point is *strictly minimal* if $D(\mathbf{z}) \cap \mathcal{V} = \{\mathbf{z}\}$. When a minimal point is not strictly minimal, one must add (or integrate) contributions from all points of $\mathcal{V} \cap T(\mathbf{z})$; this step is fairly routine.

There are only three possible types of minimal singularities [PW1, Lemma 6.1]), namely smooth points (where $\mathcal{V}$ is locally a graph of an analytic function); multiple points (where $\mathcal{V}$ is locally a union of graphs of analytic functions), and cone points (all others). It is conjectured that for all three types of points, and any $\mathbf{r} \in \boldsymbol{\kappa}(\mathbf{z})$, we indeed have

$$\log |a_{\mathbf{r}}| = \gamma(\mathbf{r}) + o(|\mathbf{r}|) = -\mathbf{r} \cdot \log |\mathbf{z}| + o(|\mathbf{r}|)\,.$$

This is proved for smooth points in [PW1] via residue integration, and the complete asymptotic series obtained. It is proved in [PW2] for multiple points under various assumptions; the fact that these do not cover all cases seems due more to taxonomical problems rather than the inapplicability of the method. The problem remains open for cone points, along with the problem of computing asymptotics.

The chief purpose of our work is to give a solution to the problem of asymptotic evaluation of coefficients that is as general as possible. An important part of this is re-derivation in a general setting of results obtainable via GF-sequence or *ad hoc* methods. Our results allow us to show that our method successfully finds asymptotics for every function in a certain large class. Familiar examples from this class include: lattice path counting, various known generating functions for polyominos and stacked balls, enumeration of Catalan trees by number of components or surjections by image cardinality (see [FS]), stopping times for certain random walks (see [LL99]), as well as the examples given in the GF-sequence papers of [Ben73] and [BR83]: ordered set partitions enumerated by number of blocks, permutations enumerated by rises, and Tutte polynomials of recursive sequences of graphs.

Nevertheless, our pursuit of this problem was also motivated by some specific applications and challenge problems. These are cases where known methods do not suffice to obtain complete asymptotic information. There is a class of tiling enumeration problems for which

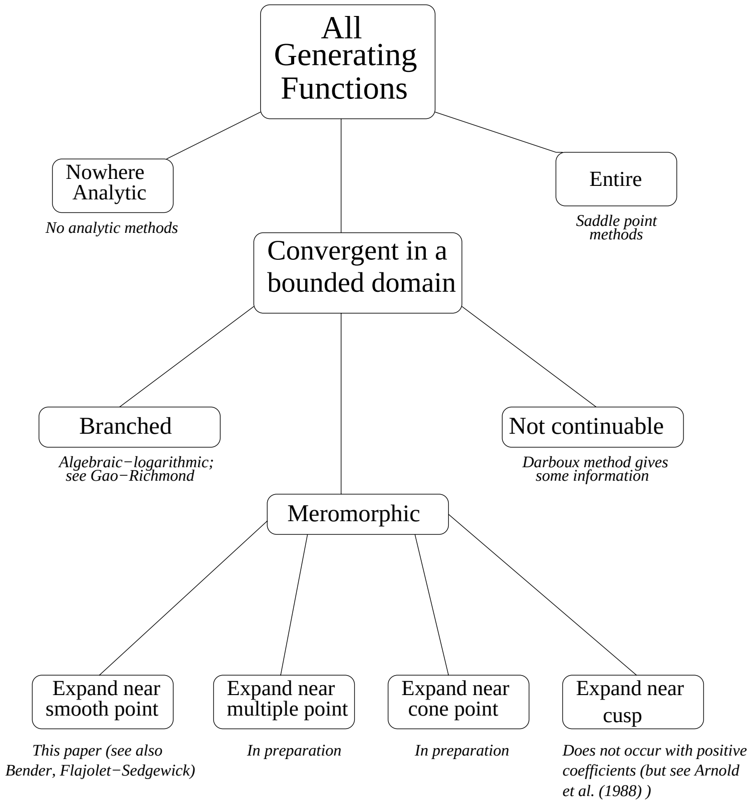


FIGURE 1. Classification of generating functions

an explicit three variable rational generating function may be obtained. This class includes the Aztec Diamond domino tilings of [CEP96]. Asymptotics in the so-called *region of fixation* are obtained from analysis of the smooth points of $\mathcal{V}$, while asymptotics in the region of positive entropy]are derived from analysis of the cone point. [CP] applies a cone point analysis to a tiling enumeration problem for which the only previous results are some pictures via simulation (`http://www.math.ohio-state.edu/ pemantle/Pix/plot.gif`). Another motivation has been to solve the general multivariable linear recursion. Depending on whether one allows forward recursion in some of the variables, one typically obtains either rational or algebraic generating functions (less well-behaved functions can also result — see [BM00] for more details). The general rational function may have any of the types of singularities mentioned above: smooth points, nodes, cones, cusps, branchpoints, etc. Even the simple rational generating function $1/(3 - 3z - w + z^2)$ requires two separate analyses in order to get asymptotics in all directions.

In the rest of this section, we describe the scope of our methods and compare them with previous work. Figure 1 depicts a classification of generating functions and illustrates the remainder of this paragraph. If a formal power series is nowhere convergent, analytic methods are useless. Among those power series converging in some neighborhood of the origin, there are three possibilities: a function may be entire, may have singularities around which analytic continuations exist, or it may be defined only on some bounded subset of $\mathbb{C}^d$. Our methods are tailored to the second class. The third class, although in some sense generic, seldom arises in any problem for which the generating function may be effectively described. Incomplete asymptotic information is available via Darboux' method; details of this method in the univariate case are given in [Hen77] and [Odl95]. The first class can and does arise frequently. Our methods are simply not equipped to handle entire functions, and systematizing the asymptotic analysis of coefficients of entire generating functions remains an important open problem.

Most existing literature on the multivariate coefficient extraction problem deals (in effect, although results are not necessarily stated in this way) with analysis near smooth pole points, whose asymptotics usually exhibit central limit behavior. There are several ways in which our treatment of the smooth case improves upon available analyses.

First, most of the existing results assume that the singular point $\mathbf{z} \in \mathcal{V}$ has positive real coordinates, and that it is strictly minimal in a sense defined in the next section. This assumption often holds when the coefficients $\{a_{\mathbf{r}}\}$ are nonnegative reals, though it will fail if, for example, there is any periodicity. The assumption always fails when the coefficients $\{a_{\mathbf{r}}\}$ have mixed signs, as is the case for example with the generating functions $(1-zw)/(1-2zw+w^2)$ and $1/(1-2zw+w^2)$ for the Chebyshev polynomials of the first and second kinds [Com74, page 50]. GF-sequence methods may be adapted to some of these situations. Indeed, the presentation of these methods by [FS, Theorem 9.7] accomplishes this adaptation in great generality. But certainly there are cases such as the rational generating function $1/(1 - z - w + \beta zw)$, where the points $\mathbf{z}$ with given moduli form a continuum and standard GF-sequence methods are not sufficient.

Second, our methods obtain automatically a full asymptotic expansion of $a_{r_1,\ldots,r_d}$ in decreasing powers of the indices r_j. This is certainly not inherent in the existing results, whose relatively short proofs involve inversion of the characteristic function (see however [Hwa95] and [Hwa96] for something in this direction). The expansion to n terms is completely effective in terms of the first n partial derivatives of $1/F$ at $\mathbf{z}$, as is the error bound.

Third, these results explicitly cover the case where the pole at $\mathbf{z}$ has order greater than 1. The behavior in this case is not according to the central limit theorem. The only existing work addressing this case is [GR92], and they require nonnegativity assumptions, as mentioned above. In the case where $F = G^k$ is an exact power, one could attempt first to solve the problem for G and then to take the k-fold convolution. This is much harder than the present approach, as may be seen by the rather involved computation in [CEP96].

Fourth, the potential for increasing the scope to new applications seems greater for contour methods than for GF-sequence methods. The contour method reduces the asymptotic problem to the problem of an oscillating integral near a singularity, which can almost certainly be done. By contrast, the GF-sequence method requires first an understanding of the sequence of $(d-1)$-dimensional generating functions arising from the given d-dimensional generating function, and then another result in order to transfer this information to asymptotics of the coefficients $a_{\mathbf{r}}$.

Fifth, although our results in the case of smooth pole points are often similar to those obtained by GF-sequence methods, our hypotheses are quite different. Our hypotheses may be universally established for bivariate functions that generate nonnegative values and are meromorphic through their domain of convergence.

Finally, we compare our method to recent results from the diagonal method. It is known [Lip88] that the diagonal sequence $a_{n,n,\ldots,n}$ of a multivariate sequence with rational generating function has a generating function satisfying a linear differential equation over rational functions. Much is known about how to compute this equation (see for example [Chy98]). If one wants asymptotics on the diagonal, or in any direction where the coordinate ratios are rational numbers with small denominators, then these methods give results that are in theory at least as good as ours. The method, unlike ours, is inherently non-uniform in the direction, so there is no hope of extending it to larger sets of directions.

2. RESULTS

To state our results as cleanly as possible, we assume that there are analytic functions G, H of $d+1 \geq 2$ variables with $F = G/H$ in the neighbourhood of the strictly minimal element $\mathbf{z} = \mathbf{1}$ of $\mathcal{V}$ (if $\mathbf{z} \neq \mathbf{1}$, a factor of $\mathbf{z}^{-\mathbf{r}}$ is introduced).

We first discuss our results, in the generic case (that is, with appropriate conditions on transversality of intersection and nonvanishing of G at $\mathbf{z}$), from a qualitative perspective. The following results are valid as $|\mathbf{r}| \to \infty$ for $\mathbf{r}$ in a certain cone $\boldsymbol{\kappa}(\mathbf{z})$ that collapses to a single ray in the smooth case. In fact $\boldsymbol{\kappa}$ can easily be described geometrically as being spanned by the outward normals to the support hyperplanes of $\log\mathcal{D}$ at $\log \mathbf{z}$.

In both smooth and multiple point cases, there is an asymptotic expansion of $a_{\mathbf{r}}$ in descending powers of $|\mathbf{r}|$. If $\mathbf{z}$ is smooth then we have Ornstein-Zernike (central limit) behaviour, $a_{\mathbf{r}} \sim C|\mathbf{r}|^{-d/2}$. By contrast, suppose $\mathbf{z}$ is a multiple point, the intersection of $n+1 \geq 2$ sheets of $\mathcal{V}$.

- If $n = d$ we have asymptotic constancy throughout the cone: $a_{\mathbf{r}} \sim C$ for some C, and the error term is exponentially small (the series has only one term).
- If $n \geq d$, then there are several subcones of $\boldsymbol{\kappa}$ on each of which we get $a_{\mathbf{r}} \sim P(\mathbf{r})$ for some polynomial of degree at most $n - d$. Thus the series is finite.
- If $n \leq d$ then asymptotics start with $|\mathbf{r}|^{n/2-d/2}$.
- Finally, suppose that our transversality conditions are dropped and that all sheets are in fact tangent at $\mathbf{z}$. Then asymptotics start with $|\mathbf{r}|^{n-d/2}$.

We now present a couple of 2-dimensional results and examples contained in [PW1, PW2], chosen because they can be stated with no extra notation and maintain the symmetry between the variables (which is broken in our analysis,, so formulae must be re-symmetrized). The numerous more complex results referred to above require the introduction of considerably more notation in order to give explicit formulae for the leading coefficient of the asymptotic expansion, and are not re-symmetrized. The quantities involved depend on the Weierstrass factorization of H near $\mathbf{z}$ and various derivatives of G and H. To the best of our knowledge, all of these are explicitly computable by symbolic algebra in appropriate commutative rings

Theorem 1. *Let $F = G/H$ be a meromorphic function of two variables, not singular at the origin. Define*

$$Q(z,w) := -w^2 H_w^2 z H_z - w H_w z^2 H_z^2 - w^2 z^2 \left(H_w^2 H_{zz} + H_z^2 H_{ww} - 2 H_z H_w H_{zw} \right).$$

Then

$$a_{r,s} \sim \frac{G(z,w)}{\sqrt{2\pi}} z^{-r} w^{-s} \sqrt{\frac{-wH_w}{sQ(z,w)}}$$

uniformly as (z,w) varies over a compact set of strictly minimal, simple poles of F on which Q and G are nonvanishing, and $(r,s) \in \boldsymbol{\kappa}(z,w)$.

Remark. Usually the expression in the radical will be positive real, as will the coefficients a_{rs}. The result is true in general, though, as long as the square root is taken to be $-(wH_w)^{-1}$ times the principal root of $(-wH_w^3)/(sQ)$. Also note that when $(r,s) \in \mathbf{dir}(z,w)$ then the expression wH_w/s is coordinate-invariant, that is, equal to zH_z/r. Thus the given expression for $a_{r,s}$ has the expected symmetry.

Example 2 (Lattice paths)**.** Let $a_{r,s}$ be the number of nearest-neighbor paths from the origin to (r,s) moving only north, east and northeast; these are sometimes called *Delannoy numbers* [Sta99, page 185]. The generating function is $F(z,w) = 1/(1-z-w-zw)$. The zero set $\mathcal{V}$ of $H = 1 - z - w - zw$ is given by $w = (1-z)/(1+z)$, and the minimal points of $\mathcal{V}$ are those where $w \in [0,1]$. With the help of relations that hold when $\mathbf{z} \in \mathcal{V}$ we may compute as follows.

$$\begin{aligned} H_z &= -1 - w \\ -zH_z &= 1 - w \\ Q &= (1-z)(1-w)(1-zw) \\ \frac{zH_z}{wH_w} &= \frac{1-w}{1-z} = \frac{1-w^2}{2w} \end{aligned}$$

with H_w and $-wH_w$ given by reversing z and w. As z varies over $[\varepsilon, 1-\varepsilon]$, the functions Q and $G := 1$ do not vanish. The minimal pair (z,w) that solves $(r,s) \in \mathbf{dir}(z,w)$ is given by $z = (\sqrt{r^2+s^2} - s)/r$ and $w = (\sqrt{r^2+s^2} - r)/s$. Theorem 1 then gives

$$\begin{aligned} a_{rs} &\sim \left(\frac{\sqrt{r^2+s^2}-s}{r}\right)^{-r} \left(\frac{\sqrt{r^2+s^2}-r}{s}\right)^{-s} \sqrt{\frac{1}{2\pi}} \sqrt{\frac{1-z}{s}\frac{1}{1-zw}} \\ &= \left(\frac{\sqrt{r^2+s^2}-s}{r}\right)^{-r} \left(\frac{\sqrt{r^2+s^2}-r}{s}\right)^{-s} \sqrt{\frac{1}{2\pi}} \sqrt{\frac{rs}{(r+s-\sqrt{r^2+s^2})^2\sqrt{r^2+s^2}}}, \end{aligned}$$

uniformly when r/s and s/r remain bounded. In particular, when $r = s = n$, this gives the following formula for the n^{th} diagonal coefficient (which may alternatively be obtained by computing the diagonal generating function $(1 - 6s + s^2)^{-1/2}$ according to the method given in [Sta99, Section 6.3]:

$$(\sqrt{2}-1)^{-2n} \sqrt{\frac{1}{2\pi} \frac{2^{-1/4}}{2-\sqrt{2}}}.$$

In the next result, a *boundedly interior* subset of a cone is one that is interior, and bounded away from the walls.

Theorem 3 (2 curves meeting transversally in 2-space)**.** *Let F be a meromorphic function of two variables, not singular at the origin, with $F(z,w) = G(z,w)/H(z,w) = \sum_{r,s} a_{rs} z^r w^s$.*

Suppose that (z,w) is a strictly minimal, double point of $\mathcal{V}$. Let $\mathcal{H}(z,w)$ denote the Hessian of H at (z,w).

Then for each boundedly interior subset K of $\boldsymbol{\kappa}(z,w)$, there is $c > 0$ such that

$$a_{rs} = z^{-r} w^{-s} \left(\frac{G(z,w)}{\sqrt{-z^2 w^2 \det \mathcal{H}(z,w)}} + O(e^{-c|(z,w)|}) \right) \qquad \textit{uniformly for } (r,s) \in K.$$

Example 4 (combinatorial application)**.** An independent sequence of random numbers uniform on $[0,1]$ is used to generate biased coin-flips: if p is the probability of heads then a number $x \le p$ means heads and $x > p$ means tails. The coins will be biased so that $p = 2/3$ for the first n flips, and $p = 1/3$ thereafter. A player desires to get r heads and s tails and is allowed to choose n. On average, how many choices of $n \le r+s$ will be winning choices?

The probability that n is a winning choice for the player is precisely

$$a_{rs} := \sum_{a+b=n} \binom{n}{a} (2/3)^a (1/3)^b \binom{r+s-n}{r-a} (1/3)^{r-a} (2/3)^{s-b}.$$

Let a_{rs} be this expression summed over n. The array $\{a_{rs}\}_{r,s\ge 0}$ is just the convolution of the arrays $\binom{r+s}{r}(2/3)^r(1/3)^s$ and $\binom{r+s}{r}(1/3)^r(2/3)^s$, so the generating function $F(z,w) :=$

$\sum a_{rs} z^r w^s$ is the product

$$F(z, w) = \frac{1}{(1 - \frac{1}{3}z - \frac{2}{3}w)(1 - \frac{2}{3}z - \frac{1}{3}w)} .$$

Applying Theorem 3 with $G \equiv 1$ and $\det \mathcal{H} = -1/9$, we see that $a_{rs} = 3$ plus a correction which is exponentially small as $r, s \to \infty$ with $r/(r + s)$ staying in any subinterval of $(1/3, 2/3)$. A purely combinatorial analysis of the sum may be carried out to yield the leading term, 3, but says nothing about the correction terms. The diagonal extraction method of [HK71] yields very precise information for $r = s$ but nothing more general in the region $1/3 < r/(r + s) < 2/3$. □

3. Comments and further work

The greatest obstacle to making all these computations completely effective lies in the location of the minimal point $\mathbf{z}$ given $\mathbf{r}$. Assuming the existence of a $\mathbf{z}(\mathbf{r})$ with $\mathbf{r} \in \boldsymbol{\kappa}(\mathbf{z})$, how may we compute $\mathbf{z}(\mathbf{r})$ and test whether it is a minimal point? Since the moduli of the coordinates of $\mathbf{z}$ are involved in the definition of minimality, this is a problem in real rather than complex computational geometry and does not appear easy.

Another natural question is whether there exists such a minimal $\mathbf{z}(\mathbf{r})$. When F generates nonnegative coefficients, the answer is generally yes. Examples show that when the coefficients have mixed signs, the answer is no. We conjecture for every direction there is a (not necessarily minimal) point $\mathbf{z} \in \mathcal{V}$ for which integration near $\mathbf{z}$ yields correct asymptotics. For example, if $G = 1$ and

$$H = (1 - (2/3)w - (1/3)z)(1 + (1/3)w - (2/3)z)$$

then the point $(3/2, 3/4)$ is not minimal but yields asymptotics in the diagonal direction; one sees this by integrating along a deformed torus rather than along $T(3/2, 3/4)$. In fact we conjecture that such a deformation always exists, but the topology seems not transparent enough to yield an easy proof.

The problem of determining asymptotics when $\widehat{\mathbf{r}}$ converges to the boundary of $\boldsymbol{\kappa}$ is dual to the problem of letting $\widehat{\mathbf{r}}$ converge to $\partial \boldsymbol{\kappa}$ from the outside. Solutions to both of these problems are necessary before we understand asymptotics "in the gaps", that is, in any region asymptotic to and containing a direction in the boundary of $\boldsymbol{\kappa}$. For example, what are the asymptotics for $a_{r, r+\sqrt{r}}$ as $r \to \infty$?

Many of our theorems rule out analysis of a minimal point $\mathbf{z}$ if one of the coordinates z_j is zero. The directions in $\boldsymbol{\kappa}(\mathbf{z})$ will always have $r_j = 0$, in which case the analysis of coefficient asymptotics reduces to a case with one fewer variable. Thus it appears no generality is lost. If we are, however, able to solve the previous problem, wherein $\mathbf{r}$ converges to $\partial \boldsymbol{\kappa}$, then we may choose to let $\mathbf{r}$ converge to something with a zero component. The problem of asymptotics when some $r_j = o(r_k)$ now makes sense and is not reducible to a previous case. Presumably these asymptotics are governed by the minimal point $\mathbf{z}$ still, but it must be sorted out which of our results persist when $z_j = 0$. Certainly the geometry near $\mathbf{z}$ has more possibilities, since it is easier to be a minimal point (it is easier to maintain $|z'_j| \geq |z_j|$ for $\mathbf{z}'$ near $\mathbf{z}$ when $z_j = 0$).

References

[Ben73] Edward A. Bender, *Central and local limit theorems applied to asymptotic enumeration*, J. Combinatorial Theory Ser. A **15** (1973), 91–111.

[Ben74] Edward A. Bender, *Asymptotic methods in enumeration*, SIAM Rev. **16** (1974), 485–515.

[BM00] M. Bousquet-Mélou and M. Petkovšek, *Linear recurrences with constant coefficients: the multivariate case*, Discrete Math. **225** (2000), 51–75.

[BR83] Edward A. Bender and L. Bruce Richmond, *Central and local limit theorems applied to asymptotic enumeration. II. Multivariate generating functions*, J. Combin. Theory Ser. A **34** (1983), no. 3, 255–265.

[BR96] Edward A. Bender and L. Bruce Richmond, *Admissible functions and asymptotics for labelled structures by number of components*, Electron. J. Combin. **3** (1996), no. 1, Research Paper 34, approx. 23 pp. (electronic).

[BR99] Edward A. Bender and L. Bruce Richmond, *Multivariate asymptotics for products of large powers with applications to Lagrange inversion*, Electron. J. Combin. **6** (1999), no. 1, Research Paper 8, 21 pp. (electronic).

[CEP96] Henry Cohn, Noam Elkies, and James Propp, *Local statistics for random domino tilings of the Aztec diamond*, Duke Math. J. **85** (1996), no. 1, 117–166.

[Chy98] F. Chyzak and B. Salvy, *Non-commutative elimination in ore algebras proves multivariate identities*, J. Symbol. Comput. **26** (1998), 187–227.

[Com74] Louis Comtet, *Advanced combinatorics*, enlarged ed., D. Reidel Publishing Co., Dordrecht, 1974.

[CP] H. Cohn and R. Pemantle, *A determination of the boundary of the fixation region for some domino tiling ensembles*, in preparation.

[FO90] Philippe Flajolet and Andrew Odlyzko, *Singularity analysis of generating functions*, SIAM J. Discrete Math. **3** (1990), no. 2, 216–240.

[FS] Philippe Flajolet and Robert Sedgewick, Analytic Combinatorics, Ch. 9, INRIA Res. Rep. 3162, 1997.

[FS93] Philippe Flajolet and Michèle Soria, *General combinatorial schemas: Gaussian limit distributions and exponential tails*, Discrete Math. **114** (1993), no. 1-3, 159–180.

[Fur67] H. Furstenberg, *Algebraic functions over finite fields*, J. Algebra **7** (1967), 271–277.

[GR92] Zhi-Cheng Gao and L. Bruce Richmond, *Central and local limit theorems applied to asymptotic enumeration. IV. Multivariate generating functions*, J. Comput. Appl. Math. **41** (1992), no. 1-2, 177–186.

[Hen77] Peter Henrici, *Applied and computational complex analysis. Vol. 2*, Wiley Interscience [John Wiley & Sons], New York, 1977.

[HK71] M. Hautus and D. Klarner, *The diagonal of a double power series*, Duke Math. J. **23** (1971), 613–628.

[Hwa95] Hsien-Kuei Hwang, *Asymptotic expansions for the Stirling numbers of the first kind*, J. Combin. Theory Ser. A **71** (1995), no. 2, 343–351.

[Hwa96] Hsien-Kuei Hwang, *Large deviations for combinatorial distributions. I. Central limit theorems*, Ann. Appl. Probab. **6** (1996), no. 1, 297–319.

[Hwa98a] Hsien-Kuei Hwang, *Large deviations of combinatorial distributions. II. Local limit theorems*, Ann. Appl. Probab. **8** (1998), no. 1, 163–181.

[Hwa98b] Hsien-Kuei Hwang, *On convergence rates in the central limit theorems for combinatorial structures*, European J. Combin. **19** (1998), no. 3, 329–343.

[Lip88] L. Lipshitz, *The diagonal of a D-finite power series is D-finite*, J. Algebra **113** (1988), 373–378.

[LL99] Michael Larsen and Russell Lyons, *Coalescing particles on an interval*, J. Theoret. Probab. **12** (1999), no. 1, 201–205.

[Odl95] A. M. Odlyzko, *Asymptotic enumeration methods*, Handbook of combinatorics, Vol. 1, 2, Elsevier, Amsterdam, 1995, pp. 1063–1229.

[Pip] N. Pippenger, *Enumeration of equicolourable trees*, SIAM J. Discrete Math. **14** (2001), 93–115.

[PW1] R. Pemantle and M. Wilson, *Asymptotics for multivariate sequences, I: smooth points of the singular variety*, J. Comb. Theory A (to appear).

[PW2] R. Pemantle and M. Wilson, *Asymptotics for multivariate sequences, II: multiple points of the singular variety*, preprint 2002.

[Sta97] Richard P. Stanley, *Enumerative combinatorics. Vol. 1*, Cambridge University Press, Cambridge, 1997.

[Sta99] Richard P. Stanley, *Enumerative combinatorics. Vol. 2*, Cambridge University Press, Cambridge, 1999.

[Wil94] Herbert S. Wilf, *generatingfunctionology*, second ed., Academic Press Inc., Boston, MA, 1994.

Department of Mathematics, Ohio State University, Columbus OH, USA
E-mail address: pemantle@math.ohio-state.edu

Department of Computer Science, University of Auckland, Private Bag 92019, Auckland, NEW ZEALAND
E-mail address: wilson@scitec.auckland.ac.nz

DISCRETE BOUNDARY-VALUE PROBLEMS (EXTENDED ABSTRACT)

MARKO PETKOVŠEK

Abstract. We consider partial linear recurrence equations in several dimensions over a domain obtained from the first orthant of the integer lattice by restricting it in one or more dimensions to an initial segment of the nonnegative integers. By means of the kernel method we obtain an explicit expression for the generating function of the solution in the case of a single restricted dimension, provided that the apex of the recurrence vanishes in all the remaining dimensions. Unlike the initial-value problem (where the domain is unrestricted in all dimensions) with apex of this type, this generating function depends rationally on the generating functions of the boundary conditions and of the right-hand side of the recurrence equation. As an example, we count lattice paths between two parallel hyperplanes with rational incline. By using the kernel method twice we also solve a discrete Dirichlet problem in two dimensions.

Résumé. Nous considérons des relations de récurrence linéaires qui définissent des suites multi-dimensionnelles. Les points de $\mathbb{N}^d$ qui indexent ces suites appartiennent à une intersection de "tranches" de la forme $0 \le x_i \le m_i$, pour un certain nombre de directions contraintes i. Grâce à la méthode du noyau, nous obtenons une expression explicite de la série génératrice de ces suites dans le cas où une seule dimension est contrainte, à condition que l'apex de la récurrence soit nul dans toutes les autres directions. Contrairement au cas où aucune dimension n'est contrainte, cette série génératrice est une fonction rationnelle des séries génératrices données par les conditions initiales et du membre de droite de la relation de récurrence. Notre approche s'applique par exemple à l'énumération des chemins situés entre deux hyperplans de pente rationnelle. En utilisant deux fois la méthode du noyau, nous résolvons aussi un problème de Dirichlet en deux dimensions.

1. Introduction and notation

Let $D = D_1 \times D_2 \times \cdots \times D_d$ where each factor D_i is either equal to $\mathbb{N}$ or to an initial segment $\{0, 1, \ldots, N_i\}$ of $\mathbb{N}$, for some $N_i \in \mathbb{N}$. In the latter case, we say that D is *restricted in dimension i*. We study multivariate generating functions

$$F(\boldsymbol{x}) = F(x_1, \ldots, x_d) = \sum_{n_1 \in D_1, \ldots, n_d \in D_d} a_{n_1,\ldots,n_d} x_1^{n_1} \cdots x_d^{n_d} = \sum_{\boldsymbol{n} \in D} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}} \tag{1}$$

whose coefficients satisfy a linear *recurrence relation* with constant coefficients

$$c_{\boldsymbol{h}_0} a_{\boldsymbol{n}+\boldsymbol{h}_0} + c_{\boldsymbol{h}_1} a_{\boldsymbol{n}+\boldsymbol{h}_1} + \cdots + c_{\boldsymbol{h}_k} a_{\boldsymbol{n}+\boldsymbol{h}_k} = b_{\boldsymbol{n}} \quad \text{for } \boldsymbol{n} \in R, \tag{2}$$

as well as *boundary conditions* of the form

$$a_{\boldsymbol{n}} = \varphi(\boldsymbol{n}), \quad \text{for } \boldsymbol{n} \in D \setminus R, \tag{3}$$

where $c_{\boldsymbol{h}_i} \in \mathbb{C} \setminus \{0\}$, $\boldsymbol{h}_0 = \boldsymbol{0}$, $H = \{\boldsymbol{h}_0, \boldsymbol{h}_1, \ldots, \boldsymbol{h}_k\} \subseteq \mathbb{Z}^d$ is the set of *shifts*, $R = \{\boldsymbol{n} \in \mathbb{N}^d;\ \boldsymbol{n} + H \subseteq D\}$ is the *range of validity* of (2), and $b : R \to \mathbb{C}$ as well as $\varphi : D \setminus R \to \mathbb{C}$ are known functions.

The *initial-value problem* where $D_i = \mathbb{N}$ for all i has been treated in [2]. There it has been shown that under certain natural conditions, the solution of the initial-value problem

Partially supported by MŠZŠ RS under grant P0-0511-0101.

exists and is unique. Here we consider the *boundary-value problem* where at least one D_i is of the form $\{0, 1, \ldots, N_i\}$. In contrast to the initial-value problem, the boundary-value problem need not be solvable, or if it is, its solution need not be unique.

Similarly to [2], we define the *apex* $\boldsymbol{p}$ of (2) to be the componentwise maximum of all the shifts. Note that $\boldsymbol{p} \geq \mathbf{0}$ because we assume that $\mathbf{0} \in H$. Our main result is the following: when $D_1 = D_2 = \cdots = D_{d-1} = \mathbb{N}$, $D_d = \{0, 1, \ldots, N\}$, $p_1 = p_2 = \cdots = p_{d-1} = 0$, and $p_d \geq 0$, the generating function $F(\boldsymbol{x})$ of the solution is a *rational function*, provided that the generating functions of the right-hand side $b_{\boldsymbol{n}}$ and of the boundary values $\varphi(\boldsymbol{n})$ are rational. The degrees of the numerator and/or denominator of $F(\boldsymbol{x})$ may, however, depend on N. We provide an explicit formula which expresses $F(\boldsymbol{x})$ in terms of some *algebraic functions*. This is analogous to, say, Binet's formula for Fibonacci numbers which expresses an integer sequence in terms of powers of an algebraic number. Applications of our result include enumeration of lattice paths between parallel hyperplanes (cf. [6, 7]).

When more than one D_i is bounded, we conjecture that the generating function $F(\boldsymbol{x})$ need not be holonomic as a function of both $\boldsymbol{x}$ and N_i. An example is furnished by the two-dimensional Gambler's Ruin problem (cf. [5, 4]).

The main tool that we use is the so-called *kernel method* which permits to solve certain systems of linear functional equations that seem to involve "too many unknowns". It works by restricting the equations to algebraic varieties on which some of the unknown terms vanish, thus providing the "missing equations" (see Section 3).

Notation. We use $\mathbb{N}$ to denote the set of nonnegative integers. We write $\boldsymbol{u} = (u_1, u_2, \ldots, u_d)$ for d-tuples of numbers or indeterminates, $\mathbf{0} = (0, 0, \ldots, 0)$, $\boldsymbol{u} \geq \boldsymbol{v}$ when $u_i \geq v_i$ for $1 \leq i \leq d$, and $\boldsymbol{u} > \boldsymbol{v}$ when $u_i > v_i$ for $1 \leq i \leq d$. The monomial $x_1^{u_1} \cdots x_d^{u_d}$ is denoted $\boldsymbol{x}^{\boldsymbol{u}}$.

2. From the recurrence relation to a functional equation

Instead of $F(\boldsymbol{x})$ as defined in (1) we consider

$$F_R(\boldsymbol{x}) = \sum_{\boldsymbol{n} \in R} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}}.$$

The two generating functions $F(\boldsymbol{x})$ and $F_R(\boldsymbol{x})$ differ only by terms whose coefficients are given explicitly by the boundary conditions:

$$F(\boldsymbol{x}) = \sum_{\boldsymbol{n} \in D} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}} = \sum_{\boldsymbol{n} \in R} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}} + \sum_{\boldsymbol{n} \in D \setminus R} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}} = F_R(\boldsymbol{x}) + \sum_{\boldsymbol{n} \in D \setminus R} \varphi(\boldsymbol{n}) \boldsymbol{x}^{\boldsymbol{n}}.$$

Let us now transform our recurrence relation into a functional equation satisfied by the generating function $F_R(\boldsymbol{x})$. Multiplying (2) by $\boldsymbol{x}^{\boldsymbol{n}}$ and summing over all $\boldsymbol{n} \in R$ we obtain

$$\begin{aligned} \sum_{\boldsymbol{n} \in R} b_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}} &= \sum_{\boldsymbol{n} \in R} \boldsymbol{x}^{\boldsymbol{n}} \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} a_{\boldsymbol{n}+\boldsymbol{h}} = \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} \boldsymbol{x}^{-\boldsymbol{h}} \sum_{\boldsymbol{n} \in R+\boldsymbol{h}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}} \\ &= \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} \boldsymbol{x}^{-\boldsymbol{h}} \left[F_R(\boldsymbol{x}) + P_{\boldsymbol{h}}(\boldsymbol{x}) - M_{\boldsymbol{h}}(\boldsymbol{x})\right] \end{aligned} \tag{4}$$

where

$$P_{\boldsymbol{h}}(\boldsymbol{x}) = \sum_{\boldsymbol{n} \in (R+\boldsymbol{h}) \setminus R} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}} = \sum_{\boldsymbol{n} \in (R+\boldsymbol{h}) \setminus R} \varphi(\boldsymbol{n}) \boldsymbol{x}^{\boldsymbol{n}} \quad \text{and} \quad M_{\boldsymbol{h}}(\boldsymbol{x}) = \sum_{\boldsymbol{n} \in R \setminus (R+\boldsymbol{h})} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}}. \tag{5}$$

Here we used the obvious identity

$$\sum_{x \in A} f(x) = \sum_{x \in B} f(x) + \sum_{x \in A \setminus B} f(x) - \sum_{x \in B \setminus A} f(x)$$

valid for any sets A and B provided that all the indicated sums exist.

Definition 1. *The* apex *of (2) is the point* $\boldsymbol{p} = (p_1, p_2, \ldots, p_d) \in \mathbb{N}^d$ *defined by*

$$p_i = \max\{h_i \,:\, \boldsymbol{h} \in H\} \quad (i = 1, 2, \ldots, d).$$

The antiapex *of (2) is the point* $\boldsymbol{q} = (q_1, q_2, \ldots, q_d) \in \mathbb{N}^d$ *defined by*

$$q_i = \min\{h_i \,:\, \boldsymbol{h} \in H\} \quad (i = 1, 2, \ldots, d).$$

Multiplying (4) by $\boldsymbol{x}^{\boldsymbol{p}}$ where $\boldsymbol{p}$ is the apex of (2) we obtain

$$Q(\boldsymbol{x})F_R(\boldsymbol{x}) = U(\boldsymbol{x}) - K(\boldsymbol{x}) \tag{6}$$

where

$$Q(\boldsymbol{x}) = \sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}}, \tag{7}$$

$$K(\boldsymbol{x}) = \sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}} P_{\boldsymbol{h}}(\boldsymbol{x}) - \sum_{\boldsymbol{n}\in R} b_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{p}+\boldsymbol{n}}, \tag{8}$$

$$U(\boldsymbol{x}) = \sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}} M_{\boldsymbol{h}}(\boldsymbol{x}), \tag{9}$$

the series $P_{\boldsymbol{h}}$ and $M_{\boldsymbol{h}}$ being given by (5).

From the definition of the apex it follows that $Q(\boldsymbol{x})$ is a polynomial in $\boldsymbol{x}$ called the *characteristic polynomial* or the *kernel* of the recursion. Note that $Q(\boldsymbol{x})$ and $K(\boldsymbol{x})$ are given directly by the coefficients and the right-hand side of the recurrence relation, and by the boundary conditions.

The functional equation (6) seems to contain not one, but two unknown functions: F_R and U. We shall show below on examples how to work with such apparently ambiguous functional equations. If $U(\boldsymbol{x})$ can be found explicitly then the generating function of the corresponding solution to (2), (3) is given by

$$F_R(\boldsymbol{x}) = \frac{U(\boldsymbol{x}) - K(\boldsymbol{x})}{Q(\boldsymbol{x})}. \tag{10}$$

3. A single restricted dimension

Theorem 1. *Let* $D = \mathbb{N}^{d-1} \times \{0, 1, \ldots, N\}$ *and* $\boldsymbol{p} = (0, \ldots, 0, p)$*. If* $K(\boldsymbol{x})$ *is a rational function of* $\boldsymbol{x}$*, and the boundary-value problem* (2), (3) *is solvable, then it has a solution whose generating function* $F_R(\boldsymbol{x})$ *is a rational function of* $\boldsymbol{x}$*.*

Proof: In this case

$$\begin{aligned} R &= \{\boldsymbol{n} \in \mathbb{N}^d;\ \forall \boldsymbol{h} \in H : (\boldsymbol{n} + \boldsymbol{h} \ge \boldsymbol{0} \wedge n_d + h_d \le N)\} \\ &= \{\boldsymbol{n} \in \mathbb{N}^d;\ \boldsymbol{n} + \boldsymbol{q} \ge \boldsymbol{0} \wedge n_d + p \le N\} \end{aligned}$$

and

$$R + \boldsymbol{h} = \{\boldsymbol{n} \in \mathbb{N}^d;\ \boldsymbol{n} + \boldsymbol{q} \ge \boldsymbol{h} \wedge n_d + p \le N + h_d\},$$

where $\boldsymbol{q}$ is the antiapex of H. In order to find $U(\boldsymbol{x})$ we wish to determine the sets $R\backslash(R+\boldsymbol{h})$ for all $\boldsymbol{h} \in H$. We distinguish three cases:

1. $h_d = 0$: In this case $\boldsymbol{h} \le \boldsymbol{0}$. If $\boldsymbol{n} \in R$ then $\boldsymbol{n} + \boldsymbol{q} \ge \boldsymbol{0} \ge \boldsymbol{h}$ and $n_d + p \le N = N + h_d$, so $\boldsymbol{n} \in R + \boldsymbol{h}$. It follows that $R \subseteq R + \boldsymbol{h}$, hence $R \setminus (R + \boldsymbol{h}) = \emptyset$.

(2) $h_d > 0$: If $\boldsymbol{n} \in R$ then $n_d + p \le N < N + h_d$ and $\boldsymbol{n} + \boldsymbol{q} \ge \boldsymbol{0}$, so $n_i + q_i \ge 0 \ge h_i$ for $i \le d-1$. Therefore $\boldsymbol{n} \notin R + \boldsymbol{h}$ if and only if $n_d + q_d < h_d$ or, equivalently, $n_d \le h_d - q_d - 1$, hence

$$R \setminus (R + \boldsymbol{h}) = \{\boldsymbol{n} \in \mathbb{N}^d;\ -q_i \le n_i \text{ for } 1 \le i \le d-1,\ -q_d \le n_d \le h_d - q_d - 1\}.$$

(3) $h_d < 0$: In this case $\boldsymbol{h} \le \boldsymbol{0}$. If $\boldsymbol{n} \in R$ then $\boldsymbol{n} + \boldsymbol{q} \ge \boldsymbol{0} \ge \boldsymbol{h}$ and $n_d + p \le N$. Therefore $\boldsymbol{n} \notin R + \boldsymbol{h}$ if and only if $n_d + p < N + h_d$ or, equivalently, $n_d \ge N - p + h_d + 1$, hence

$$R \setminus (R + \boldsymbol{h}) = \{\boldsymbol{n} \in \mathbb{N}^d;\ -q_i \le n_i \text{ for } 1 \le i \le d-1,\ N - p + h_d + 1 \le n_d \le N - p\}.$$

From this and from (9) it follows that

$$\begin{aligned} U(\boldsymbol{x}) &= \sum_{\substack{\boldsymbol{h} \in H \\ h_d > 0}} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p} - \boldsymbol{h}} \sum_{n_d = -q_d}^{h_d - q_d - 1} x_d^{n_d} f_{n_d}(x_1, \ldots, x_{d-1}) \\ &+ \sum_{\substack{\boldsymbol{h} \in H \\ h_d < 0}} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p} - \boldsymbol{h}} \sum_{n_d = N - p + h_d + 1}^{N - p} x_d^{n_d} f_{n_d}(x_1, \ldots, x_{d-1}) \end{aligned} \tag{11}$$

where

$$f_{n_d}(x_1, \ldots, x_{d-1}) = \sum_{(n_1, \ldots, n_{d-1}) \ge (q_1, \ldots, q_{d-1})} a_{n_1, \ldots, n_{d-1}, n_d} x_1^{n_1} \cdots x_{d-1}^{n_{d-1}},$$

for

$$-q_d \le n_d \le p - q_d - 1 \quad \text{or} \quad N - p + q_d + 1 \le n_d \le N - p,$$

are $p - q_d$ unknown functions.

Let $\xi_1(x_1, \ldots, x_{d-1}), \ldots, \xi_r(x_1, \ldots, x_{d-1})$ be the distinct roots of $Q(\boldsymbol{x}) = 0$ considered as an equation in x_d, with respective multiplicities $m_1, \ldots, m_r$. Then $\sum_{k=1}^r m_k = \deg_{x_d} Q(\boldsymbol{x}) = p - q_d$. Notice that all the functions F_R, Q, K, U are *polynomials* in x_d. Differentiating (6) j times w.r.t. x_d where $0 \le j < m_k$, then substituting ξ_k for x_d yields

$$\begin{aligned} &\sum_{\substack{\boldsymbol{h} \in H \\ h_d > 0}} c_{\boldsymbol{h}} x_1^{-h_1} \cdots x_{d-1}^{-h_{d-1}} \sum_{n_d = -q_d}^{h_d - q_d - 1} (n_d + p - h_d)^{\underline{j}}\, \xi_k^{n_d + p - h_d - j} f_{n_d} \\ &+ \sum_{\substack{\boldsymbol{h} \in H \\ h_d < 0}} c_{\boldsymbol{h}} x_1^{-h_1} \cdots x_{d-1}^{-h_{d-1}} \sum_{n_d = N - p + h_d + 1}^{N - p} (n_d + p - h_d)^{\underline{j}}\, \xi_k^{n_d + p - h_d - j} f_{n_d} \\ &= \frac{\partial^j K(\boldsymbol{x})}{\partial x_d^j}\Big|_{x_d = \xi_k} \quad (1 \le k \le r,\ 0 \le j \le m_k - 1), \end{aligned} \tag{12}$$

a system of $p - q_d$ linear algebraic equations for the $p - q_d$ unknown f_{n_d}. Here $x^{\underline{j}}$ denotes the falling factorial power $x(x-1)\cdots(x-j+1)$.

The coefficients of the matrix and of the right-hand side of this system are rational functions of x_i and ξ_k. Assume that the roots ξ_k of $Q(\boldsymbol{x}) = 0$ considered as an equation in x_d are all simple. From the way we constructed (12) it is clear that if (12) is solvable then it has a solution which is a rational function of x_i and ξ_k, and which is *symmetric* in the variables ξ_k. It follows that this solution is a rational function of the elementary symmetric polynomials in ξ_k (cf. [1, Ch. 14, Thm. 3.17]). These in turn are the coefficients of $Q(\boldsymbol{x})$ considered as a polynomial in x_d, hence they are polynomials in $x_1, \ldots, x_{d-1}$. Therefore the functions $f_{n_d}(x_1, \ldots, x_{d-1})$ are rational, and by (11), so are $U(\boldsymbol{x})$ and $F_R(\boldsymbol{x})$.

In the case when $Q(\boldsymbol{x})$ considered as a polynomial in x_d has multiple roots we can perturb its coefficients slightly so that the roots become simple, then take the limit as the perturbation goes to 0 in the obtained rational solution.

Remark 1. When there are either no $\boldsymbol{h} \in H$ with $h_d > 0$, or no $\boldsymbol{h} \in H$ with $h_d < 0$, only a single term appears on the right side of (11) and we can determine $U(\boldsymbol{x})$ simply by substituting the roots ξ_k for x_d in (6), then using polynomial interpolation. Hence in these two special cases the problem (2), (3) is uniquely solvable.

Example 1.: Lattice paths between two diagonal lines

Consider the problem of finding the number $a_{i,j}$ of two-dimensional lattice paths from the origin to the point (i,j) which use the steps $(1,0)$ and $(0,1)$, and always stay within the diagonal strip $\{(x,y);\ x-b+1 \le y \le x+c-1\}$ where b and c are fixed positive integers. Applying the linear transformation $(x,y) \mapsto (x+y, -x+y)$ to the plane, we see that this number equals $u_{i+j,-i+j}$ where $u_{i,j}$ is the number of lattice paths from the origin to the point (i,j) which use the steps $(1,-1)$ and $(1,1)$, and always stay within the horizontal strip $\{(x,y);\ -b+1 \le y \le c-1\}$. In turn, the affine transformation $(x,y) \mapsto (x, y+b)$ shows that $u_{i,j} = v_{i,j+b}$ where $v_{i,j}$ is the number of lattice paths from $(0,b)$ to (i,j) which use the steps $(1,-1)$ and $(1,1)$ and always stay within the horizontal strip $\{(x,y);\ 1 \le y \le b+c-1\}$. Clearly,

$$v_{i,j} = \begin{cases} v_{i-1,j-1} + v_{i-1,j+1}, & \text{when } i \ge 1 \text{ and } 1 \le j \le b+c-1, \\ \delta_{(i,j),(0,b)}, & \text{when } i = 0 \text{ or } j = 0 \text{ or } j = b+c. \end{cases}$$

This is a boundary-value problem of the type (2), (3) with $D = \mathbb{N} \times \{0, 1, \ldots, b+c\}$, $R = \{1, 2, \ldots\} \times \{1, 2, \ldots, b+c-1\}$, $H = \{(-1,-1), (-1,1)\}$ and $\boldsymbol{p} = (0,1)$. The conditions of Theorem 1 are satisfied, so we expect a rational generating function. We find that

$$Q(x,y) = y - xy^2 - x$$

and the roots of Q considered as a polynomial in y are

$$\xi_{1,2}(x) = \frac{1 \pm \sqrt{1-4x^2}}{2x}.$$

From (8) and (9),

$$K(x,y) = -xy^b(y^2+1),$$

$$U(x,y) = -xy(f_1(x) + y^{b+c} f_{b+c-1}(x)),$$

and Equation (6) is

$$Q(x,y)F_R(x,y) = xy^b(y^2+1) - xy(f_1(x) + y^{b+c} f_{b+c-1}(x)).$$

After substituting $\xi_1(x)$ and $\xi_2(x)$ for y in this equation, we obtain

$$\begin{aligned} f_1(x) &= \frac{\xi_1^c(x) - \xi_2^c(x)}{x(\xi_1^{b+c}(x) - \xi_2^{b+c}(x))}, \\ f_{b+c-1}(x) &= \frac{\xi_1^b(x) - \xi_2^b(x)}{x(\xi_1^{b+c}(x) - \xi_2^{b+c}(x))}, \end{aligned}$$

so

$$F(x,y) = F_R(x,y) + y^b = \frac{y^{b+1}}{y - xy^2 - x} - y\frac{\xi_1^c(x) - \xi_2^c(x) + y^{b+c}(\xi_1^b(x) - \xi_2^b(x))}{x(y - xy^2 - x)(\xi_1^{b+c}(x) - \xi_2^{b+c}(x))}.$$

Tracing back the transformations that we have made we find the generating function of the original problem as

$$\sum_{i,j\geq 0} a_{i,j}x^iy^j = \left(\sqrt{x/y}\right)^b F(\sqrt{xy},\sqrt{y/x}) = \frac{(2y)^c g(b) + (2x)^b g(c) - g(b+c)}{(x+y-1)g(b+c)} \tag{13}$$

where

$$g(u) = \left(1-\sqrt{1-4xy}\right)^u - \left(1+\sqrt{1-4xy}\right)^u.$$

Notice that for any fixed values of b and c, (13) is a rational function of x and y.

In a similar way we can treat the problems of lattice paths between any two parallel lines (or hyperplanes) of rational slope. The obtained generating functions are rational, in agreement with results of [6] and [7].

4. The discrete Dirichlet problem

When the domain of definition D is restricted in all dimensions, the unknown sequence has only finitely many terms and so the boundary-value problem (2), (3) reduces to simple linear algebra. However, we wish to have an explicit formula for the generating function of the solution. Even though the conditions of Theorem 1 are not satisfied, the kernel method can sometimes provide the desired explicit solution. We illustrate this by providing an explicit solution of the two-dimensional Gambler's Ruin Problem [5, 4].

Example 2.: Two-dimensional Gambler's Ruin

In the one-dimensional Gambler's Ruin Problem two players start out with i and $N-i$ dollars, respectively. At each step they toss a fair coin to decide who wins a dollar from the opponent. The game is over when one of them goes bankrupt. It is well known that the expected duration of the game is $i(N-i)$ (see [3], or almost any other textbook on probability).

In the two-dimensional variant [5] the players use two different currencies, say dollars and euros. They start out with (i dollars, j euros) and ($N-i$ dollars, $M-j$ euros), respectively. At each step they toss fair coins to decide the currency and the winner. The game is over when one of them runs out of either currency. What is the expected duration of the game?

Denote by game(i,j) the game with the first player's initial assets equal to (i,j). Assume that $1\leq i\leq N-1$ and $1\leq j\leq M-1$. Then after the first step, game(i,j) turns into one of game$(i+1,j)$, game$(i-1,j)$, game$(i,j+1)$, or game$(i,j-1)$, each with probability $1/4$. It follows that the expected duration $a_{i,j}$ of game(i,j) satisfies the recurrence equation

$$a_{i,j} = \frac{a_{i+1,j}+a_{i-1,j}+a_{i,j+1}+a_{i,j-1}}{4}+1 \qquad (1\leq i\leq N-1,\ 1\leq j\leq M-1) \tag{14}$$

and the boundary conditions

$$a_{0,j} = a_{N,j} = a_{i,0} = a_{i,M} = 0 \qquad (0\leq i\leq N,\ 0\leq j\leq M). \tag{15}$$

The unknown $a_{i,j}$, $1\leq i\leq N-1, 1\leq j\leq M-1$, can be obtained from (14), (15) by straightforward linear algebra. Instead of solving this linear system of $(N-1)(M-1)$ equations, Orr and Zeilberger [5] have shown how to obtain the values $a_{1,j}=a_{N-1,j}$ $(1\leq j\leq M-1)$ and $a_{i,1}=a_{i,M-1}$ $(1\leq i\leq N-1)$ from a system containing $O(N+M)$ equations only. Then all the remaining values $a_{i,j}$ can be computed from the recurrence (14). Here we provide explicit formulas for $a_{1,j}$ and $a_{i,1}$ using the kernel method twice in a row, at two different levels. Thus we avoid the need to solve linear systems altogether.

Writing (14) as $a_{i,j}-(a_{i+1,j}+a_{i-1,j}+a_{i,j+1}+a_{i,j-1})/4 = 1$, we have a boundary-value problem of the form (2), (3) with $D = \{0,1,\ldots,N\}\times\{0,1,\ldots,M\}$, $H =$

$\{(0,0),(1,0),(-1,0),(0,1),(0,-1)\}$, $R = \{1,2,\dots,N-1\} \times \{1,2,\dots,M-1\}$, $b_{i,j} = 1$, $\boldsymbol{p} = (1,1)$,

$$Q(x,y) = xy - \frac{1}{4}(y + x^2y + x + xy^2),$$
$$K(x,y) = -\sum_{(i,j)\in R} b_{i,j}x^{i+1}y^{j+1} = -xy\frac{x^N - x}{x-1} \cdot \frac{y^M - y}{y-1},$$
$$U(x,y) = -\frac{1}{4}xy\left(f_1(x) + f_2(y) + y^M g_1(x) + x^N g_2(y)\right)$$

where

$$f_1(x) = \sum_{i=1}^{N-1} a_{i,1}x^i, \qquad f_2(y) = \sum_{j=1}^{M-1} a_{1,j}y^j,$$
$$g_1(x) = \sum_{i=1}^{N-1} a_{i,M-1}x^i, \qquad g_2(y) = \sum_{j=1}^{M-1} a_{N-1,j}y^j.$$

As the game is symmetric with respect to the players, $a_{i,1} = a_{i,M-1}$ and $a_{1,j} = a_{N-1,j}$, so $f_1(x) = g_1(x)$ and $f_2(y) = g_2(y)$. Also, because of the zero boundary conditions, $F_R(x,y) = F(x,y)$. Thus the functional equation (6) has the form

$$Q(x,y)F(x,y) = xy\left(\frac{x^N - x}{x-1} \cdot \frac{y^M - y}{y-1} - \frac{1}{4}\left(f_1(x)(1+y^M) + f_2(y)(1+x^N)\right)\right) \tag{16}$$

where $F(x,y)$, $f_1(x)$, and $f_2(y)$ are unknown polynomials of degrees in x and y not exceeding $N-1$ and $M-1$, respectively.

To apply the kernel method to equation (16), let

$$\xi(x) = -\frac{1}{2x}\left(x^2 - 4x + 1 + (x-1)\sqrt{x^2 - 6x + 1}\right) \tag{17}$$

be the analytic root of $Q(x,y) = 0$ considered as an equation in y:

$$\xi(x) = -x - 4\,x^2 - 16\,x^3 - 68\,x^4 - 304\,x^5 - 1412\,x^6 - 6752\,x^7 - \cdots$$

One can show that $\xi(\xi(x)) = x$. Substituting $\xi(x)$ for y in (16) we have

$$f_1(x)(1+\xi^M(x)) + f_2(\xi(x))(1+x^N) = 4\frac{x^N - x}{x-1} \cdot \frac{\xi^M(x) - \xi(x)}{\xi(x) - 1}, \tag{18}$$

reducing the number of unknowns from three to two. To reduce it further, we apply the kernel method again, this time to equation (18). Write

$$\omega_N = e^{\frac{\pi i}{N}}, \qquad \omega_{k,N} = \omega_N^{2k+1} \quad (k = 0,1,\dots,N-1). \tag{19}$$

Note that $\omega_{k,N}$ are the N-th roots of -1.

Lemma 1. *Let $\xi(x)$ and $\omega_{k,N}$ be as in (17) and (19), respectively. Then:*

(i) $\omega_{k,N} \neq 1$,
(ii) $\xi(\omega_{k,N}) \neq 1$,
(iii) $\xi(\omega_{k,N})^M \neq -1$.

Proof: (i) is obvious as $\omega_{k,N}^N = -1$. The assertion $\xi(x) = 1$ is equivalent to $Q(x,1) = 0$. But $Q(x,1) = -(x-1)^2/4$, so $\xi(x) = 1$ if and only if $x = 1$. Thus (ii) follows from (i). To prove (iii), write $\varphi_{k,N} = (2k+1)\pi/(2N)$. Then $Q(\omega_{k,N}, y) = -\omega_{k,N}\left(y^2 - 2y\left(2\sin^2\varphi_{k,N} + 1\right) + 1\right)/4$, whence

$$\xi\left(\omega_{k,N}\right) = \left(\sin\varphi_{k,N} - \sqrt{\sin^2\varphi_{k,N} + 1}\right)^2.$$

As a positive real number, $\xi\left(\omega_{k,N}\right)$ is not an M^{th} root of -1. □

Substitution of $\omega_{k,N}$ for x in (18), justified by Lemma 1, yields

$$f_1(\omega_{k,N}) = \alpha_{k,N,M} \qquad (k = 0, 1, \ldots, N-1),$$

where

$$\alpha_{k,N,M} = \frac{4\,i\cot\varphi_{k,N}}{1+\xi^M(\omega_{k,N})} \cdot \frac{\xi^M(\omega_{k,N}) - \xi(\omega_{k,N})}{\xi(\omega_{k,N}) - 1} \qquad (i^2 = -1). \tag{20}$$

These N values uniquely determine the unknown polynomial $f_1(x)$. To find its coefficients explicitly we use the inversion formula

$$b_k = \sum_{i=0}^{N-1} a_i\,\omega_{k,N}^i \quad (0 \le k \le N-1) \quad \Longleftrightarrow \quad a_i = \frac{1}{N}\sum_{k=0}^{N-1} b_k\,\omega_{k,N}^{-i} \quad (0 \le i \le N-1)$$

which is verifiable by straightforward computation. Writing $f_1(x) = \sum_{i=0}^{N-1} a_i x^i$ and $b_k = f_1(\omega_{k,N})$, we obtain an explicit expression for the unknown coefficients of $f_1(x)$

$$a_{i,1} = a_i = \frac{1}{N}\sum_{k=0}^{N-1} \frac{\alpha_{k,N,M}}{\omega_{k,N}^i}. \tag{21}$$

In the same way we obtain an explicit expression for the unknown coefficients of $f_2(y)$,

$$a_{1,j} = \frac{1}{M}\sum_{k=0}^{M-1} \frac{\alpha_{k,M,N}}{\omega_{k,M}^i}. \tag{22}$$

$$f_1(x) = \frac{1}{N}\sum_{j=0}^{N-1} x^j \sum_{k=0}^{N-1} \frac{\alpha_{k,N,M}}{\omega_{k,N}^j} \tag{23}$$

where $\alpha_{k,N,M}$ and $\omega_{k,N}$ are given in (20) and (19), respectively.

By interchanging the order of summations in (23), or by using Lagrange Interpolation Formula on the data from (20), we can express $f_1(x)$ with a single summation sign:

$$f_1(x) = \frac{x^N+1}{N}\sum_{k=0}^{N-1} \frac{\alpha_{k,N,M}}{1 - x\omega_{k,N}^{-1}}.$$

In the same way we obtain the other unknown function

$$f_2(y) \;=\; \frac{1}{M}\sum_{j=0}^{M-1} y^j \sum_{k=0}^{M-1} \frac{\alpha_{k,M,N}}{\omega_{k,M}^j} \;=\; \frac{y^M+1}{M}\sum_{k=0}^{M-1} \frac{\alpha_{k,M,N}}{1 - y\omega_{k,M}^{-1}}.$$

Finally we have the following explicit expression for the entire generating function $F(x,y) = \sum_{i=0}^{N}\sum_{j=0}^{M} a_{i,j}x^i y^j$:

$$F(x,y) = \frac{xy\left(4\,\frac{x^N-x}{x-1}\,\frac{y^M-y}{y-1} - (x^N+1)(y^M+1)\left(\frac{1}{N}\sum_{k=0}^{N-1}\frac{\alpha_{k,N,M}}{1-x\omega_{k,N}^{-1}} + \frac{1}{M}\sum_{k=0}^{M-1}\frac{\alpha_{k,M,N}}{1-y\omega_{k,M}^{-1}}\right)\right)}{4\,xy - (x+y)(1+xy)}.$$

In closing we note that the values $a_{i,j}$ can be given explicitly as double trigonometric sums either using the Discrete Fourier Transform, or by direct diagonalization of the linear system (14) (cf. [4]).

References

[1] M. Artin, *Algebra*, Prentice Hall, 1991.

[2] M. Bousquet-Mélou and M. Petkovšek, Linear recurrences with constant coefficients: the multivariate case, *Discrete Math.* **225** (2000) 51–75.

[3] W. D. Feller, *An Introduction to Probability Theory and its Applications*, John Wiley & Sons, New York 1950.

[4] A. Kmet and M. Petkovšek, Gambler's ruin problem in several dimensions, *Adv. Appl. Math.* **28** (2002), pp. 107–118.

[5] C. R. Orr and D. Zeilberger, A computer algebra approach to the discrete Dirichlet problem, *J. Symb. Comp.* **18** (1994) 87–90.

[6] M. Sato, Generating functions for the number of lattice paths between two parallel lines with a rational incline, *Math. Japonica* **34** (1989) 123–137.

[7] M. Sato, Generating functions for the number of lattice paths restricted by two parallel hyperplanes, *J. Stat. Plan. Inf.* **34** (1993) 251–257.

Department of Mathematics and Mechanics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia

E-mail address: Marko.Petkovsek@fmf.uni-lj.si

A BIJECTION FOR LOOPLESS TRIANGULATIONS OF A POLYGON WITH INTERIOR POINTS

DOMINIQUE POULALHON AND GILLES SCHAEFFER

Abstract. Loopless triangulations of a polygon with k vertices in $k+2n$ triangles (with interior points and possibly multiple edges) were enumerated by Mullin in 1965, using generating functions and calculations with the quadratic method.

In this article we propose a simple bijective construction of Mullin's formula. The argument rests on *conjugation of trees*, a variation of the cycle lemma designed for planar maps. In the much easier case of loopless triangulations of the sphere ($k = 3$), we recover and prove correct an unpublished construction of the second author.

Résumé. Les triangulations sans boucles d'un polygone à k côtés en $k+2n$ triangles (avec des points intérieurs et éventuellement des arêtes multiples) ont été énumérées par Mullin en 1965, à l'aide de séries génératrices et de la méthode quadratique.

Dans cet article, nous proposons une construction bijective simple de la formule de Mullin. L'argument repose sur la *conjugaison d'arbres*, une variation sur le lemme cyclique adaptée à l'énumération des cartes planaires. Dans le cas beaucoup plus facile des triangulations ($k = 3$), nous retrouvons et démontrons une construction esquissée par le second auteur.

1. Introduction

In 1965, R.C. Mullin published the following formula for the number of planar loopless triangulations of a rooted k-gon into $k+2n$ triangles (see below for precise definitions):

$$T^*_{k,n} \;=\; |\mathcal{T}^*_{k,n}| \;=\; \frac{2^{n+2}(2k+3n-1)!(2k-3)!}{(n+1)!(2k+2n)!(k-2)!^2} \tag{1}$$

for all $k \geqslant 2$ and $n \geqslant 0$ (see [Mul65] or [GJ83, p145]), which extends the well-known formula for triangulations of a k-gon without interior points:

$$T^*_{k,-1} \;=\; |\mathcal{T}^*_{k,-1}| \;=\; \frac{(2k-4)!}{(k-1)!(k-2)!} \tag{2}$$

for all $k \geqslant 3$. By duality this formula also accounts for the number of rooted non-separable planar maps with a root vertex of degree k and $k+2n$ vertices all of degree three.

In his work, R.C. Mullin was closely following the seminal steps of W.T. Tutte in his *census* papers [Tut62a, Tut62b, Tut63]. In particular Formula (1) extends Tutte's formula

$$T_n \;=\; T^*_{3,n-2} \;=\; \frac{2^{n+1}(3n)!}{n!(2n+2)!} \tag{3}$$

for rooted loopless triangulations of the sphere with $2n$ triangles (or non-separable cubic maps with $2n$ vertices). The proof itself relies, following Tutte, on a recursive decomposition of triangulations that yields a recurrence for their number. Encoding the latter into generating functions then allows for a solution through the quadratic method and a few pages of calculus.

Ever since their discovery, efforts have been made to find derivations reflecting the elegant and simple product form of this and other formulas of Tutte for planar maps. In particular a construction based on the *conjugation of trees* principle was proposed in the second author's

PhD thesis [Sch98] for Formula (3) and a few other formulas of Tutte (all, bipartite, non-separable maps). A new generalization of both Tutte's formula and a formula of Hurwitz was also proved along these lines to enumerate planar constellations [BMS00].

However two parameter formulas for triangulations like (1) seem to resist conjugation of trees. In this article we introduce a slight variation of the family under consideration, which cardinality can be easily deduced from $T^*_{k,n}$, and that appears more suitable for bijective constructions.

In view of this family $\mathcal{T}_{k,n}$, Mullin's formula reads

$$T_{k,n} \;=\; |\mathcal{T}_{k,n}| \;=\; \frac{2^{n+2}}{2k+2n}\binom{2k-2}{k}\binom{2k+3n}{n+1}. \tag{4}$$

The purpose of the present article is to provide a bijective construction of the latter formula. A main ingredient of our construction is again the *conjugation of trees* principle, and this confirms the adequacy of this approach to the bijective enumeration of planar maps.

However the bijection involves two new ingredients with respect to the treatment of Tutte's formulas. On the one hand, a *special* vertex is introduced in the construction, that allows to account for parameter k of Mullin's formulas. On the other hand, as opposed to the case of constellations [BMS00], the inverse construction does not rely on breadth-first search. Instead, in order to deal with non-separability, one has to resort on more difficult recursive arguments.

The rest of the article is organized as follows: after Formula (4) for the cardinality of $\mathcal{T}_{k,n}$ has been proved equivalent to Formula (1) for $\mathcal{T}^*_{k,n}$, we exhibit a simple family $\mathcal{E}_{k,n}$ of trees (balanced blossom trees) that are clearly enumerated by the same formula, and we define in a few lines an application φ from $\mathcal{E}_{k,n}$ that we claim onto $\mathcal{T}_{k,n}$ (Section 2, 3). Then comes the harder part, as often with bijections, namely the proof for the unbeliever that the image of the application φ is indeed $\mathcal{T}_{k,n}$ and that it is one-to-one (Section 4 and 5).

2. The enumerative formula for rooted loopless triangulations

2.1. **Definitions around planar maps.** Let us make more precise the definitions of the objects under consideration. A (planar) map is a two-cell embedding of a connected planar graph into the oriented sphere considered up to orientation preserving homeomorphisms of the sphere. Multiple edges are allowed. The degree of a vertex or a face is the number of (sides of) edges incident to that vertex or face. A face is a k-gon if it has degree k and it is incident to k distinct vertices.

A planar map is non-separable if it contains no cut-vertex, that is to say no vertex that can be cut into two vertices (each taking part of the edges) in a way that the resulting graph would not be connected anymore.

A map is rooted if one edge is chosen and oriented. This ensures that the considered object has a trivial automorphism group. The startpoint of the root (edge) and the face on its right hand side (which is well defined since the sphere was taken oriented) are called respectively root vertex and root face. Unless explicitely mentioned, the root face is taken as infinite face when representing maps in the plane.

The dual M^* of a map M is obtained from M by putting a vertex in each face of M and an edge of M^* across each edge of M. If M is rooted, the root edge of M^* is the dual of the root edge of M, oriented in such a way that the root vertex of M^* is the dual of the root face of M. This construction is clearly involutive on unrooted maps (see Figure 1).

2.2. **Rooted loopless triangulations.** A triangulation is a planar map such that each face has degree three. We will only consider loopless triangulations, hence faces are "real"

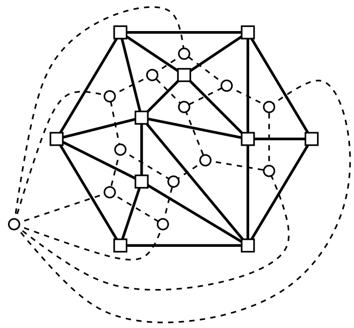


FIGURE 1. A triangulation of an hexagon and its dual

triangles, in the sense that they are 3-gons. However they are only "topological" triangles, in the sense that multiple edges are allowed so that these triangulations do not necessarily admit a representation with straight edges.

A loopless triangulation of a rooted k-gon is a planar map such that the root face is a k-gon while all other faces have degree three. A rooted triangulation of a k-gon is the same thing except that the distinguished k-gon need not be the root face. The terminology refers to the possibility, in order to draw the map in the plane, to take the k-gon as infinite face.

A loopless triangulation of a k-gon has $k+2n$ triangles for some integer $n \geqslant -1$, and hence $2k+3n$ edges and $k+n+1$ vertices (k exterior and $n+1$ interior ones). Let $\mathcal{T}_{k,n}$ be the set of rooted loopless triangulations of a k-gon into $k+2n$ triangles. Then

$$k\, T_{k,n} \;=\; 2(2k+3n)\, T^{*}_{k,n},$$

as immediately follows upon considering doubly rooted triangulations with one root on the polygon and the other anywhere: these can be regarded either as rooted loopless triangulations of a k-gon in which an edge of the k-gon is distinguished (and oriented so that the k-gon is on its right hand side), or as loopless triangulations of a rooted k-gon in which an edge is distinguished and oriented.

Hence Mullin's formula becomes

$$T_{k,n} \;=\; 2^{n+3}\frac{(2k+3n)!(2k-3)!}{k\,(n+1)!(2k+2n)!(k-2)!^2},$$

and can be rewritten as previously claimed:

$$T_{k,n} \;=\; \frac{2^{n+2}}{2k+2n}\binom{2k-2}{k}\binom{2k+3n}{n+1}.$$

This formula holds for any $k \geqslant 2$ and any $n \geqslant -1$: it specializes correctly for $k \geqslant 3$, $n=-1$, according to Formula (2); as for the degenerate case $k=2$, $n=-1$, which can only be interpreted as the case of a loop at the special vertex, it boils down to 1.

Observe that $2n\,T_n = T_{3,n-2}$: this corresponds to the fact that a map in $\mathcal{T}_{3,n-2}$ can be viewed as a rooted loopless triangulation with $2n$ triangles among which one is distinguished (the 3-gone).

2.3. **Dual family.** A cubic map is a map with all vertices of degree three, and a near-cubic map is a map with all vertices of degree three, except maybe one. Let $\mathcal{C}_n$ and $\mathcal{C}_{k,n}$ be respectively the set of non-separable cubic maps with $2n$ vertices and the set of non-separable near-cubic maps with a special vertex of degree k and $k+2n$ vertices of degree three. They are respectively the dual sets of $\mathcal{T}_n$ and $\mathcal{T}_{k,n}$.

3. The constructive census of triangulations

In this section we construct a set of simple objects counted by $T_{k,n}$ and a transformation of these objects that we claim is a bijection onto $\mathcal{T}_{k,n}$.

Terminology for trees. All the trees we are interested in are planted plane trees. In the context of planar maps, it is convenient to define a plane tree as a planar map with only one face, although this is equivalent to classical recursive definitions. Planted means that one vertex of degree one is distinguished and called the root.

We shall consider an enriched terminology for trees, with two kinds of vertices of degree one, *buds* and *leaves*, three kinds of edges, *links*, *inner edges* and *stems*, and three kinds of vertices of larger degrees, *generic*, *pathological* and *special*. Buds and leaves are always incident to stems (as opposed to links or edges) and in pictures, buds are represented by arrows. The root of a planted tree shall always be a leaf (and not a bud). This terminology reflects the very different roles played by otherwise similar items and hopefully makes things clearer once accepted...

3.1. **Planted plane trees.** The first remark is that the following binomial coefficient, taken from Formula (4),

$$A_{k,n} = \binom{2k+3n}{n+1} = \frac{1}{2k+3n+1}\binom{2k+3n+1}{1, n+1, 2k+2n-1}$$

is the number of planted plane trees with (see also Figure 4.a)

- one special vertex of degree $2k-2$,
- $n+1$ generic vertices, of degree four,
- $2k+2n$ leaves (including the root) and their $2k+2n$ stems,
- and $n+1$ inner edges connecting the generic and special vertices.

This is nothing but the classical formula for planted plane trees with given numbers of vertices of each degree ([GJ83, p113]). Let us call $\mathcal{A}_{k,n}$ the family of these trees.

Formula (4) now reads

$$T_{k,n} \;=\; \frac{2}{2k+2n}\, 2^{n+1}\binom{2k-2}{k} A_{k,n}, \tag{5}$$

and one can recognize in this formula, the appearance of the numbers of leaves, generic vertices and edges incident to the special vertex.

3.2. **Blossom trees.** Let proceed with the interpretation of the formula by considering the factor

$$B_{k,n} \;=\; 2^{n+1}\binom{2k-2}{k} A_{k,n}.$$

Since a tree A of $\mathcal{A}_{k,n}$ has $n+1$ generic vertices of degree four, the factor 2^{n+1} can be interpreted as the number of ways to select two opposite corners on each generic vertex, while the binomial factor appears as the number of ways to select $k-2$ of the $2k-2$ edges incident to the special vertex.

Given such a selection, let us apply the transformation of Figure 2.a to generic vertices and, that of Figure 2.b to the special vertex. Each generic vertex is expanded into two vertices of degree four joined by a generic link, each carrying a bud. The selected edges on the special vertex are transformed to make room for a special link and two buds attached to a pathological vertex of degree four. Observe that in these constructions buds always immediately precede links in counterclockwise direction around created vertices.

(a) Generic vertices (b) Special vertex

FIGURE 2. From trees to blossom trees

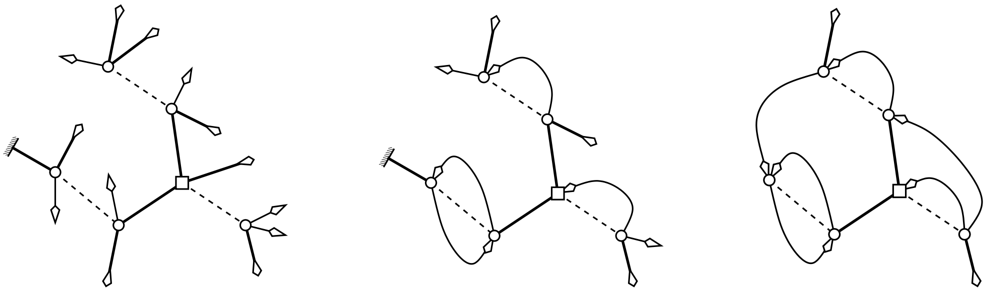


FIGURE 3. The partial closure of an unbalanced blossom tree

The set $\mathcal{B}_{k,n}$ of trees that are constructed in this manner from trees of $\mathcal{A}_{k,n}$ is of course of cardinality $B_{k,n}$. We call them *blossom trees*. By construction blossom trees are exactly the planted plane trees with (see also Figure 3 left, or Figure 4.b)

- one special vertex incident to $k-2$ special links and k edges;
- $k-2$ pathological vertices of degree four, incident to the $k-2$ previous links, each carrying two buds right before the link in counterclockwise order;
- $2n+2$ generic vertices of degree four, organized in $n+1$ pairs connected by generic links, each vertex carrying one bud right before the link in counterclockwise order;
- $2k+2n$ leaves, $2k+2n-2$ buds, and their $4k+4n-2$ stems,
- $n+1$ inner edges connecting some generic, pathological or special vertices.

Formula (4) now reads

$$T_{k,n} \;=\; \frac{2}{2k+2n}\, B_{k,n}, \tag{6}$$

making it inviting to distinguish two leaves among the $2k+2n$.

3.3. **Balanced blossom trees.** The *partial closure* of a blossom tree B consists in the following greedy procedure (see Figure 3 or Figure 4.d). Start with $\ddot{B}^{(0)} = B$, $i = 1$.

(1) Find a bud b_i and a leaf ℓ_i such that, walking from b_i to ℓ_i around the infinite face of $\ddot{B}^{(i-1)}$ in counterclockwise direction, no other bud or leaf is met.
(2) Fuse b_i, ℓ_i and their stems into an edge m_i so as to create a bounded face around the previous walk. In particular this new bounded face contains no bud or leaf.
(3) Call $\ddot{B}^{(i)}$ the resulting map and, if it still contains buds, increment i and return to Step (1).

Observe that the latter loop continues until there is no more free bud. The operation in Step (2) is called the *matching* of b and ℓ, and the resulting edge is called a *matching edge*.

The result of this partial closure is a planar map $\ddot{B} = \ddot{B}^{(2k+2n-2)}$ with $k+2n$ vertices of degree four, one special vertex of degree $2k-2$, and two remaining leaves that we call *free* in the infinite face. This map $\ddot{B}$ is independent of the exact order in which buds and leaves have been matched, (exactly like in a balanced parenthesis word, there is only a partial order of inclusion of pairs, and a greedy algorithm performing the matching has a freedom in the order it deals with incomparable pairs).

A blossom tree is called *balanced* if its root is one of the two leaves that remain free throughout partial closure. Let $\mathcal{E}_{k,n}$ be the balanced subset of $\mathcal{B}_{k,n}$. Two blossom trees are called *conjugated* if they can be obtained one from another simply by changing the root leaf. The resulting conjugacy classes of $\mathcal{B}_{k,n}$ are naturally associated with unplanted trees. Matchings between buds and leaves only depend on the conjugacy class of the blossom tree, hence we can also consider the partial closure of an unplanted tree.

Now consider a blossom tree B with root leaf r and let ℓ be one of the two free leaves of B. Taking now ℓ as root of B, a balanced blossom tree with a secondary distinguished leaf r is obtained. This yields[1]:

$$2\, B_{k,n} \;=\; (2k+2n)\, E_{k,n}$$

where $E_{k,n}$ denote the number of balanced blossom trees.

As a consequence, Formula (4) finally reads

$$T_{k,n} \;=\; E_{k,n},$$

and we are lead to seek a bijection between triangulations and balanced blossom trees.

3.4. **Case of $\mathcal{T}_n$.** A similar (but much simpler) construction provides an interpretation of Tutte's enumerative formula for the set $\mathcal{T}_n$ of loopless triangulations with $2n$ triangles, that can be rewritten in the following way:

$$T_n \;=\; \frac{2}{2n+2}\, 2^n\, \frac{1}{2n+1}\binom{3n}{n}. \tag{7}$$

The coefficient $\frac{1}{2n+1}\binom{3n}{n}$ is the number of planted plane ternary trees with n internal nodes, that is trees with n generic vertices of degree four, $n-1$ inner edges and $2n+2$ stems and leaves (including the root). The blossom trees obtained from these trees by the transformation of Figure 2.a have $2n$ generic vertices with their n links and $2n$ buds, $2n+2$ leaves, $4n+2$ stems and $n-1$ inner edges. Let $\mathcal{B}_n$ be the set of these blossom trees without special vertex. After the partial closure of any of these trees, two leaves remain unmatched, so the ratio of balanced blossom trees in $\mathcal{B}_n$ is $\frac{2}{2n+2}$. Hence the corresponding subset $\mathcal{E}_n$ has cardinality

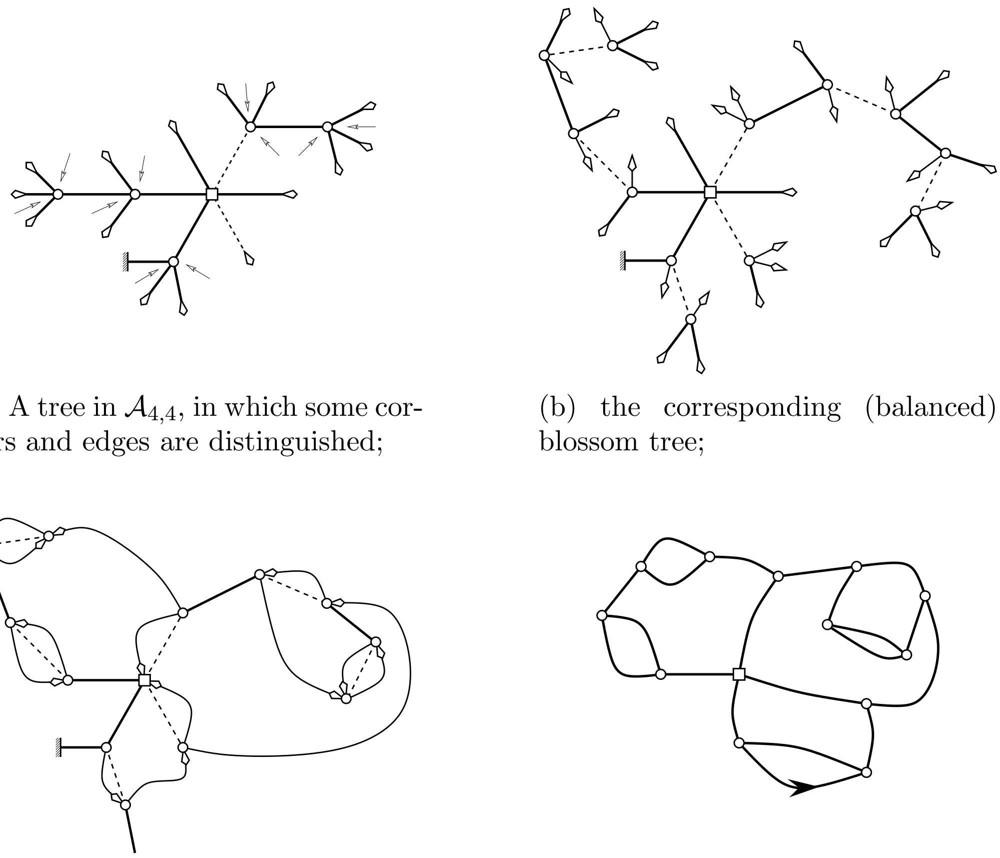
$$E_n \;=\; \frac{2}{2n+2}\,\cdot\, 2^n \,\cdot\, \frac{1}{2n+1}\binom{3n}{n} \;=\; T_n.$$

Notations. In the rest of the paper, $\mathcal{E}$ denotes the set of all balanced blossom trees (with or without special vertex), and $\mathcal{U}$ the set of all *unplanted* blossom trees. Any tree in $\mathcal{U}$ corresponds to one or two trees in $\mathcal{E}$, depending on its automorphism group.

3.5. **The complete closure.** In fact the bijection was almost already completely described. Let us define the *complete closure* φ as an application defined on the set $\mathcal{E}$. Given B a tree of $\mathcal{E}$,

(1) construct the partial closure $\ddot{B}$ of B,
(2) remove all the links and call $\overline{B}$ the result,
(3) fuse the two remaining stems of $\overline{B}$ into a root edge oriented away from the root of B,

[1]Observe that this relation is the translation for conjugacy classes of trees of the *cycle lemma* for conjugacy classes of Łukasiewicz words. This lemma, initially due to Dworetzki and Motzkin, underlies Raney's combinatorial proof of the Lagrange inversion formula [Lot97, Chap. 11]. This analogy motivates our choice of terminology.



(a) A tree in $\mathcal{A}_{4,4}$, in which some corners and edges are distinguished;

(b) the corresponding (balanced) blossom tree;

(c) its partial closure, revealing the two free leaves;

(d) and the corresponding rooted near-cubic map.

FIGURE 4. An example of complete closure

the resulting rooted planar map is $\varphi(B)$. (See Figure 4 for a complete example.)

Our main result, to be proved in the rest of the paper, is the following theorem.

Theorem 1. *The complete closure φ is a bijection from the set $\mathcal{E}_{k,n}$ (resp. $\mathcal{E}_n$) of balanced blossom trees onto the set $\mathcal{C}_{k,n}$ (resp. $\mathcal{C}_n$) of (near-)cubic maps and by duality onto the set $\mathcal{T}_{k,n}$ (resp. $\mathcal{T}_n$) of triangulations.*

The proof is twofold. First we prove that the complete closure of a tree is indeed a non-separable (near-)cubic map. Then we prove that the application is one-to-one.

4. THE CLOSURE OF A BALANCED BLOSSOM TREE IS NON-SEPARABLE

Let B be a balanced blossom tree of $\mathcal{E}_{k,n}$. In this section we prove that the complete closure $\varphi(B)$ is indeed a non-separable near-cubic map with the expected number of vertices of each kind.

The vertices of degree four of B, either generic or pathological, are incident to exactly one link. After Step (2) of the complete closure, they result in vertices of degree three. As for the special vertex, it is incident to $k-2$ links and k edges so that it yields a vertex of degree k in $\varphi(B)$. The rooted planar map $\varphi(B)$ hence contains a vertex of degree k and $k+2n$ vertices of degree three. As a consequence, $\varphi(B)$ belongs to $\mathcal{C}_{k,n}$ if and only if it is non-separable, a fact we shall now prove. Similarly, if B belongs to $\mathcal{E}_n$, $\varphi(B)$ belongs to $\mathcal{C}_n$ if and only if it is non-separable.

Observe that, since the matching of buds and leaves only depends on the conjugacy class of a blossom tree, the non-separability of the complete closure is indeed a property of the underlaying unplanted tree, and not of the balanced rooting. From now on, for the sake of convenience, we consider an unplanted blossom tree U in $\mathcal{U}$.

A preliminary observation is that any separating vertex of degree three is incident to a separating edge. It is thus sufficient to prove on the one hand that $\varphi(U)$ has no separating edge (Section 4.1 to 4.4) and on the other hand that the possible special vertex is not separating (Section 4.5).

4.1. **A preliminary lemma on the structure of blossom trees.** Consider U a blossom tree of $\mathcal{U}$ and $e = (v_1, v_2)$ an inner edge or a link of A. The *decomposition* of U at e consists in cutting e in its middle, so as to create two new leaves ℓ_1 and ℓ_2, attached by two stems e_1 and e_2 respectively to v_1 and v_2. As a result, the tree U yields two subtrees $U_1(e)$ and $U_2(e)$, respectively containing v_1 and v_2. A leaf ℓ of U is said *incoming* with respect to e if, in the partial closure of U, it is free or matched to a bud b that does not belong to the same subtree as ℓ (with respect to e). By extension, the matching edge (b, ℓ) is also called incoming with respect to e.

The following lemma is immediate upon counting leaves and buds in each subtree and considering the cyclic orders around v_1 and v_2.

Lemma 1. *Let U be a blossom tree and e an inner edge or a link of U.*

- *If e is an inner edge, then $U_1(e)$ and $U_2(e)$ are well formed blossom trees, with two more leaves than buds (including ℓ_1 and ℓ_2), and thus at least one incoming leaf each.*
- *If e is a link between two generic vertices, then $U_1(e)$ and $U_2(e)$ contain two more leaves than buds (including ℓ_1 and ℓ_2). Moreover in the partial closure of U, the bud adjacent to v_1 in $U_1(e)$ is matched with an incoming leaf of $U_2(e)$.*
- *If e is a link incident to the special vertex (assumed in $U_1(e)$), then $U_1(e)$ has four more leaves than buds and $U_2(e)$ has as many buds as leaves (including ℓ_1 and ℓ_2). As a consequence, in the closure of U, the two buds adjacent to v_2 in $U_2(e)$ are matched to two incoming leaves of $U_1(e)$, and $U_2(e)$ has at least one incoming leaf.*

4.2. **The incremental complete closure.** Let us now consider an application of the (greedy) partial closure procedure of Section 3.3 to U, resulting into the map $\ddot{U}$ through the sequences b_i, ℓ_i, m_i and $\ddot{U}^{(i)}$, for $i \geqslant 1$. Given a matching edge m, obtained from (b, ℓ), we define $e(m)$ to be the unique link incident to the vertex adjacent to b. By construction, for each link e of A there are exactly two indices $j < i$ such that $e(m_i) = e(m_j) = e$. Let us call these indices, the *dates* of e. Finally define a planar map $\underline{U}^{(i)}$ by deleting from $\ddot{U}^{(i)}$ all *generic* links with largest date less or equal to i. In other terms, $\underline{U}^{(i)}$is constructed from $\underline{U}^{(i-1)}$ by adding m_i and removing $e(m_i)$ if it is generic and the other matching edge m_j such that $e(m_j) = e(m_i)$ satisfies $j < i$.

Let finally $\underline{U} = \underline{U}^{(2k+2n-1)}$ be the resulting map. The following technical lemma precisely describes the evolution of connectedness in $\underline{U}^{(i)}$ for $i = 1, \ldots, 2k + 2n - 2$.

Lemma 2. *For all i the planar map $\underline{U}^{(i)}$ is connected. Moreover for any link or inner edge e of U, the graphs induced in $\underline{U}^{(i)}$ respectively by the vertices of $U_1(e)$ and by those of $U_2(e)$ are connected.*

Proof. The lemma is obviously true for the tree $\underline{U}^{(0)} = U$. Assume now the lemma true until $i - 1$ and consider the construction of $\underline{U}^{(i)}$ from $\underline{U}^{(i-1)}$. Let $e' = e(m_i)$ and j be the other date of e'.

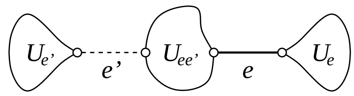


FIGURE 5. Simultaneous decomposition of U at e and e'

There is a deletion only if e' is a generic link and $j < i$. In this case, observe first that, according to Lemma 1, the matching edge m_i connects a vertex of $U_1(e')$ to a vertex of $U_2(e')$ so that by induction hypothesis $\underline{U}^{(i)}$ remains connected upon deleting e'.

Then consider another link or inner edge e and the decompositions of U at e and e': performing both decompositions yields three subtrees, U_e, $U_{ee'}$ and $U_{e'}$ where the indices refer to incidences with e and e' (see also Figure 5). In $\underline{U}^{(i-1)}$, the graphs induced respectively by U_e, $U_{e'}$, and $U_{e'} \cup U_{ee'}$ are connected by induction hypothesis. The deletion of e' does not touch the graph induced by U_e so that we only have to deal with the graph induced by $U_{e'} \cup U_{ee'}$. Since e' is generic, one of m_i or m_j is incident to the endpoint of e' in $U_{ee'}$ and has its other endpoint (the leaf) in $U_{e'}$. Since $U_{e'}$ is connected, the deletion of e' does not disconnect $U_{e'} \cup U_{ee'}$.

□

4.3. **Separating edges and generic links.**

Lemma 3. *The only separating edges in $\underline{U}$ are inner edges of U that separate the two free leaves.*

Proof. Consider a matching edge m, and let $e = e(m)$ and m' be the second matching edge with $e(m') = e$. Then Lemma 1 asserts that m and m' are incoming with respect to e. In view of Lemma 2, their respective endpoints on both sides of e can be connected to construct a cycle containing m and m'. Moreover, if e is special, the same argument provides a cycle through e and m. Hence neither matching edges nor links can be separating.

Let now e be an inner edge of A that remains separating in $\underline{U}$. Consider the decomposition of U at e. No matching edge connects a vertex of $U_1(e)$ to a vertex of $U_2(e)$, for e would not be separating (Lemma 2). In view of Lemma 1, this implies that there is one free leaf in both subtrees. □

4.4. **Matching edges are not separating in $\varphi(U)$.**

Lemma 4. *The only separating edges in $\overline{U}$ are inner edges of U that separate the two free leaves.*

Proof. If U has no special vertex, $\overline{U} = \underline{U}$, hence this lemma is equivalent to Lemma 3.

Now suppose that U has a special vertex with degree $2k-2$. In order to show that the removal of the special links from $\underline{U}$ does not make any matching edge separating, it is sufficient to prove that any two faces that are merged by removing some special links have no common matching edge.

Let us consider the $2k-2$ subtrees of U at the special vertex v, more precisely defined as the subtrees not containing v in the decomposition of U at any edge or link incident to v. We call such a subtree generic or pathological depending on whether it is attached to the special vertex by an inner edge or by a link. At any step i of the construction of Section 4.2, Lemma 2 ensures that these subtrees induce connected subgraphs of $\underline{U}^{(i)}$, and that, according to Lemma 1, any of them has at least one incoming leaf.

Given a particular ordering of the matchings in the application of the partial closure procedure, let us consider the time j at which, for the first time, an incoming edge of a subtree at v is matched. This is also the first time that a matching edge is created between

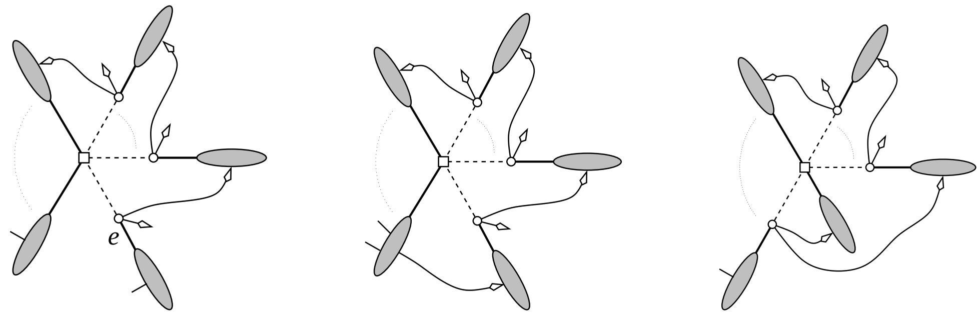


FIGURE 6. Case study for the deletion of special links.

two subtrees. Let us now consider an ordering such that j is as large as possible. In this case, at time j, all the matchings that are internal to each subtree have been performed. More precisely, with the notation of Section 4.2, the ordering is such that, for any $i < j$, b_i and ℓ_i belong to the same generic or pathological subtree at v, and, for any $i \geqslant j$, b_i and ℓ_i belong to different subtrees at v or make a complete turn around v.

Perform then the construction until Step $j-1$. As already observed, at that moment the subtrees are two by two independent. Moreover, at each pathological vertex, the bud that precedes the link in counterclockwise order is in position to be matched with the first incoming leaf of the next subtree (in counterclockwise order around the special vertex). Every such matching creates a bounded face, which cannot be affected by any further step since it does not contain generic links (Figure 6, left).

Once these $k-2$ matchings are performed, only two kinds of buds can be matched in such a way that the created bounded face contains a special link: the first unmatched bud of a generic subtree that is to be matched into a pathological one (Figure 6, middle), or the second bud of a pathological vertex that precedes a sequence of subtrees with no more unmatched leaf or bud (Figure 6, right). These matchings also create faces that will not be affected by any further step.

These different cases lead to three different ways for a sequence of faces to be merged into one face by the removal of special links. In each case, we shall argue that, as a whole, these faces do not complete a turn around v.

- In the first case (Figure 6 left), a (non-empty) sequence of bounded faces merge with the infinite face of $\underline{U}$. No bounded faces can thus perform a complete turn around v. Hence two non successive faces in the sequence share no edge, and two successive faces share a special link. In any case they do not have a matching or inner edge in common. As for the infinite face, in view of the disposition of buds, it may only be incident twice to the inner edge marked e in the figure. In this case the shaded subtree below e contains exactly one of the two free leaves so that e separates the two free leaves.
- The second case involves two generic subtrees and a non-empty sequence of pathological ones. Since $k \geqslant 2$, the two generic subtrees are distinct, and there is no complete turn.
- In the third case, there exists a pathological vertex p_1 such that its second bud is matched with a leaf that belongs to a pathological subtree attached on a pathological vertex p_2. This implies that the sequence S of subtrees that follow p_1 and precede p_2 around the special vertex has only one free leaf. In other words, this sequence contains exactly one more generic subtree than pathological ones. Hence the number of involved pathological subtrees is at least the number of involved generic subtrees

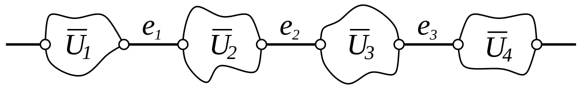


FIGURE 7. Separating inner edges are organised in a sequence in $\overline{U}$.

minus one, and at least one generic subtree at v is not involved. Finally p_1 is different from any involved pathological vertex that follows S and the complete turn is not performed.

We conclude that matching edges are not separating edges and that all separating inner edges still separate the two leaves. Since all links have been removed, the lemma is proved. □

This lemma proves that $\overline{U}$ can be described as an alternating sequence $\overline{U}_1, e_1, \overline{U}_2, \ldots, e_{p-1}, \overline{U}_p$ of submaps $\overline{U}_i$ and edges e_i (with p possibly equal to one), such that $\overline{U}_1$ and $\overline{U}_p$ carry one free leaf each, and no $\overline{U}_i$ contains a separating edge (Figure 7). As a consequence, $\varphi(U)$ has no separating edge.

4.5. **The special vertex.** Suppose that U has a special vertex. The following lemma concludes the proof that $\varphi(U)$ is non-separable.

Lemma 5. *The special vertex v is not a separating vertex of $\varphi(U)$.*

Proof. Assume that the special vertex v is separating in $\overline{U}$ (as given by Step (2) of complete closure) and consider a decomposition of $\overline{U}$ into two components $\overline{U}_1$ and $\overline{U}_2$ connected only at v. This decomposition induces a decomposition of $\ddot{U}$: special links connect v to a vertex of $\overline{U}_1$ or $\overline{U}_2$ and do not interfere; once special links are replaced, generic links appear inside bounded faces and hence inside the two components.

In turn the decomposition of $\ddot{U}$ at v induces a decomposition of the tree U into two sequences of subtrees rooted at v such that there is no matching edge from one to the other. Since Lemma 1 provides in particular an incoming leaf on the first tree of both these sequences, these leaves must be the two free leaves of U.

Returning to $\overline{U}$, we conclude that $\overline{U}_1$ and $\overline{U}_2$ each contain one free leaf. Hence v is not a cut vertex anymore after Step (3) of the complete closure. □

5. The inverse construction

In this section we define by induction on the number of edges a construction which is inverse to the complete closure.

Let us first consider the minimal cases of non-separable (near-)cubic maps with at most two vertices. The case $k = 2$, $n = -1$ is the degenerate case of the loop at a special vertex and corresponds to the tree with one special vertex of degree two. The case $n = 1$, without special vertex, is the case of a bundle of three edges between two vertices and corresponds to the unique balanced blossom tree with two generic vertices. The case $k = 3$, $n = -1$ is the case of a bundle of three edges between two vertices, one of them being special; the two different rootings of this map correspond to the two balanced rootings of the unique blossom tree with a special vertex of degree three and one single pathological vertex.

Now suppose that C is a rooted non-separable (near-)cubic map with at least three vertices among which, possibly, a special vertex of any degree, and the others of degree three. Let the root edge be oriented from a vertex v_1 to a vertex v_2, and define $\tilde{C}$ by cutting the root edge into two stems with leaves f_1 and f_2. If there exists $B \in \mathcal{E}$ such that $C = \varphi(B)$, then B is necessarily planted on leaf f_1, and reconstructing B is equivalent

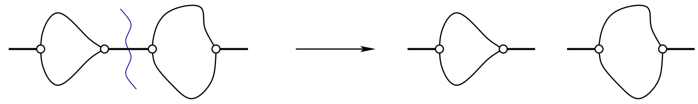


Figure 8. Induction with separating edge.

to recovering links between vertices: these links determine which vertices are generic or pathological, and which stems carry leaves or buds.

Given a map $\tilde{C}$, our strategy is to exhibit one (or more) link that exists *necessarily* in any tree U in $\mathcal{U}$ such that $\overline{U} = \tilde{C}$. Moreover we show that the decomposition of U at this link yields subtrees whose images by φ can be uniquely characterized as some strict submaps of $\tilde{C}$. The induction applied to these submaps allows to prove that φ is one-to-one.

The construction depends on whether $\tilde{C}$ is separable.

5.1. **The map $\tilde{C}$ contains a separating vertex other than v_1 or v_2.** Since C is non-separable, the map $\tilde{C}$ is organized as a chain of non-separable components between v_1 and v_2. In this section, *separating vertices* refer to separating vertices distinct from v_1 and v_2. Two cases are distinguished.

– First case: the map $\tilde{C}$ has a separating vertex v that is not the special vertex (see Figure 8). In this case v has degree three and, as already argued, there is a separating edge e. In view of the discussion of the previous section, if there is a tree U in $\mathcal{U}$ such that $\overline{U} = \tilde{C}$ then e is an inner edge of U, and in the decomposition of U at e, the leaves f_1, ℓ_1 and f_2, ℓ_2 are the free leaves of $U_1(e)$ and $U_2(e)$, (so that their partial closure are independent).

Now there is a unique way to recover such a structure. First cut e in $\tilde{C}$ into two stems e_1 and e_2 with leaves ℓ_1 and ℓ_2. The resulting two components of $\tilde{C}$ allow to recover U_1 and U_2 by induction hypothesis and the unique tree U is obtained by fusing back e_1 and e_2 between U_1 and U_2.

– Second case: the special vertex v is the only separating vertex of $\tilde{C}$ (Figure 9). Let C_1 and C_2 be the two non-separable components of $\tilde{C}$ at v. As was already analyzed in Section 4.5, if there is a tree U such that $\overline{U} = \tilde{C}$, then the links or edges incident to v in U are arranged in counterclockwise order into two successive sequences $e_1, \ldots, e_p$ with endpoints in C_1 and $e'_1, \ldots, e'_q$ with endpoints in C_2, with p and q greater or equal to two in order to avoid separating edges.

Let us prove that the subtree S of U attached to e_1 (resp. to e'_1) is reduced to a special link carrying a pathological vertex. By construction of the two sequences, the incoming leaves of S are free, so that there is at most one such leaf. In view of Lemma 1, there is exactly one. Now if e_1 is an inner edge, Lemma 1 implies that there is no matching edge leaving this subtree, and e_1 is a separating edge of $\tilde{C}$. Therefore e_1 is a special link, which by definition carries a pathological vertex. Finally the subtree cannot be bigger otherwise the pathological vertex would carry an edge and the latter would be separating in $\tilde{C}$.

Hence the tree U is decomposed at v into a special link e_1 that carries a pathological vertex, followed by the tree U_1 formed of $e_2, \ldots, e_p$ and their subtrees, by a special link e'_1 that carries a pathological vertex, and by the tree U_2 formed of $e'_2, \ldots, e'_p$ and their subtrees. Moreover U_1 and U_2 are well formed blossom trees whose free leaves are respectively matched by the buds of the two pathological vertices.

Now there is a unique way to recover such a structure. First v_1 and v_2 are identified as pathological vertices (since they carry the free leaves). Then, deleting v_1 and v_2 from their respective non-separable component yields two maps C_1 and C_2 from which U_1 and U_2 can

FIGURE 9. Induction with separating special vertex.

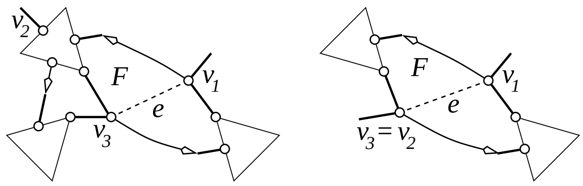


FIGURE 10. Analysis of the possible configurations in the main case of recursion.

be recovered by induction hypothesis. The unique tree U is obtained upon recreating the cyclic order around v.

5.2. **The map $\tilde{C}$ has no other separating vertex than v_1 and v_2.** A first easy case is when the special vertex v carries one of the two free leaves of $\tilde{C}$, say f_1 so that $v = v_1$. Then the analysis is exactly the same as the analysis of the case where v is the only separating vertex (second case of the previous section), with the second sequence reduced to a single leaf: v_2 is found to be pathological and upon deleting v_2 and f_1 the induction hypothesis applies to provide a unique reconstruction.

The main case is when the special vertex v is neither v_1 nor v_2. Assume, without loss of generality, that v is not on the counterclockwise path around the infinite face from v_1 to v_2. (Even if v is incident to the infinite face, it cannot appear in both path from v_1 to v_2 and back from v_2 to v_1.) Let us discuss the constraints on a tree U such that $\overline{U} = \tilde{C}$.

Observe first that v_1 cannot be a pathological vertex: even if the special vertex v is incident to the infinite face, this yields a contradiction in the way its buds are to be matched. Hence v_1 is a generic vertex of U. Let F be the bounded face incident to v_1 in $\tilde{C}$. A generic link e joins v_1 to another vertex v_3 in this face F. Let us consider the subtrees attached to v_1 and v_3 in F (Figure 10).

Since vertex v_1 is adjacent to a free leaf f_1 and a bud, it carries a unique (possibly empty) subtree, which precedes f_1 in counterclockwise order. Call this subtree S_1. According to Lemma 1 and in view of the free leaf f_1, the subtree S_1 has only one incoming leaf. The latter is therefore matched by the bud of v_3 and this matching edge is incident to both F and the infinite face. Moreover there is no other edge incident to F on the path from v_2 to v_3 along the infinite face in counterclockwise direction. Indeed this could only be an inner edge e' (for A to be connected) but then the subtree at e' containing v_2, having two incoming leaf (f_2 and the leaf matched by the bud of v_1) would have, according to Lemma 1 a bud taking part of an incoming matching edge. However this would prevent e' to be on the infinite face.

Consider next the decomposition of U at e and take $U_1(e)$ to contain v_1. In view of its previous definition, the tree S_1 is obtained from $U_1(e)$ upon deleting v_1 and it is a balanced blossom tree. On the other hand, define a tree S_2 from $U_2(e)$ as follows. First delete the bud and the stem inherited from e that are incident to v_3, so that the latter vertex has degree two. Then smooth this vertex out so as to fuse its two incident edges into one single edge e' (which may be a stem). The result is a tree $U_2(e)$ whose closure leaves e' in the infinite face, and whose free leaves are f_2 and the leaf ℓ_1 matched with the bud of v_1 in U.

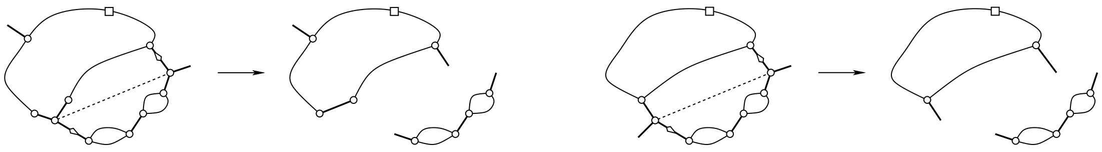


FIGURE 11. Recursive decompositions in the two cases of Figure 10.

Finally there is a unique way to recover the structure (Figure 11). First, taking F to be the bounded face incident to v_1, we dispose of a characterization of vertex v_3 as the first vertex incident to F on the path from v_2 to v_1 around the infinite face in counterclockwise direction. In particular if v_2 is incident to F then $v_3 = v_2$ (as illustrated on the right hand side of Figure 10). Second, the complete closure of the trees S_1 and S_2 are uniquely obtained as follows. Delete f_1 and its stem and cut v_1 so as to create two new leaves ℓ_1 (for the bud of v_1) and ℓ_2 (for the subtree). Detach the edge that follows v_3 along the infinite face from v_2 to v_1: this edge is also incident to F and this operation creates a leaf ℓ_3 in the same component as ℓ_2. Call this component $\tilde{C}_1$. The vertex v_3 remains of degree two and can be smoothed so as to fuse its two incident edges into one single edge e that belongs to a second component, $\tilde{C}_2$, that also contains ℓ_1 and v_2. In view of the previous analysis, the two maps $\tilde{C}_1$ and $\tilde{C}_2$ are the images of S_1 and S_2 by complete closure (upon opening the roots). By induction hypothesis, there exists exactly one couple of such trees. From S_1 and S_2 the tree U is readily recovered and the proof is complete.

REFERENCES

[BMS00] M. Bousquet-Mélou and G. Schaeffer. Enumeration of planar constellations. *Adv. in Appl. Math.*, 24(4):337–368, 2000.

[GJ83] I. P. Goulden and D. M. Jackson. *Combinatorial enumeration.* John Wiley & Sons Inc., New York, 1983.

[Lot97] M. Lothaire. *Combinatorics on words.* Cambridge University Press, Cambridge, 1997.

[Mul65] R. C. Mullin. On counting rooted triangular maps. *Canad. J. Math.*, 17:373–382, 1965.

[Sch98] G. Schaeffer. *Conjugaison d'arbres et cartes combinatoires aléatoires.* PhD thesis, U. Bordeaux I, 1998.

[Tut62a] W. T. Tutte. A census of planar triangulations. *Canad. J. Math.*, 14:21–38, 1962.

[Tut62b] W. T. Tutte. A census of slicings. *Canad. J. Math.*, 14:708–722, 1962.

[Tut63] W. T. Tutte. A census of planar maps. *Canad. J. Math.*, 15:249–271, 1963.

DOMINIQUE POULALHON, LIX, ÉCOLE POLYTECHNIQUE, PALAISEAU
E-mail address: poulalho@lix.polytechnique.fr

GILLES SCHAEFFER, CNRS – LORIA, NANCY
E-mail address: Gilles.Schaeffer@loria.fr
URL: http://www.loria.fr/~schaeffe

ON THE ASYMPTOTIC ANALYSIS OF A CLASS OF LINEAR RECURRENCES

THOMAS PRELLBERG

Abstract. Many problems in enumerative combinatorics can be expressed via linear recurrences whose generating functions satisfy a functional equation. I present the asymptotic analysis for a class of recurrences which lead to a functional equation involving a transformation with a parabolic fixed point. The method used relies heavily on analytic iteration theory. Examples given are Bell numbers, Partition lattice enumerations, and the Takeuchi recursion.

Résumé. Beaucoup de problémes en combinatorique eńumérative peuvent être exprimés par des récurrences linéaires dont les fonctions generatrices obeissent des équations fonctionelles. Je vais présenter dans cette note l'analyse asymptotique pour une classe de récurrences qui emmènent à une équation fonctionelle contenant une transformation avec point fixe parabolique. La méthode utilisée est basée sur la théorie de l'itération analytique. Les exemples discutés sont les nombres de Bell, l'énumérations de réseaux partition et la récurrence de Takeuchi.

Institut für Theoretische Physik, Technische Universität Clausthal, Arnold-Sommerfeld-Strasse, D-38678 Clausthal-Zellerfeld, Germany
E-mail address: thomas.prellberg@tu-clausthal.de

PERMUTATIONS BY NUMBERS OF ANTI-EXCEDANCES AND FIXED POINTS

FANJA RAKOTONDRAJAO

Abstract. In this paper we study the bivariated distribution of anti-excedances and fixed points on the symmetric group and give its generating function. We also study the same distribution on the alternating group and its complement, and prove combinatorially a relation between the number of odd permutations and the number of even permutations having a given number of anti-excedances and a given set of fixed points.

Résumé. Nous étudions dans ce papier la distribution de la statistique bivariée des anti-excédances et points fixes sur le groupe symétrique et nous donnons la fonction la fonction génératrice de cette distribution. Nous étudions également cette distribution sur le groupe alterné et son complémentaire, et démontrons combinatoirement une relation entre le nombre de permutations impaires et le nombre de permutations paires ayant un nombre fixé d'anti-excédances et un ensemble donné de points fixes.

1. Introduction

The distribution of fixed point distribution and the so called "Eulerian statistics" over the symmetric group have been studied in depth but always separately ([7], [2], [1]). In this paper we study the bivariated distribution of the Eulerian statistic *anti-excedances* and of the statistic of fixed points on the symmetric group. Recently, the author and Roberto Mantaci [5] have introduced and studied the numbers $a_{n,k}$ that count the derangements over n objects having k anti-excedances, this paper is a generalization of those results. We will study in this paper the numbers $a_{n,k,m}$ that count the permutations over n objects having k anti-excedances and m fixed points. We will give an exponential generating function for these numbers as well as a recursive relation defining them. We will also study the distribution of the bivariate distribution of anti-excedances and fixed points on the alternating group and its complement and prove that the number of odd permutations and the number of even permutations having k anti-excedances and having the same given set F of fixed points differ by 1 for all integers n and k and for all subset F of $[n]$. We will give a combinatorial proof of this result.

Let us denote by $[n]$ the interval $\{1, 2, \cdots, n\}$, by σ a permutation of the symmetric group $\mathcal{S}_n$ and by $\mathcal{A}_n$ the alternating group of rank n, that is, the group of all even permutations.

Definition 1.1. *We will say that $i \in [n]$ is a fixed point for σ if $\sigma(i) = i$.*

Definition 1.2. *We will say that σ presents an anti-excedance (resp. an excedance) in $i \in [n]$ if $\sigma(i) \leq i$ (resp. $\sigma(i) > i$). In this case we will say that i is an anti-exceedant (resp. an exceedant).*

We will denote by $Fix(\sigma)$ the set of the fixed points of σ, by $FIX(\sigma)$ the integer $|Fix(\sigma)|$ and by $AX(\sigma)$ the number of the anti-excedances of σ. We will denote by $\mathcal{F}_{n,m}$ the set of permutations over n objects having m fixed points and by $f_{n,m}$ the cardinality of this set, by $\mathcal{S}_{n,k}$ the set of permutations over n objects having k anti-excedances and by $A_{n,k}$ the cardinality of this set. The numbers $A_{n,k}$ are the classical Eulerian numbers (see [1], [7]). The first values of the numbers $A_{n,k}$ are given in the following table:

$A_{n,k}$						
	$k=0$	1	2	3	4	5
$n=0$	1					
1	0	1				
2	0	1	1			
3	0	1	4	1		
4	0	1	11	11	1	
5	0	1	26	66	26	1

These numbers satisfy the following recursive relation:

$$A_{n,k} = (n-k+1)A_{n-1,k-1} + kA_{n-1,k} \qquad (1 \le k \le n).$$

The Eulerian polynomials $A_n(t) = \sum_{\sigma \in \mathcal{S}_n} t^{AX(\sigma)}$ have the following generating function (see [1]):

$$A(t,u) = \sum_{n \ge 0} A_n(t) \frac{u^n}{n!} = \frac{(1-t)}{1 - t\exp((1-t)u)}.$$

The first values of the numbers $f_{n,m}$ are given in the following table:

$f_{n,m}$						
	m=0	1	2	3	4	5
$n=0$	1					
1	0	1				
2	1	0	1			
3	2	3	0	1		
4	9	8	6	0	1	
5	44	45	20	10	0	1

The numbers $f_{n,m}$ satisfy the following recursive relation (see [6])

$$f_{n,m} = f_{n-1,m-1} + (m+1)f_{n-1,m+1} + (n-1-m)f_{n-1,m} \qquad (0 \le m \le n).$$

The generating function of the fixed point distribution on the symmetric group is given by [7]:

$$F(x,u) = \sum_{n \ge 0} \sum_{\sigma \in \mathcal{S}_n} x^{FIX(\sigma)} \frac{u^n}{n!} = \frac{\exp((x-1)u)}{1-u}.$$

2. The anti-excedance and fixed point distribution on the symmetric group

We will give a recursive relation for the distribution of the bivariate statistic (AX, FIX) on the symmetric group $\mathcal{S}_n$, and will give its generating function. We will denote by $\mathcal{F}_{n,k,m}$ the set of permutations in $\mathcal{S}_{n,k}$ having m fixed points (or equivalently, the set of permutations in $\mathcal{F}_{n,m}$ having k anti-excedances) and denote $a_{n,k,m}$ the cardinality of $\mathcal{F}_{n,k,m}$. The first values of these numbers are given in the following tables for some fixed integers

m:

$m = 0$				
	k=1	2	3	4
$n = 1$	0			
2	1			
3	1	1		
4	1	7	1	
5	1	21	21	1

$m = 1$					
	k=3	4	5	6	7
$n = 1$	1				
2	0	0			
3	0	3	0		
4	0	4	4	0	
5	0	5	35	5	0

$m = 2$					
	k=2	3	4	5	
$n = 2$	1				
3	0	0			
4	0	6	0		
5	0	10	10	0	
6	0	15	105	15	0

$m = 3$					
	k=3	4	5	6	7
$n = 3$	1				
4	0	0			
5	0	10	0		
6	0	20	20	0	
7	0	35	245	35	0

Proposition 2.1. *For all integer n, one has $a_{n,n,n} = 1$.*

Proof. The identity is the only permutation of the symmetric group having n anti-excedances and n fixed points. □

Proposition 2.2. *For a fixed integer m and a fixed integer n, the numbers $a_{n,k,m}$ are symmetric, in the sense that $a_{n,k,m} = a_{n,n-k+m,m}$ for all integers $n \geq 1$ and $m \leq k \leq n$.*

Proof. The bijective map $\sigma \mapsto \sigma^{-1}$ associates a permutation σ in the set $\mathcal{F}_{n,k,m}$ with a permutation in the set $\mathcal{F}_{n,n-(k-m),m}$. □

Remark 2.3. *We have $a_{n,k,m} = 0$ if $k < 0$ or $m > k$.*

Theorem 2.4. *For all integers n, k and m such that $m \geq 0$ and $m \leq k \leq n$, one has:*

$$a_{n,k,m} = a_{n-1,k-1,m-1} + (m+1)a_{n-1,k,m+1} + (n-k)a_{n-1,k-1,m} + (k-m)a_{n-1,k,m}$$

with the initial value $a_{0,0,0} = 1$.

Proof. Notice that all permutation σ' in the symmetric group $\mathcal{S}_n$ is obtained from a permutation σ in $\mathcal{S}_{n-1}$ by multiplying σ on the left by a transposition (i, n) for an integer $i \in [n]$ and we suppose that the integer n is an exceedant for σ. Notice also that if the integer i is an exceedant (resp. an anti-exceedant) for σ, then when we multiply σ by the transposition (i, n) on the left, we create (resp. do not create) a new exceedant for σ'. When it is created, this new exceedant is the integer n itself. Now let us look for the various cases for the integer i:

(1) If $i = n$, then the transposition (i, n) is not a transposition but the 1- cycle (n) and the permutation $\sigma' = (n)\sigma$ has a new anti-exceedant fixed point, which is the integer n itself.
(2) If the integer i is an exceedant for σ, then the permutation $\sigma' = (i, n)\sigma$ has a new anti-excedant, which is the integer n itself, but does not have any new fixed point.
(3) If the integer i is a fixed point of the permutation σ, then the permutation $\sigma' = (i, n)\sigma$ does not have a new anti-exceedant but it has one fewer fixed points than the permutation σ.
(4) If the integer i is an anti-exceedant non fixed point of the permutation σ, then the permutation $\sigma' = (i, n)\sigma$ has neither a new anti-exceedant nor a new fixed point.

It follows straightforwardly that we obtain all the permutations in the set $\mathcal{F}_{n,m}$ having k anti-excedances by considering all the permutation σ indicated in the following four cases and by multiplying them by the appropriate transposition :

(1) $\sigma \in \mathcal{F}_{n-1,m-1}$ having $k-1$ anti-excedances and the only possibility for the choice of "transposition" by which multiply σ is the 1-cycle (n).
(2) $\sigma \in \mathcal{F}_{n-1,m+1}$ having k anti-excedances and there exist $m+1$ possibilities for the choice of the transposition : the transpositions (i,n) where the integer i is a fixed point of the permutation σ.
(3) $\sigma \in \mathcal{F}_{n-1,m}$ having $k-1$ anti-excedances and there exist $n-1-(k-1)$ possibilities for the choice of the transposition : the transpositions (i,n) where the integer i is an exceedant of the permutation σ.
(4) $\sigma \in \mathcal{F}_{n-1,m}$ having k anti-excedances and there exist $k-m$ possibilities for the choice of the transposition : the transpositions (i,n) where the integer i is an anti-exceedant non fixed point of the permutation σ.

□

2.1. **Generating function.** Let us denote by $A(t,x,u)$ the exponential generating function of the distribution of the bivariate statistic $(AX,\ FIX)$ on the symmetric groups, that is, the function $A(t,x,u)$ defined by :

$$A(t,x,u) = \sum_{n\geq 0}\sum_{\sigma\in\mathcal{S}_n} t^{AX(\sigma)}x^{FIX(\sigma)}\frac{u^n}{n!}.$$

Proposition 2.5. *The generating function $A(t,x,u)$ of the distribution of the bivariate statistic $(AX,\ FIX)$ on the symmetric groups satisfies the following differential equation :*

$$xtA = (1-tu)\frac{\partial A}{\partial u} + t(t-1)\frac{\partial A}{\partial t} + (x-1)\frac{\partial A}{\partial x}$$

with the initial conditions :

$$A(t,1,u) = \frac{1-t}{1-t\exp\left((1-t)u\right)}$$

and $A(1,1,0)=0$.

Proof. The differential equation can be easily derived from the recurrence relation given in Theorem 2.4. The initial condition :

$$A(t,1,u) = \frac{1-t}{1-t\exp\left((1-t)u\right)}$$

is due to the fact that when we set $x=1$ the resulting formal series is the well-known generating function of the Eulerian polynomials and when we set $x=1, t=1$ and $u=0$ we obtain the first value of the numbers $a_{n,k,m}$. □

Theorem 2.6. *The generating function*

$$A(t,x,u) = \sum_{n\geq 0}\sum_{\sigma\in\mathcal{S}_n} t^{AX(\sigma)}x^{FIX(\sigma)}\frac{u^n}{n!}$$

of the distribution of the bivariate statistic $(AX,\ FIX)$ on the symmetric group has the following closed form

$$A(t,x,u) = \frac{(1-t)\exp\left((x-1)tu\right)}{1-t\exp\left((1-t)u\right)}.$$

Proof. The function $\dfrac{(1-t)\exp((x-1)tu)}{1-t\exp((1-t)u)}$ satisfies the differential equation given in the previous proposition, as well as the initial conditions. □

3. Anti-excendances, fixed points and parity

R. Mantaci introduced in [3] the numbers $P_{n,k}$ and $D_{n,k}$ that count respectively the cardinality of the set $\mathcal{A}_{n,k}$ of even permutations having k anti-excedances, and the cardinality of the set $\mathcal{S}_{n,k} \setminus \mathcal{A}_{n,k}$. These numbers satisfy the following relations:

$$P_{n,k} = P_{n-1,k-1} + kD_{n-1,k} + (n-k)D_{n-1,k-1}$$
$$D_{n,k} = D_{n-1,k-1} + kP_{n-1,k} + (n-k)P_{n-1,k-1}$$

for all integers n and $1 \leq k \leq n$ with $P_{0,0} = 1$ and $D_{0,0} = 0$.
Let us denote by:

- $p_{n,m}$ the cardinality of the set of even permutations in $\mathcal{F}_{n,m}$,
- $i_{n,m}$ the cardinality of the set of odd permutations in $\mathcal{F}_{n,m}$.

The following tables report the first values of the numbers $p_{n,m}$ and $i_{n,m}$:

$p_{n,m}$						
	$m=0$	1	2	3	4	5
$n=0$	1					
1	0	1				
2	0	0	1			
3	2	0	0	1		
4	3	8	0	0	1	
5	24	15	20	0	0	1

$i_{n,m}$						
	$m=0$	1	2	3	4	5
$n=0$	0					
1	0	0				
2	1	0	0			
3	0	3	0	0		
4	6	0	6	0	0	
5	20	30	0	10	0	0

Proposition 3.1. *The numbers $p_{n,m}$ and $i_{n,m}$ satisfy the following relations:*

$$p_{n,m} = p_{n-1,m-1} + (m+1)i_{n-1,m+1} + (n-m-1)i_{n-1,m}$$
$$i_{n,m} = i_{n-1,m-1} + (m+1)p_{n-1,m+1} + (n-m-1)p_{n-1,m}$$

for all integers n and $0 \leq m \leq n$ with $p_{0,0} = 1$ and $i_{0,0} = 0$.

Proof. We will use the same idea as in Theorem 2.4. Suppose $i \neq n$. If σ is an even (resp. odd) permutation of $\mathcal{S}_{n-1}$, by multiplying $\sigma \in \mathcal{S}_{n-1}$ by the transposition (i, n) we obtain an odd (resp. even) permutation σ' in $\mathcal{S}_n$. When $i = n$, we obtain a permutation having the same parity. □

We denote by

- $P_{n,k,m}$ the cardinality of $\mathcal{A}_{n,k,m}$, the set of even permutations having k anti-excedances and m fixed points,
- $D_{n,k,m}$ the cardinality of $(\mathcal{S}_{n,k} \bigcap \mathcal{F}_{n,m}) \setminus \mathcal{A}_{n,k,m}$, the set of odd permutations having k anti-excedances and m fixed points.

The following tables report the first values of these numbers for some fixed integers m:

$m = 0$

$P_{n,k,0}$				
	$k=1$	2	3	4
$n=1$	0			
2	0			
3	1	1		
4	0	3	0	
5	1	11	11	1

$D_{n,k,0}$					
	$k=1$	2	3	4	5
$n=1$	0				
2	1				
3	0	0			
4	1	4	1		
5	0	10	10	0	

$m = 1$

$P_{n,k,1}$				
	$k=1$	2	3	4
$n=1$	1			
2	0			
3	0	0		
4	0	4	4	
5	0	0	15	0

$D_{n,k,1}$				
	$k=1$	2	3	4
$n=1$	0			
2	0			
3	0	3		
4	0	0	0	
5	0	5	10	5

$m = 2$

$P_{n,k,2}$			
	$k=2$	3	4
$n=2$	1		
3	0		
4	0	0	
5	0	10	10

$D_{n,k,2}$			
	$k=2$	3	4
$n=2$	0		
3	0		
4	0	6	
5	0	0	0

We have the following results:

Proposition 3.2. *For all positive integers n, k and m such that $m \geq 0$ and $m \leq k \leq n$, one has:*

$$P_{n,k,m} = P_{n-1,k-1,m-1} + (m+1)D_{n-1,k,m+1} + (n-k)D_{n-1,k-1,m} + (k-m)D_{n-1,k,m}$$
$$D_{n,k,m} = D_{n-1,k-1,m-1} + (m+1)P_{n-1,k,m+1} + (n-k)P_{n-1,k-1,m} + (k-m)P_{n-1,k,m}$$

with the initial conditions $P_{0,0,0} = 1$ and $D_{0,0,0} = 0$.

Proof. The process described in the Theorem 2.4 to prove the recursive formula for the numbers $a_{n,k,m}$ allows to construct an odd permutation of $\mathcal{S}_n$ starting from an even one of $\mathcal{S}_{n-1}$ and vice-versa, when $i \neq n$. In the case $i = n$, the parity remains the same. □

Let F be a subset of $[n]$ such that $|F| = m$. We denote by

- $\mathcal{F}_{n,F}$ the set of permutations having F as set of fixed points,
- $D_{n,k,F}$ the cardinality of the set $\mathcal{F}_{n,F} \bigcap (\mathcal{S}_{n,k} \setminus \mathcal{A}_{n,k})$
- $P_{n,k,F}$ the cardinality of the set $\mathcal{F}_{n,F} \bigcap \mathcal{A}_{n,k}$,
- $i_{n,F}$ the cardinality of the set $(\mathcal{S}_n \setminus \mathcal{A}_n) \bigcap \mathcal{F}_{n,F}$
- $p_{n,F}$ the cardinality of the set $\mathcal{A}_n) \bigcap \mathcal{F}_{n,F}$

We have the following theorem:

Theorem 3.3. *For all positive integers n, k and m such that $m \geq 0$ and $m \leq k \leq n$ and for all subset F of the interval $[n]$ such that $|F| = m$, one has:*

$$D_{n,k,F} - P_{n,k,F} = (-1)^{n-m}.$$

Proof. A combinatorial proof of this result is the subject of the next separate section of this paper. □

Corollary 3.4. *For all integers n, k and m with $m \geq 0$ and $m \leq k \leq n$ and for all subset $F \subset [n]$ such that $|F| = m$, one has:*

(1) $i_{n,F} - p_{n,F} = (-1)^{n-m}(n-m-1)$

(2) $i_{n,m} - p_{n,m} = (-1)^{n-m}(n-m-1)\binom{n}{m}$

(3) $D_{n,k,m} - P_{n,k,m} = (-1)^{n-m}\binom{n}{m}$

(4) $D_{n,k} - P_{n,k} = (-1)^{n-k+1}\binom{n-1}{k-1}$

Proof. Notice that:

(1)
$$\mathcal{F}_{n,F} = \bigsqcup_{k=m+1}^{n-1} \mathcal{S}_{n,k} \bigcap \mathcal{F}_{n,F}.$$

(2)
$$\mathcal{F}_{n,m} = \bigcup_{\substack{F \subset [n] \\ |F|=m}} \mathcal{F}_{n,F}.$$

(3)
$$\mathcal{S}_{n,k,m} = \bigcup_{\substack{F \subset [n] \\ |F|=m}} \mathcal{F}_{n,F} \bigcap \mathcal{S}_{n,k}.$$

(4)
$$\mathcal{S}_{n,k} = \bigsqcup_{m=0}^{k-1} \mathcal{F}_{n,m} \bigcap \mathcal{S}_{n,k}$$

and
$$\sum_{m=0}^{k-1} (-1)^{n-m}\binom{n}{m} = (-1)^{n-k+1}\binom{n-1}{k-1}.$$

The last identity is a simple combinatorial exercice. □

3.1. Generating functions of the numbers $P_{n,k,m}$ and $D_{n,k,m}$.

Proposition 3.5. *The generating function*

$$P(t,x,u) = \sum_{n \geq 0} \sum_{\sigma \in \mathcal{A}_n} t^{AX(\sigma)} x^{FIX(\sigma)} \frac{u^n}{n!}$$

of the distribution of the bivariate statistic (AX, FIX) on the set of even permutations has the closed form:

$$P(t,x,u) = \frac{1}{2}\left\{\frac{\exp\left((x-1)tu\right) - t\exp\left((xt-1)u\right)}{1-t} + \frac{(1-t)\exp\left((x-1)tu\right)}{1-t\exp\left((1-t)u\right)}\right\}.$$

The generating function

$$D(t,x,u) = \sum_{n \geq 0} \sum_{\substack{\sigma \in \mathcal{S}_n \\ \sigma \notin \mathcal{A}_n}} t^{AX(\sigma)} x^{FIX(\sigma)} \frac{u^n}{n!}$$

of the distribution of the bivariate statistic $(AX,\ FIX)$ *on the set of odd permutations has the closed form :*

$$D(t,x,u)=\frac{1}{2}\left\{\frac{\exp\left((x-1)tu\right)-t\exp\left((xt-1)u\right)}{t-1}+\frac{(1-t)\exp\left((x-1)tu\right)}{1-t\exp\left((1-t)u\right)}\right\}.$$

Proof. The series $\sum_{n\geq 0}\sum_{m=0}^{n}\sum_{k=m+1}^{n}(-1)^{n-m}\binom{n}{m}t^k x^m\frac{u^n}{n!}$ has the closed form :

$$\frac{t\exp\left((xt-1)u\right)-\exp\left((x-1)tu\right)}{1-t}.$$

Therefore, the two functions $P(t,x,u)$ and $D(t,x,u)$ are solutions of the system:

$$\begin{cases} P(t,x,u)+D(t,x,u)=A(t,x,u)=\dfrac{(1-t)\exp\left((x-1)tu\right)}{1-t\exp\left((1-t)u\right)}\\ D(t,x,u)-P(t,x,u)=\dfrac{t\exp((xt-1)u)-\exp\left((x-1)tu\right)}{1-t}\end{cases}$$

$\square$

4. Combinatorial proof of $D_{n,k,F}-P_{n,k,F}=(-1)^{n-m}$

Let Φ be a map of $\mathcal{S}_n$ onto itself defined as follows.
For each permutation σ, let (i,j) be the couple (if it exists) such that :

- (p_0) the integers i and j are not fixed points for the permutation σ,
- (p_1) the integer i is the smallest exceedant of σ such that $\sigma^2(i)\neq j$,
- (p_2) $\sigma(i)<j$,
- (p_3) the integers $\sigma(i)$ and j are two consecutive non fixed points of the permutation σ, that is, if k is any integer such that $\sigma(i)<k<j$, then $\sigma(k)=k$.

The permutation $\sigma'=\Phi(\sigma)$ is obtained by multiplying the permutation σ by $(\sigma(i),j)$ on the right, that is, $\sigma'=\sigma(\sigma(i),j)$. In other terms, the permutation σ' is obtained from σ by exchanging the letters $\sigma(i)$ and j in the word $\sigma(1)\dots\sigma(n)$.

For example, if we take $\sigma=3\,2\,4\,1\,5\,7\,8\,6$ then $(i,j)=(3,6)$ and $\sigma'=3\,2\,6\,1\,5\,7\,8\,4$.

Definition 4.1 (*Critical permutations*)**.** *Let* $F=\{f_1,f_2,\cdots,f_m\}$ *a subset of* $[n]$ *and* $\{d_1,d_2,\cdots,d_{n-m}\}=[n]\setminus F$ *with* $d_1<d_2<\cdots<d_{n-m}$*. Consider the following permutations in the set* $\mathcal{F}_{n,F}$*:*

$$\Pi_{n,F,i}=(f_1)(f_2)\cdots(f_m)(d_i\ d_{i-1}\cdots d_1\ d_{i+1}\ d_{i+2}\cdots d_{n-m})$$

for all integer $i=1,\cdots,n-m$*. We will call these "critical permutations" and denote by* $\mathcal{K}_{n,F}$ *the set of the permutations* $\Pi_{n,F,i}$ *for* $i=1,\cdots,n-m$ *and by* $\mathcal{K}_{n,m}=\bigcup_{\substack{F\subset[n]\\|F|=m}}\mathcal{K}_{n,F}$

Example 4.2. For $n=9$ and $F=\{1,3,4,8\}$. We have $m=4$ and

- $\Pi_{9,F,1}=(1)(3)(4)(8)(2\ 5\ 6\ 7\ 9)$
- $\Pi_{9,F,2}=(1)(3)(4)(8)(5\ 2\ 6\ 7\ 9)$
- $\Pi_{9,F,3}=(1)(3)(4)(8)(6\ 5\ 2\ 7\ 9)$
- $\Pi_{9,F,4}=(1)(3)(4)(8)(7\ 6\ 5\ 2\ 9)$
- $\Pi_{9,F,5}=(1)(3)(4)(8)(9\ 7\ 6\ 5\ 2)$

Remark 4.3. *For all positive integer* i *and for all subset* $F\subset[n]$ *such that* $|F|=m$*, one has* $AX(\Pi_{n,F,i})=m+i$*.*

Proposition 4.4. *The map Φ is not defined on the elements of the set $\mathcal{K}_{n,m}$.*

Proof. The critical permutations $\Pi_{n,F,i}$ $(i = 1, \ldots, n-m)$ are the only permutations for which the map Φ is not defined, because there does not exist a couple (i, j) satisfying all the properties. □

Proposition 4.5. *The map Φ preserves the set of fixed points, that is, an integer $\ell \in [n]$ is a fixed point for the permutation σ if and only if it is a fixed point for the permutation $\Phi(\sigma)$.*

Proof. Notice that the two integers $\sigma(i)$ and j that need to be exchanged in σ to compute $\Phi(\sigma)$ are both non fixed points and we have $i < \sigma(i) < j$. After the exchange, we have $\sigma'(i) = \sigma^2(i) = j \neq i$ and $\sigma'(\sigma^{-1}(j)) = \sigma(i) \neq \sigma^{-1}(j)$ because $\sigma^2(i) \neq j$ and $\sigma(i) < j$. □

Proposition 4.6. *The map Φ changes the parity of a given permutation, that is, if σ is an even permutation then $\Phi(\sigma)$ is an odd permutation and vice-versa.*

Proof. The action of Φ consists in multiplying a permutation by a transposition. This operation changes the parity. □

Proposition 4.7. *The map preserves the set of anti-excedances of all permutations, that is, the permutation σ has an anti-excedance in an integer ℓ of the set $[n]$ if and only if $\Phi(\sigma)$ has an anti-excedance in ℓ.*

Proof. The two integers $\sigma(i)$ and j that need to be exchanged in σ to compute $\Phi(\sigma)$ are non fixed points. The exceedant i is both an exceedant for σ and for $\Phi(\sigma)$. Let us look at the possible cases for the integer $\sigma^{-1}(j)$:

(1) if $\sigma^{-1}(j)$ is an exceedant then $\sigma'^{-1}(j)$ is an exceedant as well, since $\sigma^{-1}(j) < \sigma(i) < j$.
(2) if $\sigma^{-1}(j) < j$ (that is, if $\sigma'^{-1}(j)$ is an anti- exceedant), then $\sigma^{-1}(j) < \sigma(i)$.
    Suppose that $\sigma^{-1}(j) > \sigma(i)$, then $j > \sigma^{-1}(j) > \sigma(i)$ and $\sigma^{-1}(j)$ would be the smallest non fixed point greater than $\sigma(i)$, which contradicts the hypothesis that j is the smallest non fixed point greater than $\sigma(i)$.

□

Corollary 4.8. *The map Φ preserves the number of fixed points of a given permutation, as well as the number of anti-excedances.*

Theorem 4.9. *The map Φ is a bijection on the set $\mathcal{F}_{n,F} \setminus \mathcal{K}_{n,F}$ onto itself.*

Proof. Notice that if Φ is defined on a permutation σ, then Φ is also defined on $\sigma' = \Phi(\sigma)$, because it is impossible to obtain a critical permutation as image of another permutation. Notice that the two integers that need to be exchanged in σ' to compute $\Phi(\sigma')$ are the same as the two integers that need to be exchanged in σ to compute $\sigma' = \Phi(\sigma)$ from σ. Therefore, if τ is the transposition such that $\Phi(\sigma) = \sigma\tau$ then $\Phi(\Phi(\sigma)) = \Phi(\sigma\tau) = \sigma\tau\tau = \sigma$. Therefore, Φ is an involution and hence is a bijection. □

Theorem 4.10. *For all integers n, k and m such that $m \geq 0$ and $m \leq k \leq n$ and for all subset F of $[n]$ such that $|F| = m$, one has*

$$D_{n,k,F} - P_{n,k,F} = (-1)^{n-m}.$$

Proof. For all integer $k = m+1, \ldots, n$ and for all subset F of $[n]$ such that $|F| = m$, there exists a unique permutation $\Pi_{n,F,k-m}$ of the set $\mathcal{S}_{n,k}$ which is an element of the set $\mathcal{K}_{n,F}$. Furthermore, this permutation $\Pi_{n,F,k-m}$ is always an even permutation if and only if the integer $n-m$ is odd and vice-versa. □

Corollary 4.11. *The map Φ is a bijection on the set $\mathcal{F}_{n,m} \setminus \mathcal{K}_{n,m}$ onto itself.*

Corollary 4.12. *For all integers n, k and m such that $0 \leq m \leq k \leq n$, one has*

$$D_{n,k,m} - P_{n,k,m} = (-1)^{n-m}\binom{n}{m}.$$

References

[1] D. Foata, M. P. Schïutzenberger, *Théorie Géometrique des Polynômes Eulériens*, Lect. Notes in Math. **138**, Springer-Verlag, Berlin, (1970).

[2] D. Knuth, *The Art of Computer Programming*, **3**, Sorting and Searching 2nd Ed., Addison-Wesley, Reading, (1981).

[3] R. Mantaci, *Sur la distribution des anti-excédances dans le groupe symétrique et dans ses sous-groupes*, Theoret. Comp. Science A, **117**, pp 243-253, (1993).

[4] R. Mantaci, *Binomial Coefficients and Anti-exceedances of Even Permutations: A Combinatorial Proof* , Jour. of Comb. Theory, Series A, Vol.**63**, No. 2, (July 1993).

[5] R. Mantaci, F. Rakotondrajao, *Exceedingly deranging !*, Proc. Conf. "Formal Power Series and Algebraic Combinatorics" (FPSAC) 2001, Tempe, Arizona, USA, (2001).

[6] F. Rakotondrajao, Ph. D. thesis, Université d'Antananarivo, Madagascar, November 1999.

[7] J. Riordan, *An introduction to combinatorial Analysis* , New York, J. Wiley, (1958).

Département de Mathématiques et Informatique, Université d' Antananarivo, 101 Antananarivo, Madagascar

E-mail address: `frakoton@syfed.refer.mg`

ASSOCIATION SCHEMES BASED ON ISOTROPIC SUBSPACES, PART I

MICHAEL Q. RIECK

Abstract. The subspaces of a given dimension in a finite classical polar space form the points of an association scheme. When the dimension is zero, this is the scheme of the collinearity graph of the space. At the other extreme, when the dimension is maximal, it is the scheme of the corresponding dual polar graph. These extreme cases have been thoroughly studied. In this article, the general case is examined and a detailed computation of the intersection numbers of these association schemes is initiated.

Résumé. Les sous-espaces d'une dimension donnée dans un espace polaire fini classique forment les points d'un schéma d'association. Quand la dimension est zéro, cela constitue le schéma du graphe collinéaire de l'espace. A l'autre extrème, quand la dimension est maximale, c'est le schéma du graphe duel polaire correspondant. On a déjà étudié de façon approfondie ces cas extrèmes. Dans cet article on examinera le cas général, afin d'inaugurer un calcul détaillé des nombres d'intersection de ces schémas d'association.

1. Introduction

Grassman graphs and dual polar graphs are two well-known classes of distance-regular graphs. The intersection numbers and eigenvalues of these and other distance-regular graphs with "classical parameters" are recorded in [4, Chapter 9]. In the case of the dual polar graphs, they were determined by D. Stanton in [14]. In his Ph.D. dissertation [10], the author of the present article used a number of formulas from [7] to give an alternative derivation of the intersection numbers found in [14]. Also in [10], adjacency in a graph which the author calls a *hyperbolic partner graph* was observed to be part of an association scheme, although this fact is an easy consequence of Witt's Extension Theorem. Actually, there are classes of such graphs and related association schemes. These association schemes include those of the dual polar graphs as well as the collinearity graphs of finite classical polar spaces, and are rather evident generalizations of these. They can also be viewed as a type of analogue of the Grassman graph association schemes, which in turn are "q-analogues" of Johnson graph association schemes.

The association schemes to be considered are defined as follows. Fix an N-dimensional vector space V over a finite field $\mathrm{GF}(q)$. In order to simplify the discussion, it will be assumed that q is odd. Equip this space with a non-degenerate form $\langle \cdot, \cdot \rangle$ that is symmetric bilinear, alternating bilinear or Hermitian. Let d be the Witt index of this form (*i.e.* the dimension of any maximal isotropic subspace), and assume that $d \geq 1$.

Fix an integer m between 1 and d. The points of the association scheme will be the isotropic (*i.e.* totally singular) m-subspaces of V.

If U and U' are isotropic m-subspaces of V, and if $\dim(U \cap U') = m - k$, and $\dim(U^\perp \cap U') = m - \gamma$, then we will we say that these two subspaces are (k, γ) - *associates.* $U^\perp$ is defined to be $\{ v \in V \mid (\forall u \in U) \langle u, v \rangle = 0 \}$. Clearly $0 \leq \gamma \leq k \leq m$ here since $U \subseteq U^\perp$. It is not hard to see that $\dim(U \cap U'^\perp)$ will also equal $m - \gamma$ as follows. $(U \cap U'^\perp)^\perp = U^\perp + U'$

$\mu = 0, \nu = 1$ orthogonal D_d $N = 2d$		
$\mu = 0, \nu = \frac{1}{2}$ unitary ${}^2A_{2d-1}$ $N = 2d$	$\mu = \frac{1}{2}, \nu = \frac{1}{2}$ orthogonal B_d $N = 2d+1$	
$\mu = 0, \nu = 0$ symplectic C_d $N = 2d$	$\mu = \frac{1}{2}, \nu = 0$ unitary ${}^2A_{2d}$ $N = 2d+1$	$\mu = 1, \nu = 0$ orthogonal ${}^2D_{d+1}$ $N = 2d+2$

FIGURE 1. Parameters for the six geometry types

has dimension $(N-m)+m-(m-\gamma) = N-(m-\gamma)$, so $U \cap U'^{\perp}$ has dimension $m-\gamma$. The ordered pair (k,γ) will be written using the following unorthodox notation: k_γ. Thus for example, 5_2 really means (5,2). We will speak of k_γ - associates, rather than (k,γ) - associates. Two isotropic subspaces that are k_γ - associates are a distance k apart in the Grassman graph whose vertices are the m-subspaces of V.

Let $\mathcal{N}_m$ denote the collection of all isotropic m-subspaces of V. Let $\mathcal{R}_{m,k_\gamma}$ be the collection of all ordered pairs of isotropic m-subspaces of V that are k_γ-associates. If m is implicit, then $\mathcal{R}_{k_\gamma}$ will be written instead of $\mathcal{R}_{m,k_\gamma}$. The claim of course is that with $\mathcal{N}_m$ as the set of "points", the relations $\mathcal{R}_{k_\gamma}$ $(0 \le \gamma \le k \le m)$ form an association scheme. It is an immediate consequence of Witt's Extension Theorem that when the isometry group of form $\langle\,\cdot\,,\,\cdot\,\rangle$ acts (in the canonical way) on $\mathcal{N}_m$, this action is generously transitive and the orbitals are in fact the relations $\mathcal{R}_{k_\gamma}$. It follows that these relations do indeed form an association scheme. This is the same reasoning typically used to establish that Grassman graphs and dual polar graphs are distance-transitive (cf. [4, Theorems 9.3.1, 9.3.3, 9.4.3]).

There are actually six distinct types of forms on V ("geometries") to consider. If the form is alternating, then V is said to have a *symplectic geometry* or a geometry of type C_d. Here $N = 2d$. If the form is symmetric, then one speaks of V having an *orthogonal geometry*. Here if n is odd, then n will equal $2d+1$ and the geometry is said to be of type B_d. But if n is even, then it is possible that either $N = 2d$ or $N = 2d+2$, and one speaks of geometries of type D_d or ${}^2D_{d+1}$, accordingly. Finally, if the form is Hermitian, then V is said to have a *unitary geometry*. If N is even, then $N = 2d$, and the geometry is said to be of type ${}^2A_{2d-1}$, while if N is odd, then $N = 2d+1$, and the geometry is said to be of type ${}^2A_{2d}$. Following notation introduced in [10], let $\mu = \frac{1}{2}N - d$ and let ν be such that $\mu+\nu$ equals 0, $\frac{1}{2}$, 1 in the symplectic, unitary, orthogonal cases, respectively. The parameters for the six types of geometries are shown in Figure 1.

We will require some additional terminology and notation. $\mathcal{L}(V)$ will denote the lattice of all subspaces of V. Given $U \in \mathcal{L}(V)$ (*i.e.* a subspace of V), the *radical* of U, denoted $\mathrm{rad}(U)$, is simply the subspace $U \cap U^{\perp}$, which is necessarily isotropic. Note that U is isotropic if and only if $U = \mathrm{rad}(U)$ ($U \subseteq U^{\perp}$). When U is isotropic, the quotient space $U^{\perp}/U$ inherits the form on V by defining $\langle u+U, v+U\rangle = \langle u, v\rangle$, and this inherited form is non-degenerate and produces the same type of geometry on $U^{\perp}/U$ as on V. Also, $W \in \mathcal{L}(V)$ is called *coisotropic* if $W^{\perp} \subseteq W$. Also, for $A, B \in \mathcal{L}(V)$ with $B \subseteq A$, $[A : B]$ will denote $\dim(A/B)$ ($= \dim(A)$ - $\dim(B)$).

In the discussion that follows, standard notation for *q-shifted factorials* and *q-binomial coefficients* will be used. Thus

$$(a;q)_m = \prod_{j=0}^{m-1}(1-aq^j)\,, \qquad (a,b;q)_m = (a;q)_m(b;q)_m\,,$$

$$\begin{bmatrix} n \\ m \end{bmatrix} = \frac{(q^{n-m+1};q)_m}{(q;q)_m} = \frac{(q;q)_n}{(q;q)_m\,(q;q)_{n-m}}.$$

The latter is the number of m-subspaces of an n-dimensional GF(q)-vector space. The number of these that are isotropic, with respect to an implied form, will be denoted

$$\begin{bmatrix} n \\ m \end{bmatrix}_0$$

This value will be computed in the next section.

Define $p_{i_\alpha j_\beta k_\gamma}$ to be the number of ordered triples (U, U', U'') of isotropic m-subspaces such that $(U, U') \in \mathcal{R}_{k_\gamma}$, $(U, U'') \in \mathcal{R}_{j_\beta}$ and $(U', U'') \in \mathcal{R}_{i_\alpha}$. Taking $p^{k_\gamma}_{i_\alpha j_\beta} = p_{i_\alpha j_\beta k_\gamma}/|\mathcal{R}_{k_\gamma}|$, note that this is the number of isotropic m-subspaces U'' such that $(U, U'') \in \mathcal{R}_{j_\beta}$ and $(U', U'') \in \mathcal{R}_{i_\alpha}$, given any $(U, U') \in \mathcal{R}_{k_\gamma}$. These are the intersection numbers of the association scheme. The goal of this paper is to compute $p^{1_0}_{i_\alpha j_\beta}$ as a first step towards computing all of the intersection numbers and eigenvalues for the association scheme ($\mathcal{N}_m$, $\{\, \mathcal{R}_{k_\gamma} \,\}$).

A question related to the computation of $p^{k_\gamma}_{i_\alpha j_\beta}$, which will be called the "lattice problem", is the following. Consider the sublattice $\mathcal{M}$ of $\mathcal{L}(V)$ generated by 0 and V, and the isotropic m-subspaces U and U', and the coisotropic subspaces $U^{\perp}$ and $U'^{\perp}$. Since $U \subseteq U^{\perp}$ and $U' \subseteq U'^{\perp}$, when 0 and V are removed from $\mathcal{M}$, the resulting lattice, in the generic case, is the lattice with 18 nodes considered in [3, Exercise 3.7.6]. $\mathcal{M}$ is depicted in Figure 2. We now ask, how many isotropic m-subspaces U'' intersect each of the nodes of $\mathcal{M}$ in some prescribed number (depending on the node) of dimensions? In general this is a very complicated enumeration problem. However, if its solution was known, then $p^{k_\gamma}_{i_\alpha j_\beta}$ could be computed by simply summing such counts. To see this, and still assuming that $(U, U') \in \mathcal{R}_{k_\gamma}$, note that $p^{k_\gamma}_{i_\alpha j_\beta}$ is just the number of isotropic m-subspaces U'' satisfying $\dim(U \cap U'') = m - j$, $\dim(U' \cap U'') = m - i$, $\dim(U^{\perp} \cap U'') = m - \beta$, $\dim(U'^{\perp} \cap U'') = m - \alpha$. So by summing the counts from all of the lattice problem enumerations that impose these four conditions on U'', one would obtain $p^{k_\gamma}_{i_\alpha j_\beta}$.

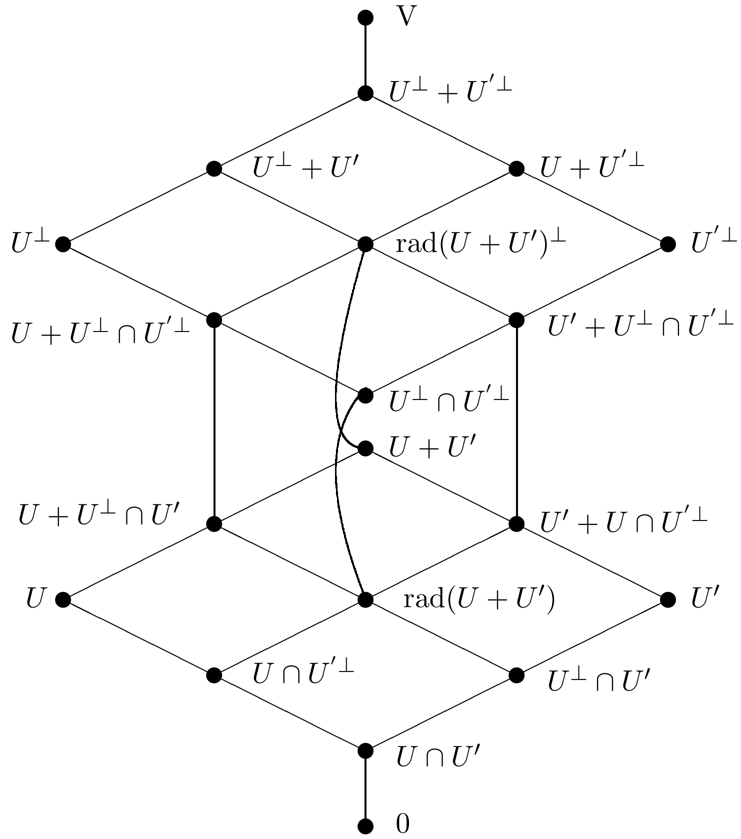


FIGURE 2. The lattice $\mathcal{M}$ generated by $0, V, U, U', U^{\perp}$ and $U'^{\perp}$ (generic case)

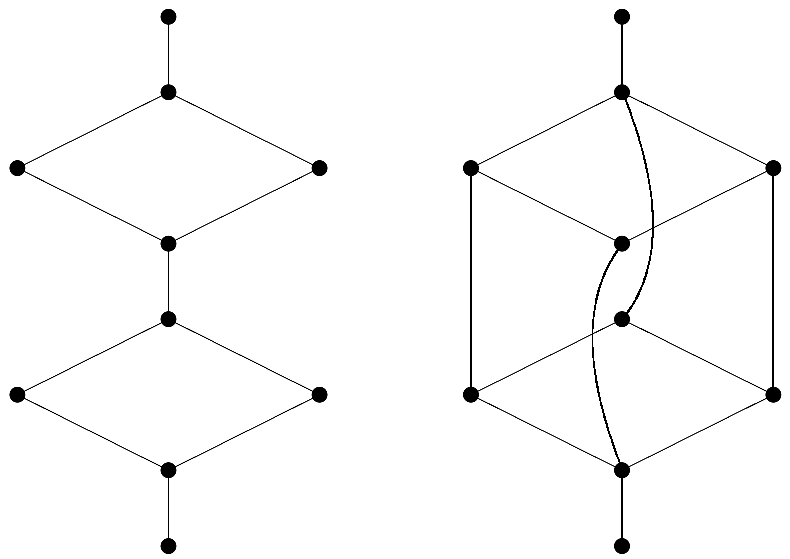
Although the lattice problem seems to be difficult in general, we will consider a couple special cases that are more tractable. The first of these is the case where $U+U' \subseteq U^{\perp} \cap U'^{\perp}$, which means that $\gamma = 0$. Here the lattice $\mathcal{M}$ is considerably simpler than the lattice in Figure 2. It is depicted in Figure 3(a). A subcase of this special case is handled in Section 3, and this is used to compute $p^{1_0}_{i_\alpha j_\beta}$.

Another special case is the case where $U \cap U' = \mathrm{rad}(U + U')$. Here $\gamma = k$ ($U^{\perp} \cap U' = U \cap U'$). This case is depicted in Figure 3(b). A subcase of this special case leads to the computation of $p^{1_1}_{i_\alpha j_\beta}$, which will be computed in a follow-up article.

Remark 1. *By reasoning along the lines used in the theory of distance-regular graphs* (*cf.* [4, Section 4.1]), *it may be seen that together the formulas for* $p^{1_0}_{i_\alpha j_\beta}$ *and* $p^{1_1}_{i_\alpha j_\beta}$ *provide enough information to produce a recursion formula for computing the general intersection numbers* $p^{k_\gamma}_{i_\alpha j_\beta}$.

2. SOME ENUMERATIVE LEMMAS

Most of the lemmas in this section are concerned with counting the number of isotropic subspaces that satisfy certain conditions. They provide the basic tools for handling the enumeration problems discussed in the introduction, at least in special cases. The first two



(a) $U + U' \subseteq U^{\perp} \cap {U'}^{\perp}$ (b) $U \cap U' = \mathrm{rad}(U + U')$

FIGURE 3. Two special cases of $\mathcal{M}$

lemmas concern a reduction that will be routinely applied in the next section. Both are straightforward to check, and the proof of the first one is omitted.

Lemma 1. *Fix an isotropic subspace Z of V. Let $Z^{\perp}/Z$ inherit the form on V. Let $\varphi : \mathcal{L}(V) \to \mathcal{L}(Z^{\perp}/Z)$ by $\varphi(X) = (Z + (Z^{\perp} \cap X))/Z$. This mapping preserves containment and the $\perp$ operation. That is, $\varphi(X^{\perp}) = \varphi(X)^{\perp}$ for all subspaces X of V. However, it does not in general preserve the lattice operations ($+$ and $\cap$).*

Lemma 2. *Fix an isotropic m-subspace Z of V. Let $Z^{\perp}/Z$ inherit the form on V. Let $\varphi : \mathcal{L}(V) \to \mathcal{L}(Z^{\perp}/Z)$ by $\varphi(X) = (Z+(Z^{\perp}\cap X))/Z$. Fix $A, B_1, ..., B_j \in \mathcal{L}(V)$. Consider the restriction of φ to the collection of isotropic $(m+k)$ - subspaces X, containing Z, contained in $Z + A$, and such that $X \cap B_{\alpha} \subseteq Z$ for all $\alpha = 1, ..., j$. This restriction is a one-to-one correspondence with the isotropic k-subspaces of $\varphi(A)$ that trivially intersect each $\varphi(B_{\alpha})$ $(\alpha = 1, ..., j)$.*

Proof. For a given X (as above), note that $Z \subseteq X \subseteq X^{\perp} \subseteq Z^{\perp}$. So, $\varphi(X) = X/Z$ and $\varphi(X) \cap \varphi(B_{\alpha}) = [X \cap (Z + (Z^{\perp} \cap B_{\alpha}))]/Z = (Z + (X \cap B_{\alpha}))/Z = \varphi(X \cap B_{\alpha})$. This is clearly zero since $X \cap B_{\alpha} \subseteq Z$. Also, $\dim(\varphi(X)) = k$ and $\varphi(X) \subseteq \varphi(Z + A) = \varphi(A)$. So $\varphi(X)$ is as claimed. Conversely, any isotropic k-subspace W of $\varphi(A)$ that trivially intersects each $\varphi(B_{\alpha})$ $(\alpha = 1, ..., j)$ lifts to a unique isotropic $(m + k)$ - subspace Y of $Z + A$, containing Z, and such that $\varphi(Y) = W$. Note that Y is necessarily in $Z^{\perp}$, so that $\varphi(Y) = Y/Z$. The properties that Y is required to satisfy can be immediately verified by reversing the above argument. $\square$

The following enumerative lemma is known, and in fact is [14, Proposition 4.2] and [4, Lemma 9.4.1]. The statement and proof in [4] is followed by a list of citations to articles in which this result was originally produced for the individual geometry types.

Lemma 3.

$$\begin{bmatrix} N \\ m \end{bmatrix}_0 = \frac{(q^{\frac{1}{2}N-\mu-m+1}, -q^{\frac{1}{2}N-\nu-m+1}; q)_m}{(q;q)_m} = \frac{(q^{\frac{1}{2}N-\mu}, -q^{\frac{1}{2}N-\nu}; q^{-1})_m}{(q;q)_m}$$

Proof. Using the notation in [12], the number of isotropic vectors not contained in, and orthogonal to, a given isotropic j-subspace is, by [12, Theorem 3.2],

$$\left[\begin{array}{c|c} N;\tau;q & 0 \\ \hline 0\cdots 0 & 0 \\ \vdots\ \ \vdots & \vdots \\ \underbrace{0\cdots 0}_{j\times j} & 0 \end{array}\right]$$

$$= \quad q^j\,[N-2j\,;\,\tau\,;\,q \mid 0] \quad = \quad q^j(q^{\frac{1}{2}N-\mu-j}-1)(q^{\frac{1}{2}N-\nu-j}+1).$$

It follows that the number sought is

$$\prod_{j=0}^{m-1} \frac{q^j(q^{\frac{1}{2}N-\mu-j}-1)(q^{\frac{1}{2}N-\nu-j}+1)}{q^m-q^j} = \prod_{j=0}^{m-1} \frac{(q^{\frac{1}{2}N-\mu-j}-1)(q^{\frac{1}{2}N-\nu-j}+1)}{q^{m-j}-1} =$$

$$\prod_{i=1}^{m} \frac{(q^{\frac{1}{2}N-\mu-m+i}-1)(q^{\frac{1}{2}N-\nu-m+i}+1)}{q^i-1} = \frac{(q^{\frac{1}{2}N-\mu-m+1}, -q^{\frac{1}{2}N-\nu-m+1}; q)_m}{(q;q)_m}.$$

□

The next two lemmas enumerate particular isotropic k-subspaces. These will be applied repeatedly in the next section, in combination with Lemma 2, in order to count the number of ways to extend a given isotropic subspace to a larger one, subject to certain restrictions.

Lemma 4. *The number of isotropic k-subspaces of V that trivially intersect a given coisotropic subspace of codimension m is*

$$\begin{bmatrix} m \\ k \end{bmatrix} \cdot q^{k(N-m-\frac{1}{2}k+\frac{1}{2}-\mu-\nu)}.$$

Proof. Let W denote the coisotropic subspace. First consider the special case where $k=m$. An isotropic m-subspace X that trivially intersects W is a complement of W. That is, $V = X \oplus W$. Note that $\dim(\mathrm{rad}(W)) = \dim(W^\perp) = m$. Now, by the modularity of $\mathcal{L}(V)$, $X^\perp = X^\perp \cap (X \oplus W) = X \oplus (X^\perp \cap W)$. So $\dim(X^\perp \cap W) = N-2m$. It follows that $X^\perp \cap W$ is a maximal nondegenerate subspace of W. Now, $X \cap W^\perp \subseteq X \cap W = 0$. So $X \cap W^\perp = 0$, and this is the unique maximal non-degenerate subspace of X. X and W are therefore pseudo-orthogonal complements in the sense defined in [11]. Conversely, every pseudo-orthogonal complement of W is isotropic. The result sought, in the special case $k=m$, then follows by [13, Theorem 5]. (The γ there equals $1-2\mu-2\nu$ here).

For general k, note first that a given suitable k-subspace Z is contained in multiple isotropic m-subspaces that trivially intersect W. To find the number of such, consider the mapping $\varphi : \mathcal{L}(V) \to \mathcal{L}(Z^\perp/Z)$ given by $\varphi(X) = (Z + (Z^\perp \cap X))/Z$. The coisotropic subspace $\varphi(W) = (Z \oplus (Z^\perp \cap W))/Z \cong Z^\perp \cap W = (Z \oplus W^\perp)^\perp$, which has dimension $N-m-k$. $Z^\perp/Z$ has dimension $N-2k$. By the first part of this proof, the number of isotropic

$(m-k)$ - subspaces of $Z^{\perp}/Z$ that trivially intersect $\varphi(W)$ is $q^{(m-k)(N-\frac{3}{2}m-\frac{1}{2}k+\frac{1}{2}-\mu-\nu)}$. By Lemma 2, this is also the number of isotropic m-subspaces Y of V that contain Z and have $Y \cap W \subseteq Z$. But $Y \cap W \subseteq Z$ if and only if $Y \cap W = 0$. Now note that each isotropic m-subspace that trivially intersects W contains $\left[{m \atop k}\right]$ isotropic k-subspaces (that

trivially intersect W). The result sought, for general k, now follows by double counting suitable pairs (Z, Y), one time selecting Z first,

the other time selecting Y first.

□

Lemma 5. *Consider a collection of coisotopic hyperplanes $H_1, H_2, ..., H_k$ of V in general position, and having the property that the intersection of any subset of these is coisotropic. The number of isotropic 1-subspaces of V that trivially intersect each of these hyperplanes is*

$$q^{N-k-\mu-\nu}(q-1)^{k-1}$$

Proof. For $S \subseteq \{1, 2, ..., k\}$, let $X_S = \bigcap_{j \in S} H_j$. Since the hyperplanes are in general position, $\dim(X_S) = N - |S|$. From Lemmas 3 and 4, it can be seen that the number of isotropic 1-subspaces contained in X_S is

$$\left[\begin{array}{c} N \\ 1 \end{array} \right]_0 - \left[\begin{array}{c} |S| \\ 1 \end{array} \right] \cdot q^{N-|S|-\mu-\nu} = \frac{1 - q^{\frac{1}{2}N-\mu} + q^{\frac{1}{2}N-\nu} - q^{N-|S|-\mu-\nu}}{1-q}$$

Applying the Principle of Inclusion/Exclusion, the desired count is

$$\sum_{j=0}^{k} (-1)^j \binom{k}{j} \left[\frac{1 - q^{\frac{1}{2}N-\mu} + q^{\frac{1}{2}N-\nu} - q^{N-j-\mu-\nu}}{1-q} \right]$$

$$= \sum_{j=0}^{k} (-1)^j \binom{k}{j} \left[\frac{q^{N-j-\mu-\nu}}{q-1} \right]$$

$$= \frac{q^{N-\mu-\nu}}{q-1} \sum_{j=0}^{k} \binom{k}{j} (-q^{-1})^j = q^{N-k-\mu-\nu}(q-1)^{k-1}.$$

□

The simple form of the count in Lemma 5 correctly suggests that a more intuitive proof is possible. With $S = \{1, 2, ..., k\}$, let Y be a pseudo-orthogonal complement of X_S. Y is isotropic and a hyperbolic partner of $\mathrm{rad}(X_S)$, as defined in [12, Section 2]. The decomposition $V = Y \oplus \mathrm{rad}(X_S) \oplus (Y \oplus \mathrm{rad}(X_S))^{\perp}$ can be used to obtain an intuitive proof. This involves considering "matched bases", as in the proof of [11, Proposition 1], for Y and $\mathrm{rad}(X_S)$, and an arbitrary basis for $(Y \oplus \mathrm{rad}(X_S))^{\perp}$. The details are left to the interested reader.

The next lemma is also known, and is a special case of [14, Proposition 4.10] and [4, Lemma 9.4.2]. It provides the

sizes of the binary relations involved in the association scheme.

Lemma 6.

$$\left|\mathcal{R}_{k_\gamma}\right| = \begin{bmatrix} N \\ m \end{bmatrix}_0 \begin{bmatrix} m \\ k \end{bmatrix} \begin{bmatrix} N-2m \\ k-\gamma \end{bmatrix}_0 \begin{bmatrix} k \\ \gamma \end{bmatrix} \cdot q^{k^2+\gamma(N-2m-2k+\frac{3}{2}\gamma+\frac{1}{2}-\mu-\nu)}$$

$$= \frac{(q^{\frac{1}{2}N-\mu}, -q^{\frac{1}{2}N-\nu}; q^{-1})_{m+k-\gamma}}{(q;q)_{m-k}\,(q;q)_\gamma\,(q;q)_{k-\gamma}^2} \cdot \; q^{k^2+\gamma(N-2m-2k+\frac{3}{2}\gamma+\frac{1}{2}-\mu-\nu)}$$

Proof. Fix an isotropic m-subspace U. There are $\left[{N \atop m}\right]_0$ choices this. Now consider selecting an isotropic m-subspace U' such that $(U,U') \in \mathcal{R}_{k_\gamma}$, by selecting this a portion at a time, beginning with the selection of an isotropic $(m-k)$ - subspace Z of U to play the role of $U \cap U'$. There are $\left[{m \atop k}\right]$ choices for Z.

After selecting Z, a subspace Y to play the role of $U^\perp \cap U'$ needs to be selected. This will be accomplished indirectly by selecting an isotropic subspace Y/Z of $Z^\perp/Z$. This selection must trivially intersect the k-subspace U/Z. It must also be a subspace of $U^\perp/Z$. By Lemma 2, these conditions are necessary and
sufficient. Y/Z must therefore project injectively into the quotient space $(U^\perp/Z)/(U/Z)$ ($\cong U^\perp/U$). $\dim(U^\perp/U) = N-2m$, and Y/Z must have dimension $k-\gamma$. So the number of choices for the isotropic subspace of $(U^\perp/Z)/(U/Z)$ is $\left[{N-2m \atop k-\gamma}\right]_0$. Each such choice lifts to $q^{k(k-\gamma)}$ choices for Y/Z, and each of these uniquely determines Y. So, given a choice for
Z, there are $\left[{N-2m \atop k-\gamma}\right]_0\, q^{k(k-\gamma)}$ choices for Y.

After selecting Y, the rest of U' needs to be selected. This will be done by selecting an isotropic γ-subspace U'/Y of $Y^\perp/Y$, which has dimension $N-2m+2\gamma$. Again by Lemma 2, the selection must trivially intersect the coisotropic subspace $(Y^\perp \cap U^\perp)/Y = (Y+U)^\perp/Y$, which has the k-dimensional radical $(Y+U)/Y$ ($\cong U/(U\cap Y) = U/Z$), and so has codimension k. By Lemma 4, the number of choices for this selection is $\left[{k \atop \gamma}\right]\, q^{\gamma(N-2m+2\gamma-k-\frac{1}{2}\gamma+\frac{1}{2}-\mu-\nu)}$.

By multiplying the counts for the various portions, we see that the number of choices for U' is

$$\begin{bmatrix} m \\ k \end{bmatrix} \begin{bmatrix} N-2m \\ k-\gamma \end{bmatrix}_0 \begin{bmatrix} k \\ \gamma \end{bmatrix} \cdot q^{k^2+\gamma(N-2m-2k+\frac{3}{2}\gamma+\frac{1}{2}-\mu-\nu)} =$$

$$\frac{(q;q)_m\,(q^{\frac{1}{2}N-m-\mu}, -q^{\frac{1}{2}N-m-\nu}; q^{-1})_{k-\gamma}}{(q;q)_{m-k}\,(q;q)_\gamma\,(q;q)_{k-\gamma}^2} \cdot q^{k^2+\gamma(N-2m-2k+\frac{3}{2}\gamma+\frac{1}{2}-\mu-\nu)}$$

Multiply this by the number of choices for U, namely $\left[{N \atop m}\right]_0$, expand and simplify to get the stated result.

□

3. The computation of $p^{1_0}_{i_\alpha j_\beta}$

In order to compute $p^{1_0}_{i_\alpha j_\beta}$, fix two isotropic m subspaces U and U', and assume that these intersect in
$m-1$ dimensions and that they are orthogonal to each other. So $(U,U') \in \mathcal{R}_{1_0}$, and necessarily $1 \le m \le d-1$. Notice that $U+U' \subseteq U^\perp \cap U'^\perp$ and so the lattice $\mathcal{M}$ reduces

to the one in Figure 3(a). The goal of this section is to count the number of isotropic m-subspaces U'' satisfying

1. $\dim(U \cap U'') = m - j$,
2. $\dim(U' \cap U'') = m - i$,
3. $\dim(U^\perp \cap U'') = m - \beta$,
4. $\dim(U'^\perp \cap U'') = m - \alpha$,

where $0 \leq \alpha \leq i \leq m$ and $0 \leq \beta \leq j \leq m$.

However, in order to effect a systematic enumeration of these subspaces, four additional conditions will be imposed on U''. These additional conditions will later be relaxed by summing over the four new

parameters to be introduced. Specifically, fix numbers h, e, η, ε and impose on U'' the following additional requirements:

5. $\dim(U \cap U' \cap U'') = m - h$,
6. $[(U + U') \cap U'' : U \cap U'' + U' \cap U''] = e$,
7. $\dim(U^\perp \cap U'^\perp \cap U'') = m - \eta$,
8. $[(U^\perp + U'^\perp) \cap U'' : U^\perp \cap U'' + U'^\perp \cap U''] = \varepsilon$.

Thus, the goal is then to solve the lattice problem mentioned in the introduction, but only for the special case $k = 1$ and $\gamma = 0$. The various parameters are required to satisfy certain conditions in order that a U'' as described above exists. Some of these conditions are given in the next lemma, while others will be addressed in Theorem 1.

Lemma 7. *If a subspace U'' exists that satisfies the above eight conditions, then the parameters $h, i, j, e, \eta, \alpha, \beta$ and ε must satisfy all of the following relationships:*

9. *h is between 1 and m, equals i or $i + 1$, and equals j or $j + 1$;*
10. *η is between 0 and m, equals*
α or $\alpha + 1$, and equals β or $\beta + 1$;
11. *e and ε are each either 0 or 1;*
12. $0 \leq i + j - h - e - \eta \leq d - m - 1$;
13. $0 \leq \alpha + \beta - \eta - \varepsilon < h$;
14. $i - \alpha \leq d - m$;
15. $j - \beta \leq d - m$;
16. $0 \leq h - 1 - \alpha - \beta + \eta + \varepsilon \leq d - m$.

Proof. The first three items in this list are straightforward to check. For example, $e = [(U + U') \cap U'' : U \cap U'' + U' \cap U''] \leq [(U + U') \cap U'' : U \cap U''] \leq [U + U' : U] = [U' : U \cap U'] = 1$. The first inequality in item 12 follows from the fact that the dimension of $(U + U') \cap U''$ is required to be $(m - i) + (m - j) - (m - h) + e = m - i - j + h + e$, and so $0 \leq [U^\perp \cap U'^\perp \cap U'' : (U + U') \cap U''] = i + j - h - e - \eta$. The second inequality in item 12 will be proved below.

Now, $\dim[(U^\perp + U'^\perp) \cap U''] = (m - \alpha) + (m - \beta) - (m - \eta) + \varepsilon = m - \alpha - \beta + \eta + \varepsilon \leq m$, so $\alpha + \beta - \eta - \varepsilon \geq 0$. Since $\dim(U \cap U' \cap U'') = m - h$, $\alpha + \beta - \eta - \varepsilon \leq h$. Suppose that equality holds here, *i.e.* suppose that $(U^\perp + U'^\perp) \cap U'' = U \cap U' \cap U''$. Let $\iota = [U \cap U' : U \cap U' \cap U''] = [U^\perp + U'^\perp + U''^\perp : U^\perp + U'^\perp]$. Then $\iota \geq [U^\perp + U'^\perp + U'' :$

$U^\perp + U'^\perp] = [U'' : (U^\perp + U'^\perp) \cap U''] = [U'' : U \cap U' \cap U'']$. Therefore $\dim(U'') = \dim(U \cap U' \cap U'') + [U'' : U \cap U' \cap U''] \leq (m-1-\iota)+\iota < m$, which is impossible. Therefore $\alpha + \beta - \eta - \varepsilon < h$. This establishes item 13.

To establish item 14, note that $U' + (U'^\perp \cap U'')$ is isotropic and has dimension $\dim(U')$ + $\dim(U'^\perp \cap U'')$ - $\dim(U' \cap U'') = m + (m-\alpha) - (m-i) = m + i - \alpha$, which cannot exceed d. Item 15 requires the obvious adjustment to the proof of item 14. The second inequality in item 12 is shown by means of the same sort of reasoning, but relies on the fact that $(U+U') + (U^\perp \cap U'^\perp \cap U'')$ is isotropic, so $\dim[(U+U') + (U^\perp \cap U'^\perp \cap U'')] = (m+1)+(m-\eta)-(m-i-j+h+e) = m+1+i+j-h-e-\eta \leq d$. Item 16 likewise results from the isotropic nature of $(U \cap U') + (U^\perp + U'^\perp) \cap U''$, which has dimension $(m-1)+(m-\alpha-\beta+\eta+\varepsilon)-(m-h) = m+h-1-\alpha-\beta+\eta+\varepsilon$. This cannot exceed d. But, $\dim(U \cap U' \cap U''^\perp) = N - \dim((U^\perp + U'^\perp) + U'') = N-(N-m+1)-m+(m-\alpha-\beta+\eta+\varepsilon) = m-1-\alpha-\beta+\eta+\varepsilon$, so that $[U \cap U' \cap U''^\perp : U \cap U' \cap U''] = h-1-\alpha-\beta+\eta+\varepsilon \geq 0$. □

Henceforth, it will be assumed that the parameters satisfy the conditions listed in Lemma 7. The following theorem is concerned with counting the number of possible U''. In this enumeration, $q \wedge x$ denotes q^x.

Theorem 1. *The number of isotropic m-subspaces U'' satisfying conditions 1 through 16 (above) is*

$$\frac{(q;q)_{m-1}\,(q^{N/2-\mu-m-1}, -q^{N/2-\nu-m-1}; q^{-1})_{i+j-h-e-\eta}}{(q;q)_{m-h}\,(q;q)_{\alpha+\beta-\eta-\varepsilon}\,(q;q)_{h-1-\alpha-\beta+\eta+\varepsilon}\,(q;q)_{i+j-h-e-\eta}} \cdot (q-1)^{e+\varepsilon-e\varepsilon}$$

$$\begin{gathered}\cdot\, q \wedge \{\, i(h-i) + j(h-j) + (h-1)e + (i+j-h-e-\eta)(i+j-h-e+1) \\ +\,(N-2m-\mu-\nu-i+\alpha+\beta)(\eta-\alpha) + (N-2m-\mu-\nu-j+\alpha+\beta)(\eta-\beta) \\ +\,(\eta-\alpha)(\eta-\beta) + (N-2m-\mu-\nu-1-h+2\eta+e)\varepsilon \\ +\,(\alpha+\beta-\eta-\varepsilon)(N-2m-\mu-\nu-h+3[\alpha+\beta-\eta-\varepsilon+1]/2)\,\}\,,\end{gathered}$$

provided that at least one of the following five conditions is also satisfied:

- *A.* $\varepsilon = e = 0$ *and* $h = i = j$;
- *B.* $\varepsilon = e = 0$ *and* $\eta = \alpha = \beta$;
- *C.* $\varepsilon = e = 0$, $h = i$ *and* $\eta = \alpha$;
- *D.* $\varepsilon = e = 0$, $h = j$ *and* $\eta = \beta$;
- *E.* $h = i = j$ *and* $\eta = \alpha = \beta$.

In all other cases, the count is zero.

Proof. To count the number of suitable U'', we will consider selecting such a subspace a portion at a time, similar to the way in which U' was selected in Lemma 6. The restrictions stated in Lemma 7 will be assumed. It will also be assumed that i and j are both nonzero, the case where one of these is zero being trivial to check. However, we will not impose any of the five additional conditions, listed at the bottom of Theorem 1, at least not until the end of this proof.

The nodes of $\mathcal{M}$ will be taken in sequence, in a non-decreasing order, and at each stage the number of ways to select the intersection of U'' with the current node, given the previously selected portion of U'', will be determined. To begin this process, consider the selection of $U \cap U' \cap U''$, which must be an $(m-h)$-subspace of the $(m-1)$-subspace $U \cap U'$. The number of choices for this is simply

$$\begin{bmatrix} m-1 \\ m-h \end{bmatrix} = \begin{bmatrix} m-1 \\ h-1 \end{bmatrix} = \frac{(q;q)_{m-1}}{(q;q)_{h-1}\,(q;q)_{m-h}} \tag{1}$$

Now assume that a subspace Z_1 has been selected to play the role of $U \cap U' \cap U''$. Define a mapping between lattices of subspaces as follows:

$$\varphi_1 : \mathcal{L}(V) \to \mathcal{L}(Z_1^{\perp}/Z_1),\ \varphi_1(X) = (Z_1 + Z_1^{\perp} \cap X)/Z_1\ (\cong (Z_1^{\perp} \cap X)/(Z_1 \cap X))$$

In order to select the $U \cap U''$ portion of U'', it suffices by Lemma 2 to select a suitable subspace of the h-subspace $\varphi_1(U)\ (= U/Z_1\,)$ to play the role of $\varphi_1(U \cap U'')\ (= (U \cap U'')/Z_1\,)$, which is required to have dimension $h-j$, and to trivially intersect $\varphi_1(U \cap U')\ (= (U \cap U')/Z_1\,)$. $\varphi_1(U)$ and $\varphi_1(U \cap U')$ have dimensions h and $h-1$, respectively. If $h = j$, then there is nothing to select, and U'' will be required to satisfy $U \cap U'' = Z_1$. But if $h = j+1$, then the number of choices is $\left[{h \atop 1}\right] - \left[{h-1 \atop 1}\right] = q^{h-1} = q^j$. So in general, assuming as we are that h is j or $j+1$, the number of choices for $\varphi_1(U \cap U'')$ is simply

$$q^{j(h-j)} \tag{2}$$

Next, assume that a subspace Z_2 has been selected to play the role of $U \cap U''$ and let $\varphi_2 : \mathcal{L}(V) \to \mathcal{L}(Z_2^{\perp}/Z_2)$ be defined analogous to the definition of φ_1. The next step in the process is to select a subspace of $Z_2 + U'$ to play the role of $U \cap U'' + U' \cap U''$, which must have dimension $(m-j) + (m-i) - (m-h) = m + h - i - j$. Now $\dim(Z_2) = m - j$, and by Lemma 2, there is a bijection between $\{X \in \mathcal{L}(V) \mid X \text{is isotropic}, \dim(X) = m+h-i-j, Z_2 \leq X \leq Z_2 + U', X \cap U \leq Z_2\}$ and the set of isotropic $(h-i)$ - subspaces of $\varphi_2(U')$ that trivially intersect $\varphi_2(U)$. Note that $\varphi_2(U') = \varphi_2(Z_2 + U') \cong U'/(U \cap U' \cap U'')$, which has dimension h. Selecting the desired subspace of $Z_2 + U'$ amounts to selecting an isotropic $(h-i)$ - subspace of $\varphi_2(U')$ that trivially intersects $\varphi_2(U)$. But $\varphi_2(U) \cap \varphi_2(U') = [Z_2^{\perp} \cap (Z_2 + U) \cap (Z_2 + U')]/Z_2 = [Z_2^{\perp} \cap U \cap (Z_2 + U')]/Z_2 = [Z_2^{\perp} \cap (Z_2 + U \cap U')]/Z_2 = \varphi_2(U \cap U') \cong (U \cap U')/(U \cap U' \cap U'')$, which has dimension $h-1$. Note that subspaces of $\varphi_2(U')$ that trivially intersect $\varphi_2(U \cap U')$ also trivially intersect $\varphi_2(U)$. So, similar to the previous count, and assuming h is i or $i+1$, the number of selections for $\varphi_2(U' \cap U'')$ is seen to be

$$q^{i(h-i)} \tag{3}$$

Continuing to the next node of $\mathcal{M}$, assume that a subspace Z_3 has been selected to play the role of $U \cap U'' + U' \cap U''$. If $e = 0$, then this subspace will also be $(U + U') \cap U''$. But if $e = 1$, then $(U + U') \cap U''$ must be one dimension larger. However, for this to be the case, it is necessary for h, i and j to all be equal. To see why, observe that $[(U + U') \cap U'' : U \cap U''] \leq [U + U' : U] = [U' : U \cap U'] = 1$. Similarly, $[(U + U') \cap U'' : U' \cap U''] \leq 1$. So unless $U \cap U'' = U' \cap U''$, it must be the case that $(U + U') \cap U'' = U \cap U'' + U' \cap U''$.

When $h = i = j$ and $e = 1$, and so $Z_1 = Z_2 = Z_3$, the problem of selecting $(U+U')\cap U''$, again by Lemma 2, amounts to selecting a one-dimensional subspace inside the isotropic space $(U+U')/Z_1$ that trivially intersects each of the two hyperplanes U/Z_1 and U'/Z_1. The number of such selections is $\left[{h+1 \atop 1}\right] - 2\left[{h \atop 1}\right] + \left[{h-1 \atop 1}\right] = (q-1)q^{h-1}$.

When $e = 1$, but h, i and j are not all equal, there are of course no suitable choices for $(U+U')\cap U''$. Rather than imposing a further restriction on the parameters, at this point, this possibility will be allowed, and will be handled by saying that in general the number of choices for $(U+U')\cap U''$ is

$$\left[(q-1)q^{h-1}\cdot 0^{2h-i-j}\right]^e \tag{4}$$

where 0^0 is to be interpreted as 1. The "zero notation" here is simply a bookkeeping convenience, and essentially means that either $h = i = j$ or $e = 0$ is required in order to have a nonzero count.

Assume next that some subspace Z_4 has been selected to serve as $(U+U')\cap U''$, and let φ_4 have the evident meaning. Z_4 has dimension $\dim(Z_3) + e = m + h + e - i - j$. Now $U^\perp \cap U'^\perp \cap U''$ can be selected by selecting $\varphi_4(U^\perp \cap U'^\perp \cap U'')\,(= (U^\perp \cap U'^\perp \cap U'')/Z_4\,)$, which must be isotropic and have dimension $(m-\eta)-(m+h+e-i-j) = i+j-h-e-\eta$. Next, note that $\varphi_4(V)\,(=\ Z_4^\perp/Z_4\,)$ has dimension $N - 2m - 2h - 2e + 2i + 2j$, and $\varphi_4(U+U')\,(= (U+U')/Z_4\,)$ has dimension $(m+1)-(m+h+e-i-j) = i+j-h-e+1$. So $\varphi_4(U^\perp\cap U'^\perp)\,(= (U^\perp\cap U'^\perp)/Z_4\,)$ has dimension $N-2m-2h-2e+2i+2j-(i+j-h-e+1) = N-2m-h-e+i+j-1$.

The selection of $\varphi_4(U^\perp\cap U'^\perp\cap U'')$ must be made inside the coisotropic subspace $\varphi_4(U^\perp\cap U'^\perp)$, whose radical is $\varphi_4(U+U')$. The selection must also trivially intersect this latter subspace, and so must project injectively into the quotient space $\varphi_4(U^\perp \cap U'^\perp)/\varphi_4(U + U')$ $(\cong (U^\perp \cap U'^\perp)/(U+U')\,)$. This quotient space has dimension $N-2m-2$. Each isotropic $(i+j-h-e-\eta)$ - subspace of this quotient space lifts to $q^{(i+j-h-e-\eta)(i+j-h-e+1)}$ choices for $\varphi_4(U^\perp\cap U'^\perp\cap U'')$. The total number of selections for this latter subspace, and so for $U^\perp\cap U'^\perp\cap U''$ given the particular Z_4, is therefore

$$\left[\begin{matrix} N-2m-2 \\ i+j-h-e-\eta \end{matrix}\right]_0 \cdot q^{(i+j-h-e-\eta)(i+j-h-e+1)} \tag{5}$$

$$= \frac{(q^{\frac{1}{2}N-m-1-\mu}, -q^{\frac{1}{2}N-m-1-\nu}; q^{-1})_{i+j-h-e-\eta}}{(q;q)_{i+j-h-e-\eta}} \cdot q^{(i+j-h-e-\eta)(i+j-h-e+1)}$$

Now suppose that $Z_5 = U^\perp\cap U'^\perp\cap U''$ has been selected, and define φ_5 in the usual way. $U^\perp\cap U''$ can then be selected by selecting $\varphi_5(U^\perp\cap U'')\,(= (U^\perp\cap U'')/Z_5\,)$, which must have dimension $(m-\beta)-(m-\eta) = \eta-\beta$. This selection must be made inside the coisotropic subspace $\varphi_5(U^\perp)$, whose radical is $\varphi_5(U)\,(\cong U/(U\cap U'')\,)$. It is straightforward to check that $\dim(\varphi_5(V)) = N-2m+2\eta$, $\dim(\varphi_5(U)) = j$, $\dim(\varphi_5(U^\perp)) = N-2m+2\eta-j$. By Lemma 2, it is further necessary and sufficient that the selection of $\varphi_5(U^\perp\cap U'')$ trivially intersect the coisotropic subspace $\varphi_5(U^\perp\cap U'^\perp)\,(= \varphi_5(U^\perp)\cap\varphi_5(U'^\perp)\,)$, which has radical $\varphi_5(U+U')\,(\cong (U+U')/[(U+U')\cap U'']\,)$. Again, it is straightforward to check that

$\dim(\varphi_5(U+U')) = i+j-h-e+1$ and $\dim(\varphi_5(U^\perp \cap U'^\perp)) = N-2m+2\eta-i-j+h+e-1$.

Each suitable choice for $\varphi_5(U^\perp \cap U'')$ projects injectively into the quotient space $\varphi_5(U^\perp)/\varphi_5(U)$ and must trivially intersect $\varphi_5(U^\perp \cap U'^\perp)/\varphi_5(U)$. Note that $[\varphi_5(U^\perp) : \varphi_5(U^\perp \cap U'^\perp)] = [\varphi_5(U+U') : \varphi_5(U)] = i-h-e+1$. This is either 0 or 1. When it is 1, that is, when $i = h$ and $e = 0$, and when $\eta = \beta + 1$, then it is necessary to select a one-dimensional isotropic subspace of $\varphi_5(U^\perp)/\varphi_5(U)$ that trivially intersects the coisotopic hyperplane $\varphi_5(U^\perp \cap U'^\perp)/\varphi_5(U)$. Assume that this is the case for the moment. Note that $\dim[\varphi_5(U^\perp)/\varphi_5(U)] = N-2m-2j+2\eta$. By Lemma 4, the number of such 1-subspaces is $q^{N-2m-2j+2\eta-1-\mu-\nu}$. Each of these lifts to q^j ($= \dim(\varphi_5(U))$) possible choices for $\varphi_5(U^\perp \cap U'')$. Putting this together, we see in general that the number of choices for $U^\perp \cap U''$, given a particular Z_5, is

$$\left[q^{N-2m-j+2\eta-1-\mu-\nu} \cdot 0^{h-i+e}\right]^{\eta-\beta}$$

which, because η is either β or $\beta+1$, equals

$$\left[q^{N-2m-j+\alpha+\beta-\mu-\nu} \cdot 0^{h-i+e}\right]^{\eta-\beta} \cdot q^{(\eta-\alpha)(\eta-\beta)} \tag{6}$$

Next, assume that $Z_6 = U^\perp \cap U''$ has been selected. $U^\perp \cap U'' + U'^\perp \cap U''$ can then be selected by selecting $\varphi_6(U'^\perp \cap U'')$, where φ_6 is defined as usual. Now $\varphi_6(U'^\perp \cap U'')\,(= (U^\perp \cap U'' + U'^\perp \cap U'')/(U^\perp \cap U'') \cong (U'^\perp \cap U'')/(U^\perp \cap U'^\perp \cap U''))$ must have dimension $(m-\alpha)-(m-\eta) = \eta-\alpha$, and will be selected in a manner similar to that used previously to select Z_6.

Observe that $\varphi_6(U') = (Z_6 + Z_6^\perp \cap U')/Z_6 \cong (Z_6^\perp \cap U')/(U' \cap U'')$, which has dimension $i-\eta+\beta$, because $\dim(Z_6^\perp \cap U') = N$ - $\dim(Z_6+U'^\perp) = N$ - $\dim(Z_6)$ - $\dim(U'^\perp) + \dim(Z_5)$ $= N-(m-\beta)-(N-m)+(m-\eta) = m-\eta+\beta$. Now, $\dim(\varphi_6(V)) = N-2m+2\beta$, so $\dim[\varphi_6(U'^\perp)/\varphi_6(U')] = N-2m-2i+2\eta$.

Note too that $\varphi_6(U)+\varphi_6(U') = \varphi_6(U+U') = [Z_6+Z_6^\perp \cap (U+U')]/Z_6 \cong [Z_6^\perp \cap (U+U')]/Z_4$. This has dimension $i+j-h-e-\eta+\beta+1$, since $\dim[Z_6^\perp \cap (U+U')] = N$ - $\dim[Z_6+(U^\perp \cap U'^\perp)]$ $= N$ - $\dim(Z_6)$ - $\dim(U^\perp \cap U'^\perp) + \dim(U^\perp \cap Z_5) = N-(m-\beta)-(N-m-1)+(m-\eta) = m-\eta+\beta+1$, and since $\dim(Z_4) = m+h+e-i-j$. It follows that $[\varphi_6(U'^\perp) : \varphi_6(U^\perp) \cap \varphi_6(U'^\perp)] = [\varphi_6(U'^\perp) : \varphi_6(U^\perp \cap U'^\perp)] = [\varphi_6(U+U') : \varphi_6(U')] = j-h-e+1$, which is either 0 or 1, and is 1 only when $h = j$ and $e = 0$.

By the same reasoning that was applied to select Z_6, it can now be seen that the number of choices for $U^\perp \cap U'' + U'^\perp \cap U''$, given the choice for Z_6, is

$$\left[q^{N-2m-i+\alpha+\beta-\mu-\nu} \cdot 0^{h-j+e}\right]^{\eta-\alpha} \tag{7}$$

since $\dim[\varphi_6(U'^\perp)/\varphi_6(U')] = N-2m-2i+2\eta$ and $\dim(\varphi_6(U')) = i-\eta+\beta$.

So now assume that $Z_7 = U^\perp \cap U'' + U'^\perp \cap U''$ has been selected. If $\varepsilon = 0$ then this is also $(U^\perp + U'^\perp) \cap U''$. But if $\varepsilon = 1$, then $(U^\perp + U'^\perp) \cap U''$ needs to be selected to be one dimension larger. However, similar to the situation when selecting $(U+U') \cap U''$, this is only possible when $\eta = \alpha = \beta$, since $[U^\perp + U'^\perp : U^\perp]$ and $[U^\perp + U'^\perp : U'^\perp]$ are both one. Assume for the time being that $\eta = \alpha = \beta$ and $\varepsilon = 1$. Thus $Z_5 = Z_6 = Z_7$. Consider selecting the isotropic 1-subspace $\varphi_5((U^\perp + U'^\perp) \cap U'')$, which must be chosen inside $\varphi_5(U^\perp + U'^\perp)$, and, by Lemma 2, must trivially intersect both $\varphi_5(U^\perp)$ and $\varphi_5(U'^\perp)$, and any such selection lifts to an acceptable selection for $(U^\perp + U'^\perp) \cap U''$.

Recall that $\dim(\varphi_5(V)) = N - 2m + 2\eta$. Notice that $\varphi_5(U) + \varphi_5(U') = \varphi_5(U + U') \cong (U + U')/[(U + U') \cap U'']$, which has dimension $i + j - h - e + 1$. Also, $\varphi_5(U \cap U') \cong (U \cap U')/(U \cap U' \cap U'')$, which has dimension $h - 1$. However, $\dim[\varphi_5(U) \cap \varphi_5(U')] = \dim[\varphi_5(U)] + \dim[\varphi_5(U')]$ - $\dim[\varphi_5(U+U')] = j+i-(i+j-h-e+1) = h+e-1$. So in general, $\varphi_5(U\cap U')$ and $\varphi_5(U)\cap\varphi_5(U')$ are not the same, and in fact $[\varphi_5(U)\cap\varphi_5(U') : \varphi_5(U\cap U')] = e$. It follows that $[\varphi_5(U^\perp + U'^\perp) : \varphi_5(U^\perp) + \varphi_5(U'^\perp)] = e$.

Also note that $[\varphi_5(U^\perp + U'^\perp) : \varphi_5(U^\perp)] = [\varphi_5(U) : \varphi_5(U \cap U')] = j - h + 1$ and $[\varphi_5(U^\perp + U'^\perp) : \varphi_5(U'^\perp)] = [\varphi_5(U') : \varphi_5(U \cap U')] = i - h + 1$. Clearly $h = i = j$ is required in order to be able to select the required 1-subspace. So assume this as well for the time being. It is straightforward now to check that $\varphi_5(U^\perp) = \varphi_5(U'^\perp)$ if and only if $\varphi_5(U) = \varphi_5(U')$ if and only if $e = 1$.

In the $e = 1$ subcase, we are dealing with the same sort of enumeration problem considered earlier. Specifically, we must select an isotropic 1-subspace in the $(N-2m+2\eta-2h+2)$ - dimensional quotient space $\varphi_5(U^\perp + U'^\perp)/\varphi_5(U \cap U')$, one that trivially intersects the corresponding coisotropic hyperplane $\varphi_5(U^\perp)/\varphi_5(U \cap U')$. This then needs to be lifted to one of q^{h-1} choices for $\varphi_5((U^\perp + U'^\perp) \cap U'')$. By Lemma 4 or 5, the number of choices for $(U^\perp + U'^\perp) \cap U''$ is $q^{N-2m+2\eta-2h+2-1-\mu-\nu} \cdot q^{h-1} = q^{N-2m-h+2\eta-\mu-\nu}$. The $e = 0$ subcase is similar but requires trivially intersecting two coisotropic hyperplanes, $\varphi_5(U^\perp)/\varphi_5(U \cap U')$ and $\varphi_5(U'^\perp)/\varphi_5(U \cap U')$. By Lemma 5, the number of choices for $(U^\perp + U'^\perp) \cap U''$ is $(q - 1)\, q^{N-2m-h+2\eta-1-\mu-\nu}$.

Combining the subcases, we see that in general the number of choices for $(U^\perp+U'^\perp)\cap U''$, given a particular choice for Z_7, is

$$\left[(q-1)^{1-e} \cdot q^{N-2m-h+2\eta+e-1-\mu-\nu} \cdot 0^{2h-i-j+2\eta-\alpha-\beta}\right]^{\varepsilon} \tag{8}$$

As expected, suppose next that $Z_8 = (U^\perp + U'^\perp) \cap U''$ has been selected, and let φ_8 be the evident mapping.

The final task in selecting U'' is to extend Z_8 to all of U''. This can be accomplished by selecting $\varphi_8(U'')\,(= U''/Z_8\,)$, which must have dimension $\alpha + \beta - \eta - \varepsilon$, since $\dim(Z_8) = \dim(Z_7) + \varepsilon = (m - \beta) + (m - \alpha) - (m - \eta) + \varepsilon = m + \eta - \alpha - \beta + \varepsilon$. It must also trivially intersect the coisotropic subspace $\varphi_8(U^\perp + U'^\perp)$. The radical of this coisotropic subspace is $\varphi_8(U\cap U')\,(= (Z_8 + U\cap U')/Z_8 \cong (U\cap U')/(U\cap U'\cap U'')\,)$, which has dimension $(m - 1) - (m - h) = h - 1$. Note that $\dim(\varphi_8(V)) = N - 2m - 2\eta + 2\alpha + 2\beta - 2\varepsilon$. By Lemma 4, the number of choices for $\varphi_8(U'')$ is

$$\begin{bmatrix} h-1 \\ \alpha+\beta-\eta-\varepsilon \end{bmatrix} \cdot q^{(\alpha+\beta-\eta-\varepsilon)[(N-2m-2\eta+2\alpha+2\beta-2\varepsilon)-(h-1)-\frac{1}{2}(\alpha+\beta-\eta-\varepsilon)+\frac{1}{2}-\mu-\nu]}$$

$$(9) \qquad = \frac{(q;q)_{h-1}}{(q;q)_{\alpha+\beta-\eta-\varepsilon}\,(q;q)_{h-\alpha-\beta+\eta+\varepsilon-1}} \cdot q^{(\alpha+\beta-\eta-\varepsilon)[N-2m-h+\frac{3}{2}(\alpha+\beta-\eta-\varepsilon+1)-\mu-\nu]}$$

The total number of allowable choices for U'' is of course obtained by multiplying the various counts ((1) through (9)) associated with the various stages in the above process used to select this subspace. Furthermore, each of the five conditions (A through E) listed in the theorem is sufficient to ensure that the exponents in all of the "zero notations" are zero. Conversely, it is straightforward to check that at least one of these conditions must hold in order for all of these exponents to be zero.

□

The next corollary offers a slightly different expression of the count in Theorem 1, in an effort to produce a more uniform approach to enumerating the values of $p^{10}_{i_\alpha j_\beta}$ for various cases, some of which are considered in Corollary 2.

Corollary 1. *The count in Theorem 1 can also be expressed as*

$$\frac{(q;q)_{m-1}\,(q^{d-m-1},-q^{d'-m-1};q^{-1})_{g-\kappa-r-\rho-e}}{(q;q)_{m-f-r}\,(q;q)_{\lambda-\rho-\varepsilon}\,(q;q)_{f-\lambda-1+r+\rho+\varepsilon}\,(q;q)_{g-\kappa-r-\rho-e}} \cdot (q-1)^{e+\varepsilon-e\varepsilon} \cdot q\wedge$$

$$\{\,[fg+g+d\kappa+d'\kappa-2m\kappa-2g\kappa+g\lambda-f\lambda+\kappa^2+\tfrac{1}{2}\lambda^2-\kappa+\tfrac{3}{2}\lambda]+4r\rho+r(f-g+2\kappa-2\lambda)$$
$$+\rho(d+d'-2m+f-3g+3\kappa-2\lambda)+e\varepsilon+e(f-2g+\kappa-1)+\varepsilon(2\kappa-3\lambda-1)\,\}\,,$$

assuming that at least one of the five conditions in Theorem 1 holds, and so either $e=\varepsilon=0$ *or* $r=\rho=0$*, where*

$$r = \begin{cases} 0 & \text{if } h=i=j \\ 1 & \text{otherwise,} \end{cases} \qquad \rho = \begin{cases} 0 & \text{if } \eta=\alpha=\beta \\ 1 & \text{otherwise,} \end{cases}$$

and where $d = N/2-\mu$, $d' = N/2-\nu$, $f=\min\{i,j\}$, $g=\max\{i,j\}$, $\kappa=\min\{\alpha,\beta\}$ *and* $\lambda=\max\{\alpha,\beta\}$.

Proof. First observe that $h=f+r$, $i+j-h=g-r$, $\eta=\kappa+\rho$ and $\alpha+\beta-\eta=\lambda-\rho$. The formula given in Theorem 1 is easily converted to the one given here, except that a portion of the power of q there needs to be handled in a special way. Specifically, $-i(\eta-\alpha)-j(\eta-\beta)$ can be seen to equal $(g-r)(\lambda-\kappa-2\rho)$, by the following argument. In order for $\eta-\alpha$ to be nonzero, it must be that $h=j$, since one of the conditions in Theorem 1 is presumably satisfied. In this case, $i=i+j-h=g-r$. Likewise, in order for $\eta-\beta$ to be nonzero, it must be that $j=g-r$. So $-i(\eta-\alpha)-j(\eta-\beta) = -(g-r)(2\eta-\alpha-\beta) = (g-r)(\lambda-\kappa-2\rho)$.

□

For given values of i,j,α and β, the numbers $p^{10}_{i_\alpha j_\beta}$ can now be computed by summing the count in Corollary 1 over the allowable values of r,e,ρ and ε, each of which must be either 0 or 1. The case where $0<\alpha=\beta<i=j<m$ involves the maximum number of terms (six), because here each of e,ε,r and ρ can be either 0 or 1, restricted only by the requirements that r and ρ cannot both be 1, and that either $e=\varepsilon=0$ or $r=\rho=0$. The need for these restrictions can be seen from the restrictions listed in Theorem 1 applied in the context that $i=j$ and $\alpha=\beta$. The other cases for $p^{10}_{i_\alpha j_\beta}$ involve fewer than six terms, and in fact if $i\neq j$ or $\alpha\neq\beta$, then there is just a single term to consider. Most of the numbers $p^{10}_{i_\alpha j_\beta}$ shall now be given explicitly in the next corollary and a subsequent remark.

Corollary 2. *Assuming that* $0 < \alpha < i < m$,

$$p^{1_0}_{i_\alpha i_\alpha} = \frac{(q;q)_{m-1}\,(q^{d-m-1}, -q^{d'-m-1}; q^{-1})_{i-\alpha-1}}{(q;q)_{m-i}\,(q;q)_\alpha\,(q;q)^2_{i-\alpha}} \cdot q^{i^2-i+d\alpha+d'\alpha-2m\alpha-2i\alpha+\frac{3}{2}\alpha^2-\frac{1}{2}\alpha-1} \cdot X,$$

where

$$\begin{gathered} X = q^{d+d'-2m}\,(2q^{i+2\alpha+1} - q^{i+\alpha+1} - q^{3\alpha+1} - q^{3\alpha} + q^{2\alpha}) \\ + (q^{d-m} - q^{d'-m})\,(q^{2i+\alpha+1} - q^{i+2\alpha} - q^{i+\alpha+1} + q^{i+\alpha}) \\ + q^{m+2i+1} - q^{m+i+\alpha+1} - q^{3i+1} - q^{3i} + 2q^{2i+\alpha} + q^{i+\alpha+1} - q^{i+\alpha}). \end{gathered}$$

Also,

$$p^{1_0}_{i_\alpha i_{\alpha-1}} = \frac{(q;q)_{m-1}\,(q^{d-m-1}, -q^{d'-m-1}; q^{-1})_{i-\alpha}}{(q;q)_{m-i}\,(q;q)_{\alpha-1}\,(q;q)^2_{i-\alpha}} \cdot q^{i^2+i+d\alpha+d'\alpha-2m\alpha-2i\alpha+\frac{3}{2}\alpha^2-\frac{1}{2}\alpha-1},$$

$$p^{1_0}_{i_\alpha (i-1)_\alpha} = \frac{(q;q)_{m-1}\,(q^{d-m-1}, -q^{d'-m-1}; q^{-1})_{i-\alpha-1}}{(q;q)_{m-i}\,(q;q)_\alpha\,(q;q)^2_{i-\alpha-1}} \cdot q^{i^2+d\alpha+d'\alpha-2m\alpha-2i\alpha+\frac{3}{2}\alpha^2+\frac{3}{2}\alpha-1},$$

$$p^{1_0}_{i_\alpha (i-1)_{\alpha-1}} = \frac{(q;q)_{m-1}\,(q^{d-m-1}, -q^{d'-m-1}; q^{-1})_{i-\alpha-1}}{(q;q)_{m-i}\,(q;q)_{\alpha-1}\,(q;q)_{i-\alpha}\,(q;q)_{i-\alpha-1}} \cdot q^{i^2+d\alpha+d'\alpha-2m\alpha-2i\alpha+\frac{3}{2}\alpha^2+\frac{1}{2}\alpha-1}.$$

Therefore,

$$p_{i_\alpha i_\alpha 1_0} = \frac{(q^d, q^{d'}; q^{-1})_{m+i-\alpha}}{(1-q)^2\,(q;q)_{m-i}\,(q;q)_\alpha\,(q;q)^2_{i-\alpha}} \cdot q^{i^2-i+d\alpha+d'\alpha-2m\alpha-2i\alpha+\frac{3}{2}\alpha^2-\frac{1}{2}\alpha} \cdot X,$$

$$p_{i_\alpha i_{\alpha-1} 1_0} = \frac{(q^d, q^{d'}; q^{-1})_{m+i-\alpha+1}}{(1-q)^2\,(q;q)_{m-i}\,(q;q)_{\alpha-1}\,(q;q)^2_{i-\alpha}} \cdot q^{i^2+i+d\alpha+d'\alpha-2m\alpha-2i\alpha+\frac{3}{2}\alpha^2-\frac{1}{2}\alpha},$$

$$p_{i_\alpha (i-1)_\alpha 1_0} = \frac{(q^d, q^{d'}; q^{-1})_{m+i-\alpha}}{(1-q)^2\,(q;q)_{m-i}\,(q;q)_\alpha\,(q;q)^2_{i-\alpha-1}} \cdot q^{i^2+d\alpha+d'\alpha-2m\alpha-2i\alpha+\frac{3}{2}\alpha^2+\frac{3}{2}\alpha},$$

$$p_{i_\alpha (i-1)_{\alpha-1} 1_0} = \frac{(q^d, q^{d'}; q^{-1})_{m+i-\alpha}}{(1-q)^2\,(q;q)_{m-i}\,(q;q)_{\alpha-1}\,(q;q)_{i-\alpha}\,(q;q)_{i-\alpha-1}} \cdot q^{i^2+d\alpha+d'\alpha-2m\alpha-2i\alpha+\frac{3}{2}\alpha^2+\frac{1}{2}\alpha}.$$

Proof. Using Corollary 1 to compute $p^{1_0}_{i_\alpha i_\alpha}$, we obtain

$$\sum_{r,e,\rho,\varepsilon} \frac{(q;q)_{m-1}\,(q^{d-m-1}, -q^{d'-m-1}; q^{-1})_{i-\alpha-r-\rho-e}}{(q;q)_{m-i-r}\,(q;q)_{\alpha-\rho-\varepsilon}\,(q;q)_{i-\alpha-1+r+\rho+\varepsilon}\,(q;q)_{i-\alpha-r-\rho-e}} \cdot (q-1)^{e+\varepsilon-e\varepsilon} \cdot q\wedge$$

$$\{\, (i^2+i+d\alpha+d'\alpha-2m\alpha-2i\alpha+\tfrac{3}{2}\alpha^2+\tfrac{1}{2}\alpha)+4r\rho+\rho(u+v-\alpha)+e\varepsilon-e(i-\alpha+1)-\varepsilon(\alpha+1)\,\}.$$

This can be rewritten in the form

$$\sum_{r,e,\rho,\varepsilon} \frac{(q;q)_{m-1}\,(q^{d-m-1}, -q^{d'-m-1}; q^{-1})_{i-\alpha-1}}{(q;q)_{m-i}\,(q;q)_\alpha\,(q;q)^2_{i-\alpha}} \cdot q^{i^2+i+d\alpha+d'\alpha-2m\alpha-2i\alpha+\frac{3}{2}\alpha^2+\frac{1}{2}\alpha}$$

$$\cdot \Big\{ \big[(1-q^{d-m-i+\alpha})(1+q^{d'-m-i+\alpha})\big]^{1-r\vee\rho\vee e}\,(1-q^\alpha)^{\rho\vee\varepsilon}\,(1-q^{m-i})^r\,(1-q^{i-\alpha})^{1+e-\varepsilon}$$

$$\cdot\, (q-1)^{e\vee\varepsilon} \;\cdot\; q^{4r\rho+\rho(u+v-\alpha)+e\varepsilon-e(i-\alpha+1)-\varepsilon(\alpha+1)} \Big\},$$

where the operation $\vee$ is defined by $\sigma \vee \tau = \sigma + \tau - \sigma\tau$. Symbolic manipulation software, such as Mathematica[1] and Maple[2], can then be used to quickly establish that the sum of the quantity in curly braces here, over the allowable values of r, e, ρ and ε, equals $q^{-2i-\alpha-1}$ times X.

The numbers $p^{1_0}_{i_\alpha i_{\alpha-1}}, p^{1_0}_{i_\alpha (i-1)_\alpha}$ and $p^{1_0}_{i_\alpha (i-1)_{\alpha-1}}$ can likewise be computed using Corollary 1. Because of the restrictions imposed by Theorem 1 on the parameters $(r, \rho, e, \varepsilon)$, only one term is nonzero in each of these cases. The only nonzero term in the computation of $p^{1_0}_{i_\alpha i_{\alpha-1}}$ occurs when $\rho = 1$, $r = e = \varepsilon = 0$. The only nonzero term in the computation of $p^{1_0}_{i_\alpha (i-1)_\alpha}$ occurs when $r = 1$, $\rho = e = \varepsilon = 0$. The only nonzero term in the computation of $p^{1_0}_{i_\alpha (i-1)_{\alpha-1}}$ occurs when $r = \rho = 1$, $e = \varepsilon = 0$. The remaining claims follows from Lemma 6 and the fact that $p_{i_\alpha j_\beta 1_0} = |\mathcal{R}_{1_0}|\, p^{1_0}_{i_\alpha j_\beta}$.

□

Remark 2. *Using the values from Corollary 2, it is straightforward, possibly with the aid of symbolic manipulation software, to check that*

$$\sum_{j_\beta} p^{1_0}_{i_\alpha j_\beta} = p^{1_0}_{i_\alpha i_\alpha} + p^{1_0}_{i_\alpha i_{\alpha-1}} + p^{1_0}_{i_\alpha i_{\alpha+1}} + p^{1_0}_{i_\alpha (i-1)_\alpha} + p^{1_0}_{i_\alpha (i+1)_\alpha} + p^{1_0}_{i_\alpha (i-1)_{\alpha-1}} + p^{1_0}_{i_\alpha (i+1)_{\alpha+1}}$$

$$= \frac{(q;q)_{m-1}\,(q^{d-m-1}, -q^{d'-m-1}; q^{-1})_{i-\alpha-1}}{(q;q)_{m-i}\,(q;q)_\alpha\,(q;q)^2_{i-\alpha}} \cdot q^{i^2-i+d\alpha+d'\alpha-2m\alpha-2i\alpha+\frac{3}{2}\alpha^2-\frac{1}{2}\alpha-1}$$

$$\cdot \Big\{ q^{d+d'-2m}(2q^{i+2\alpha+1} - q^{i+\alpha+1} - q^{3\alpha+1} - q^{3\alpha} + q^{2\alpha})$$

$$+ (q^{d-m} - q^{d'-m})(q^{2i+\alpha+1} - q^{i+2\alpha} - q^{i+\alpha+1} + q^{i+\alpha})$$

$$+ (q^{m+2i+1} - q^{m+i+\alpha+1} - q^{3i+1} - q^{3i} + 2q^{2i+\alpha} + q^{i+\alpha+1} - q^{i+\alpha})$$

$$+ (1-q^{d-m-i+\alpha})(1+q^{d'-m-i+\alpha})(1-q^\alpha)q^{2i} + (1-q^{i-\alpha})^2 q^{d+d'-2m+3\alpha+1}$$

$$+ (1-q^{i-\alpha})^2 q^{i+2\alpha} + (1-q^{d-m-i+\alpha})(1+q^{d'-m-i+\alpha})(1-q^{m-i})q^{3i+1}$$

$$+ (1-q^\alpha)(1-q^{i-\alpha})q^{i+\alpha} + (1-q^{m-i})(1-q^{i-\alpha})q^{d+d'-2m+i+2\alpha+1} \Big\}$$

$$= \frac{(q;q)_m\,(q^{d-m}, -q^{d'-m}; q^{-1})_{i-\alpha}}{(q;q)_{m-i}\,(q;q)_\alpha\,(q;q)^2_{i-\alpha}} \cdot q^{i^2+d\alpha+d'\alpha-2m\alpha-2i\alpha+\frac{3}{2}\alpha^2+\frac{1}{2}\alpha} = \Big|\mathcal{R}_{i_\alpha}\Big| \Big/ \begin{bmatrix} N \\ m \end{bmatrix}_0 .$$

This is as required, since the summation must count all of the isotropic m-subspaces U'' for which $(U', U'') \in \mathcal{R}_{i_\alpha}$, given an isotropic m-subspace U'.

[1] *Mathematica* is a registered trademark of Wolfram Research, Inc.
[2] *Maple* is a registered trademark of Waterloo Maple Inc.

Notice that Corollary 2 and Remark 2 pertain to the intersection numbers $p^{1_0}_{i_\alpha j_\beta}$ for which $0 < \alpha < i < m$. It is staightforward to use Theorem 1 or Corollary 1 and to reason as in the proof of Corollary 2, in order to compute the remaining intersection numbers. Such computations can then be double checked along the lines of Remark 2. Some of these remaining intersection numbers are as follows, as the reader can verify.

Remark 3. *Assume that $0 < i < m$. Then*

(1) *the only nonzero cases of $p^{1_0}_{i_0 j_\beta}$ are*

$$p^{1_0}_{i_0 i_0} = \frac{(q;q)_{m-1}(q^{d-m-1}, -q^{d-m-1}; q^{-1})_{i-1}}{(q;q)_{m-i}(q;q)_i(q;q)_{i-1}} \cdot q^{i^2-1}$$

$$\cdot \Big[(1-q^{d-m-1})(1+q^{d'-m-1})q^{i+1} + (1-q^{m-i})q^{i+1} + (1-q^i)(q-1)\Big],$$

$$p^{1_0}_{i_0 i_1} = \frac{(q;q)_{m-1}(q^{d-m-1}, -q^{d-m-1}; q^{-1})_{i-1}}{(q;q)_{m-i}(q;q)^2_{i-1}} \cdot q^{d+d'-2m+i^2-i},$$

$$p^{1_0}_{i_0 (i+1)_0} = \frac{(q;q)_{m-1}(q^{d-m-1}, -q^{d-m-1}; q^{-1})_{i}}{(q;q)_{m-i-1}(q;q)^2_{i}} \cdot q^{i^2+2i},$$

$$p^{1_0}_{i_0 (i+1)_1} = \frac{(q;q)_{m-1}(q^{d-m-1}, -q^{d-m-1}; q^{-1})_{i-1}}{(q;q)_{m-i-1}(q;q)_i(q;q)_{i-1}} \cdot q^{d+d'-2m+i^2},$$

$$p^{1_0}_{i_0 (i-1)_0} = \frac{(q;q)_{m-1}(q^{d-m-1}, -q^{d-m-1}; q^{-1})_{i-1}}{(q;q)_{m-i}(q;q)^2_{i-1}} \cdot q^{i^2-1} \quad \textit{and}$$

(2) *the only nonzero cases of $p^{1_0}_{i_i j_\beta}$ are*

$$p^{1_0}_{i_i i_i} = \frac{(q;q)_{m-1}}{(q;q)_{m-i}(q;q)_{i-1}} \cdot (q-1) \cdot q^{di+d'i-2mi+\frac{1}{2}i^2+\frac{1}{2}i-1},$$

$$p^{1_0}_{i_i (i+1)_i} = \frac{(q;q)_{m-1}}{(q;q)_{m-i-1}(q;q)_{i}} \cdot q^{di+d'i-2mi+\frac{1}{2}i^2+\frac{3}{2}i},$$

$$p^{1_0}_{i_i i_{i-1}} = \frac{(q;q)_{m-1}}{(q;q)_{m-i}(q;q)_{i-1}} \cdot q^{di+d'i-2mi+\frac{1}{2}i^2+\frac{1}{2}i-1}.$$

As a final thought, it might be noted that the partial results obtained above concerning the intersection numbers of the association schemes being discussed might yield partial results concerning the eigenvalues of these association schemes. Presumably, for each $\mathcal{R}_{i_\alpha}$, the Bose-Mesner algebra of the association scheme discussed in this article has a corresponding "natural" idempotent E_{i_α}. Let F_{i_α} denote its rank. Of course each $\mathcal{R}_{i_\alpha}$ has a corresponding adjacency matrix A_{i_α} in the Bose-Mesner algebra. The eigenvalue matrix $(P_{i_\alpha j_\beta})$ is defined by $A_{i_\alpha} = \Sigma_{j_\beta} P_{i_\alpha j_\beta} E_{j_\beta}$, and its entries (the eigenvalues) are known to be related to the intersection numbers by the formula $p_{i_\alpha j_\beta k_\gamma} = \Sigma_{t_\omega} P_{i_\alpha t_\omega} P_{j_\beta t_\omega} P_{k_\gamma t_\omega} F_{t_\omega}$. Perhaps it is possible at this point to gain some information about the eigenvalues from a knowledge of the numbers $p_{i_\alpha j_\beta 1_0}$.

Acknowledgements

The author is grateful to the referees for several helpful suggestions that improved the clarity and completeness of this article. The translation of the abstract into French was provided by Professor Susan Hanson at Drake University, and initiated by Professor Marylin Mell, and for this, the author is also very appreciative.

References

[1] Artin, E. , *Geometric Algebra*, Interscience Publishers (Wiley), New York, 1957.

[2] Bannai, E., and Ito, T., *Algebraic Combinatorics I, Association Schemes*, Benjamin/Cummings Lecture Notes Series **58**, London, 1984.

[3] Birkhoff, G., *Lattice Theory*, 3rd ed., American Mathematical Society, Providence, 1967.

[4] Brouwer, A. E., Cohen, A. M., Neumaier, A., *Distance-Regular Graphs*, Springer-Verlag, New York, 1989.

[5] Dembowski, P., *Finite Geometries*, Springer-Verlag, Berlin, 1968.

[6] Dieudonné, J., *La Géométrie des Groupes Classiques*, 2nd ed., Springer-Verlag, Berlin, 1963.

[7] Fine, N. J., *Basic Hypergeometric Series and Applications*, Mathematical Surveys and Monographs **27**, AMS, Providence, 1988.

[8] Jacobson, N., *Basic Algebra, Volume One*, 2nd. ed., W. H. Freeman, New York, 1985.

[9] Kaplansky, I., *Linear Algebra and Geometry*, Chelsea, New York, 1969.

[10] Rieck, M. Q., *Ph.D. Dissertation: Pseudo-Orthogonal Complementary Subspaces and Hyperbolic Partner Graphs, U. of South Florida, 1996.*

[11] Rieck, M. Q., "Totally isotopic subspaces, subspace complements and generalized inverses", *Linear Algebra Appl.* **251**(1997) 239–248.

[12] Rieck, M. Q., "Counting configurations of vectors in a finite vector space with an orthogonal, symplectic or unitary geometry", *JCMCC* **34**(2000) 23–32.

[13] Rieck, M. Q., "Maximal orthogonality and pseudo-orthogonality with applications to generalized inverses", *Linear Algebra Appl.* **315**(2000) 155–173.

[14] Stanton, D., "Some q-Krawtchouk polynomials on Chevalley groups", *Amer. J. Math.* **102**(1980) 625–662.

Department of Mathematics and Computer Science, Drake University, 2507 University Avenue, Des Moines, Iowa 50311-4505, USA

E-mail address: Michael.Rieck@Drake.edu

A SCHENSTED INSERTION FOR TENSOR POWERS OF THE WEIL REPRESENTATION

TOM ROBY

Abstract. We give the first known example of a Robinson-Schensted type insertion for a class of infinite dimensional representations. The tensor powers of the Weil representation W of $\mathfrak{sp}(2n,\mathbb{C})$ decompose as a direct sum of certain highest weight $\mathfrak{sp}(2n,\mathbb{C})$-modules $L_k(\lambda)$ tensored with some corresponding finite dimensional irreducible $O(k,\mathbb{C})$-module V_λ. Concentrating only on the $\mathfrak{sp}(2n,\mathbb{C})$-module structure, we can see this as iterating the following decomposition:

$$L_k(\lambda) \otimes W \simeq \bigoplus_{\mu} L_{k+1}(\mu)\,. \tag{2}$$

If k is sufficently large relative to n and λ, then the module $L_k(\lambda)$ and all the terms on the right hand side belong to the holomorphic discrete series and our insertion algorithm allows us to give a weighted bijection proving the formal character identity corresponding to the decomposition (2). It seems likely that a correspondence along these lines can be given to combinatorially explain similar identities of formal characters for small k as well. We give a such a correspondence when $n = 2$, but this remains open for general n.

Résumé. On donne le premier exemple d'une insertion de type Robinson-Schensted pour une classe representations de dimension infinies. La puissance tensorielle de la representation de Weil decomposee en somme directe de certains $\mathfrak{sp}(2n,\mathbb{C})$-modules de highest weight est tensoree avec des $O(k,\mathbb{C})$-modules V_λ irreducibles de dimensions finies. En etudiant seulement les structures des $\mathfrak{sp}(2n,\mathbb{C})$-module on peut obtenir ces resultats par iteration de la decomposition suivante

$$L_k(\lambda) \otimes W \simeq \bigoplus_{\mu} L_{k+1}(\mu)\,. \tag{2}$$

Si k est suffisament grand par rapport a n et λ alors le module $L_k(\lambda)$ et tous les termes de droite appartiennent a la serie discrete holomorphe et notre algorithm d'insertion permet de donner une weighted bijection demontrant ainsi l'indentite du caractere formel correspondant a la decomposition (2). Il semblerait probable qu'une correspondence similaire pourait etre donnee de facon a donner une expliquation combinatoire sur les identites des caracteres formels pour petit k. Ceci est a l'heure actuelle un projet. On donne un tel correspondence pour $n = 2$, mais le problème reste ouvert pour $n > 2$.

1. Introduction

Our goal is to extend the notion of tableaux, insertion and the Robinson-Schensted correspondence (or the R-S correspondence for short) to a class of infinite-dimensional representations of the Lie algebra $\mathfrak{g} = \mathfrak{sp}(2n,\mathbb{C})$. (Here $n \geq 2$ is a fixed positive integer; see § 2 for definitions of this and terms to follow.) The Weil representation of $\mathfrak{g}$ is an infinite-dimensional representation which can be realized on the space of polynomials $W = \mathbb{C}[x_1, x_2, \ldots, x_n]$ using multiplication operators and certain simple differential operators. We can reformulate results of M. Kashiwara and M. Vergne ([KV]) (see also the work of R. Howe [How]) to show that the centralizer of the action of $\mathfrak{g}$ on the k-th tensor power

The author gratefully acknowledges support from National Science Foundation Grant #DMS–9353149, the University of Wisconsin, Madison and California State University, Hayward.

$W^{\otimes k}$ of W is given by an action of the orthogonal group $O(k,\mathbb{C})$, and that they constitute a dual pair, namely the space $W^{\otimes k}$ decomposes into a multiplicity-free sum:

$$W^{\otimes k} \simeq \bigoplus_{\lambda} L_k(\lambda) \otimes V_\lambda, \tag{1}$$

where each $L_k(\lambda)$ is an irreducible $\mathfrak{g}$-module, each V_λ is an irreducible $O(k,\mathbb{C})$-module (using the notation in [KV]), and the summation is for $\lambda \in \mathcal{L}_k$ where $\mathcal{L}_k = \{\, \lambda \in \mathbb{Y} \mid l(\lambda) \leq n,\ \lambda'_1 + \lambda'_2 \leq k \,\}$. Here $\mathbb{Y}$ denotes Young's lattice, the lattice of all partitions of nonnegative integers (ordered by inclusion). These $L_k(\lambda)$ all turn out to be irreducible highest weight modules (for a fixed choice of Cartan subalgebra $\mathfrak{h}$ and Borel subalgebra $\mathfrak{b}$). The highest weight of $L_k(\lambda)$ is

$$\begin{aligned} \Lambda(k,\lambda) &= -(\frac{k}{2}+\lambda_n)\epsilon_1 - (\frac{k}{2}+\lambda_{n-1})\epsilon_2 - \ldots - (\frac{k}{2}+\lambda_1)\epsilon_n \\ &= -\frac{k}{2}\omega_n - \lambda_n\epsilon_1 - \lambda_{n-1}\epsilon_2 - \cdots - \lambda_1\epsilon_n\,, \end{aligned}$$

where $\omega_n = \epsilon_1 + \epsilon_2 + \cdots + \epsilon_n$.

The highest weights $\Lambda(k,\lambda)$ obtained in this manner from (1) for all $k \in \mathbb{Z}_{\geq 0}$ lie in (but do not fill) a cone $C_0 = \{\, \sum_{i=1}^n a_i\epsilon_i \in \widetilde{Q} \mid 0 > a_1 \geq a_2 \geq \cdots \geq a_n \,\}$ in the lattice $\widetilde{Q} = \{\, \sum_{i=1}^n a_i\epsilon_i \mid a_i \in \frac{1}{2}\cdot\mathbb{Z}\ (1 \leq i \leq n),\ \text{and}\ a_i - a_{i+1} \in \mathbb{Z}\ (1 \leq i \leq n-1) \,\}$. Since the V_λ are finite-dimensional $O(k,\mathbb{C})$-modules, it is possible to describe their behavior in combinatorial terms. R. King and B. Wybourne exploited this connection to obtain some combinatorial results for the infinite-dimensional representations $L_k(\lambda)$ ([KW]).

Our model is the following version of the classical R-S correspondence which depicts the decomposition of the k-th tensor power of the natural representation of $\mathfrak{g}l(n,\mathbb{C})$:

$$[1,n]^k \xrightarrow{\sim} \coprod_{\substack{\lambda\vdash k \\ l(\lambda)\leq n}} \mathrm{CST}^{(n)}(\lambda) \times \mathrm{SYT}(\lambda).$$

Here $\mathrm{CST}^{(n)}(\lambda)$ (resp. $\mathrm{SYT}(\lambda)$) denotes the set of column strict or semistandard tableaux (resp. standard tableaux) of shape λ with entries from $[1,n] = \{1,2,\ldots,n\}$. This bijection is constructed as a repetition of a procedure called *row insertion*, $\mathrm{CST}^{(n)}(\lambda) \times [1,n] \xrightarrow{\sim} \coprod_{\substack{\mu \dot{\supset} \lambda \\ l(\mu)\leq n}} \mathrm{CST}^{(n)}(\mu)$, which depicts the decomposition of the tensor product $V_\lambda \otimes V_\square \simeq \bigoplus_\mu V_\mu$ (where V_λ and V_μ are here irreducible finite-dimensional $\mathfrak{g}l(n,\mathbb{C})$-modules, $V_\square$ is the space of natural representation of $\mathfrak{g}l(n,\mathbb{C})$, and μ runs over partitions such that μ/λ is one box and $l(\mu) \leq n$). By analogy, our correspondence should consist of a repetition of a bijection reflecting the decomposition of the tensor product

$$L_k(\lambda) \otimes W \simeq \bigoplus_{\mu} L_{k+1}(\mu), \tag{2}$$

where μ runs over partitions in $\mathcal{L}_{k+1}$ such that μ/λ is a horizontal strip.

The first ingredient we need is a set $\mathrm{SIST}(\lambda)$ of certain tableaux which, with a notion of weight depending on k, gives the weight generating function equal to the formal character of $L_{\Lambda(k,\lambda)}$. Here, for any $\Lambda \in \mathfrak{h}^*$, L_Λ denotes the (unique) irreducible highest weight $\mathfrak{g}$-module with highest weight Λ.

This has been done for a subclass of these $\mathfrak{g}$-modules in [TY], namely for those having highest weights in a smaller cone $C_1 = \{\, \sum_{i=1}^n a_i\epsilon_i \in C_0 \mid -n > a_1 \,\}$, in other words those in the holomorphic discrete series, in which case L_Λ is actually a generalized Verma module:

If we put $\mathfrak{k} = \mathfrak{gl}(n, \mathbb{C})$ embedded in $\mathfrak{g}$ in a suitable fashion, then all weights in C_0 are $\mathfrak{k}$-dominant integral (with respect to the Borel subalgebra $\mathfrak{b} \cap \mathfrak{k}$ of $\mathfrak{k}$). Therefore, if $\Lambda \in C_0$, we have a finite-dimensional irreducible $\mathfrak{k}$-module F_Λ with highest weight Λ. Let $\mathfrak{p}$ denote the parabolic subalgebra of $\mathfrak{g}$ containing $\mathfrak{b}$ and with Levi part $\mathfrak{k}$. Then we can form the generalized Verma module $N_\Lambda = U(\mathfrak{g}) \otimes_{U(\mathfrak{p})} F_\Lambda$, where F_Λ is regarded as a $\mathfrak{p}$-module by defining the action of the nilpotent radical of $\mathfrak{p}$ to be trivial. If $\Lambda \in C_0 - C_1$, L_Λ is the irreducible quotient of N_Λ (still possibly isomorphic to N_Λ for some such Λ).

It can be easily seen that, in (2), if $L_k(\lambda)$ has highest weight in C_1, then so does every summand on the right hand side. Our result in § 3 presents a weight-preserving bijection

$$\mathrm{SIST}(\lambda) \times \mathbb{Z}^n_{\geq 0} \xrightarrow{\sim} \coprod_{\mu} \mathrm{SIST}(\mu)\,, \tag{3}$$

where μ runs through partitions in $\mathcal{L}_{k+1}$ such that μ/λ is a horizontal strip. Here, the generating function of $\mathbb{Z}^n_{\geq 0}$, with a proper definition of weights, gives the formal character of W, and that of $\mathrm{SIST}(\lambda)$ similarly gives the formal character of $L_k(\lambda) = L_{\Lambda(k,\lambda)}$ as stated above; so this bijection models the decomposition (2) for the case $\Lambda(k, \lambda) \in C_1$.

Unlike the case of finite-dimensional representations, an equality of formal characters is in general far from sufficient to draw out any conclusion on the decomposition of a tensor product. In our case, however, it is known that the tensor product decomposes as a direct sum of irreducible highest weight modules. (See [Kob].)

If we start from an $L_k(\lambda)$ in the holomorphic discrete series, we can repeat our procedure describing (2) and depict the decomposition of $L_k(\lambda) \otimes W^{\otimes k'}$. However, this is not enough to depict the whole decomposition of the tensor powers of W, since W itself (the very starting point) is a direct sum of two irreducible modules whose highest weights are not in C_1. (Recall that we have assumed that $n \geq 2$. If $n = 1$, then all $N_{\Lambda(k,\lambda)}$ in consideration are irreducible, so that (3) covers all desired cases.)

For $n = 2$, which is the smallest value of n having reducible $N_{\Lambda(k,\lambda)}$, we also fill this gap (§ 4). Namely we find subsets $\mathrm{SIST}(k, \lambda)$ of $\mathrm{SIST}(\lambda)$, for any (k, λ) with $\lambda \in \mathcal{L}_k$, whose weight generating function gives the formal character of $L_k(\lambda)$, including the case where $\Lambda(k, \lambda) \in C_0 - C_1$. We also present a modification of the bijection (3) which depicts the decomposition (2) for the case $\Lambda(k, \lambda) \in C_0 - C_1$. So for $n = 2$, we have associated the whole decomposition of the tensor powers of W with combinatorial bijections. For general n, we conjecture the existence of such subsets $\mathrm{SIST}(k, \lambda)$ and modified bijections that work for the cases where $\Lambda(k, \lambda) \in C_0 - C_1$.

The author is grateful to I. Terada for suggesting this problem and providing patient, persistent support. We are indebted to K. Nishiyama, T. Kobayashi and T. Oshima for discussion and clarification of many points concerning these infinite dimensional representations.

2. The Weil representation and its modules

We assume throughout the paper that $n \geq 2$ to avoid the degenerate case $n = 1$.

Definition 2.1 Fix the skew-symmetric form on $\mathbb{C}^{2n}$ given as $\langle v, w\rangle = v^t J w$, where J is the $2n \times 2n$ matrix given as follows:

$$J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}.$$

The *symplectic Lie algebra* $\mathfrak{g} = \mathfrak{sp}(2n, \mathbb{C})$ is then given by

$$\mathfrak{sp}(2n, \mathbb{C}) = \{X \in \mathfrak{gl}(2n, \mathbb{C}) | X^t J + JX = 0\}.$$

We may consider $\mathfrak{gl}(n,\mathbb{C})$ as living inside $\mathfrak{sp}(2n,\mathbb{C})$ via the following embedding

$$A \mapsto \begin{pmatrix} A & 0 \\ 0 & -A^t \end{pmatrix}.$$

We fix a Cartan subalgebra $\mathfrak{h}$ and a Borel subalgebra $\mathfrak{b}$ containing $\mathfrak{h}$ in the manner read from Fig. 1 and its footnote. Let $\epsilon_i \in \mathfrak{h}^*$ be such that $\epsilon_i(E_{jj} - E_{n+j,n+j}) = \delta_{ij}$. The set of corresponding simple roots is is $\{\epsilon_i - \epsilon_{i+1}\,(1 \le i < n),\ 2\epsilon_n\}$.

Note that $\mathfrak{gl}(n,\mathbb{C})$ shares a Cartan subalgebra $\mathfrak{h}$ with $\mathfrak{sp}(2n,\mathbb{C})$. Also, $\mathfrak{gl}(n,\mathbb{C}) \cap \mathfrak{b}$ is a Borel subalgebra of $\mathfrak{gl}(n,\mathbb{C})$.

Definition 2.2 The *Weil representation* ρ of $\mathfrak{sp}(2n,\mathbb{C})$ on $W = \mathbb{C}[x_1, x_2, \ldots, x_n]$ is given explicitely in Figure 1 by defining how various basis elements act on the polynomial ring. It extends, up to a twist by an automorphism of $\mathfrak{sp}(2n,\mathbb{C})$, the simpler action of $\mathfrak{gl}(n,\mathbb{C})$ given by sending $E_{ij} \mapsto x_i\partial_j$, where $\partial_j := \frac{\partial}{\partial x_j}$.

W has two independent highest weight vectors, namely 1 (weight $-\frac{1}{2}\omega_n$) and x_n (weight $-\frac{1}{2}\omega_n - \epsilon_n$). $U(\mathfrak{g}) \cdot 1$ (resp. $U(\mathfrak{g}) \cdot x_n$) is the space of all polynomials of even total degree (resp. odd total degree). We have $W = U(\mathfrak{g}) \cdot 1 \oplus U(\mathfrak{g}) \cdot x_n$, and that these two submodules are irreducible. In the notation defined below, they are $L_{\Lambda(1,(0))}$ and $L_{\Lambda(1,\Box)}$. These two highest weights belong to $C_0 - C_1$. (See Definition 2.6).

Definition 2.3 Let Λ be a weight for $\mathfrak{g} = \mathfrak{sp}(2n,\mathbb{C})$, i.e, $\Lambda \in \mathfrak{h}^*$. Let $\mathfrak{p}$ be the parabolic subalgebra which contains $\mathfrak{b}$ and whose Levi part is $\mathfrak{k} = \mathfrak{gl}(n,\mathbb{C})$ (thought of as embedded in $\mathfrak{sp}(2n,\mathbb{C})$ by the map described immediately after Definition 2.1). Λ is called $\mathfrak{k}$-*dominant integral* if $\Lambda = \sum_{i=1}^n c_i\epsilon_i$, $c_i - c_{i+1} \in \mathbb{Z}_{\ge 0}$ $(1 \le i \le n-1)$.

We define the following four modules.

- L_Λ = the unique irreducible highest weight $\mathfrak{g}$-module with highest weight Λ (for all $\Lambda \in \mathfrak{h}^*$)
- M_Λ := the Verma module for $\mathfrak{g}$ of highest weight Λ (for all $\Lambda \in \mathfrak{h}^*$).
- F_Λ =the finite dimensional irreducible $\mathfrak{k}$-module with highest weight Λ (for $\mathfrak{k}$-dominant integral Λ).
- N_Λ =the generalized Verma module $U(\mathfrak{g}) \otimes_{U(\mathfrak{p})} F_\Lambda$ for $\mathfrak{g}$, where F_Λ is regarded as a $U(\mathfrak{p})$-module by defining the action of the nilpotent radical of $\mathfrak{p}$ to be trivial.

Definition 2.4 For $\lambda \in \mathbb{Y}$ with $l(\lambda) \le n$ and $k \in \mathbb{Z}_{\ge 0}$, we put:

$$\Lambda(k,\lambda) = -\frac{k}{2}(\epsilon_1 + \epsilon_2 + \cdots + \epsilon_n) - \lambda_n\epsilon_1 - \lambda_{n-1}\epsilon_2 - \cdots - \lambda_1\epsilon_n.$$

$$= -\sum_{i=1}^{n}(\lambda_{n-i+1} + \frac{k}{2})\epsilon_i$$

Definition 2.5 We define the following lattice $\widetilde{Q}$ in $\mathfrak{h}^*$, which will contain all the weights we consider:

$$\widetilde{Q} = \left\{ \sum_{i=1}^{n} a_i\epsilon_i \mid a_i \in \frac{1}{2}\cdot\mathbb{Z}\ (1 \le i \le n),\ \text{and}\ a_i - a_{i+1} \in \mathbb{Z}\ (1 \le i \le n-1) \right\}$$

Unlike the usual weight lattice for $\mathfrak{sp}(2n,\mathbb{C})$ (see, e.g., [Hum]), which is just the one freely generated by the ϵ_i's, here we also need to include those half-integral linear combinations for which every coefficient a_i is not an integer.

Definition 2.6 We define two cones in the lattice $\widetilde{Q}$ as follows

$$C_0 = \{\, a_1\epsilon_1 + a_2\epsilon_2 + \ldots + a_n\epsilon_n \in \widetilde{Q} \mid 0 > a_1 \ge a_2 \ge \cdots \ge a_n \,\}.$$

$$C_1 = \{\, a_1\epsilon_1 + a_2\epsilon_2 + \ldots + a_n\epsilon_n \in \widetilde{Q} \mid -n > a_1 \ge a_2 \ge \cdots \ge a_n \,\}.$$

FIGURE 1. The Weil representation for $\mathfrak{s}p(2n,\mathbb{C})$.

Root	Element A	$n=2$ Example	Operator $\rho(A)$
Cartan SA	$E_{ii}-E_{n+i,n+i}$	p $\begin{pmatrix}1&0&0&0\\0&0&0&0\\0&0&-1&0\\0&0&0&0\end{pmatrix}$	$-x_i\partial_i-\frac{1}{2}$
$\epsilon_i-\epsilon_j$ for $i<j$	$E_{ij}-E_{n+j,n+i}$	$\begin{pmatrix}0&1&0&0\\0&0&0&0\\0&0&0&0\\0&0&-1&0\end{pmatrix}$	$x_j\partial_i$
$-\epsilon_i+\epsilon_j$ for $i<j$	$E_{ji}-E_{n+i,n+j}$	$\begin{pmatrix}0&0&0&0\\1&0&0&0\\0&0&0&-1\\0&0&0&0\end{pmatrix}$	$x_i\partial_j$
$\epsilon_i+\epsilon_j$ for $i\neq j$	$E_{i,n+j}+E_{j,n+i}$	$\begin{pmatrix}0&0&0&1\\0&0&1&0\\0&0&0&0\\0&0&0&0\end{pmatrix}$	$\sqrt{-1}\partial_i\partial_j$
$-\epsilon_i-\epsilon_j$ for $i\neq j$	$E_{n+i,j}+E_{n+j,i}$	$\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&1&0&0\\1&0&0&0\end{pmatrix}$	$\sqrt{-1}x_ix_j$
$2\epsilon_i$	$E_{i,n+i}$	$\begin{pmatrix}0&0&1&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}$	$\sqrt{-1}\partial_i^2$
$-2\epsilon_i$	$E_{n+i,i}$	$\begin{pmatrix}0&0&0&0\\0&0&0&0\\1&0&0&0\\0&0&0&0\end{pmatrix}$	$\sqrt{-1}x_i^2$

In this table, the positive and negative root vectors alternate $+,-,+,-,+,-$ after the Cartan subalgebra entry. There is a Lie algebra automorphism of $\mathfrak{sp}(2n,\mathbb{C})$ which exchanges E_β and $E_{-\beta}$ for all positive roots β, and which act as -1 on $\mathfrak{h}$. This means that there is another version of the Weil representation given by composing this automorphism with ρ. It is this version which extends the representation $E_{ij}\mapsto x_i\partial_j$ of $\mathfrak{g}l(n,\mathbb{C})$.

Facts 2.7. *We have the following facts concerning the modules L_Λ.*

(1) *([KV] and [How]) For $k\in\mathbb{Z}_{>0}$, we put $\mathcal{L}_k=\{\,\lambda\in\mathbb{Y}\mid l(\lambda)\le n \text{ and } \lambda'_1+\lambda'_2\le k\,\}$, where λ'_j denotes the length of the jth column of the Young diagram of λ. Then we have*

$$W^{\otimes k}\simeq\bigoplus_{\lambda\in\mathcal{L}_k}L_k(\lambda)\otimes V_\lambda$$

where $L_k(\lambda)\simeq L_{\Lambda(k,\lambda)}$ and V_λ is a finite-dimensional irreducible $O(k,\mathbb{C})$-module.

(2) $C_1\subset\{\,\Lambda(k,\lambda)\mid\lambda\in\mathcal{L}_k,\ k\in\mathbb{Z}_{>0}\,\}\subset C_0$. *(This is immediate. Set $k=2n$ or $k=2n+1$.)*

(3) *If $\Lambda \in \widetilde{Q}$, L_Λ is in the holomorphic discrete series if and only if $\Lambda \in C_1$. Moreover, in this case L_Λ is isomorphic to N_Λ.*

These conditions make it clear that, for k sufficiently large ($\geq 2n$), all the irreducible $\mathfrak{sp}(2n,\mathbb{C})$-modules appearing in the decomposition of $W^{\otimes k}$ are holomorphic discrete series modules. On the other hand, the two irreducible constituents of W (see after Definition 2.2) are not in the holomorphic discrete series.

Lemma 2.8. *The formal character of the Weil representation W is given as the sum of all monomials in t_1^{-1}, t_2^{-1}, ..., t_n^{-1} with an additional factor*

$$\operatorname{ch} W = \sum_{c\in\mathbb{Z}^n_{\geq 0}} t^{-c-(\frac{1}{2},\frac{1}{2},\dots,\frac{1}{2})} = \frac{(t_1t_2\cdots t_n)^{-\frac{1}{2}}}{\prod_{i=1}^n(1-t_i^{-1})}.$$

Namely, if we define the $\frac{1}{2}$-*adjusted weight* of a monomial x^c or $c \in \mathbb{Z}^n_{\geq 0}$ to be $c - \frac{1}{2}\omega_n$, then the weight generating function of $\mathbb{Z}^n_{\geq 0}$ for the $\frac{1}{2}$-adjusted weight equals the formal character of W.

Lemma 2.9.

$$\operatorname{ch} N_{\Lambda(k,\lambda)} = \frac{\operatorname{ch} F_{\Lambda(k,\lambda)}}{\prod_{1\leq i\leq j\leq n}(1-t_i^{-1}t_j^{-1})} = \frac{(t_1t_2\cdots t_n)^{-\frac{k}{2}} s_\lambda(t_1^{-1},\dots,t_n^{-1})}{\prod_{1\leq i\leq j\leq n}(1-t_i^{-1}t_j^{-1})}$$

where s_λ is the Schur function. Therefore, we can compute the character of $L_k(\lambda) = L_{\Lambda(k,\lambda)} = N_{\Lambda(k,\lambda)}$ if $\Lambda(k,\lambda) \in C_1$.

The following definitions of semi-infinite symplectic tableaux and weight are from [TY].

Definition 2.10 Fix a positive integer n, and let $\lambda \in \mathbb{Y}$ have length $\leq n$. Let Γ_n denote the totally ordered set $\{1 < \bar{1} < 2 < \bar{2} < \cdots < n < \bar{n}\}$. The *semi-infinite complementary shape* to λ is the following convex subset of $\mathbb{Z}\times\mathbb{Z}$:

$$S_\lambda = \{(i,j) | 1 \leq i \leq n,\ -\infty < j_i < -\lambda_{n+1-i}\}$$

A *semi-infinite* $\mathfrak{sp}(2n,\mathbb{C})$ *tableau* of *complementary shape* λ is a map $T : S_\lambda \to \Gamma_n$, satisfying the following conditions:

(1) For each $i \in [n]$, there exists $N_i \in \mathbb{Z}$ such that $T(i,j) = i$ whenever $j < N_i$.
(2) Each row is weakly increasing: $T(i,j) \leq T(i,j+1)$ whenever both are defined.
(3) Each column is strictly increasing: $T(i,j) < T(i+1,j)$ whenever both are defined.

Note that the first two conditions imply that $T(i,j) \geq i$, which is part of the definition of (finite) symplectic tableaux used by, e.g., [KE], [Ber], and [KT] [Ber]).

We write $\operatorname{sch}(T) = \lambda$ to indicate that the semi-infinite tableau T has semi-infinite complementary shape λ, and denote the set of all SIST's of that shape by $\mathrm{SIST}(\lambda)$.

Definition 2.11 Define the *crude weight* of $T \in \mathrm{SIST}(\lambda)$ to be $\sum_{i=1}^n c_i\epsilon_i$ ($\in \mathfrak{h}^*$), which we abbreviate as $(c_1, c_2, \dots, c_n) \in \mathbb{Z}^n$ where

$$c_i := -(\text{number of } \bar{i}\text{'s in } T) - (\text{ number of columns of } T \text{ not containing } i) - \lambda_n.$$

The idea is that, as for the usual symplectic tableaux, the letter i (resp. $\bar{i}$) has weight ϵ_i (resp. $-\epsilon_i$), and that the reference tableau (whose crude weight is defined to be zero) is the one of complementary shape $\lambda = \emptyset$ with all entries in the ith row equal to i. For example, the crude weight of the tableau in Figure 2 is $(-13,-15,-13,-14)$. Also, for each $k \in \mathbb{Z}_{>0}$ we define the $\frac{k}{2}$-*adjusted weight* of T to be (the crude weight of T) $-\frac{k}{2}\epsilon_1 - \frac{k}{2}\epsilon_2 - \cdots - \frac{k}{2}\epsilon_n$. For example, the $\frac{7}{2}$-adjusted weight of the tableau in Figure 2 is $(-\frac{33}{2}, -\frac{37}{2}, -\frac{33}{2}, -\frac{35}{2})$.

FIGURE 2. A Semi-infinite symplectic tableau of complementary shape (9,9,4,1)

$\cdots$	1	1	1	1	1	1	1	1	1	$\bar{1}$	$\bar{1}$	$\bar{1}$	$\bar{1}$	$\bar{1}$	2	3
$\cdots$	2	2	2	2	$\bar{2}$	$\bar{2}$	$\bar{2}$	3	3	3	3	3	$\bar{3}$			
$\cdots$	3	3	3	$\bar{3}$	$\bar{3}$	$\bar{3}$	$\bar{3}$	4								
$\cdots$	4	4	4	4	4	$\bar{4}$	$\bar{4}$	$\bar{4}$								

Lemma 2.12 ([TY]). *If $\lambda \in \mathbb{Y}$ with $l(\lambda) \leq n$ and $k \in \mathbb{Z}_{>0}$, then the weight generating function of $\mathrm{SIST}(\lambda)$ for the $\frac{k}{2}$-adjusted weight equals the formal character of $N_{\Lambda(k,\lambda)}$. In particular, if $\Lambda(k,\lambda) \in C_1$, this also equals the formal character of $L_{\Lambda(k,\lambda)}$.*

3. INSERTION

Using the duality (2.7 (1)) and some knowledge of $O(k,\mathbb{C})$-modules, it is not difficult to see the decomposition (2) holds. The goal of this section is to give an insertion scheme which shows that the formal characters of both sides of (2) are equal.

Definition 3.1 We define an insertion operator

$$I_i : \mathrm{SIST}(\lambda) \longrightarrow \coprod_{\mu} \mathrm{SIST}(\mu)$$

where μ will turn out to be a shape such that μ/λ is a single box as follows. Remove the rightmost i in the ith row, making it a hole. Start sliding by jeu de taquin moving letters in the northwest direction, until a hole is left on the boundary of the SIST. (Or equivalently one could regard the slide starting from infinite left in the ith row.) The resulting tableau is a SIST because the sliding process preserves semistandardness. μ/λ is one box since the tableau loses one box of its shape at its outer corner.

In what follows, the *weight* of an ordered pair is the sum of the weights of its components.

Lemma 3.2. *Fix $\lambda \in \mathbb{Y}$ with $l(\lambda) \leq n$. Then the map*

$$\mathrm{SIST}(\lambda) \times [1,n] \to \coprod_{\mu} \mathrm{SIST}(\mu)$$

defined by $(T,i) \mapsto I_i(T)$ is a weight-preserving bijection, where μ runs over partitions $\mu \in \mathbb{Y}$ with $l(\mu) \leq n$ and such that μ/λ is one box, and the SIST's are assigned their crude weights, and the weight of $i \in [1,n]$ is defined to be $-\epsilon_i$.

Proof. This map is well-defined because given any hole in a column-strict tableaux there is a unique jeu de taquin slide starting at that hole to the boundary of the tableaux. The candidate for the inverse map is the reverse slide, which moves letters in the southeast direction, starting from the hole at μ/λ. (So in order to reverse this map, one must know λ as well as μ.) Any such reverse slide must eventually converge to some single row i, and the output of this inverse candidate will be the pair consisting of the resulting tableau and the letter i. That this gives the inverse map follows in the same way that it does for the case of finite jeux de taquin slides. The forward map forces one i to be lost in the tableaux, so by the definition of crude weight the map is weight-preserving. $\square$

Example 3.3 Consider $I_1(T)$ where $T =$

$$\begin{array}{ccccccc} \cdots & 1 & 1 & \bar{1} & 2 & \bar{2} & 3 \\ \cdots & 2 & 2 & 2 & \bar{2} & \bar{3} & \\ \cdots & 3 & 3 & \bar{3} & \bar{3} & & \end{array} .$$

We show the sliding step by step:

$$\begin{array}{ccccccc} \cdots & 1 & \ & \bar{1} & 2 & \bar{2} & 3 \\ \cdots & 2 & 2 & 2 & \bar{2} & \bar{3} & \\ \cdots & 3 & 3 & \bar{3} & \bar{3} & & \end{array} \longrightarrow \begin{array}{ccccccc} \cdots & 1 & \bar{1} & \ & 2 & \bar{2} & 3 \\ \cdots & 2 & 2 & 2 & \bar{2} & \bar{3} & \\ \cdots & 3 & 3 & \bar{3} & \bar{3} & & \end{array} \longrightarrow \begin{array}{ccccccc} \cdots & 1 & \bar{1} & 2 & 2 & \bar{2} & 3 \\ \cdots & 2 & 2 & \ & \bar{2} & \bar{3} & \\ \cdots & 3 & 3 & \bar{3} & \bar{3} & & \end{array} \longrightarrow$$

$$\begin{array}{ccccccc} \cdots & 1 & \bar{1} & 2 & 2 & \bar{2} & 3 \\ \cdots & 2 & 2 & \bar{2} & \ & \bar{3} & \\ \cdots & 3 & 3 & \bar{3} & \bar{3} & & \end{array} \longrightarrow \begin{array}{ccccccc} \cdots & 1 & \bar{1} & 2 & 2 & \bar{2} & 3 \\ \cdots & 2 & 2 & \bar{2} & \bar{3} & \bar{3} & \\ \cdots & 3 & 3 & \bar{3} & \ & & \end{array} .$$

$I_1(T)$ is the final tableau in the sequence above, and μ/λ is the leftmost empty box in row 3.

Definition 3.4 Let I_i^m denote the operator which repeats I_i m times, and for $c \in \mathbb{Z}^n_{\geq 0}$, set $I^c(T) = I_n^{c_n} \circ I_{n-1}^{c_{n-1}} \circ \cdots I_1^{c_1}(T)$. In other words, $I^c(T)$ is the map which successively slides c_1 1's to the left in T, c_2 2's to the left in T, and so on.

The main theorem of this section follows. It gives an analogue of Schensted's correspondence which shows the effect of tensoring a *stable* module $L_k(\lambda)$ (e.g., when $k \geq 2n$) with one copy of W.

Theorem 3.5. *Let $\lambda \in \mathbb{Y}$ with $l(\lambda) \leq n$. Then the map*

$$\mathrm{SIST}(\lambda) \times \mathbb{Z}^n_{\geq 0} \xrightarrow{\sim} \coprod_{\mu} \mathrm{SIST}(\mu)$$

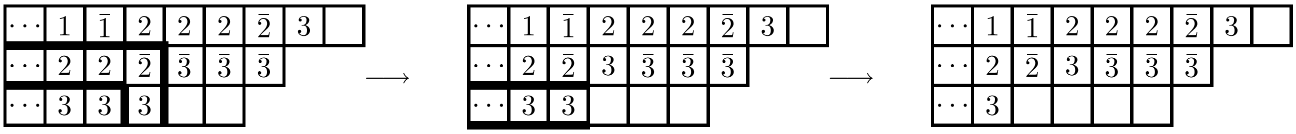
defined by $(T, c) \mapsto I^c(T)$ is a weight-preserving bijection, where μ runs over partitions $\mu \in \mathbb{Y}$ with $l(\mu) \leq n$ and such that μ/λ is a horizontal strip, and the SIST's and the n-tuples are assigned their crude weights.

Example 3.6 Consider $I^{(3,1,1)}(T)$ where T is the following tableau of weight $(-5, -5, -7)$:

$$\begin{array}{ccccccccc} \cdots & 1 & 1 & 1 & 1 & \bar{1} & 2 & \bar{2} & 3 \\ \cdots & 2 & 2 & 2 & 2 & \bar{2} & \bar{3} & & \\ \cdots & 3 & 3 & 3 & \bar{3} & \bar{3} & & & \end{array} .$$

We show the effect of successively inserting $1, 1, 1, 2, 3$, and indicate the sliding path by highlighting the sides of the squares involved. The sliding between the second and third tableaux is the same as that shown in Example 3.3 above.

$$\begin{array}{ccccccccc} \cdots & 1 & 1 & 1 & 1 & \bar{1} & 2 & \bar{2} & 3 \\ \cdots & 2 & 2 & 2 & 2 & \bar{2} & \bar{3} & & \\ \cdots & 3 & 3 & 3 & \bar{3} & \bar{3} & & & \end{array} \longrightarrow \begin{array}{ccccccccc} \cdots & 1 & 1 & 1 & \bar{1} & 2 & \bar{2} & 3 & \ \\ \cdots & 2 & 2 & 2 & 2 & \bar{2} & \bar{3} & & \\ \cdots & 3 & 3 & 3 & \bar{3} & \bar{3} & & & \end{array} \longrightarrow \begin{array}{ccccccccc} \cdots & 1 & 1 & \bar{1} & 2 & 2 & \bar{2} & 3 & \ \\ \cdots & 2 & 2 & 2 & \bar{2} & \bar{3} & \bar{3} & & \\ \cdots & 3 & 3 & 3 & \bar{3} & \ & & & \end{array} \longrightarrow$$



Note that the final tableau has weight $(-8, -6, -8)$, as it should.

Before beginning the proof of Theorem 3.5, we note the following

Lemma 3.7. *Let P be a skew tableau, and let C_1, C_2, …, C_l be a sequence of squares such that C_i is an inner (resp. outer) cocorner of the skew shape* $\operatorname{shape}(P) \cup C_1 \cup \cdots \cup C_{i-1}$*. Recursively define P_i by $P_0 = P$ and P_i to be the result of the sliding which moves letters northwest (resp. southeast) performed on P_{i-1} starting from the hole at C_i. Let D_i be the square lost at the end of the sliding from P_{i-1} to P_i. If the C_i form a horizontal strip from right to left (resp. left to right), then the D_i also form a horizontal strip from right to left (resp. left to right).*

Proof. By a result of M. Haiman ([Hai], Lemma 2.7), we can exchange the role of the skew tableau which the sliding process actually moves and the skew tableau which designates the order of the starting positions of the sliding. Therefore it suffices to show that, if a sequence of slides is performed on a horizontal strip tableau numbered from left to right, then the result is also a horizontal strip tableau numbered from left to right. This is a well known fact. □

of Theorem 3.5. First we show that the result of $I^c(T)$ has shape complementary to μ with μ/λ being a horizontal strip. We can cut off the infinite stable part of T (leaving some of its stable part) to obtain a finite skew tableau T^0 whose inner shape is such that we can put a horizontal strip, adjacent to the inner shape of T^0, with c_i boxes in row i. Performing slides (moving letters northwest) along this horizontal strip, right to left, has the same effect as I^c, and by Lemma 3.7 we know that μ/λ is a horitontal strip (actually created from right to left).

Next we define a candidate for the inverse map. Let U be a SIST which belongs to the left-hand side. We can cut off the stable part, similarly to the above, in such a way that the lengths of adjacent rows of the inner shape differ by at least $|\mu/\lambda|$. We perform slides (moving letters southeast) along the horizontal strip μ/λ from right to left. By Lemma 3.7, the inner boundary of U moves southeast by a horizontal strip (vacation being created from right to left). Denote by c_i the number of boxes in this horizontal strip lying in row i. Then we see that if we iterate the inverse map of Lemma 3.2 to U in such a way that the outer shape loses μ/λ from right to left, then the letters obtained are c_n n's, c_{n-1} $(n-1)$'s, …, c_1 1's, in this order. The output of our inverse candidate is defined to be the pair formed by the SIST obtained at the end of this iteration of the inverse to Lemma 3.2, and the n-tuple $(c_1, c_2, \ldots, c_n)$.

Now from the fact that map described as the inverse candidate in the proof of Lemma 3.2 is really the inverse, our map (in Theorem 3.5) and our inverse candidate (in this proof) are also inverses to each other. It preserves weights because it is an iteration of the weight-preserving map in Lemma 3.2 (and because of the definition of crude weights of n-tuples).

□

Corollary 3.8. *Let $k \in \mathbb{Z}_{\geq 2}$, and suppose that $\lambda \in \mathcal{L}_k$ and $\Lambda(k, \lambda) \in C_1$. Then all μ appearing on the right-hand side of Theorem 3.4 belong to $\mathcal{L}_{k+1}$, and for those μ we have $\Lambda(k+1, \mu) \in C_1$. Moreover, the map in Theorem 3.4 is also weight-preserving if we assign*

$\frac{k}{2}$*-adjusted weights to* $\mathrm{SIST}(\lambda)$, $\frac{1}{2}$*-adjusted weights to* $\mathbb{Z}^n_{\geq 0}$, *and* $\frac{k+1}{2}$*-adjusted weights to* $\mathrm{SIST}(\mu)$. *Thus interpreted, this bijection gives the formal character identity for* (2).

Proof. If $\Lambda(k,\lambda) \in C_1$, then $\Lambda(k+1,\mu) \in C_1$ for any $\mu \supset \lambda$. The condition $\mu \in \mathcal{L}_{k+1}$ is more subtle. $l(\mu) \leq n$ is assured from the beginning (Theorem 3.5), so we only have to check if $\mu'_1 + \mu'_2 \leq k+1$ holds. If $k+1 \geq 2n$, this is satisfied automatically. So suppose $k+1 \leq 2n-1$, or equivalently $\frac{k}{2} \leq n-1$. Now the condition $\Lambda(k,\lambda) \in C_1$ implies $\lambda_n + \frac{k}{2} > n$, so that we have $\lambda_n \geq 2$. This implies $\lambda'_1 + \lambda'_2 = 2n$, so that if $\lambda \in \mathcal{L}_k$, we must have $2n \leq k$, which is a contradiction. Therefore the case $k+1 \leq 2n-1$ never occurs, and we have $\mu \in \mathcal{L}_{k+1}$.

Since the map of Theorem 3.4 is weight-preserving for crude weights, it is also weight-preserving for the adjusted weights because the adjustments match. Using Lemma 2.7 and 2.9, and comparing with (2), we see that the weight generating function identity (for the adjusted weights) resulting from this bijection coincides with the formal character identity for (2). □

Definition 3.9 Following R. Proctor's notion of "N-orthogonal tableaux" [Pro], we propose a skew semistandard tableau of shape μ/λ with entries from $[k+1, l] = \{k+1, k+2, \ldots, l\}$ to be called a *skew* $O(l)/O(k)$*-tableau* if it satisfies what he calls the q-orthogonal condition for $q = k,\ k+1,\ \ldots,\ l$ in the following sense. For such q, let $\lambda^{(q)}$ denote the shape containing λ and all letter $\leq q$. Thus $\lambda = \lambda^{(k)} \subset \lambda^{(k+1)} \subset \cdots \subset \lambda^{(l)} = \mu$, and the differences of adjacent terms are horizontal strips. The q-orthogonal condition requires that the sum of the first two columns of $\lambda^{(q)}$ is at most q. Let $O(l)/O(k)(\mu/\lambda)$ denote the set of all $O(l)/O(k)$-tableaux of shape μ/λ.

Corollary 3.10. *Let* $k \in \mathbb{Z}_{\geq 2}$, *and let* $\lambda \in \mathcal{L}_k$. *Also assume that* $\Lambda(k,\lambda) \in C_1$. *Iterating the map in Theorem 3.4, we obtain a weight-preserving bijection*

$$\mathrm{SIST}(\lambda) \times \left(\mathbb{Z}^n_{\geq 0}\right)^{k'} \xrightarrow{\sim} \coprod_{\mu} \mathrm{SIST}(\mu) \times O(k+k')/O(k)(\mu/\lambda)$$

where μ *ranges over all partitions with* $l(\mu) \leq n$. *The SIST-component of the output is the result of inserting (by Lemma 3.2)* k' n*-tuples, and the skew-tableau component of the output is the one that records the complementary shapes of SIST's after each insertion of an* n*-tuple, namely the shape* $\lambda^{(k+i)}$ *is the shape complementary to the SIST after inserting the* i*th* n*-tuple.*

Example 3.11 Suppose $n = k = 3$ and $\lambda = (3,1)$. Successively inserting the monomials $t_1^2 t_2$, $t_1 t_2^2 t_3$, and t_2 into the following tableau of weight $(-3,-5,-2)$ proceeds as shown:

$$\begin{array}{|c|c|c|c|c|}\hline \cdots & 1 & 1 & \bar{1} & 3 \\ \hline \cdots & 2 & \bar{2} & \bar{2} & * \\ \hline \cdots & 3 & * & * & * \\ \hline \end{array} \quad \xrightarrow{(2,1,0)} \quad \begin{array}{|c|c|c|c|c|c|}\hline \cdots & 1 & \bar{1} & \bar{2} & \bar{2} & 3 \\ \hline \cdots & 2 & 3 & \cdot & \cdot & * \\ \hline \cdots & 3 & \cdot & * & * & * \\ \hline \end{array}$$

$$\begin{array}{|c|c|c|c|c|c|}\hline \cdots & 1 & \bar{1} & \bar{2} & \bar{2} & 3 \\ \hline \cdots & 2 & 3 & * & * & * \\ \hline \cdots & 3 & * & * & * & * \\ \hline \end{array} \quad \xrightarrow{(1,2,1)} \quad \begin{array}{|c|c|c|c|c|c|c|c|c|}\hline \cdots & 1 & 1 & 1 & \bar{1} & \bar{2} & \bar{2} & 3 & \cdot \\ \hline \cdots & 2 & 2 & 3 & 3 & 3 & * & * & * \\ \hline \cdots & 3 & \cdot & \cdot & \cdot & * & * & * & * \\ \hline \end{array}$$

$$\begin{array}{|c|c|c|c|c|c|c|c|c|}\hline \cdots & 1 & 1 & 1 & \bar{1} & \bar{2} & \bar{2} & 3 & * \\ \hline \cdots & 2 & 2 & 3 & 3 & 3 & * & * & * \\ \hline \cdots & 3 & * & * & * & * & * & * & * \\ \hline \end{array} \quad \xrightarrow{(0,1,0)} \quad \begin{array}{|c|c|c|c|c|c|c|c|c|}\hline \cdots & 1 & 1 & 1 & \bar{1} & \bar{2} & \bar{2} & 3 & * \\ \hline \cdots & 2 & 3 & 3 & 3 & \cdot & * & * & * \\ \hline \cdots & 3 & * & * & * & * & * & * & * \\ \hline \end{array}$$

The $*$'s keep track of the complementary shape at the start of each insertion, while the $\cdot$'s indicate the horizontal strip that is added by each inserted monomial. The final complementary shape here is $\mu = (7, 3, 1)$, and the final skew-orthogonal recording tableau of shape λ/μ is

$*$	$*$	$*$	4	5	5	5
$*$	4	4	6			
5						

4. The complete correspondence when $n = 2$

The previous section's insertion map gives the decomposition of tensoring with W only for those modules $L_{\Lambda(k,\lambda)}$ whose characters are the same as for the corresponding generalized Verma module $N_{\Lambda(k,\lambda)}$. This is always true when k is sufficiently large relative to n, but not for small values of k.

However, using the duality of these modules with the $O(k, \mathbb{C})$ modules, it is possible to describe $\operatorname{ch} L_{\Lambda(k,\lambda)}$ as an alternating sum of terms of the form $\operatorname{ch} N_{\Lambda(k,\mu)}$. For example, if we define $\hat{L}(k, \lambda) := \operatorname{ch} \hat{L}_{\Lambda(k,\lambda)}$ and $\hat{N}(k, \lambda) := \operatorname{ch} N_{\Lambda(k,\lambda)}$, then we have the following expressions.

For $n = 1$ all the modules are *stable*, i.e., $\hat{L}(k, \lambda) = \hat{N}(k, \lambda)$ for all k.

For $n = 2$ we have

$$\hat{L}(1, \emptyset) = \hat{N}(1, \emptyset) - \hat{N}(1, (2, 2)) \tag{1}$$

$$\hat{L}(1, \Box) = \hat{N}(1, \Box) - \hat{N}(1, (2, 1)) \tag{2}$$

$$\hat{L}(2, (m)) = \hat{N}(2, (m)) - \hat{N}(2, (m, 2)) \tag{3}$$

In this section we use the above relations to define a subset of SIST tableaux, $\mathrm{SIST}(k, \lambda)$ which gives the weight generating function for $\hat{L}(k, \lambda)$. Then we give the following weight-preserving bijections f_k for $k = 0, 1, 2$:

$$\mathbb{Z}^2_{\geq 0} \xrightarrow{f_0} \mathrm{SIST}(1, \emptyset) \coprod \mathrm{SIST}(1, \Box) \tag{4}$$

$$\mathrm{SIST}(1, \emptyset) \times \mathbb{Z}^2_{\geq 0} \xrightarrow{f_1} \coprod_{r \geq 0} \mathrm{SIST}(2, (r)) \tag{5}$$

$$\mathrm{SIST}(1, \Box) \times \mathbb{Z}^2_{\geq 0} \xrightarrow{f_1} \mathrm{SIST}(2, (1, 1)) \coprod_{r \geq 1} \mathrm{SIST}(2, (r)) \tag{6}$$

$$\mathrm{SIST}(2, (s)) \times \mathbb{Z}^2_{\geq 0} \xrightarrow{f_2} \coprod_{r} \left[\mathrm{SIST}((s + r)) \coprod \mathrm{SIST}((s + r, 1)) \right] \quad (s \geq 2) \tag{7}$$

In all cases, $\operatorname{sch} f_i(T) - \operatorname{sch}(T)$ is a horizontal strip. By the last step we have reached the $k = 3$ stage, which is stable, so our usual definition of SIST gives the correct weight generating function, and the insertion of Section 3 works thereafter. For $k \geq 2$, the shapes $(1, 1)$, $\emptyset$, and $\Box$ are stable, so there is no need for a special f_2 bijection to handle these cases.

4.1. Definition of $\mathrm{SIST}(k, \lambda)$.

Definition 4.1 We call $T \in \mathrm{SIST}(\emptyset)$ $\substack{22\\22}$*-deletable* if it contains the configuration $\substack{22\\22}$, in which case removing the rightmost such configuration and shoving left will leave a valid SIST.

We call $T \in \mathrm{SIST}(\square)$ $\frac{2}{\bar{2}}$-*deletable* if it contains a 2 in row 1, a $\bar{2}$ in row 2, and if it remains a SIST when the rightmost 2 in $\mathrm{Row}_1\, T$ and the rightmost $\bar{2}$ in $\mathrm{Row}_2\, T$ are removed and any elements to the right of the deleted 2 are shoved one box left along the top row.

We call $T \in \mathrm{SIST}((s))$ $2\bar{2}$-*deletable* if $\mathrm{Row}_1\, T$ contains adjacent entries $2\bar{2}$, and if it remains a SIST when this pair is removed and any elements to the right of the deleted $\bar{2}$ are shoved two boxes left along the top row.

In our representation of tableaux, we will use $\cdot$'s to indicate cells that have been vacated due to sliding, as opposed to cells that are part of the complementary shape, which we indicate by $*$. We use $\times$ to indicate a cell that might be either a $\cdot$ or a $*$. To highlight that a certain configuration must occur as the rightmost columns of a tableau, we make the right boarder bold. In the bijections that follow, it will be necessary to keep track of the original complementary shape of the tableau as sliding proceeds.

Example 4.2 In the examples that follow, the tableaux on the left are deletable, and the ones on the right are not.

$$\begin{array}{|c|c|c|c|c|c|}\hline \cdots & 1 & \bar{1} & 2 & \bar{2} & \bar{2} \\ \hline \cdots & 2 & 2 & \cdot & \cdot & \cdot \\ \hline \end{array} \hookrightarrow \begin{array}{|c|c|c|c|c|c|}\hline \cdots & 1 & \bar{1} & \bar{2} & \cdot & \cdot \\ \hline \cdots & 2 & 2 & \cdot & \cdot & \cdot \\ \hline \end{array} \quad \text{BUT NOT} \quad \begin{array}{|c|c|c|c|c|}\hline \cdots & 1 & 2 & \bar{2} & \bar{2} \\ \hline \cdots & 2 & \bar{2} & \cdot & \cdot \\ \hline \end{array}$$

$$\begin{array}{|c|c|c|c|}\hline \cdots & 1 & \bar{1} & 2 \\ \hline \cdots & 2 & \bar{2} & * \\ \hline \end{array} \hookrightarrow \begin{array}{|c|c|c|c|}\hline \cdots & 1 & \bar{1} & \cdot \\ \hline \cdots & 2 & \cdot & * \\ \hline \end{array} \quad \text{BUT NOT} \quad \begin{array}{|c|c|c|c|c|}\hline \cdots & 1 & 1 & \bar{1} & 2 \\ \hline \cdots & 2 & 2 & 2 & * \\ \hline \end{array}$$

$$\begin{array}{|c|c|c|c|c|}\hline \cdots & 1 & 1 & 2 & \bar{2} \\ \hline \cdots & 2 & \bar{2} & \bar{2} & * \\ \hline \end{array} \hookrightarrow \begin{array}{|c|c|c|c|c|}\hline \cdots & 1 & 1 & \bar{2} & \cdot \\ \hline \cdots & 2 & \bar{2} & \cdot & * \\ \hline \end{array} \quad \text{BUT NOT} \quad \begin{array}{|c|c|c|c|c|}\hline \cdots & 1 & 1 & \bar{1} & \bar{1} \\ \hline \cdots & 2 & \bar{2} & \bar{2} & * \\ \hline \end{array}$$

$$\begin{array}{|c|c|c|c|c|c|}\hline \cdots & 1 & 1 & 2 & 2 & 2 \\ \hline \cdots & 2 & \bar{2} & \bar{2} & \bar{2} & \bar{2} \\ \hline \end{array} \hookrightarrow \begin{array}{|c|c|c|c|c|c|}\hline \cdots & 1 & 1 & 2 & \cdot & \cdot \\ \hline \cdots & 2 & \bar{2} & \bar{2} & \cdot & \cdot \\ \hline \end{array} \quad \text{BUT NOT} \quad \begin{array}{|c|c|c|c|c|c|}\hline \cdots & 1 & 1 & \bar{1} & \bar{1} & 2 \\ \hline \cdots & 2 & \bar{2} & \bar{2} & \bar{2} & \bar{2} \\ \hline \end{array}$$

A simple argument shows that $T \in \mathrm{SIST}(\square)$ is $\frac{2}{\bar{2}}$-deletable if and only if $2 \in \mathrm{Row}_1\, T$ and $\bar{2} \in \mathrm{Row}_2\, T$. Similarly, $T \in \mathrm{SIST}(\emptyset)$ is $\frac{22}{\bar{2}\bar{2}}$-deletable if and only if $\mathrm{Row}_1\, T$ contains at least two 2's. Finally, $T \in \mathrm{SIST}((s))$, where $s \geq 2$ is $2\bar{2}$-deletable if and only if $\mathrm{Row}_1\, T$ contains a $2\bar{2}$ pair which lies above empty cells in $\mathrm{Row}_2\, T$; in other words, the pair of elements lie in cells $T(1, -i)$ and $T(1, -i+1)$, where $2 \leq i \leq s$.

Definition 4.3 For $n = 2$, we define $\mathrm{SIST}(k, \lambda)$ as follows:

$$\begin{aligned} \mathrm{SIST}(1, \emptyset) &:= \{T \in \mathrm{SIST}(\emptyset) \mid T \text{ is not } \tfrac{22}{\bar{2}\bar{2}}\text{-deletable}\} \\ \mathrm{SIST}(1, \square) &:= \{T \in \mathrm{SIST}(\square) \mid T \text{ is not } \tfrac{2}{\bar{2}}\text{-deletable}\} \\ \mathrm{SIST}(2, (s)) &:= \{T \in \mathrm{SIST}((s)) \mid T \text{ is not } 2\bar{2}\text{-deletable}\}, \text{ for } s \geq 2. \end{aligned}$$

For all other pairs (k, λ), $\mathrm{SIST}(k, \lambda) := \mathrm{SIST}(\lambda)$.

Before giving the algorithms that define the maps f_i and their inverses, we need to define some operators that replace a given SIST with a different one of the same weight. Each step of f_i will consist of a normal sliding step, possibly augmented by a transformation. In order to keep the transformation straight, we will use the following notation.

- $S \mapsto T$ OR $S \overset{t_i}{\mapsto} T$ will indicate a slide that is part of a single step of the algorithm.

- $T \hookrightarrow U$ indicates a transformation other than a slide that is performed immediately after a slide as part of the same step of the algorithm.
- $S \longrightarrow U$ indicates a single step of the algorithm, including any nonslide transformation that may occur. If the entire step consists soley of an t_i slide, we write: $S \xrightarrow{t_i} U$.
- For the reverse algorithms, we use the same general conventions, but write the extracted monomial in square brackets above the arrow, e.g., $S \xrightarrow{[t_i]} U$, or $S \overset{[t_i]}{\mapsto} T$.

The first operation will replace $\frac{2}{2}$-deletable tableaux with elements of $\mathrm{SIST}(1,\square)$

Definition 4.4 Let $T \in \mathrm{SIST}(\square)$ with $T(1,-1) = 2$ and $T(2,-2) = \bar{2}$. Define $\mathrm{RaiseBar}\, T$ to be the tableau obtained from T by changing the $(1,-1)$-entry to $\bar{2}$, and changing the leftmost $\bar{2}$ in $\mathrm{Row}_2\, T$ to a 2. (In general, this may not be a SIST, because column-strictness may fail for the entry that was lowered, but as we use it below, we always get another SIST.) Inversely, if $T \in \mathrm{SIST}(\square)$ satisfies $T(1,-1) = \bar{2}$, then $\mathrm{LowerBar}\, T$ is defined to be the tableau obtained from T by changing the $(1,-1)$-entry to 2, and changing the rightmost 2 in $\mathrm{Row}_2\, T$ to a $\bar{2}$. (This operation always gives a well-defined SIST of the same shape.)

Definition 4.5 Let T be a tableau with $\mathrm{sch}\, T = (s)$ for some $s \geq 2$, and $i = 1$ or 2. To *explode* an $i\bar{i}$ pair in row i, means to place the horizontally adjacent entries $i\bar{i}$ in row i, immediately to the right of the rightmost i in row i. If $i = 1$, the added entries displace two entries to the cells immediately below. If $i = 2$, they displace the entries to the right of the explosion two place to the right. (While this procedure might not yield a valid SIST in general, it always will in the contexts where we apply it.)

Definition 4.6 Now define the map f_0 as follows. Slide according to the monomial until a shape other than $\emptyset$ or $\square$ is reached. (Call this a *shape violation* for f_0.) This will necessarily imply $\mathrm{sch}\, T = (2)$ because sliding always yields a horizontal strip. If the violation occurs while sliding t_i, then explode (E) an $i\bar{i}$ pair in row i. Continue until the entire monomial has been used. Finally, apply the RaiseBar (RB) operator if the result is $\frac{2}{2}$-deletable of $\mathrm{sch}(\square)$.

Example 4.7 Start with the monimial $t_1^3 t_2^2$. We get the following sequence of insertions:

$$\begin{array}{|c|c|}\hline \cdots & 1 \\ \hline \cdots & 2 \\ \hline\end{array} \xrightarrow{t_1} \begin{array}{|c|c|c|}\hline \cdots & 1 & 2 \\ \hline \cdots & 2 & \cdot \\ \hline\end{array} \overset{t_1}{\mapsto} \begin{array}{|c|c|c|c|}\hline \cdots & 1 & 2 & 2 \\ \hline \cdots & 2 & \cdot & \cdot \\ \hline\end{array} \overset{(E)}{\hookrightarrow} \begin{array}{|c|c|c|c|}\hline \cdots & 1 & 1 & \bar{1} \\ \hline \cdots & 2 & 2 & 2 \\ \hline\end{array} \xrightarrow{t_1} \begin{array}{|c|c|c|c|}\hline \cdots & 1 & \bar{1} & 2 \\ \hline \cdots & 2 & 2 & \cdot \\ \hline\end{array} \overset{t_2}{\mapsto}$$

$$\begin{array}{|c|c|c|c|}\hline \cdots & 1 & \bar{1} & 2 \\ \hline \cdots & 2 & \cdot & \cdot \\ \hline\end{array} \overset{(E)}{\hookrightarrow} \begin{array}{|c|c|c|c|}\hline \cdots & 1 & \bar{1} & 2 \\ \hline \cdots & 2 & 2 & \bar{2} \\ \hline\end{array} \xrightarrow{t_2} \begin{array}{|c|c|c|c|}\hline \cdots & 1 & \bar{1} & 2 \\ \hline \cdots & 2 & \bar{2} & \cdot \\ \hline\end{array} \overset{(RB)}{\hookrightarrow} \begin{array}{|c|c|c|c|}\hline \cdots & 1 & \bar{1} & \bar{2} \\ \hline \cdots & 2 & 2 & \cdot \\ \hline\end{array}$$

Note that the output of this algorithm will always have complementary shape $\emptyset$ or $\square$ because anytime we reach a tableaux T with $\mathrm{sch}\, T = (2)$ it gets replaced by (E) with a tableaux U with $\mathrm{sch}\, U = \emptyset$. Since (RB) only applies to tableaux of $\mathrm{sch} = \square$, it will never take place immediately after (E), only after a slide.

The following lemma is easily shown by direct calculation.

Lemma 4.8. *The tableau T created from the monomial $t_1^j t_2^l$ by f_0 can be described as follows.*

a) *If j and l are even, then $\mathrm{Row}_1\, T$ consists of $j/2$ $\bar{1}$'s, and $\mathrm{Row}_2\, T$ consists of $l/2$ $\bar{2}$'s, and $\mathrm{sch}\, T = \emptyset$.*
b) *If j is even and l is odd, then $\mathrm{Row}_1\, T$ consists of $j/2$ $\bar{1}$'s and $\mathrm{Row}_2\, T$ consists of $(l-1)/2$ $\bar{2}$'s and $\mathrm{sch}\, T = \square$.*

c) *If j is odd and l is even and positive, then* $\mathrm{Row}_1 T$ *consists of* $(j-1)/2$ $\bar{1}$*'s followed by a* $\bar{2}$*, and* $\mathrm{Row}_2 T$ *consists of* $\frac{l}{2}-1$ $\bar{2}$*'s, and* $\mathrm{sch}\, T = \square$*. (This is the only case that* RaiseBar *is applied.) The special case j is odd and $l = 0$ gives the same result, except that $T(1,-1) = 2$ instead of a $\bar{2}$.*

d) *If j and l are odd, then* $\mathrm{Row}_1 T$ *consists of* $(j-1)/2$ $\bar{1}$*'s, followed by a* 2*, and* $\mathrm{Row}_2 T$ *consists of* $(l+1)/2$ $\bar{2}$*'s, and* $\mathrm{sch}\, T = \emptyset$*.*

Theorem 4.9. *The map f_0 is a weight-preserving bijection:*

$$\mathbb{Z}^2_{\geq 0} \xrightarrow{f_0} \mathrm{SIST}(1,\emptyset) \coprod \mathrm{SIST}(1,\square)$$

Proof. Define a map $g_0 : \mathrm{SIST}(1,\emptyset) \coprod \mathrm{SIST}(1,\square) \longrightarrow \mathbb{Z}^2_{\geq 0}$ by setting $g_0(T) = \mathrm{wt}(T)$, the weight of the tableaux T. We claim that f_0 and g_0 are inverses. If we start with a monomial m in $\mathbb{Z}^2_{\geq 0}$ and compute $T = f_0(m)$, then $\mathrm{wt}(T) = m$; for this would be the result of normal sliding, and it is easy to check that performing Explode or RaiseBar does not change the weight of a tableau. Hence, $g_0(f_0(m)) = m$. Conversely, suppose we start with an arbitrary $T \in \mathrm{SIST}(1,\emptyset) \coprod \mathrm{SIST}(1,\square)$. Let $w = \mathrm{wt}(T)$. If $\mathrm{sch}\, T = \emptyset$, then it is easy to see that T is the unique tableau of weight w which is not $\begin{smallmatrix}2&2\\2&2\end{smallmatrix}$-deletable. The only possible columns in T are $\begin{smallmatrix}1\\2\end{smallmatrix}$, $\begin{smallmatrix}\bar{1}\\2\end{smallmatrix}$, $\begin{smallmatrix}1\\\bar{2}\end{smallmatrix}$, $\begin{smallmatrix}\bar{1}\\\bar{2}\end{smallmatrix}$, and $\begin{smallmatrix}2\\\bar{2}\end{smallmatrix}$, which have respective weights $(0,0)$, $(-2,0)$, $(0,-2)$, $(-2,-2)$, and $(-1,-1)$. Since T can contain either columns of $\begin{smallmatrix}\bar{1}\\2\end{smallmatrix}$ or columns of $\begin{smallmatrix}1\\\bar{2}\end{smallmatrix}$ but not both, each possible weight for a tableau of sch $= \emptyset$ can be made uniquely as a integral linear combination of these weights which uses the column $\begin{smallmatrix}2\\\bar{2}\end{smallmatrix}$ at most once. Now Lemma 4.8 shows that $f_0(w)$ is not $\begin{smallmatrix}2&2\\\bar{2}&\bar{2}\end{smallmatrix}$-deletable, so we must have $T = f_0(w) = f_0(g_0(T))$.

If $\mathrm{sch}\, T = \square$, then the following similar but slightly more complicated argument shows that there is a unique way to make a tableau of weight $w(T)$, which is not $\begin{smallmatrix}2\\\bar{2}\end{smallmatrix}$-deletable. Here we can no longer use $\begin{smallmatrix}2\\\bar{2}\end{smallmatrix}$ columns, but we must use exactly one of the following columns to insure that the total degree is odd: $\begin{smallmatrix}1\\\cdot\end{smallmatrix}$, $\begin{smallmatrix}\bar{1}\\\cdot\end{smallmatrix}$, $\begin{smallmatrix}2\\\cdot\end{smallmatrix}$, or $\begin{smallmatrix}\bar{2}\\\cdot\end{smallmatrix}$, which have respective weights $(0,-1)$, $(-2,-1)$, $(-1,0)$, and $(-1,-2)$.

In the first case, the only possible columns are $\begin{smallmatrix}1\\2\end{smallmatrix}$, $\begin{smallmatrix}1\\\bar{2}\end{smallmatrix}$, and $\begin{smallmatrix}1\\\cdot\end{smallmatrix}$, giving a unique way to attain any weight of the form $(0,-l)$, for l odd and positive.

In the second, the possible nonzero weight columns are $\begin{smallmatrix}\bar{1}\\2\end{smallmatrix}$, $\begin{smallmatrix}1\\\bar{2}\end{smallmatrix}$, $\begin{smallmatrix}\bar{1}\\\bar{2}\end{smallmatrix}$ and $\begin{smallmatrix}\bar{1}\\\cdot\end{smallmatrix}$. As in the $\mathrm{sch}\, T = \emptyset$ case, only one of the first two types of columns can appear; thus, omitting the final column $\begin{smallmatrix}\bar{1}\\\cdot\end{smallmatrix}$, we have a unique way to attain any even weight. So this case gives a unique way to construct a tableau whose weight is of the form $(-j,-l)$ for j even and positive, l odd and positive.

In the third case, since our tableau cannot have a $\bar{2}$ in row 2, we may only have columns $\begin{smallmatrix}1\\2\end{smallmatrix}$, $\begin{smallmatrix}\bar{1}\\2\end{smallmatrix}$, and $\begin{smallmatrix}2\\\cdot\end{smallmatrix}$, giving a unique way to attain any weight of the form $(-j,0)$, for j odd and positive. In the fourth case, we find that since T can contain either columns of $\begin{smallmatrix}\bar{1}\\2\end{smallmatrix}$ or columns of $\begin{smallmatrix}1\\\bar{2}\end{smallmatrix}$ but not both, and no columns $\begin{smallmatrix}2\\\bar{2}\end{smallmatrix}$, that the single $\begin{smallmatrix}\bar{2}\\\cdot\end{smallmatrix}$ column contributes $(-1,-2)$ to the weight, and the other columns contribute a unique way of writing any weight with both coordinates even, yielding a unique way to attain any weight of the form $(-j,-l)$, for j odd, l even and positive. ∎

Definition 4.10 Define the map f_1 by iterating the following procedure. We start with a tableau $T \in \mathrm{SIST}(1,\emptyset)$ or $\mathrm{SIST}(1,\square)$ and a monomial m. Perform one slide according to the monomial. If we obtain a tableau T of the following form, we make one of the replacements shown, the *first* one that applies. Recall our convention that $\cdot$'s indicate cells that have been vacated due to sliding, while $*$ indicate cells that are part of the complementary shape of the original tableaux.

- (**Explode**) $\operatorname{sch} T = \mu = (2,1)$; we call this a *shape violation* for f_1, since $\mu_1' + \mu_2' > k = 2$. To repair this shape violation, we Explode a 2 in $\mathrm{Row}_1\, T$ and a $\bar{2}$ in $\mathrm{Row}_2\, T$, to wit, we make the following replacement (which makes the resulting tableau $\frac{2}{\bar{2}}$-deletable):

$$\begin{array}{|c|c|}\hline \bar{2} & \cdot \\ \hline \cdot & * \\ \hline \end{array} \overset{(E1)}{\hookrightarrow} \begin{array}{|c|c|}\hline 2 & \bar{2} \\ \hline \bar{2} & * \\ \hline \end{array} \quad \text{OR} \quad \begin{array}{|c|c|}\hline a & \cdot \\ \hline \cdot & * \\ \hline \end{array} \overset{(E2)}{\hookrightarrow} \begin{array}{|c|c|}\hline a & 2 \\ \hline \bar{2} & * \\ \hline \end{array}, \quad \text{for } a \leq 2.$$

  Then continue sliding according to the monomial.
- (**BarTur**) T has a *symplectic violation*, i.e., T is $2\bar{2}$-deletable. In this case, T must contain one of the two configurations below, to which we apply the following BarTur operation:

$$\begin{array}{|c|c|}\hline 2 & \bar{2} \\ \hline \cdot & * \\ \hline \end{array} \overset{(B1)}{\hookrightarrow} \begin{array}{|c|c|}\hline 2 & \cdot \\ \hline \bar{2} & * \\ \hline \end{array} \quad \text{OR} \quad \begin{array}{|c|c|}\hline 2 & \bar{2} \\ \hline \cdot & \cdot \\ \hline \end{array} \overset{(B2)}{\hookrightarrow} \begin{array}{|c|c|}\hline 2 & 2 \\ \hline \bar{2} & \bar{2} \\ \hline \end{array}.$$

  (These make the resulting tableau $\frac{2}{\bar{2}}$-deletable or $\frac{22}{\bar{2}\bar{2}}$-deletable.) Then continue sliding according to the monomial.

Continue this procedure until the entire monomial has been used, and there are no further shape or symplectic violations. At each step at most one BarTur or Explode may occur before the next slide happens.

Lemma 4.11. *The procedure above gives a well-defined map*

$$\mathrm{SIST}(1, \emptyset) \times \mathbb{Z}^2_{\geq 0} \xrightarrow{f_1} \coprod_{r \geq 0} \mathrm{SIST}(2, (r))$$

Proof. Since the original tableau T has $\operatorname{sch} T = \emptyset$, the only transformation we could apply is $(B2)$. Suppose that the algorithm described gives a sequence of slides:

$$T = T_0 \longrightarrow T_1 \longrightarrow T_2 \longrightarrow \cdots \longrightarrow T_m = f_1(T)\,.$$

We will show by induction that each T_i has the following properties:

A) $\operatorname{sch} T_i$ is a horizontal strip (possibly empty), created from right to left in the bottom row of T_i.
B) T_i is a SIST.
C) T_i is not $2\bar{2}$-deletable.

T_0 clearly enjoys these properties. Suppose T_i has these properties. If no BarTur is involved in the move $T_i \longrightarrow T_{i+1}$, then by the results of Section 3, T_{i+1} also satisfies (A–C). So now assume that the move involves BarTuring so that we have $T_i \mapsto U \hookrightarrow T_{i+1}$, where U is the tableau after sliding, before BarTuring. We have

$$T_i \mapsto \begin{array}{|c|c|c|c|c|}\hline \cdots & a & 2 & \bar{2} & \cdots \\ \hline \cdots & b & \cdot & \cdot & \cdots \\ \hline \end{array} \hookrightarrow \begin{array}{|c|c|c|c|c|}\hline \cdots & a & 2 & 2 & \cdots \\ \hline \cdots & b & \bar{2} & \bar{2} & \cdots \\ \hline \end{array}$$

Say that the rightmost 2 in $\mathrm{Row}_1\, U$ occurs in column j. We claim that the cell $U(2, j-1)$ (marked as containing b) could not be empty. For if the slide $T_i \mapsto U$ caused this cell to vacate, T_i itself would have been $2\bar{2}$-deletable, containing the configuration $\begin{smallmatrix} 2 & \bar{2} \\ \cdot & \cdot \end{smallmatrix}$ in columns j and $j+1$. On the other hand, if this configuration was caused by the slide, then either the 2 or $\bar{2}$ must have ascended from the second row, where it would have been resting in the middle of the the horizontal shape of $\operatorname{sch} T_i$, contradicting (A) for T_i. We conclude that $b = 2$ or $\bar{2}$, and that $\operatorname{sch} U$ is a horizontal strip starting at column j. So $\operatorname{sch} T_{i+1}$ is a horizontal strip starting at column $j+2$, all of whose cells lie in the bottom row. Now (A) follows by induction since we have just taken the leftmost cell away from $\operatorname{sch} T_i$. T_{i+1} is a SIST by inspection, and is not $2\bar{2}$-deletable since $T_{i+1}(1,k) = U(1,k) = \bar{2}$ for $k > j+1$. $\square$

Lemma 4.12. *The procedure above gives a well-defined map*

$$\mathrm{SIST}(1,\square) \times \mathbb{Z}^2_{\geq 0} \xrightarrow{f_1} \coprod_{r\geq 1} \mathrm{SIST}(2,(r)) \coprod \mathrm{SIST}(2,(1,1))$$

Proof. Suppose we have a sequence of insertions as above, and assume that T_i satisfies the following conditions:

A) $\operatorname{sch} T_i = (1,1)$ or (r).
B) T_i is a SIST.
C) T_i is not $2\bar{2}$-deletable.

We wish to show that the same properties hold for T_{i+1}.

If $T_i \longrightarrow T_{i+1}$ involves no Explodes or BarTurs, then (A–B) are automatic, and we cannot get a $2\bar{2}$-deletable tableaux, which would force us to BarTur.

Now suppose we need to BarTur: $T_i \longrightarrow U \hookrightarrow T_{i+1}$, where the rightmost 2 in $\mathrm{Row}_1 T$ occurs in column $-j$. As in the lemma above we find that $U(2,-j-1)$ is nonempty. We get two cases, depending on whether $j = 2$ or $j > 2$:

$$\begin{array}{|c|c|c|c|}\hline \cdots & a & 2 & \bar{2} \\ \hline \cdots & b & \cdot & * \\ \hline\end{array} \overset{(B1)}{\hookrightarrow} \begin{array}{|c|c|c|c|}\hline \cdots & a & 2 & \cdot \\ \hline \cdots & b & \bar{2} & * \\ \hline\end{array} \quad (j=2) \qquad \text{OR} \qquad \begin{array}{|c|c|c|c|c|}\hline \cdots & a & 2 & \bar{2} & \cdots \\ \hline \cdots & b & \cdot & \cdot & \cdots \\ \hline\end{array} \overset{(B2)}{\hookrightarrow} \begin{array}{|c|c|c|c|c|}\hline \cdots & a & 2 & 2 & \cdots \\ \hline \cdots & b & \bar{2} & \bar{2} & \cdots \\ \hline\end{array}$$

Since we reached U by sliding without Explode, we know that $\operatorname{sch} T_i = (r)$, some $r \geq 2$ (not $(1,1)$), so after BarTuring, we get $\operatorname{sch} T_{i+1} = (r-1)$ or if $j = r = 2$, $\operatorname{sch} T_{i+1} = (1,1)$ In either case no $\begin{smallmatrix} 2 \\ \cdot \end{smallmatrix}$ column appears, whence (C). By inspection, T_{i+1} is a SIST.

Now suppose we need to Explode: $T_i \longrightarrow U \hookrightarrow T_{i+1}$ has two cases:

$$\begin{array}{|c|c|c|c|}\hline \cdots & b & \bar{2} & \cdot \\ \hline \cdots & c & \cdot & * \\ \hline\end{array} \overset{(E1)}{\hookrightarrow} \begin{array}{|c|c|c|c|}\hline \cdots & b & 2 & \bar{2} \\ \hline \cdots & c & \bar{2} & * \\ \hline\end{array} \qquad \text{OR} \qquad \begin{array}{|c|c|c|c|}\hline \cdots & b & a & \cdot \\ \hline \cdots & c & \cdot & * \\ \hline\end{array} \overset{(E2)}{\hookrightarrow} \begin{array}{|c|c|c|c|}\hline \cdots & b & a & 2 \\ \hline \cdots & c & \bar{2} & * \\ \hline\end{array}$$

Since T_i had no shape violation, $U(2,-3) = c$ is not empty. It is clear that $\operatorname{sch} T_{i+1} = \square$. Since $b < c \leq \bar{2}$ implies $b \leq 2$, we get that T_{i+1} is a SIST in the first case. In the second it follows from the condition $a \leq 2$. Finally, we see that T_{i+1} is not $2\bar{2}$-deletable because there is only one empty cell in the second row.

$\square$

Definition 4.13 Within a SIST, we use the symbol $\times$ to represent a cell which may contain either a $\cdot$ or $*$. We define a map g_1 by iterating the following procedure. We start with $T \in \mathrm{SIST}(2,(s))$ or $\mathrm{SIST}((1,1))$ and a complementary shape $\lambda = \emptyset$ or $\square$. Before performing a single reverse slide into the complementary shape, we make one of the following replacements (the first one that applies, if any), *provided the resulting tableau is a SIST.*

- (**BarTurBack**)

$$\begin{array}{|c|c|}\hline 2 & \cdot \\ \hline \bar{2} & * \\ \hline\end{array} \overset{(BB1)}{\hookrightarrow} \begin{array}{|c|c|}\hline 2 & \bar{2} \\ \hline \cdot & * \\ \hline\end{array} \qquad \text{OR} \qquad \begin{array}{|c|c|}\hline 2 & 2 \\ \hline \bar{2} & \bar{2} \\ \hline\end{array} \overset{(BB2)}{\hookrightarrow} \begin{array}{|c|c|}\hline 2 & \bar{2} \\ \hline \cdot & \cdot \\ \hline\end{array}$$

- (**Implode**)

$$\begin{array}{|c|c|}\hline a & 2 \\ \hline \bar{2} & * \\ \hline\end{array} \overset{(I2)}{\hookrightarrow} \begin{array}{|c|c|}\hline a & \cdot \\ \hline \cdot & * \\ \hline\end{array}, \text{ for } a \leq 2\,. \qquad \text{OR} \qquad \begin{array}{|c|c|c|}\hline a & 2 & \bar{2} \\ \hline b & \bar{2} & * \\ \hline\end{array} \overset{(I1)}{\hookrightarrow} \begin{array}{|c|c|c|}\hline a & \bar{2} & \cdot \\ \hline b & \cdot & * \\ \hline\end{array}, \quad \text{for } a < 2,$$

Then perform a single reverse slide into the leftmost cell of the complementary shape of the resulting tableau, which extracts a term t_i. The cases overlap for tableaux ending $\begin{smallmatrix} 2 & 2 & 2 \\ \bar{2} & \bar{2} & * \end{smallmatrix}$ and $\begin{smallmatrix} 2 & 2 & \cdot \\ \bar{2} & \bar{2} & * \end{smallmatrix}$.

The procedure ends when a tableau $T \in \mathrm{SIST}(1, \lambda)$ is reached to which none of the replacements above applies. The output is the pair (T, m), where m is the product of the terms t_i extracted at each step.

Note that the proviso that the immediate result of replacement be a SIST only applies to $(BB2)$; e.g.,

$$\begin{array}{|c|c|c|}\hline 2 & 2 & 2 \\ \hline \bar{2} & \bar{2} & \bar{2} \\ \hline \end{array} \quad \overset{(BB2)}{\hookrightarrow} \quad \begin{array}{|c|c|c|}\hline 2 & 2 & \bar{2} \\ \hline \bar{2} & \cdot & \cdot \\ \hline \end{array} \qquad \text{NOT} \qquad \begin{array}{|c|c|c|}\hline 2 & \bar{2} & 2 \\ \hline \cdot & \cdot & \bar{2} \\ \hline \end{array}$$

Lemma 4.14. *The procedure above gives a well-defined map*

$$\coprod_{r \geq 1} \mathrm{SIST}(2, (r)) \coprod \mathrm{SIST}(2, (1, 1)) \xrightarrow{g_1} \coprod_{\lambda = \emptyset, \square} \mathrm{SIST}(1, \lambda) \times \mathbb{Z}^2_{\geq 0}$$

Proof. Given a pair (S, λ) where S is a SIST in the preimage of g_1 and $\lambda = \emptyset$ or $\square$ is the complementary shape to slide into, let

$$S = S_0 \longrightarrow S_1 \longrightarrow S_2 \longrightarrow \cdots \longrightarrow S_m$$

represent the sequence of SISTs obtained by the procedure. We will show by induction that each S_i has the following properties:

A) $\mathrm{sch}\, S_i \backslash \lambda$ is a horizontal strip (possibly empty).
B) S_i is in $\mathrm{SIST}(2, \mathrm{sch}\, S_i)$.
C) $n(S_i) := |\,\mathrm{sch}\, S_i \backslash \lambda| + 2|\{\bar{2} \in S_i\}|$ is a strictly decreasing function of i.

The function n gives an upper bound on the number of steps remaining in the algorithm, since each implosion requires a $\bar{2}$ and leads to two more sliding steps.

Note we always have a well-defined case to apply to our tableau: say $S_i \hookrightarrow U \mapsto S_{i+1}$. To show (C), note that each BarTurBack and Implode transformation reduces by one the number of $\bar{2}$'s and increases $|\,\mathrm{sch}\, S_i \backslash \lambda|$ by two, which the following reverse slide reduces by one. Hence, $n(S_{i+1}) = n(S_i) - 1$. Since each BarTurBack and Implode transformation preserves (A) and (B), as does sliding, it follows by induction that each S_i satisfies all three properties.

The g_1 procedure must finish after finitely many steps because of property (C). We claim that the final output is in $\mathrm{SIST}(1, \lambda)$. If $\lambda = \square$, then we need that the final SIST S_m is not $\frac{2}{\bar{2}}$-deletable. If it were, then it would contain a $\frac{2}{\bar{2}}$ column, forcing one of the Implode or BarTurBack cases to apply. But then S_m is not the last SIST of the procedure. On the other hand, if $\lambda = \emptyset$ and S_m is $\frac{22}{\bar{2}\bar{2}}$-deletable, then S_m falls into the third BarTurBack case, and sliding would have continued. $\square$

The following technical lemma will help us show that the maps f_1 and g_1 are inverses.

Lemma 4.15. *In applying the g_1 map step by step, we extract all t_2 terms before extracting any t_1 terms.*

Proof. The penultimate paragraph of the proof of Theorem 3.5 showed that this holds for each step of g_1 consisting of a pure reverse slide. Thus, it suffices to show the following: at no point in the g_1 algorithm do we have the sequence of moves

$$R \overset{[t_1]}{\mapsto} S \overset{(\tau)}{\hookrightarrow} T \overset{[t_2]}{\mapsto} U$$

where (τ) is one of the allowed replacements. We show this by contradiction in each case.

In case $(BB2)$ we would have:

$$\begin{array}{|c|c|c|c|c|c|c|}\hline \cdots & a_2 & a_1 & 2 & 2 & c_1 & \cdots \\ \hline \cdots & b_3 & b_2 & b_1 & \bar{2} & c_2 & \cdots \\ \hline \end{array} \overset{[t_1]}{\mapsto} \begin{array}{|c|c|c|c|c|c|c|}\hline \cdots & a_3 & a_2 & a_1 & 2 & 2 & \cdots \\ \hline \cdots & b_3 & b_2 & b_1 & \bar{2} & \bar{2} & \cdots \\ \hline \end{array} \overset{(BB2)}{\hookrightarrow} \begin{array}{|c|c|c|c|c|c|c|}\hline \cdots & a_3 & a_2 & a_1 & 2 & \bar{2} & \cdots \\ \hline \cdots & b_3 & b_2 & b_1 & \cdot & \cdot & \cdots \\ \hline \end{array} \overset{[t_2]}{\mapsto} U$$

where the final step implies the inequalities $b_1 > 2$ and $b_{i+1} > a_i$ for $i \geq 2$ in order for the t_2 reverse slide to take place entirely along row two. These inequalities force the t_1 reverse slide to have the form shown, entirely across the top row until the last column shown explicitely; the column $\substack{c_1 \\ c_2}$ must be of the form $\substack{\bar{2} \\ \cdot}$, $\substack{\bar{2} \\ *}$, or $\substack{\cdot \\ *}$. Since $b_1 = \bar{2}$, we find the tableau R has the configuration $\begin{smallmatrix} 2 & 2 \\ \bar{2} & \bar{2} \end{smallmatrix}$, and $(BB2)$ should have been applied to the rightmost such, rather than a pure t_1 reverse slide. This contradicts the definition of g_1.

Following the same line of reasoning, in case $(BB1)$ we would have:

$$\begin{array}{|c|c|c|c|c|c|}\hline \cdots & a_2 & a_1 & 2 & \bar{2} & \cdot \\ \hline \cdots & b_3 & b_2 & b_1 & \cdot & * \\ \hline \end{array} \overset{[t_1]}{\mapsto} \begin{array}{|c|c|c|c|c|c|}\hline \cdots & a_3 & a_2 & a_1 & 2 & \cdot \\ \hline \cdots & b_3 & b_2 & b_1 & \bar{2} & * \\ \hline \end{array} \overset{(BB2)}{\mapsto} \begin{array}{|c|c|c|c|c|c|}\hline \cdots & a_3 & a_2 & a_1 & 2 & \bar{2} \\ \hline \cdots & b_3 & b_2 & b_1 & \cdot & * \\ \hline \end{array} \overset{[t_2]}{\mapsto} U.$$

The leftmost tableau R could not have occurred in the g_1 process since it is not a valid SIST, nor a configuration that arises from one of the g_1 transformations.

Following the same line of reasoning, in case $(I2)$ we would have:

$$\begin{array}{|c|c|c|c|c|}\hline \cdots & a_2 & a & 2 & \cdot \\ \hline \cdots & b_3 & b_2 & \bar{2} & * \\ \hline \end{array} \overset{[t_1]}{\mapsto} \begin{array}{|c|c|c|c|c|}\hline \cdots & a_3 & a_2 & a & 2 \\ \hline \cdots & b_3 & b_2 & \bar{2} & * \\ \hline \end{array} \overset{(I2)}{\mapsto} \begin{array}{|c|c|c|c|c|}\hline \cdots & a_3 & a_2 & a & \cdot \\ \hline \cdots & b_3 & b_2 & \cdot & * \\ \hline \end{array} \overset{[t_2]}{\mapsto} U.$$

The leftmost tableau R could not have occurred in the g_1 process since it is not the output of any g_1-transformation, and the g_1 process would have applied $(BB1)$ rather than a simple reverse slide.

Finally, one can never have $(I1)$ followed by $[t_2]$ since

$$\begin{array}{|c|c|c|c|c|}\hline \cdots & a_2 & a & 2 & \bar{2} \\ \hline \cdots & b_2 & b & \bar{2} & * \\ \hline \end{array} \overset{(I1)}{\mapsto} \begin{array}{|c|c|c|c|c|}\hline \cdots & a_2 & a & \bar{2} & \cdot \\ \hline \cdots & b_2 & b & \cdot & * \\ \hline \end{array} \overset{[t_2]}{\mapsto} U.$$

would force b to land below $\bar{2}$, which is impossible. $\square$

Lemma 4.16. *For $\lambda = \emptyset$ or $\square$ the procedures f_1 and g_1 are inverses, to wit:*

$$g_1 \circ f_1(T, M) = \mathrm{id} \colon \ \mathrm{SIST}(1, \lambda) \times \mathbb{Z}^2_{\geq 0} \ \textit{and} \ f_1 \circ g_1 = \mathrm{id} \colon \coprod_{r \geq 1} \mathrm{SIST}(2, (r)) \coprod \mathrm{SIST}(2, (1, 1))$$

Proof. To show that $g_1 \circ f_1(T, M) = \mathrm{id}$, we must prove that at each step of the g_1 algorithm we recreate the same sequence of shapes that f_1 must have created to reach a given tableau T. In other words, even if T might have arisen *a priori* by some other combination of slides and f_1-transformations, we will show by tracing backwards the impossibility that f_1 could have created T in any other way. This will insure that g_1 correctly undoes the f_1 procedure step by step. In general T may arise from (1) a single t_1 or t_2 slide, or possibly the empty slide if $T \in \mathrm{SIST}(1, \lambda)$, (2) from an $(E1)$ or $(E2)$ explosion preceded immediately by a slide, or (3) from a $(B1)$ or $(B2)$ BarTur preceded immediately by slide.

Case $(BB1)$: Suppose T contains the configuration $\begin{smallmatrix} 2 & \cdot \\ \bar{2} & * \end{smallmatrix}$. By inspection, T is not the result of $(E1)$ or $(E2)$. If it came from a pure slide, then we must have:

$$\begin{array}{|c|c|c|c|c|}\hline \cdots & a_2 & a_1 & 2 & \cdot \\ \hline \cdots & b_2 & b_1 & \bar{2} & * \\ \hline \end{array} \overset{t_1}{\longleftarrow} \begin{array}{|c|c|c|c|}\hline \cdots & a_2 & a_1 & 2 \\ \hline \cdots & b_1 & \bar{2} & * \\ \hline \end{array} \overset{(E2)}{\hookleftarrow} \begin{array}{|c|c|c|c|}\hline \cdots & a_2 & a_1 & \cdot \\ \hline \cdots & b_1 & \cdot & * \\ \hline \end{array} \overset{t_2}{\longleftarrow}$$

where a quick inspection reveals that the middle tableau could only have come from $(E2)$. Since $a_i < b_i$ for all i, the preceding slide must be an t_2. But now we have an f_1 path in which an t_2 slide preceded an t_1 slide, contradicting the definition of f_1. Alternatively, if it came from $(B2)$ preceded by a slide: $T \overset{(B2)}{\hookleftarrow} T' \xleftarrow{t_i} U$, then $\operatorname{sch} U = (2,1)$, which is impossible for a tableau in the image of f_1. Hence, T must have been created in the last step by a $(B1)$ transfomration.

Case $(BB2)$**:** This case is the most involved because the $(B2)$ and $(BB2)$ transformations can occur in columns other than the rightmost two. This requires us to consider several subcases.

Case $\lambda = \emptyset$*:* In this case the only columns occuring to the right of the rightmost $\begin{smallmatrix} 2 & 2 \\ \bar{2} & \bar{2} \end{smallmatrix}$ (which is the only columns where $(BB2)$ is allowed) are of the form $\begin{smallmatrix} \bar{2} \\ \cdot \end{smallmatrix}$. Tracing back, we find that at each stage the only possibilites are that we applied $(B2)$ or a pure t_1 slide. Since we are assuming by way of contradiction that T was not created by $(B2)$, the last step was an t_1 slide. If any move as we trace back was $(B2)$, then it must have been preceeded by an t_2 slide since $a_{i+s} \leq a_i < b_i$ for any $s \in \mathbb{Z}_{\geq 0}$. But we must eventually apply $(B2)$ since there are a finite number of $\begin{smallmatrix} \bar{2} \\ \cdot \end{smallmatrix}$ columns to the right of $\begin{smallmatrix} 2 & 2 \\ \bar{2} & \bar{2} \end{smallmatrix}$, with one $\cdot$ being filled with each t_1 slide we trace back. Eventually we reach $\operatorname{sch} = \emptyset$, and the final two columns are $\begin{smallmatrix} 2 & 2 \\ \bar{2} & \bar{2} \end{smallmatrix}$. Therefore, we reach the contradiction of having an t_2 slide precede and t_1 slide.

Case $\lambda = \square$*:* Let the negative integer k denote the rightmost column in which $\begin{smallmatrix} 2 \\ \bar{2} \end{smallmatrix}$ appears. The only possible columns to the right of this are $\begin{smallmatrix} \bar{2} \\ \cdot \end{smallmatrix}$, $\begin{smallmatrix} \bar{2} \\ * \end{smallmatrix}$, or $\begin{smallmatrix} \cdot \\ * \end{smallmatrix}$. In particular, when $k = -2$, we know that T ends with: $\begin{smallmatrix} 2 & 2 & \bar{2} \\ \bar{2} & \bar{2} & * \end{smallmatrix}$ or $\begin{smallmatrix} 2 & 2 & \cdot \\ \bar{2} & \bar{2} & * \end{smallmatrix}$. We already showed that the latter case must come from $(B1)$, so we consider the former, tracing back:

$$\begin{array}{|c|c|c|c|c|c|} \hline \cdots & a_2 & a_1 & 2 & 2 & \bar{2} \\ \hline \cdots & b_2 & b_1 & \bar{2} & \bar{2} & * \\ \hline \end{array} \overset{(E1)}{\hookleftarrow} \begin{array}{|c|c|c|c|c|c|} \hline \cdots & a_2 & a_1 & 2 & \bar{2} & \cdot \\ \hline \cdots & b_2 & b_1 & \bar{2} & \cdot & * \\ \hline \end{array} \xleftarrow{t_1} \begin{array}{|c|c|c|c|c|c|} \hline \cdots & a_3 & a_2 & a_1 & 2 & \cdot \\ \hline \cdots & b_2 & b_1 & \bar{2} & \bar{2} & * \\ \hline \end{array} \overset{(B1)}{\hookleftarrow}$$

where the last $(B1)$ is forced by an earlier case, and must be preceded by an t_2 slide by the usual inequalities, yielding the same t_2 slide preceding an t_1 slide contradiction as before.

To handle the general case where $k < -2$ we note that as in the $\lambda = \emptyset$ case, until we slide into $(2, -2)$ the only possible backwards steps are $(B2)$ preceded by an t_1 slide, or a pure t_1 slide. Since we want to show that $(BB2)$ is correct, we suppose by way of contradiction that T was immediately created by a pure t_1 slide. The arguments of the last two paragraphs show that we eventually perform the $(BB2)$ operation preceded by an t_2 slide, which is a contradiction. This finishes the $(BB2)$ case.

In the $(I1)$ and $(I2)$ cases, it is clear by inspection that such a tableau could only arise from $(E1)$ or $(E2)$ respectively. This completes the proof that $g_1 \circ f_1(T, M) = \mathrm{id}$.

To show that $f_1 \circ g_1 = \mathrm{id}$ is much easier. By definition, g_1 is a sequence of moves each of which is a reverse slide, or a transformation followed by a reverse slide. Since each f_1-move is a slide or a slide followed by a transformation, and since sliding is reversible by the results of Section 3, we know that the initial slide of f_1 correctly reconstructs the tableau T created by the final reverse slide of g_1. We claim that if T falls into one of the cases calling for a transformation in the definition of f_1, that this T could only have arisen from its opposite transformation in g_1.

Case $(E1)$*:* If T ends with $\begin{smallmatrix} \bar{2} & \cdot \\ \cdot & * \end{smallmatrix}$, we claim that it must have arisen from $(I1)$. Indeed, since T contains a symplectic violation, it could not be the end result of any step of g_1 (Lemma 4.14). By inspection this configuration can not be created by $(BB1)$ or $(I2)$. If it were created by $(BB2)$, the symplectic violation would remain. The claim follows.

The $(E2)$, $(B1)$, and $(B2)$ cases follow similarly. The only new idea is that for the latter two cases, T is $2\bar{2}$-deletable, which is why it could not be the end result of any step of g_1. $\square$

Definition 4.17 Define the map f_2 by iterating the following procedure. We start with a tableau in $\mathrm{SIST}(2,(s))$ with $s \geq 2$ and a monomial m. Perform one slide according to the monomial. If we obtain a tableau T which contains the configuration $\substack{\cdot\ \cdot\\ *\ *}$ (a shape violation) then explode a $2\bar{2}$ pair in $\mathrm{Row}_1\, T$. The new $2\bar{2}$ pair must be placed just to the right of the rightmost 2 in $\mathrm{Row}_1\, T$, possibly displacing a string of $\bar{2}$'s to the right.

Lemma 4.18. *For $s \geq 2$ the procedure above gives a well-defined map:*

$$\mathrm{SIST}(2,(s)) \times \mathbb{Z}^2_{\geq 0} \xrightarrow{f_2} \coprod_r \Big[\mathrm{SIST}((s+r)) \coprod \mathrm{SIST}((s+r,1)) \Big] .$$

Proof. Suppose we have a sequence of steps $T = T_0 \longrightarrow T_1 \longrightarrow \cdots \longrightarrow T_n = f_2(T)$ as described above, and assume that T_j satisfies the following condition:

$$T_j \in \mathrm{SIST}((s+r)) \coprod \mathrm{SIST}((s+r,1) \quad (*)$$

We wish to show that the same condition holds for T_{j+1}.

If the step $T_j \longrightarrow T_{j+1}$ involves no $(E3)$ then the condition holds by familiar properties of ordinary sliding. Now suppose we have $T_j \overset{t_i}{\mapsto} U \overset{(E2)}{\hookrightarrow} T_{j+1}$. The configuration $\substack{\cdot\ \cdot\\ *\ *}$ must occur in the two rightmost columns since U was obtained by sliding from a SIST. Further, these are the only columns of the form $\substack{\cdot\\ *}$ because if there were more than two, then U was created by a slide that left in hole a column $k < -2$, whence the last two columns of T_j were empty, contradicting the condition. So when we apply $(E3)$ to U,we get $\mathrm{sch}\, T_{j+1} = (s)$. It is easy to see that T_{j+1} is a SIST since Exploding $2\bar{2}$ in the first row preserves inequalities along the first row and column strictness. So T_{j+1} satisfies $(*)$. This argument also shows that all the explosions happen in the initial stages of f_1: a pair of t_1 slides entirely across the top row is followed by $(E3)$, returning the tableau to the initial shape. Once a slide descends into the second row, no more explosions can occur and the complementary shape monotonically increases.

Hence, by induction, f_2 is well defined. $\square$

Definition 4.19 Define the map g_2 by iterating the following procedure. Given an integer $s \geq 2$ and a tableaux in $\mathrm{SIST}((s+r)) \coprod \mathrm{SIST}((s+r,1)$ for some $r \in \mathbb{Z}_{\geq 0}$, we check whether the $\mathrm{sch}\, T = (s)$. If not, then perform a single reverse slide into $\mathrm{sch} = (s)$. If so, and if T is $2\bar{2}$-deletable, then implode a $2\bar{2}$ pair in row 1 of T (shoving any elements to the right of the implosion leftwards two spaces) and perform a single reverse slide along $\mathrm{Row}_1\, T$. The procedure terminates when $\mathrm{sch}\, T = (s)$ and T is not $2\bar{2}$-deletable, and the monomial extracted is the product of the terms extracted in each sliding step.

Lemma 4.20. *For $s \geq 2$ the procedure above gives a well-defined map:*

$$\coprod_r \Big[\mathrm{SIST}((s+r)) \coprod \mathrm{SIST}((s+r,1)) \Big] \xrightarrow{g_2} \mathrm{SIST}(2,(s)) \times \mathbb{Z}^2_{\geq 0} .$$

Proof. Since only ordinary sliding is applied until $\mathrm{sch}\, T = (s)$, the procedure is clearly well-defined and always gives a SIST. At this point, if T is $2\bar{2}$-deletable, then it must contain a $2\bar{2}$ pair that lies above empty cells in (s); hence we can implode $2\bar{2}$, leaving a valid SIST, then reverse slide, giving a tableau of $\mathrm{sch} = (s,1)$. (In the next step, an ordinary reverse slide will be applied to this resulting tableau.)

To show that the procedure terminates, note that the number of implosions is bounded above by the minimum of the number of 2's in $\mathrm{Row}_1 T$ and the number of $\bar{2}$'s in $\mathrm{Row}_1 T$. Once we reach $\mathrm{sch} = (s)$, all the remaining reverse slides take place entirely across the top row. The final output will be in $\mathrm{SIST}(2, (s))$ since the procedure continues until it reaches a tableau that is not $2\bar{2}$-deletable. ☐

Lemma 4.21. *For $s \geq 2$ the procedures f_2 and g_2 are inverses, to wit:*

$$g_2 \circ f_2(T, M) = \mathrm{id}\colon\ \mathrm{SIST}(2,(s)) \times \mathbb{Z}^2_{\geq 0} \ \textit{and}\ f_2 \circ g_2 = \mathrm{id}\colon\ \coprod_r \Big[\mathrm{SIST}((s+r)) \coprod \mathrm{SIST}((s+r,1))\Big]$$

Proof. The key idea is to note that whereas all the f_2 explosions take place as the initial steps of f_2 and reset the complementary shape to be (s), after which all steps are ordinary slides, that g_2 exactly reverses this, reverse sliding until it gets to complementary shape (s), then undoing all the initial explosions. ☐

This completes the full version of Schensted's correspondence for the Weil representation in the case $n = 2$, and leaves the case of general n as a tantilizing open problem, for which the techniques developed here may be useful.

References

[Ber] A. Berele, "A Schensted-Type Correspondence for the Sympectic Group," *J. Comb. Thy. A* **43** (1986), 320–328.

[Hai] M. Haiman, "Dual equivalence with applications, including a conjecture of Proctor," Discrete Math **99** (1992), 79–113.

[How] R. Howe, Remarks on classical invariant theory, *Trans. A.M.S.* 313 (1989), 539–570.

[Hum] J.E. Humphreys, *Introduction to Lie algebras and representation theory*, Springer-Verlag, New York, 1972.

[KE] R. C. King and N. G. I. El-Sharkaway, "Standard Young tableaux and weight multiplicities of the classical Lie groups. *J. Phys. A* **16** (1983), 3153–3177.

[KK] S. V. Kerov and A. N. Kirillov, "Combinatorics of rational representations of the group $GL(n, \mathbb{C})$." (Russian) *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **200** (1992), *Kraev. Zadachi Mat. Fiz. Smezh. Voprosy Teor. Funktsii.* **24**, 83–90, 188 [Math Rev 94a:05214].

[Kob] T. Kobayashi, "Discrete series representations for the orbit spaces arising from two involutions of real reductive spaces," preprint.

[KT] K. Koike and I. Terada, "Young diagrammatic methods for the restriction of representations of complex classical Lie groups to reductive subgroups of maximal rank," *Adv. Math.* **79** (1990), 104–135.

[KV] M. Kashiwara and M. Vergne, "On the Segal-Shale-Weil representations and harmonic polynomials, " *Inv. Math* **44** (1978), 1–47.

[KW] R. C. King and B. G. Wybourne, "Holomorphic discrete series and harmonic series unitary irreducible representations of non-compact Lie groups: $Sp(n, \mathbb{R})$, $U(p, q)$ and $SO^*(2n)$," *J. Phys. A* **18** (1985), 3113–3139.

[Pro] R. Proctor, "Young tableaux, Gelfand patterns, and branching rules for classical groups," *Journal of Algebra* **164** (1994), 299–360.

[TY] I. Terada and H. Yaoko, "Semistandard updown tableaux and formal characters of certain induced highest weight modules for $\mathfrak{sp}(2n, \mathbb{C})$," in preparation.

Department of Math & Computer Science, California State University, 25800 Carlos Bee Blvd., Hayward, CA 94542-3092

E-mail address: troby@csuhayward.edu

TWO INVOLUTIONS ON VERTICES OF ORDERED TREES

SEUNGHYUN SEO AND HEESUNG SHIN

Abstract. We will give two natural involutions on the set of pointed ordered trees. These involutions strongly rely on the fact that the pointed ordered trees are embedded in the plane. Using these involutions, we provide combinatorial bijections between the terminal and the internal vertices of ordered trees, and then we deal with some counting problems on the pointed ordered trees.

Résumé. Nous donnons deux involutions naturelles sur l'ensemble des arbres pointés planaires. Ces deux involutions reposent fortement sur le fait que les arbres planaires pointés sont plongés dans le plan. En utilisant ces involutions, nous donnons des bijections entre les sommets internes et externes des arbres planaires. Nous donnons des résultats sur certains de leurs énumérations.

1. Introduction

An *ordered tree* is a rooted tree in which children of each vertex are ordered. A *pointed ordered tree* is an ordered tree one of whose vertex is pointed, where 'pointed' means distinguishing one vertex. For an ordered tree $\mathbf{o}$ and a vertex v in $\mathbf{o}$, let $\mathbf{o}^v$ denote the pointed ordered tree with v pointed. Let $\mathcal{O}_n$ be the set of all ordered trees with n edges and $\mathcal{O}_n^\bullet$ be the set of all pointed ordered trees with n edges. Figure 1 shows all the elements of $\mathcal{O}_3$ with roots at the top.

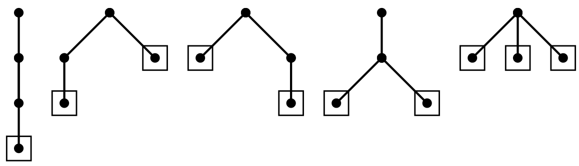


Figure 1. Ordered trees with 3 edges

A pointed ordered tree $\mathbf{o}^v$ is called *terminal*, if v is a *leaf* in $\mathbf{o}^v$, otherwise, *internal*. Let $\mathcal{O}_n^-$ be the set of all terminal pointed ordered trees in $\mathcal{O}_n^\bullet$ and $\mathcal{O}_n^+$ the set of all internal pointed ordered trees. Then these two sets have the same cardinality, as illustrated in Figure 1, where there are the same number of boxed and unboxed vertices.

Many researchers proved $|\mathcal{O}_n^+| = |\mathcal{O}_n^-|$ with various combinatorial methods. Dasarathy and Yang [3] gave combinatorial explanations by Knuth natural correspondence [7, pp. 332–333]. Chauve [2], Deutsch [5], and Seo [8] introduced proofs in Dyck path version independently. These proofs used binary trees, Dyck paths, or some other objects, but no proof has been built directly on ordered trees.

In this paper, we will give two *natural* bijections on $\mathcal{O}_n^\bullet$. These bijections strongly rely on the fact that pointed ordered trees are embedded in the plane. We will analyze the set $\mathcal{O}_n^\bullet$ and find some interesting formulas.

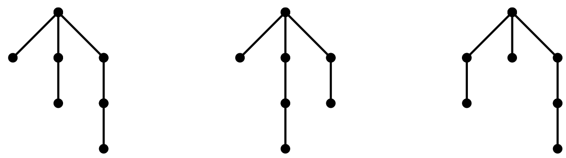


FIGURE 2. Trees, Plane Trees, Ordered Trees

2. Preliminary

We adopt conventional terminology on trees as in [1], [6]. Let $\mathcal{T}_n$ be the set of all trees with n edges and $\mathcal{T}_n^\bullet$ the set of pointed trees with n edges. Every tree can be embedded in a plane without edge-crossing. A tree embedded in the plane is called a *plane tree*. Two plane trees are considered to be the same, if they can be made identical by an orientation preserving homeomorphism defined on the plane. Let $\mathcal{P}_n$ be the set of all plane trees with n edges and $\mathcal{P}_n^\bullet$ the set of all pointed plane trees with n edges.

An *ordered tree* is a plane tree which has a distinguished vertex (called the *root*) and a distinguished edge (called the *leftmost edge*) which is incident with the root. A tree consisting of one vertex is considered to be an ordered tree. This definition of ordered tree is equivalent to the original definition which mentioned in previous section. Two ordered trees are the same, if there is an orientation preserving homeomorphism which preserves the root and the leftmost edge. Let $\mathcal{O}_n$ be the set of all ordered trees with n edges and $\mathcal{O}_n^\bullet$ the set of all pointed ordered trees with n edges.

In Figure 2, we illustrate three different ordered trees with roots at the top. As plane trees, two are the same, the second and the third, and all three are the same as trees.

Let $\mathbf{o}$ be an ordered tree. For each vertex v of $\mathbf{o}$, let $\mathrm{p}(v)$ be the unique path from the root to v. For the non-root vertex v, the *parent* of v is the neighbor of v in $\mathrm{p}(v)$, and any neighbor of v in $\mathbf{o}$, which is not the parent of v, is called a *child* of v. A vertex v in an ordered tree $\mathbf{o}$ is called a *terminal vertex or leaf*, if it has no child, and an *internal vertex*, otherwise. The vertex in the ordered tree consisting of one vertex is considered as a terminal vertex. We call $\mathbf{o}^v$ *terminal*, if v is a terminal vertex in $\mathbf{o}$, and *internal*, if v is an internal vertex. Let $\mathcal{O}_n^+$ denote the set of all internal pointed ordered trees, and let $\mathcal{O}_n^-$ denote the set of all terminal pointed ordered trees, i.e. $\mathcal{O}_n^- = \mathcal{O}_n^\bullet \setminus \mathcal{O}_n^+$. A vertex w in an ordered tree $\mathbf{o}$ is called a *descendant* of a vertex v, if $\mathrm{p}(v) \subsetneq \mathrm{p}(w)$ holds.

Given a pointed ordered tree $\mathbf{o}^v$, let $\mathrm{d}(\mathbf{o}^v)$, the *descendant subtree of* $\mathbf{o}^v$, denote the induced subtree of $\mathbf{o}^v$ consisting of all descendants of v and v itself as a root, and let $\bar{\mathrm{d}}(\mathbf{o}^v)$, the *non-descendant subtree of* $\mathbf{o}^v$, denote the induced subtree of $\mathbf{o}^v$ consisting of the vertices which are not descendants of v.

Let $\mathbf{o}^v$ be a pointed ordered tree in which v is not the root of $\mathbf{o}$, and u be the parent of v. Define $\mathrm{l}(\mathbf{o}^v)$ (resp. $\mathrm{r}(\mathbf{o}^v)$), the *left (resp. right)-descendant subtree of* $\mathbf{o}^v$ by the subtree of $\mathrm{d}(\mathbf{o}^u)$ consisting of u as a root and all descendants of u to the left (resp. right) of the edge $\{u, v\}$. Finally, let $\bar{\mathrm{l}}(\mathbf{o}^v)$ (resp. $\bar{\mathrm{r}}(\mathbf{o}^v)$) be the union of $\mathrm{r}(\mathbf{o}^v)$ (resp. $\mathrm{l}(\mathbf{o}^v)$), $\bar{\mathrm{d}}(\mathbf{o}^u)$, $\mathrm{d}(\mathbf{o}^v)$ and the edge $\{u, v\}$ (see Figure 3). In fact, $\mathrm{d}(\mathbf{o}^v)$ (resp. $\mathrm{l}(\mathbf{o}^v)$, $\mathrm{r}(\mathbf{o}^v)$) is the edge complement of $\bar{\mathrm{d}}(\mathbf{o}^v)$ (resp. $\bar{\mathrm{l}}(\mathbf{o}^v)$, $\bar{\mathrm{r}}(\mathbf{o}^v)$).

3. Involutions L and R on pointed ordered trees

In this section, we assume that $n > 0$. So if $\mathbf{o}^v \in \mathcal{O}_n^+$, then v has at least one child, and if $\mathbf{o}^v \in \mathcal{O}_n^-$, then v has a parent.

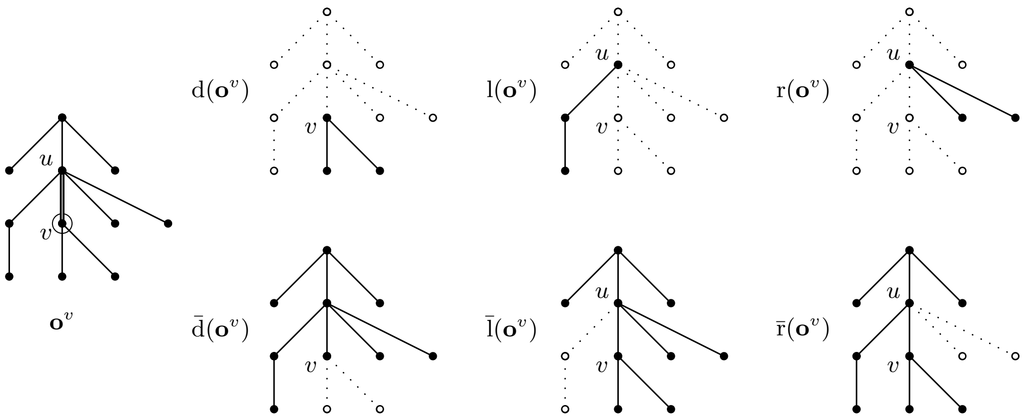


FIGURE 3. Decomposition of $\mathbf{o}^v$

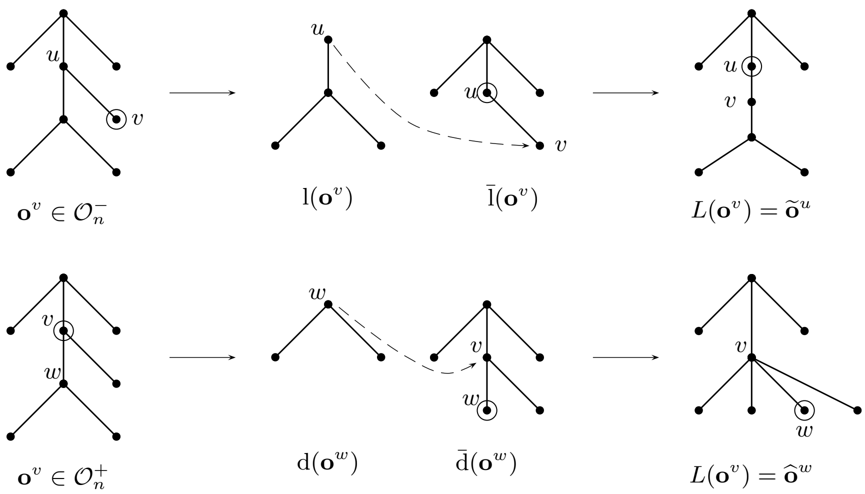


FIGURE 4. L map on $\mathcal{O}_n^-$ and $\mathcal{O}_n^+$

We can define a mapping $L^{\mp} : \mathcal{O}_n^- \to \mathcal{O}_n^+$ as follows (Figure 4): Given $\mathbf{o}^v \in \mathcal{O}_n^-$, where u is the parent of v, we first cut $\mathrm{l}(\mathbf{o}^v)$ from $\mathbf{o}^v$, and paste it to $\bar{\mathrm{l}}(\mathbf{o}^v)$ by identifying $u \in \mathrm{l}(\mathbf{o}^v)$ and $v \in \bar{\mathrm{l}}(\mathbf{o}^v)$. Let $\widetilde{\mathbf{o}}^u$ be the resulting pointed ordered tree. Note that u is an internal vertex in $\widetilde{\mathbf{o}}^u$, so $\widetilde{\mathbf{o}}^u \in \mathcal{O}_n^+$. Set $L^{\mp}(\mathbf{o}^v) = \widetilde{\mathbf{o}}^u$.

Conversely, we can also define a mapping $L^{\pm} : \mathcal{O}_n^+ \to \mathcal{O}_n^-$ as follows: Given $\mathbf{o}^v \in \mathcal{O}_n^+$, where w is the leftmost child of v, we first cut $\mathrm{d}(\mathbf{o}^w)$ from $\mathbf{o}^v$, and paste it to $\bar{\mathrm{d}}(\mathbf{o}^w)$ by identifying $w \in \mathrm{d}(\mathbf{o}^w)$ and $v \in \bar{\mathrm{d}}(\mathbf{o}^w)$, and put $\mathrm{d}(\mathbf{o}^w)$ on the left of the edge $\{v, w\} \in \bar{\mathrm{d}}(\mathbf{o}^w)$. Let $\widehat{\mathbf{o}}^w$ be the resulting pointed ordered tree. Note that w is a terminal vertex in $\widehat{\mathbf{o}}^w$, so $\widehat{\mathbf{o}}^w \in \mathcal{O}_n^-$. Set $L^{\pm}(\mathbf{o}^v) = \widehat{\mathbf{o}}^w$. Now define $L : \mathcal{O}_n^\bullet \to \mathcal{O}_n^\bullet$ by

$$L(\mathbf{o}^v) = \begin{cases} L^{\pm}(\mathbf{o}^v), & \text{if } \mathbf{o}^v \in \mathcal{O}_n^+, \\ L^{\mp}(\mathbf{o}^v), & \text{if } \mathbf{o}^v \in \mathcal{O}_n^-. \end{cases}$$

Figure 5 shows how L maps $\mathcal{O}_3^-$ to $\mathcal{O}_3^+$.

By replacing left with right in the definition of L, we can define $R : \mathcal{O}_n^\bullet \to \mathcal{O}_n^\bullet$ similarly.

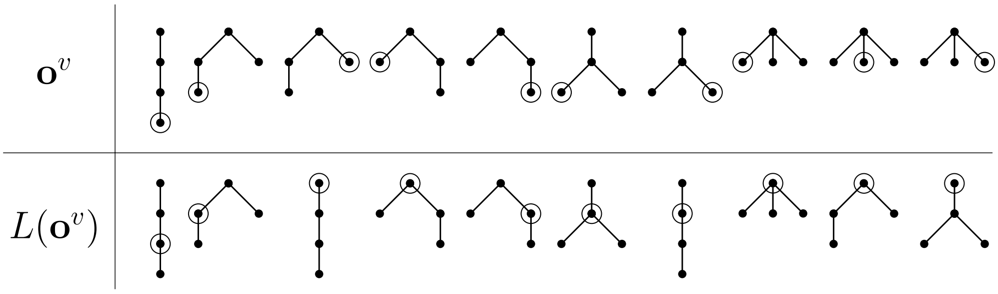


FIGURE 5. Correspondence in $\mathcal{O}_3^\bullet$ by L

Theorem 1. *The maps L and R are* involutions *in $\mathcal{O}_n^\bullet$ with $L(\mathcal{O}_n^-) = \mathcal{O}_n^+$ and $R(\mathcal{O}_n^-) = \mathcal{O}_n^+$. So L and R are bijections from $\mathcal{O}_n^-$ to $\mathcal{O}_n^+$.*

Proof. It suffices to show that $(L^\mp)^{-1} = L^\pm$. Let $\mathbf{o}^v$ be an arbitrary tree in $\mathcal{O}_n^-$, u be the parent of v, and $\widetilde{\mathbf{o}}^u$ be the image of $\mathbf{o}^v$ under $L^\mp$. Then, from the definition of $L^\mp$, we obtain $\mathrm{l}(\mathbf{o}^v) = \mathrm{d}(\widetilde{\mathbf{o}}^v)$ and $\bar{\mathrm{l}}(\mathbf{o}^v) = \bar{\mathrm{d}}(\widetilde{\mathbf{o}}^v)$. Observe that v is the leftmost child of u in $\widetilde{\mathbf{o}}^u$. Let $L^\pm(\widetilde{\mathbf{o}}^u) = \widehat{\mathbf{o}}^v$. Then, from the definition of $L^\pm$, we obtain $\mathrm{d}(\widetilde{\mathbf{o}}^v) = \mathrm{l}(\widehat{\mathbf{o}}^v)$ and $\bar{\mathrm{d}}(\widetilde{\mathbf{o}}^v) = \bar{\mathrm{l}}(\widehat{\mathbf{o}}^v)$. So we get $\mathrm{l}(\mathbf{o}^v) = \mathrm{l}(\widehat{\mathbf{o}}^v)$ and $\bar{\mathrm{l}}(\mathbf{o}^v) = \bar{\mathrm{l}}(\widehat{\mathbf{o}}^v)$, which yield $\mathbf{o}^v = \widehat{\mathbf{o}}^v$, and so $L^\pm \circ L^\mp$ is the identity. □

Moreover, from Theorem 1, we can easily get the following result. The statement about the average level previously appeared in [4].

Corollary 2. *For $\mathbf{o}^v \in \mathcal{O}_n^\bullet$, define the* level *of $\mathbf{o}^v$, denoted by $\rho(\mathbf{o}^v)$, to be the number of edges in the path $\mathrm{p}(v)$. If $\rho(\mathbf{o}^v) = m$, then*

$$\rho(L(\mathbf{o}^v)) = \rho(R(\mathbf{o}^v)) = \begin{cases} m+1, & \text{if } \mathbf{o}^v \text{ is internal,} \\ m-1, & \text{if } \mathbf{o}^v \text{ is terminal.} \end{cases}$$

Consequently, the average level of terminal $\mathbf{o}^v$'s is greater than the average level of internal $\mathbf{o}^v$'s by 1.

4. Various Enumerations on $\mathcal{O}_n^\bullet$

4.1. A group action on $\mathcal{O}_n^\bullet$.

For simplicity, we define L and R on $\mathcal{O}_0^\bullet$ to be the identity map. Let $\mathbf{G}$ be the group generated by L and R with composition as the operation. Since L and R are involutions, $\mathbf{G}$ has the following presentation.

$$\mathbf{G} = \langle L, R : L^2 = 1,\ R^2 = 1 \rangle.$$

The group $\mathbf{G}$ acts on $\mathcal{O}_n^\bullet$ by

$$G \cdot \mathbf{o}^v = G(\mathbf{o}^v) \quad \text{for all } \mathbf{o}^v \in \mathcal{O}_n^\bullet \text{ and all } G \in \mathbf{G}.$$

From now on, we write $G\mathbf{o}^v$ for $G(\mathbf{o}^v)$. $\mathcal{O}_n^\bullet$ is partitioned into G-orbits, we are interested in finding *the number of distinct* $\mathbf{G}$*-orbits* and *the size of each orbit.*

Clearly, every orbit has an even number of elements and exactly half of them are in $\mathcal{O}_n^+$. For any element $\mathbf{o}^v$ in $\mathcal{O}_n^\bullet$, let $[\mathbf{o}^v]$ denote the $\mathbf{G}$-orbit of $\mathbf{o}^v$, and set $[\mathbf{o}^v]^+ = [\mathbf{o}^v] \cap \mathcal{O}_n^+$ and $[\mathbf{o}^v]^- = [\mathbf{o}^v] \cap \mathcal{O}_n^-$. Since $[\mathbf{o}^v]^+$ and $[\mathbf{o}^v]^-$ are equinumerous, $|[\mathbf{o}^v]| = 2|[\mathbf{o}^v]^+|$. Let $\mathbf{H}$ be the subgroup of $\mathbf{G}$ generated by RL, i.e. $\mathbf{H} = \langle RL \rangle$. If $\mathbf{o}^v$ is chosen from $\mathcal{O}_n^+$, then we can easily see that

$$[\mathbf{o}^v]^+ = \{H\mathbf{o}^v \mid H \in \mathbf{H}\}.$$

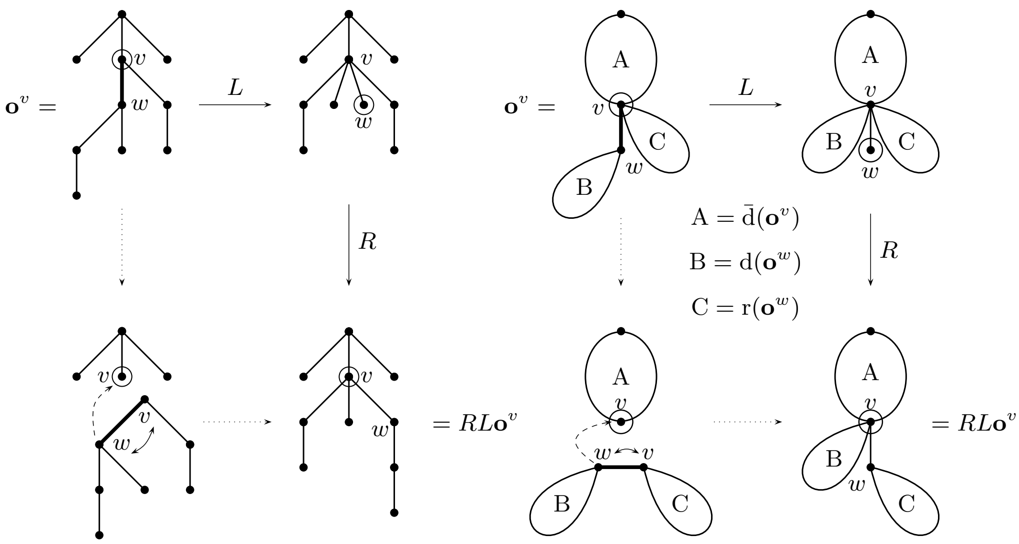


FIGURE 6. How RL acts on a given $\mathbf{o}^v \in \mathcal{O}_n^+$

And the number of $\mathbf{H}$-orbits in $\mathcal{O}_n^+$ equals the number of $\mathbf{G}$-orbits in $\mathcal{O}_n^\bullet$. So instead of the $\mathbf{G}$-action on $\mathcal{O}_n^\bullet$, we will discuss the $\mathbf{H}$-action on $\mathcal{O}_n^+$. The following proposition shows how RL acts on $\mathbf{o}^v \in \mathcal{O}_n^+$ more intuitively.

Proposition 3. *For each $\mathbf{o}^v \in \mathcal{O}_n^+$, $RL\mathbf{o}^v$ is obtained as follows: Let w be the leftmost child of v. First cut $\mathrm{d}(\mathbf{o}^v)$ from $\mathbf{o}^v$, exchange labels v and w in $\mathrm{d}(\mathbf{o}^v)$, and finally identify v in $\mathrm{d}(\mathbf{o}^v)$ and the original v in $\bar{\mathrm{d}}(\mathbf{o}^v)$. The resulting pointed ordered tree is $RL\mathbf{o}^v$.*

Proof. This is apparent from the definition of L and R. See Figure 6. □

4.2. **Enumeration of H-orbits.**

The Proposition 3 enable us to compute the number of distinct $\mathbf{H}$-orbits in $\mathcal{O}_n^+$. Recall that $\mathcal{P}_n$ is the set of all plane trees with n edges. Let p_n denote the cardinality of $\mathcal{P}_n$.

Theorem 4. *Let orb_n be the number of distinct $\mathbf{H}$-orbits in $\mathcal{O}_n^+$. Then,*

$$orb_n = p_n + \sum_{k=1}^{n-1} \binom{2k-1}{k} p_{n-k}. \tag{4.1}$$

Proof. Let U be the natural map from $\mathcal{O}_n$ to $\mathcal{P}_n$ by just *forgetting* the root and the initial edge. By Proposition 3, $\bar{\mathrm{d}}(RL\mathbf{o}^v) = \bar{\mathrm{d}}(\mathbf{o}^v)$ and $U(\mathrm{d}(RL(\mathbf{o}^v))) = U(\mathrm{d}(\mathbf{o}^v))$. So two orbits $[\mathbf{o}^v]^+$ and $[\check{\mathbf{o}}^w]^+$ are different if and only if $\bar{\mathrm{d}}(\mathbf{o}^v) \neq \bar{\mathrm{d}}(\check{\mathbf{o}}^w)$ or $U(\mathrm{d}(\mathbf{o}^v)) \neq U(\mathrm{d}(\check{\mathbf{o}}^w))$. Assume that $\bar{\mathrm{d}}(\mathbf{o}^v)$ has k edges. Since v has no descendants in $\bar{\mathrm{d}}(\mathbf{o}^v)$, $\bar{\mathrm{d}}(\mathbf{o}^v)$ is an element of $\mathcal{O}_k^-$, and clearly $U(\mathrm{d}(\mathbf{o}^v))$ is an element of $\mathcal{P}_{n-k}$. Then the number of distinct orbits is the sum

$$\sum_{k=0}^{n-1} |\mathcal{O}_k^-| \cdot |\mathcal{P}_{n-k}|.$$

Now (4.1) follows, since $|\mathcal{O}_k| = c_k = \frac{1}{k+1}\binom{2k}{k}$, where c_k denotes the k-th Catalan number, and for $k > 0$, $|\mathcal{O}_k^-| = \frac{1}{2}(k+1)|\mathcal{O}_k| = \binom{2k-1}{k}$, and $|\mathcal{O}_0^-| = 1$. □

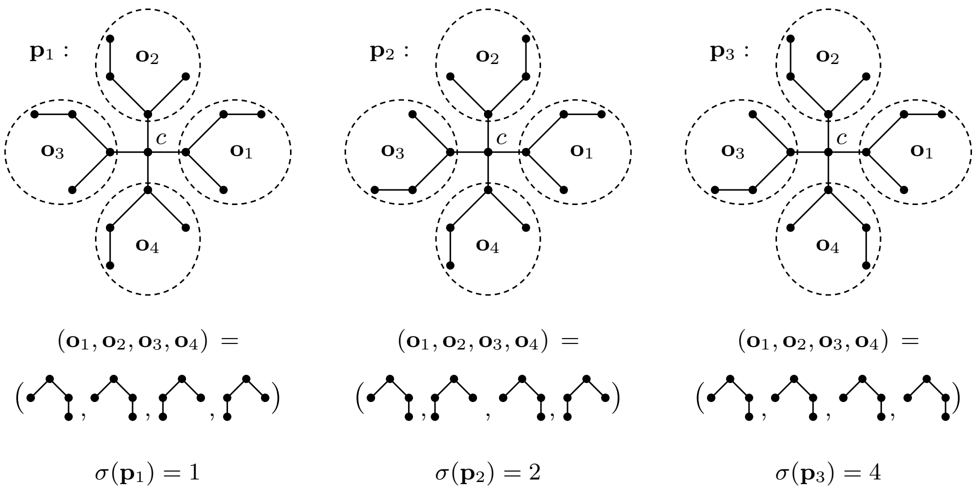


FIGURE 7. $\sigma(\mathbf{p})$: When $c(\mathbf{p})$ is a vertex.

Let $\mathcal{P}(x)$ denote the ordinary generating function for p_n, and $\mathcal{O}(x)$ for Catalan number c_n. Then by dissymmetry Theorem for trees [1, Theorem 4.1.1],

$$\mathcal{P}(x) = 1 + \sum_{n\geq 1} \frac{\phi(n)}{n} \log \frac{1}{1 - x^n \mathcal{O}(x^n)} + \frac{x}{2} \left(\mathcal{O}(x^2) - \mathcal{O}^2(x)\right)$$

and

$$p_n = \frac{1}{2n} \sum_{d|n} \phi\left(\frac{n}{d}\right) \binom{2d}{d} - \frac{1}{2} c_n + \frac{1}{2} \chi_{\text{odd}}(n)\, c_{\frac{n-1}{2}}, \tag{4.2}$$

where ϕ is the Euler's totient function and χ_{odd} is the characteristic function of odd integers. From (4.1) and (4.2), we can have the summation form of orb_n, but we can not find a simple formula. The sequence $\{orb_n\}_{n=0}^{\infty}$ starts with 1, 1, 2, 6, 18, 60, 210, 754, 2766, 10280, 38568, ... , and it does not appear in On-Line Encyclopedia of Integer Sequences [9].

4.3. **Counting the cardinality of an orbit.**

Given $\mathbf{o}^v \in \mathcal{O}_n^+$, we compute the size of $[\mathbf{o}^v]^+$. For a plane tree $\mathbf{p}$ having at least one edge, we define $c(\mathbf{p})$, the *center* of $\mathbf{p}$, which appears in [1], as follows :

(1) Delete all the leaves in $\mathbf{p}$.
(2) Iterate (1), until the resulting plane tree has at most one internal vertex.
(3) If the resulting plane tree has no internal vertex (one edge), then the edge is $c(\mathbf{p})$.
(4) If the resulting tree has a unique internal vertex (star shape), then the internal vertex is $c(\mathbf{p})$.

If we delete[1] $c(\mathbf{p})$ from $\mathbf{p}$, $\mathbf{p} \setminus c(\mathbf{p})$ is decomposed into several components. In fact, each component is an ordered tree whose root is a neighbor of the center. We fix one component $\mathbf{o}_1$ and label other components counterclockwise $\mathbf{o}_2, \ldots, \mathbf{o}_i$, where i is the number of components of $\mathbf{p} \setminus c(\mathbf{p})$. Let j be the smallest positive integer such that $(\mathbf{o}_{j+1}, \ldots, \mathbf{o}_{j+i}) = (\mathbf{o}_1, \ldots, \mathbf{o}_i)$, where indices are read in mod i. Define $\sigma(\mathbf{p})$ the *symmetry number of* $\mathbf{p}$ to be the value i/j (see Figures 7, 8).

The symmetry number plays an important role in obtaining the size of an orbit as follows:

[1]If $c(\mathbf{p})$ is a vertex, then it is a vertex deletion and if $c(\mathbf{p})$ is an edge, then it is an edge deletion.

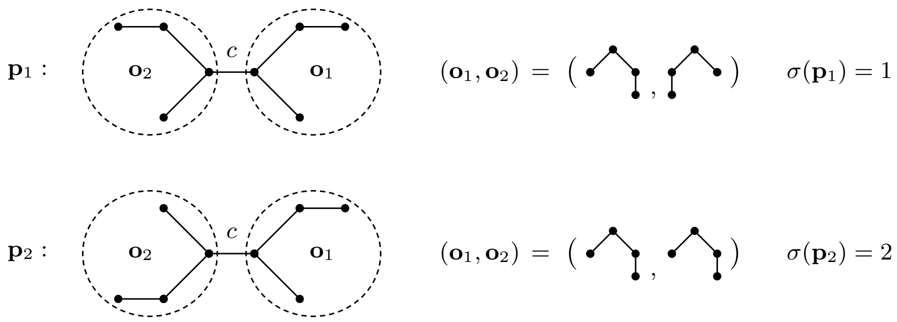


FIGURE 8. $\sigma(\mathbf{p})$: When $c(\mathbf{p})$ is an edge.

Theorem 5. *Given $\mathbf{o}^v \in \mathcal{O}_n^+$, the cardinality of $[\mathbf{o}^v]^+$ is*

$$|[\mathbf{o}^v]^+| = \frac{2\,\epsilon(\mathbf{p})}{\sigma(\mathbf{p})}, \tag{4.3}$$

where $\mathbf{p} = U(\mathrm{d}(\mathbf{o}^v))$, and $\epsilon(\mathbf{p})$ is the number of edges in $\mathbf{p}$.

Proof. Since $\bar{\mathrm{d}}(RL\mathbf{o}^v) = \bar{\mathrm{d}}(\mathbf{o}^v)$ and $U(\mathrm{d}(RL\mathbf{o}^v)) = U(\mathrm{d}(\mathbf{o}^v))$, the size of $[\mathbf{o}^v]^+$ is determined by $\mathbf{p}$. More precisely, the size of $[\mathbf{o}^v]^+$ equals the number of ways of identifying a vertex in $w \in \mathbf{p}$ with $v \in \bar{\mathrm{d}}(\mathbf{o}^v)$. For each vertex w in $\mathbf{p}$, since $\mathbf{p}$ is embedded in the plane, we have $\deg(w)$ distinct ways of identifying w with $v \in \mathrm{d}(\mathbf{o}^v)$. So, if we allow repetition, the number of all possible ways of attaching $\mathbf{p}$ to $\bar{\mathrm{d}}(\mathbf{o}^v)$ is $\sum_{w\in\mathbf{p}} \deg(w) = 2\epsilon(\mathbf{p})$. But by the definition of the symmetry number, each pattern occurs exactly $\sigma(\mathbf{p})$ times. This yields (4.3). □

REFERENCES

[1] F. Bergeron, G. Labelle and P. Leroux, *Combinatorial Species and Tree-like Structures*, Encyclopedia of Mathematics and its Applications vol. 67, Cambridge University Press, 1998.

[2] C. Chauve, *Half of the nodes of Catalan trees are leaves*, personal communication.

[3] B. Dasarathy and C. Yang, A transformation on ordered trees, *Comput. J.*, **23**, No. 2 (1980), 161–164.

[4] N. Dershowitz and S. Zaks, Applied tree enumerations, in *Lecture Notes in Computer Science*, vol. 112, 180–193, Springer-Verlag, Berlin, 1981.

[5] E. Deutsch, An involution on Dyck paths and its consequences, *Discrete Math.*, **204** (1999), 163–166.

[6] F. Harary and E. Palmer, *Graphical Enumeration*, Academic Press, New York, 1973.

[7] D. E. Knuth, *The Art of Computer Programming*, vol. 1 , Fundamental Algorithms, 2nd ed., Addison-Wesley, Reading, Massachusetts, 1973.

[8] S. Seo, A pairing of vertices in ordered trees, *Discrete Math.*, **241** (2001), 471–477.

[9] N. J. A. Sloane, *On-Line Encyclopedia of Integer Sequences*, `http://www.research.att.com/~njas/sequences/`.

DEPARTMENT OF MATHEMATICS, KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY, TAEJON, 305-701, REPUBLIC OF KOREA

E-mail address: `shseo@math.kaist.ac.kr, chattest@math.kaist.ac.kr`

PERIODIC DE BRUIJN TRIANGLES: EXACT AND ASYMPTOTIC RESULTS

B. SHAPIRO, M. SHAPIRO, AND A. VAINSHTEIN

Abstract. We study the distribution of the number of permutations with a given periodic up-down sequence w.r.t. the last entry, find exponential generating functions and prove asymptotic formulas for this distribution.

Résumé. Nous étudions la distribution du nombre de permutations ayant une suite périodique donnée de montées–descentes par rapport leur derniére entrée. Nous en trouvons les séries génératrices et montrons des formules asymptotiques pour cette distribution.

Let $\sigma = (\sigma_1, \dots, \sigma_n)$ be a permutation of length n. We associate with σ its *up-down sequence* (sometimes called the *shape* of σ, or the *signature* of σ) $\mathcal{P}(\sigma) = (p_1, \dots, p_{n-1})$, which is a binary vector of length $n-1$ such that $p_i = 1$ if $\sigma_i < \sigma_{i+1}$ and $p_i = 0$ otherwise. During the last 120 years, many authors have studied the number $\natural_n^{\mathcal{P}}$ of all permutations of length n with a given up-down sequence $\mathcal{P}$. Apparently, for the first time this problem was investigated by D. Andrè [An1, An2], who considered the so-called alternating (or up-down) permutations corresponding to the sequence $\mathcal{P} = (1,0,1,0,\dots) = (10)^*$ and proved that the exponential generating function for the number of such permutations is equal to $\tan x + \sec x$. In [An3] he proved that this number grows asymptotically as $2n!(2/\pi)^{n+1}$.

A general approach to this problem was suggested by MacMahon (see [MM]). This approach leads to determinantal formulas for $\natural_n^{\mathcal{P}}$, rediscovered later by Niven [Ni] from very basic combinatorial considerations. For the relations of this approach to the representation theory of the symmetric group, and for its generalizations, see [Fo, St, BW].

Another, purely combinatorial approach to the same problem was suggested by Carlitz [Ca11]. His general recursive formula for $\natural_n^{\mathcal{P}}$ is rather difficult to use. However, he managed to obtain explicit expressions for the corresponding exponential generating functions for certain *periodic* cases, that is for up-down sequences of the form $\mathcal{P} = (p)^*$, where p is a binary vector of a fixed length called the *period* of $\mathcal{P}$. In [Ca11, Ca2] he considered the case $\mathcal{P} = (1^k0)^*$ and expressed the corresponding generating function via the *Olivier functions of the kth order*

$$\varphi_{k,i}(x) = \sum_{j=0}^{\infty} \frac{x^{jk+i}}{(jk+i)!}, \qquad 0 \le i \le k-1.$$

Another case, $\mathcal{P} = (1^2 0^2)^*$, was considered in [CS1, CS2] and solved via Olivier functions of the fourth order. It follows that asymptotically $\natural_n^{\mathcal{P}}$ in this case grows as $4n!(2/\gamma)^{n+1}$, where $\gamma = 3.7502\dots$ is the smallest positive solution of the equation $\cos t \cosh t + 1 = 0$.

The general periodic problem was solved completely in [CGJN]. As in the two particular cases mentioned above, the answer is expressed via Olivier functions. The techniques used involves matrix Riccati equations, and is rather complicated.

An additional dimension in the problem was introduced by Entringer [En] who studied the distribution of the alternating permutations by the last entry. He observed that the number $\natural_{i,j}$ of alternating permutations of length i whose last entry equals j satisfy the following equations:

$$\begin{aligned} \natural_{i,j} &= \natural_{i,j-1} + \natural_{i-1,j-1}, \quad \natural_{i,1} = 0, \qquad i = 2k, k > 0, \\ \natural_{i,j} &= \natural_{i,j+1} + \natural_{i-1,j}, \qquad \natural_{i,i} = 0, \qquad i = 2k+1, k > 0, \end{aligned} \tag{1}$$

with $\natural_{1,1} = 1$. These equations can be represented graphically as the following triangle displayed on Figure 1.

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 0 & & 1 & & & \\ & & 1 & & 1 & & 0 & & \\ & 0 & & 1 & & 2 & & 2 & \\ 5 & & 5 & & 4 & & 2 & & 0 \end{array}$$

FIGURE 1. The Entringer triangle

Each even row of the triangle starts with 0, and an entry in such a row is equal to the sum of its *left* neighbors in the current and in the previous rows. Similarly, each odd row (except for the first one) ends with 0, and an entry in such a row is equal to the sum of its *right* neighbors in the current and in the previous rows.

The Entringer triangle was studied by many authors. In particular, Arnold [Ar1, Ar2] gave an interpretation of the entries of this triangle in terms of real polynomials with real critical values. Besides, he considered the exponential generating function

$$A(x, y) = \sum_{i \geq 1} \sum_{j=1}^{i} (-1)^{i(i-1)/2} \natural_{i,j} \frac{x^{i-j} y^{j-1}}{(i-j)!(j-1)!}$$

and proved that $A(x, y) = e^y / \cosh(x + y)$. In fact, $A(x, y)$ is the generating function of the *signed Entringer triangle*, which is obtained from the ordinary one by reversing signs in each ith row, where i equals 2 or 3 modulo 4. Observe that the entries $\tilde{\natural}_{i,j}$ of the signed Entringer triangle satisfy relations

$$\tilde{\natural}_{i,j} = \tilde{\natural}_{i,j-1} + \tilde{\natural}_{i-1,j-1} \tag{2}$$

with boundary conditions $\tilde{\natural}_{i,1} = 0$ for $i = 2k$, $\tilde{\natural}_{i,i} = 0$ for $i = 2k+1$, $k > 0$, $\tilde{\natural}_{1,1} = 1$. General triangles satisfying relation (2) with arbitrary boundary conditions were first studied more than 120 years ago by Seidel [Se]. In particular, he proved that the ratio of exponential generating functions for the numbers on the right and on the left sides of such a triangle equals e^x. More recently such triangles where studied, from the combinatorial point of view, in [DV, Du1, Du2]. In particular, it is proved in [DV] that the exponential generating function for a Seidel triangle is equal to $e^y F(x+y)$, where $F(x)$ is the corresponding function for the left side of the triangle.

The case of general up-down sequences was addressed by de Bruijn [dB] (see also [Vi] for another version of the same result). Let $\natural^{\mathcal{P}}_{i,j}$ be the number of permutations of length i whose last entry equals j and whose up-down sequence equals $\mathcal{P} = (p_1, p_2, \dots)$. He proved that these numbers satisfy the following equations:

$$\sharp^{\mathcal{P}}_{i,j} = \sharp^{\mathcal{P}}_{i,j-1} + \sharp^{\mathcal{P}}_{i-1,j-1}, \quad \sharp^{\mathcal{P}}_{i,1} = 0, \qquad \text{if } p_{i-1} = 1$$
$$\sharp^{\mathcal{P}}_{i,j} = \sharp^{\mathcal{P}}_{i,j+1} + \sharp^{\mathcal{P}}_{i-1,j}, \quad \sharp^{\mathcal{P}}_{i,i} = 0, \qquad \text{if } p_{i-1} = 0.$$

with $\sharp^{\mathcal{P}}_{1,1} = 1$. Evidently, for $\mathcal{P} = (10)^*$ one gets the Entringer relations (1). As before, these equations can be represented graphically as a triangle, and the direction in which one has to advance along the rows of the triangle is governed by the sequence $\mathcal{P}$. We call this triangle the *de Bruijn triangle* corresponding to the up-down sequence $\mathcal{P}$. A de Bruijn triangle is said to be *periodic* if the corresponding up-down sequence is periodic.

The *signed de Bruijn triangle* is obtained from the ordinary de Bruijn triangle by multiplying its ith row by $\varepsilon_i = (-1)^{p_1+p_2+\cdots+p_{i-1}+i-1}$, $i \geq 2$. The corresponding exponential generating function is defined by

$$F^{\mathcal{P}}(x,y) = \sum_{i\geq 1}\sum_{j=1}^{i} \varepsilon_i \sharp^{\mathcal{P}}_{ij} \frac{x^{i-j}y^{j-1}}{(i-j)!(j-1)!}.$$

Let $\mathcal{P}$ be a periodic up-down sequence with period p of length m, and let $i_1 < i_2 < \cdots < i_r$ be the locations of zeros in p. Without loss of generality we assume that $i_r = m$ (otherwise we consider instead of $\mathcal{P}$ the up-down sequence $\bar{\mathcal{P}} = (\bar{p})^*$, where $\bar{p}_i = 1 - p_i$ for $1 \leq i \leq m$; evidently, the de Bruijn triangle for $\bar{\mathcal{P}}$ is obtained from that for $\mathcal{P}$ by the reflection in the main diagonal).

Theorem 1. *The exponential generating function of the signed periodic de Bruijn triangle corresponding to the up-down sequence $\mathcal{P}$ is given by $F^{\mathcal{P}}(x,y) = e^y f^{\mathcal{P}}(x+y)$, where*

$$f^{\mathcal{P}}(t) = \frac{\det \bar{M}^{\mathcal{P}}(t)}{\det M^{\mathcal{P}}(t)}$$

and $M^{\mathcal{P}}(t)$ and $\bar{M}^{\mathcal{P}}(t)$ are $r \times r$ matrices

$$M^{\mathcal{P}}(t) = \begin{pmatrix} \varphi_{m,0} & t^{i_1}\varphi_{m,n-i_1} & t^{i_2}\varphi_{m,n-i_2} & \cdots & t^{i_{r-1}}\varphi_{m,n-i_{r-1}} \\ t^{-i_1}\varphi_{m,i_1} & \varphi_{m,0} & t^{i_1}\varphi_{m,n-i_1} & \cdots & t^{i_{r-2}}\varphi_{m,n-i_{r-2}} \\ t^{-i_2}\varphi_{m,i_2} & t^{-i_1}\varphi_{m,i_1} & \varphi_{m,0} & \cdots & t^{i_{r-3}}\varphi_{m,n-i_{r-3}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t^{-i_{r-1}}\varphi_{m,i_{r-1}} & t^{-i_{r-2}}\varphi_{m,i_{r-2}} & t^{-i_{r-3}}\varphi_{m,i_{r-3}} & \cdots & \varphi_{m,0} \end{pmatrix}$$

and

$$\bar{M}^{\mathcal{P}}(t) = \begin{pmatrix} 1 & t^{i_1} & t^{i_2} & \cdots & t^{i_{r-1}} \\ t^{-i_1}\varphi_{m,i_1} & \varphi_{m,0} & t^{i_1}\varphi_{m,n-i_1} & \cdots & t^{i_{r-2}}\varphi_{m,n-i_{r-2}} \\ t^{-i_2}\varphi_{m,i_2} & t^{-i_1}\varphi_{m,i_1} & \varphi_{m,0} & \cdots & t^{i_{r-3}}\varphi_{m,n-i_{r-3}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t^{-i_{r-1}}\varphi_{m,i_{r-1}} & t^{-i_{r-2}}\varphi_{m,i_{r-2}} & t^{-i_{r-3}}\varphi_{m,i_{r-3}} & \cdots & \varphi_{m,0} \end{pmatrix}$$

with $\varphi_{m,j} = \varphi_{m,j}(t)$.

The proof of the theorem follows easily from the above mentioned properties of Seidel triangles and an evident identity $e^x = \sum_{i=0}^{m-1} \varphi_{m,i}(x)$. As a corollary we get the result of [CGJN] cited above.

Moreover, the same techniques allows to obtain generating functions for other Seidel triangles with periodic boundary conditions, such as the triangle for Genocchi numbers (see [DV]). It can be also extended to pairs of Seidel triangles with periodic boundary conditions, such as Arnold triangles $L(\beta)$ and $R(\beta)$ for Springer numbers (see [Ar2, Du2]), thus recovering several combinatorial results obtained in [Sp, Ar2].

Let us now consider the asymptotic behavior of the numbers $\sharp^{\mathcal{P}}_{i,j}$. It was observed without a proof in [Ar2, p. 18] that the even rows of the ordinary Entringer triangle approximate, after an appropriate normalization, the function $\sin x$ on the interval $[0,\pi/2]$, while the odd rows approximate $\cos x$. Exact statements with the first two correction terms can be found in [SY].

We generalize this result to arbitrary periodic de Bruijn triangles. Consider the space $\mathcal{L}_2$ of all L_2-functions on the interval $[0,\pi/2]$. Let $\mathfrak{S}_k$ be the linear space of all sequences of real numbers of length k. We define the standard inclusion $i_k\ \mathfrak{S}_k \hookrightarrow \mathcal{L}_2$ as the linear map sending a sequence $\{a_1,\dots,a_k\}$ to the function whose value equals a_j on the interval $[\frac{\pi(j-1)}{2k},\frac{\pi j}{2m})$.

For any l, $0\le l\le m-1$, let $\mathcal{N}^{\mathcal{P}}_{k,l}$ denote the sequence $\{\sharp^{\mathcal{P}}_{kn+l,j}\}\in\mathfrak{S}_{km+l}$, and let $\xi^{\mathcal{P}}_{k,l}=ci_{km+l}(\mathcal{N}^{\mathcal{P}}_{k,l})$, where c is a normalizing coefficient depending on k and l and chosen in such a way that $\int_0^{\pi/2}\xi^{\mathcal{P}}_{k,l}(\theta)d\theta=1$.

Theorem 2. *For any l, $0\le k\le m-1$, the sequence of functions $\xi^{\mathcal{P}}_{k,l}$ converges in the L_2-metric to the normalized first eigenfunction $\Xi^{\mathcal{P}}_l$ of the two-point spectral problem*

$$z^{(m)}=(-1)^r\lambda^m z$$

with m homogeneous boundary conditions

$$\begin{array}{ll} z^{(i)}(0)=0 & \text{if } p_{l+i-1}=1,\\ z^{(i)}(\pi/2)=0 & \text{if } p_{l+i-1}=0,\end{array}$$

where r is the number of zeros in the period p.

The first eigenvalue of the above spectral problem is the minimal positive solution of the equation

$$\det\tilde{M}^{\mathcal{P}}(\varepsilon\lambda\pi/2)=0,$$

where ε is the mth primitive root of $(-1)^r$ and

$$\tilde{M}^{\mathcal{P}}(t)=\begin{pmatrix}\varphi_{m,0} & \varphi_{m,i_2-i_1} & \varphi_{m,i_3-i_1} & \cdots & \varphi_{m,i_r-i_1}\\ \varphi_{m,i_1-i_2} & \varphi_{m,0} & \varphi_{m,i_3-i_2} & \cdots & \varphi_{m,i_r-i_2}\\ \varphi_{m,i_1-i_3} & \varphi_{m,i_2-i_3} & \varphi_{m,0} & \cdots & \varphi_{m,i_r-i_3}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \varphi_{m,i_1-i_r} & \varphi_{m,i_2-i_r} & \varphi_{m,i_3-i_r} & \cdots & \varphi_{m,0}\end{pmatrix}$$

with $\varphi_{m,j}=\varphi_{m,j}(t)$.

The eigenfunction $\Xi^{\mathcal{P}}_l$ is normalized by the condition $\int_0^{\pi/2}\Xi^{\mathcal{P}}_l(\theta)d\theta=1$, and the indices are understood modulo m.

In particular, for the Entringer numbers $\sharp_{i,j}$ one gets the sine law of [SY]:

$$\begin{cases}\lim_{k\to\infty,\ \frac{j}{2k+1}\to u}\sharp_{2k+1,j}=2(2k)!\left(\frac{2}{\pi}\right)^{2k+1}\cos\frac{\pi u}{2},\\ \lim_{k\to\infty,\ \frac{j}{2k}\to u}\tilde{\sharp}_{2k,j}=2(2k-1)!\left(\frac{2}{\pi}\right)^{2l}\sin\frac{\pi u}{2}.\end{cases}$$

The starting point of this research was the result by the first and the third author that the numbers $\sharp^{\mathcal{P}}_{i,j}$ for $\mathcal{P}=(1^20^2)^*$ arise naturally in counting real rational functions of a certain type, see [SV]. The authors are grateful to Max–Planck–Institut für Mathematik, Bonn for its hospitality in September 2000 and to the Royal Institute of Technology, Stockholm for the financial support of the visit of A. V. to Stockholm in July-August 2001. Sincere thanks goes to Prof. Harold Shapiro from The Royal Institute of Technology for his help with functional analysis in the proof of Theorem 2.

References

[An1] D. André, Développments de $\sec x$ et de $\tan x$, *C.R. Heb.* **88** (1879), 965–967.

[An2] D. André, Sur les permutations alternées, *J. Math. Pures Appl.* **7** (1881), 167–184.

[An3] D. André, Probabilité pour qu'une permutation donnée de n lettres soit une permutation alternée, *C.R. Heb.* **97** (1883), 983–984.

[Ar1] V. Arnold, Bernoulli–Euler updown numbers associated with functions singularities, their combinatorics and arithmetics, *Duke Math. J.* **63** (1991), 537–555.

[Ar2] V. Arnold, The calculus of snakes and the combinatorics of Bernoulli, Euler and Springer numbers of Coxeter groups, *Russ. Math. Surv.* **47** (1992), 1–51.

[BW] A. Björner and M. Wachs, Generalized quotients in Coxeter groups, *Trans. Amer. Math. Soc.* **308** (1988), 1–37.

[Ca11] L. Carlitz, Permutations with prescribed pattern, *Math. Nachr.* **58** (1973), 31–53.

[Ca2] L. Carlitz, Generating functions for a special class of permutations, *Proc. Amer. Math. Soc.* **47** (1975), 251–256.

[CS1] L. Carlitz and R. Scoville, Enumeration of rises and falls by position, *Discr. Math.* **5** (1973), 45–59.

[CS2] L. Carlitz and R. Scoville, Generating functions for certain types of permutations, *J. Comb. Theory A* **18** (1975), 262–275.

[CGJN] C. Collins, I. Goulden, D. Jackson, and O. Nierstrasz, A combinatorial application of matrix Riccati equations and their q-analogue, *Discr. Math.* **36** (1981), 139–153.

[dB] N. de Bruijn, Permutations with given ups and downs, *Nieuw Arch. Wisk.* **18** (1970), 61–65.

[Du1] D. Dumont, Matrices d'Euler–Seidel, *Sem. Lothar. Comb.* **5** (1981), B05c.

[Du2] D. Dumont, Further triangles of Seidel–Arnold type and continued fractions related to Euler and Springer numbers, *Adv. Appl. Math.* **16** (1995), 275–296.

[DV] D. Dumont and G. Viennot, A combinatorial interpretation of the Seidel generation of Genocchi numbers, *Ann. Discr. Math.* **6** (1980), 77–87.

[En] R. Entringer, A combinatorial interpretation of the Euler and Bernoulli numbers, *Nieuw Arch. Wisk.* **14** (1966), 241–246.

[Fo] H. Foulkes, Enumeration of permutations with prescribed up-down and inversion sequences, *Discr. Math.* **15** (1976), 235–252.

[MM] P. MacMahon, Second memoir on the compositions of numbers, *Philos. Trans. Roy. Soc. London (A)* **207** (1908), 65–134.

[Ni] I. Niven, A combinatorial problem of finite sequences, *Nieuw Arch. Wisk.* **16** (1968), 116–123.

[Se] L. Seidel, Über eine einfache Entstehungsweise der Bernoullischen Zahlen und einiger verwandten Reihen, *Sitzungsber. Münch. Akad.* **4** (1877), 157–187.

[Sp] T. Springer, Remarks on a combinatorial problem, *Nieuw Arch. Wisk.* **19** (1971), 30–36.

[St] R. Stanley, Binomial posets, Möbius inversion, and permutation enumeration, *J. Comb. Theory A*
20 (1976), 336–356.

[SY] G. Simons and Y-C. Yao, Asymptotic Sine laws arising from alternating random permutations and sequences, *Random Structures and Algorithms* **8** (1996), 75–95.

[SV] B. Shapiro and A. Vainshtein, Counting real rational functions with all real critical values, (in preparation).

[Vi] G. Viennot, Permutations ayant une forme donne, *Discrete Math.* **26** (1979), 279–284.

Department of Mathematics, University of Stockholm, S-10691, Sweden
E-mail address: `shapiro@matematik.su.se`

Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA
E-mail address: `mshapiro@math.msu.edu`

Department of Mathematics and Department of Computer Science, University of Haifa, Mount Carmel, 31905 Haifa, Israel
E-mail address: `alek@cs.haifa.ac.il`

INEQUALITIES IN PRODUCTS OF MINORS OF TOTALLY NONNEGATIVE MATRICES

MARK SKANDERA

ABSTRACT. Let $\Delta_{I,I'}$ be the minor of a matrix which corresponds to row set I and column set I'. We give a characterization of the inequalities of the form

$$\Delta_{I,I'}\Delta_{K,K'} \leq \Delta_{J,J'}\Delta_{L,L'}$$

which hold for all totally nonnegative matrices. This generalizes a recent result of Fallat, Gekhtman, and Johnson.

RÉSUMÉ. Soit $\Delta_{I,I'}$ le mineur d'une maitrice qui correspond a l'ensemble I des files et l'ensemble I' des colonnes. Nous donons une caractérisation des inégalités de la forme

$$\Delta_{I,I'}\Delta_{K,K'} \leq \Delta_{J,J'}\Delta_{L,L'}$$

qui sont vrais pour chacque maitrice totalement positive. Ce résultat généralise un résultat récent de Fallat, Gekhtman, et Johnson.

1. Introduction

Let A be an $n \times n$ matrix and let I and I' be subsets of $[n] = \{1, \ldots, n\}$. We define $\Delta_{I,I'}$, the (I, I') *minor* of A, to be the determinant of the submatrix of A corresponding to rows I and columns I'. A matrix is called *totally nonnegative* if each of its minors is nonnegative. While this definition may be applied to nonsquare matrices, we will restrict our attention to *square* totally nonnegative matrices. It is easy to see that the concatenation of a row or column of zeros to a totally nonnegative matrix introduces no negative minors.

One setting in which totally nonnegative matrices arise is in the counting of paths in directed graphs. Let us define a *planar network of order* n to be a planar acyclic directed graph $G = (V, E)$ in which $2n$ vertices are distinguished as n sources and n sinks. We will assume that all sources and sinks are boundary vertices, labeled cyclically (counterclockwise) as $s_1, \ldots, s_n, t_n, \ldots, t_1$. We will use S and T to denote the sources and sinks of a planar network, and S_I and T_I to denote the subsets of sources and sinks corresponding to an index set I,

$$S_I = \{s_i \,|\, i \in I\},$$
$$T_I = \{t_i \,|\, i \in I\}.$$

In figures we will draw sources on the left of a planar network and sinks on the right. The orientations of edges will be understood to be from left to right. (See Figure 1.1.)

Let $G = (V, E)$ be a planar network of order n and let each edge e of G be labeled by a positive real *weight* w_e. Define the weight of a *path* from s to t to be product of weights of edges along this path, and define the *weighted path matrix* of G to be the matrix $A = [a_{ij}]$, where a_{ij} is the sum of weights of all paths from source s_i to sink t_j. Such a weighted path matrix is always totally nonnegative. In particular, the minors of such a matrix have an interpretation in terms of *families* $\pi = (\pi_1, \ldots, \pi_n)$ of paths in G from sources to sinks.

Supported by the NSF.

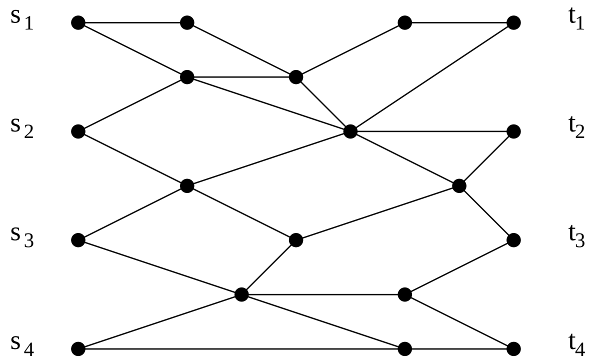


FIGURE 1.1. A planar network of order 4

Defining the weight of a path family to be the product of weights of its n paths, we have the following result.

Theorem 1.1. *Let A be the weighted path matrix of a planar network G. The minor $\Delta_{I,I'}$ of A is equal to the sum of weights of all nonintersecting path families which connect the sources indexed by I to the sinks indexed by I'. In particular, A is totally nonnegative.*

The first proofs of this fact were given by Karlin and MacGregor [7], and by Lindström [8]. Since then, Theorem 1.1 has been used to prove that matrices arising in various situations are totally nonnegative. (See for example [5] and [6].)

By results of Whitney [12], Loewner [9], Cryer [2], and Brenti [1], the converse of Theorem 1.1 is true as well. This result was first stated in [1].

Theorem 1.2. *Every $n \times n$ totally nonnegative matrix A is the weighted path matrix of a planar network G of order n.*

While the planar network G in Theorem 1.2 is not uniquely determined, it is easy to see that G may be chosen so that each source (sink) has indegree zero and outdegree one (outdegree zero and indegree one). Furthermore, it may always be chosen to be of a canonical form [1]. (See also [4].)

Some recent interest in totally nonnegative matrices involves polynomial functions in n^2 variables $\{x_{ij} \,|\, i,j \in [n]\}$ which evaluate to nonnegative numbers whenever we set $x_{ij} = a_{ij}$ for a totally nonnegative matrix $A = [a_{ij}]$ of size at least $n \times n$. Let us call such polynomials *totally nonnegative*. In particular, Lusztig [10] has shown that the elements of the dual canonical basis of the coordinate ring of GL_n are totally nonnegative. In order to better understand this basis, which currently has no simple description, one might hope to characterize all totally nonnegative polynomials, or at least to study some subset of these polynomials.

One such subset, discovered recently by Fallat et. al. [3], may be described in terms of principal minors. An example is the totally nonnegative polynomial

$$\Delta_{\{1,3\}\{1,3\}}\Delta_{\{2,4\}\{2,4\}} - \Delta_{\{1,4\}\{1,4\}}\Delta_{\{2,3\}\{2,3\}}. \tag{1.1}$$

In other words, the inequality

$$\Delta_{\{1,4\}\{1,4\}}\Delta_{\{2,3\}\{2,3\}} \leq \Delta_{\{1,3\}\{1,3\}}\Delta_{\{2,4\}\{2,4\}} \tag{1.2}$$

holds for all totally nonnegative matrices of size at least 4×4. This and sixteen similar inequalities are shown as a poset P in Figure 1.2. Each element of the form I-K in P

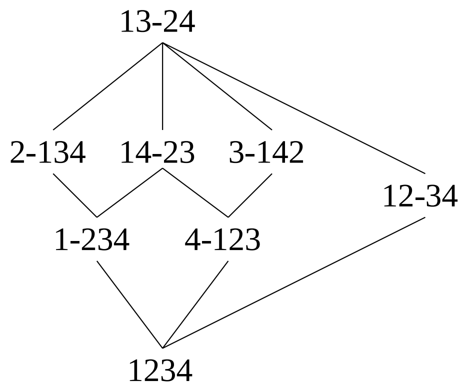


FIGURE 1.2. A partial order on partitions of $\{1, 2, 3, 4\}$ into at most two blocks.

represents the product $\Delta_{I,I}\Delta_{K,K}$, and the minimal element represents the determinant $\Delta_{\{1,2,3,4\}}$. The relation $<_P$ is defined by I-$K <_P J$-L whenever the inequality

$$\Delta_{I,I}\Delta_{K,K} \leq \Delta_{J,J}\Delta_{L,L}$$

holds for all totally nonnegative matrices. Thus the relation 14-23 $<_P$ 13-24 represents the inequality (1.2).

This poset raises the question of finding a more general class of inequalities in products of minors of totally nonnegative matrices. Let us consider products of *non-principal* minors of the form $\Delta_{I,I'}\Delta_{\overline{I},\overline{I'}}$, where $\overline{I} = [n] \smallsetminus I$.

Question 1.1. What conditions on four subsets I, I', J, J' of $[n]$ imply the inequality

$$\Delta_{I,I'}\Delta_{\overline{I},\overline{I'}} \leq \Delta_{J,J'}\Delta_{\overline{J},\overline{J'}} \tag{1.3}$$

for all totally nonnegative matrices of size at least $n \times n$?

By Theorem 1.2, we may interpret the products of minors which occur in the inequality (1.3) in terms of families of paths in planar networks. The combinatorial interpretation of the product $\Delta_{I,I'}\Delta_{\overline{I},\overline{I'}}$ is quite simple.

Observation 1.3. *Let A be an $n \times n$ totally nonnegative matrix, and let G be any planar network whose weighted path matrix is A. Then the product $\Delta_{I,I'}\Delta_{\overline{I},\overline{I'}}$ of minors of A is equal to the weighted sum of all path families $\pi = (\pi_1, \ldots, \pi_n)$ in G with the following properties.*

1. *Each path connects a source in S_I to a sink in $T_{I'}$ or a source in $S_{\overline{I}}$ to a sink in $T_{\overline{I'}}$.*
2. *The paths from S_I to $T_{I'}$ are pairwise vertex disjoint, as are the paths from $S_{\overline{I}}$ to $T_{\overline{I'}}$.*

We will refer to the combination of the two conditions in Observation 1.3 as a *binary crossing rule* or more specifically as *the (I, I') crossing rule.* We will refer to source-to-sink paths and to families of these simply as *paths* and *path families.*

Using Observation 1.3 we may reformulate Question 1.1.

Question 1.2. What conditions on four subsets I, I', J, J' of $[n]$ imply that for each planar network G of order n, the weighted sum of path families in G which obey the (I, I') crossing rule is less than or equal to the weighted sum of path families in G which obey the (J, J') crossing rule?

In Section 2 we will examine a special case of Question 1.2 which will lead to a proof of our main theorem in Section 3. This theorem, which generalizes recent results of Fallat et. al. [3], characterizes all inequalities of the form

$$\Delta_{I,I'}\Delta_{\overline{I},\overline{I'}} \leq \Delta_{J,J'}\Delta_{\overline{J},\overline{J'}}$$

which hold for all totally nonnegative matrices. A corollary provides a combinatorial interpretation of the corresponding totally nonnegative polynomials

$$\Delta_{J,J'}\Delta_{\overline{J},\overline{J'}} - \Delta_{I,I'}\Delta_{\overline{I},\overline{I'}}.$$

In Section 4 we will show that our main result essentially characterizes even the more general class of inequalities of the form

$$\Delta_{I,I'}\Delta_{K,K'} \leq \Delta_{J,J'}\Delta_{L,L'}$$

which hold for all totally nonnegative matrices.

2. Path families which cover a planar network

A path family π in a planar network G will not in general use all of the edges of G. We will say that π *covers* G if it does use all of the edges. Since the weight of a path family which covers G is equal to the product of all edge weights in G (with multiplicities for multiply covered edges), we may calculate the weighted sum of path families which cover G simply by *counting* path families which cover G. This suggests the following specialization of Question 1.2.

Question 2.1. What conditions on four subsets I, I', J, J' of $[n]$ imply that for each planar network G of order n, the number of path families which *cover* G and obey the (I, I') crossing rule is less than or equal to the number of path families which cover G and obey the (J, J') crossing rule?

Questions 1.2 and 2.1 are more closely than one might imagine: in fact they have the same answer. The easiest way to see this is to consider three related families of graphs and maps between these.

Let $\mathcal{G}_1(n)$ be the family of planar networks of order n which may be expressed as a union of n source-to-sink paths, no three of which share a vertex. These are precisely the planar networks which can be covered by path families which obey a binary crossing rule. Note that the vertices in such planar networks have indegree and outdegree bounded by two.

Let $\mathcal{G}_2(n)$ be the family of undirected graphs with at least n connected components which are paths, and arbitrarily many connected components which are cycles. To define our first map $\phi : \mathcal{G}_1(n) \to \mathcal{G}_2(n)$, let $G = (V, E)$ be a graph in $\mathcal{G}_1$, and create $\phi(G) = G' = (V', E')$ as follows.

(1) For each vertex $x \in V$ which has indegree and/or outdegree two, create vertices x^- and x^+ in V'. Otherwise create vertex x in V'.
(2) For each edge (x, y) in E, create the unique edge (x^+, y^-) or (x, y^-) or (x^+, y) or (x, y) which can be defined in E'.

Figure 2.1 shows a planar network G and Figure 2.2 shows the graph $\phi(G)$. As we have used S and T to refer to the sources and sinks in G, we will again use S and T to refer to the corresponding vertices in $\phi(G)$.

It is not hard to see that the components of $\phi(G)$ are of three types: paths whose two endvertices are sources or sinks (there are necessarily n of these), paths containing no source or sink, and cycles. Examination of these components immediately gives a formula for the number of path families which cover G and obey a binary crossing rule.

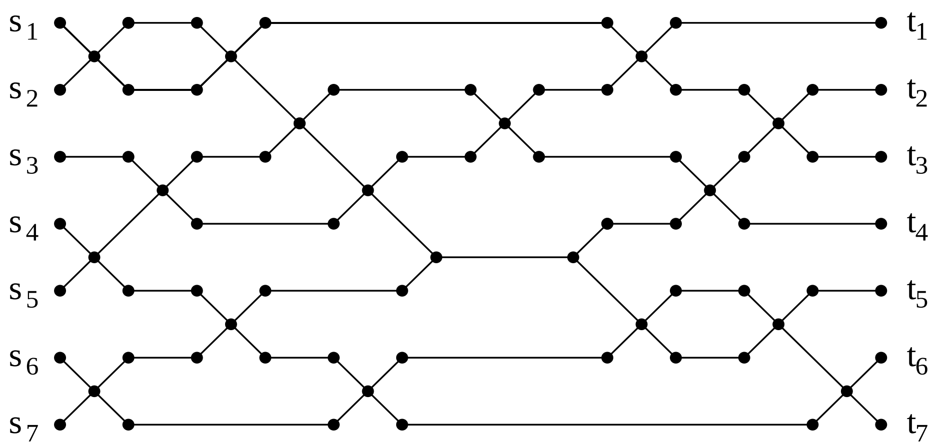


FIGURE 2.1. A planar network G which is a union of 7 paths.

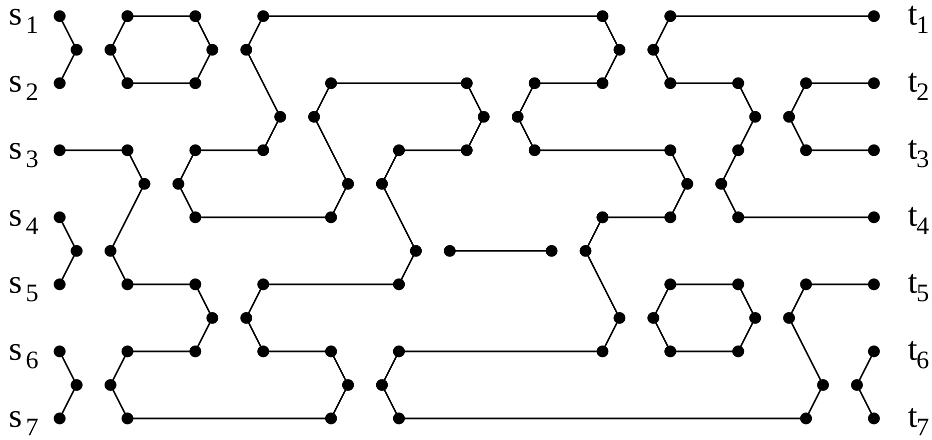


FIGURE 2.2. The graph $\phi(G)$.

Proposition 2.1. *Let G be a planar network in $\mathcal{G}_1(n)$, let k be the number of cyclic components of $\phi(G)$, and let I, I' be two subsets of $[n]$. Then the number of path families which cover G and obey the (I, I') crossing rule is 2^k if no path component of $\phi(G)$ contains two vertices in $S_I \cup T_{\overline{I'}}$. It is zero otherwise.*

Proof. Let G be a graph in $\mathcal{G}_1(n)$. Counting the number of path families which cover G and obey the (I, I') crossing rule is equivalent to counting the edge colorings of G which satisfy the following conditions.

(1) Each edge incident upon a source in S_I or a sink in $T_{I'}$ is colored blue; edges incident upon the remaining sources and sinks are colored red.
(2) Each edge in the graph must be colored red, blue, or red and blue.
(3) For each vertex v which is not a source or a sink, the number of red (blue) paths entering v must equal the number of red (blue) paths leaving v.

These rules ensure that $|I|$ nonintersecting blue paths lead from S_I to $T_{I'}$, $n-|I|$ nonintersecting red paths lead from $S_{\overline{I}}$ to $T_{\overline{I'}}$, and the union of these n paths is equal to the edge set of G.

Consider any component H of $\phi(G)$ and let F be the corresponding set of edges in G. If H is a path containing no vertex in $S \cup T$, then each edge in F is colored both red and blue. If H is a cycle, then F may be colored in two ways, each of which is consistent with the above rules. Now suppose H is a path containing two vertices in $S \cup T$. If exactly one of these two vertices belongs to $S_I \cup T_{I'}$, then there is exactly one coloring of F which is consistent with the above rules, otherwise there is none. □

Proposition 2.1 implies a surprising relationship between the numbers of path families which cover a planar network and obey one of two different crossing rules.

Corollary 2.2. *Let G be a planar network of order n and let I, I', J, J' be subsets of $[n]$. If the number of path families which cover G and obey the (I, I') crossing rule is not equal to the number of path families which cover G and obey the (J, J') crossing rule, then one of these numbers is zero.*

Thus, the comparison of weighted sums of path families which cover a planar network G and obey two different crossing rules reduces to a problem of determining the *existence* of such families. By Proposition 2.1, this in turn reduces to the problem of deciding if n paths in the graph $\phi(G)$ determine a perfect matching of $S_I \cup T_{\overline{I'}}$ with $S_{\overline{I}} \cup T_{I'}$. We therefore introduce a third class of graphs which records only this information.

Let $\mathcal{G}_3(2n)$ be the family of graphs on $2n$ vertices labeled $1, \ldots, 2n$, and n edges which can be drawn so that if the vertices lie in increasing order on a horizontal line, then the edges form n noncrossing arcs above them. It is well known that the cardinality of $\mathcal{G}_3(2n)$ is C_n, the nth Catalan number. (See [11, p. 222].) To define a second map $\psi : \mathcal{G}_1(n) \to \mathcal{G}_3(2n)$, let $G = (V, E)$ be a graph in $\mathcal{G}_1(n)$, let $G' = (V', E')$ be $\phi(G)$ and create $\psi(G) = G'' = (V'', E'')$ as follows.

(1) For each source s_i in V create vertex i in V''.
(2) For each sink t_i in V create vertex $2n+1-i$ in V''.
(3) For each path component of G' with end vertices in $S \cup T$, connect the corresponding two vertices in V'' with an arc.

Figures 2.3 and 2.4 show two embeddings of the graph $\psi(G)$, where G is the graph in Figure 2.1.

The map ψ allows us to state a simple criterion for comparing path families which cover a given graph in $\mathcal{G}_1(n)$ and obey two different crossing rules.

Observation 2.3. *Let G be a graph in $\mathcal{G}_1(n)$, let I, I', J, J' be subsets of $[n]$, and define the sets*

$$I'' = I \cup \{2n+1-i \,|\, i \in \overline{I'}\}, \qquad \overline{I''} = [2n] \smallsetminus I'',$$
$$J'' = J \cup \{2n+1-j \,|\, j \in \overline{J'}\}, \qquad \overline{J''} = [2n] \smallsetminus J''.$$

More path families which cover G obey the (I, I') crossing rule than the (J, J') crossing rule if and only if $\psi(G)$ is a perfect matching of I'' with $\overline{I''}$ and it is not a perfect matching of J'' with $\overline{J''}$.

It is easy to show that the map ψ is surjective. One proof of this is given by the following map $\rho : \mathcal{G}_3(2n) \to \mathcal{G}_1(n)$, which identifies a canonical representative in the preimage $\psi^{-1}(H)$ of each graph H in $\mathcal{G}_3(2n)$.

(1) Draw a vertical representation of H and label the vertices $s_1, \ldots, s_n, t_n, \ldots, t_1$ as in Figure 5(a).

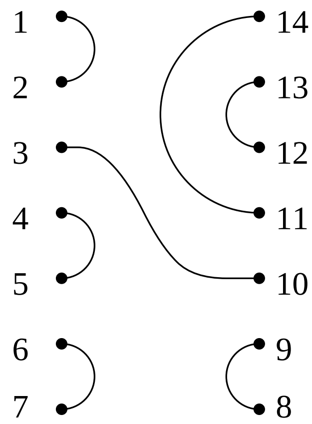


FIGURE 2.3. The graph $\psi(G)$, drawn similarly to $\phi(G)$.

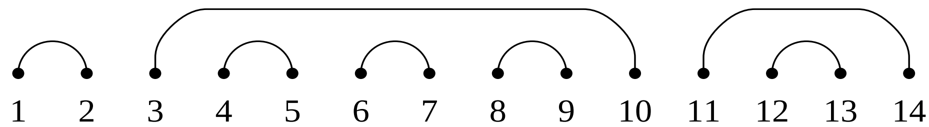
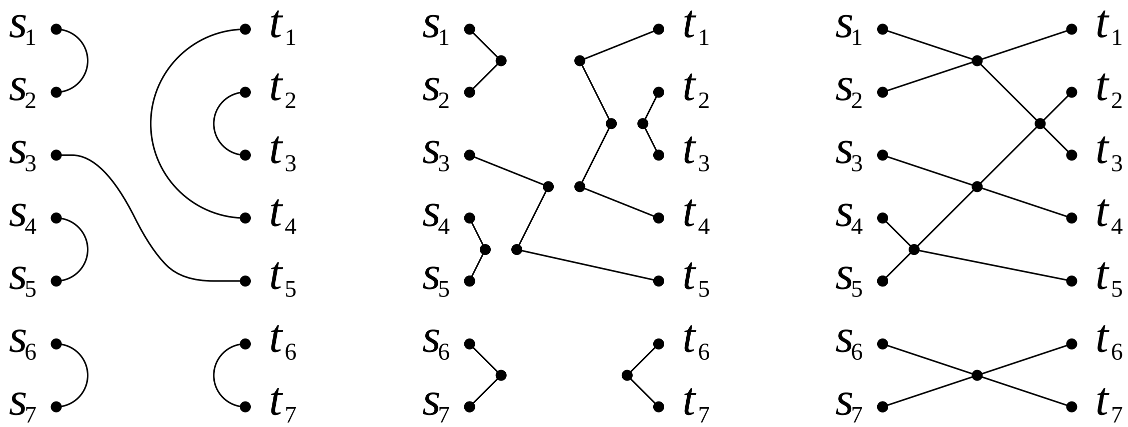


FIGURE 2.4. The graph $\psi(G)$, drawn with vertices on a horizontal line and edges above this line.



(a) H (b) (c) $\rho(H)$

FIGURE 2.5. A graph H in $\mathcal{G}_3(14)$ and the creation of the graph $\rho(H)$ in $\mathcal{G}_2(7)$.

(2) For each edge e of H whose endvertices have indices which differ by k, replace e by a zig-zag path of $k+1$ straight edges and k new vertices as in Figure 5(b).
(3) Identify pairs of new vertices as in Figure 5(c) so that each vertex in the resulting graph has degree one or four.

It is easy to see that the graph created in step 2 is $\phi(\rho(H))$, and that all pairs of vertices which are identified in step 3 belong to two different components of $\phi(\rho(H))$. This important property proves the equivalence of Question 2.1 to Questions 1.1 and 1.2.

Lemma 2.4. *Let H be a graph in $\mathcal{G}_3(2n)$, and let I, I' be subsets of $[n]$. If $\rho(H)$ contains any path family which obeys the (I, I') crossing rule, then this path family covers $\rho(H)$ and is unique.*

Proof. Let π be a path family in $\rho(H)$ which obeys the (I, I') crossing rule but does not cover $\rho(H)$, and color the edges of $\rho(H)$ red and/or blue as in the proof of Proposition 2.1. Construct the graph $\phi(\rho(H))$, and color its edges as those of $\rho(H)$ are colored. All components of $\phi(\rho(H))$ are paths, some containing uncolored edges. For each component containing at least one uncolored edge, consider the vertex labels of the corresponding arc in H, and choose a component C which minimizes the difference between these two labels. By the choice of C and the fact that no component of $\phi(\rho(H))$ contains two vertices x^+ and x^- for any vertex x of $\rho(H)$, we have that some edge of C is uncolored, one of its neighboring edges is colored twice, the next edge is uncolored, etc. This implies that a terminal edge of C is colored twice or not at all, contradicting our assumption that the edge coloring determines a path family in $\rho(H)$ which obeys the (I, I') crossing rule. $\square$

3. MAIN THEOREM

The maps ϕ and ψ defined in Section 2 reduce the problem of comparing products $\Delta_{I,I'}\Delta_{\overline{I},\overline{I'}}$, $\Delta_{J,J'}\Delta_{\overline{J},\overline{J'}}$ of minors in $n \times n$ totally nonnegative matrices to a problem of examining perfect matchings in graphs on $2n$ vertices. We will now demonstrate that this problem reduces to examining subintervals of $[2n]$ and their intersections with the two sets I'', J'' defined in Observation 2.3. In particular, we will define an interval $[b_1, b_2] = \{b_1, b_1 + 1, \dots, b_2\}$ to be *even* if its cardinality is even, and we will show that the inequality

$$\Delta_{I,I'}\Delta_{\overline{I},\overline{I'}} \leq \Delta_{J,J'}\Delta_{\overline{J},\overline{J'}}$$

holds for all totally nonnegative matrices if and only if the sets I'', J'' satisfy the inequality

$$\max\{|B \cap I''|, |B \smallsetminus I''|\} \geq \max\{|B \cap J''|, |B \smallsetminus J''|\} \tag{3.1}$$

for all even subintervals B of $[2n]$.

Let us first consider the circumstances under which the inequality (3.1) fails to hold for some subinterval B of $[2n]$.

Lemma 3.1. *Let I'' and J'' be two subsets of $[2n]$ and let $B = [b_1, b_2]$ be a minimal subinterval of $[2n]$ which satisfies*

$$\max\{|B \cap I''|, |B \smallsetminus I''|\} < \max\{|B \cap J''|, |B \smallsetminus J''|\}. \tag{3.2}$$

Then the numbers b_1 and b_2 both belong to J'', or both belong to $\overline{J''}$, while exactly one of these numbers belongs to I. Furthermore, the cardinalities $|[b_1+1, b_2-1] \cap I''|$, $|[b_1+1, b_2-1] \smallsetminus I''|$, $|[b_1 + 1, b_2 - 1] \cap J''|$, and $|[b_1 + 1, b_2 - 1] \smallsetminus J''|$ are all equal to $\frac{1}{2}(|B| - 2)$. In particular, B is even.

Proof. Omitted. $\square$

Armed with this technical lemma, we may now state and prove our main result.

Theorem 3.2. *Let I, I', J, J' be subsets of $[n]$ and define the subsets I'', J'' of $[2n]$ by*

$$I'' = I \cup \{2n + 1 - i \mid i \in \overline{I'}\},$$

$$J'' = J \cup \{2n + 1 - j \mid j \in \overline{J'}\}.$$

Then the following statements are equivalent.

(1) In each totally nonnegative matrix of size at least $n \times n$, the minors $\Delta_{I,I'}$, $\Delta_{\overline{I},\overline{I'}}$, $\Delta_{J,J'}$, and $\Delta_{\overline{J},\overline{J'}}$ satisfy

$$\Delta_{I,I'}\Delta_{\overline{I},\overline{I'}} \leq \Delta_{J,J'}\Delta_{\overline{J},\overline{J'}}.$$

(2) In each planar network G of order n, the weighted sum of path families in G which obey the (I, I') crossing rule is less than or equal to that of the path families in G which obey the (J, J') crossing rule.

(3) For each planar network $G = (V, E)$ in which the edges are weighted by formal variables $w_1, \dots, w_m$, the difference between the weighted sum of path families in G which obey the (J, J') crossing rule and that of path families in G which obey the (I, I') crossing rule is a subtraction-free polynomial in $w_1, \dots, w_m$.

(4) For each planar network G of order n, the number of path families which cover G and obey the (I, I') crossing rule is zero, or is equal to the number of path families which cover G and obey the (J, J') crossing rule.

(5) For each even subinterval B of $[2n]$, the sets I'' and J'' satisfy

$$\max\{|B \cap I''|, |B \smallsetminus I''|\} \geq \max\{|B \cap J''|, |B \smallsetminus J''|\}. \tag{3.3}$$

(6) For each subinterval B of $[2n]$, the sets I'' and J'' satisfy (3.3).

Proof. ($1 \Leftrightarrow 2$) Follows from Observation 1.3.
($3 \Rightarrow 2$) Obvious.
($4 \Rightarrow 3$) The difference between the weighted sums of path families in G which obey the two binary crossing rules can be expressed as a sum of differences, over all subnetworks of G which can be covered by path families which obey at least one of the two rules. Assume as in (4) that the difference corresponding to each subnetwork is a monomial in $w_1, \dots, w_m$ with coefficient $(a - a)$ or $(a - 0)$ for some nonnegative integer a. Clearly the sum of these is a subtraction-free polynomial in $w_1, \dots, w_m$.
($5 \Rightarrow 4$) Let G be a planar network, and let $k(I, I'), k(J, J')$ be the numbers of path families which cover G and obey the (I, I') crossing rule and the (J, J') crossing rule, respectively. If $k(I, I')$ is greater than $k(J, J')$, then $k(J, J')$ must be equal to zero by Corollary 2.2. Thus $\psi(G)$ is a perfect matching of I'' with $\overline{I''}$ and not a perfect matching of J'' with $\overline{J''}$. Let $B = [b_1, b_2]$ be a minimal subinterval of $[2n]$ such that (b_1, b_2) is an edge in $\psi(G)$ and b_1, b_2 both belong to J'' or both belong to $\overline{J''}$. By the minimality of B we have

$$\max\{|B \cap I''|, |B \smallsetminus I''|\} = \frac{b_2 - b_1 + 1}{2} < \frac{b_2 - b_1 + 3}{2} = \max\{|B \cap J''|, |B \smallsetminus J''|\},$$

which contradicts (5).
($6 \Rightarrow 5$) Obvious.
($2 \Rightarrow 6$) Suppose that (6) is false. Let $B = [b_1, b_2]$ be a minimal subinterval of $[2n]$ which satisfies

$$\max\{|B \cap I''|, |B \smallsetminus I''|\} < \max\{|B \cap J''|, |B \smallsetminus J''|\}.$$

Without loss of generality, assume that we have

$$|B \smallsetminus J''| < |B \smallsetminus I''| \leq |B \cap I''| < |B \cap J''|.$$

By Lemma 3.1, the cardinality of B is even, and exactly half of the elements of the subinterval $[b_1 + 1, b_2 - 1]$ belong to I''. Furthermore, both b_1 and b_2 belong to J'' while only one of these belongs to I''.

Create a graph H in $\mathcal{G}_3(2n)$ by drawing $2n$ distinct points labeled $1, \dots, 2n$ on a horizontal line, and by connecting them with n nonintersecting arcs above the line as follows.

(1) Connect b_1 to b_2.

(2) Below the arc (b_1, b_2) and above the horizontal line, draw arcs to create a noncrossing perfect matching of $[b_1+1, b_2-1] \cap I''$ with $[b_1+1, b_2-1] \smallsetminus I''$.
(3) Above the arc (b_1, b_2) and above the horizontal line, draw arcs to create a noncrossing perfect matching of $([2n] \smallsetminus B) \cap I''$ with $([2n] \smallsetminus B) \smallsetminus I''$.

Clearly H is a matching of I'' with $\overline{I''}$, but not a matching of J'' with $\overline{J''}$. Now consider the graph $G = \rho(H)$ in $\mathcal{G}_1(n)$. By Lemma 2.4, exactly one path family in G obeys the (I, I') crossing rule and none obeys the (J, J') crossing rule. This contradicts (2). □

As an immediate corollary of Theorem 3.2, we obtain a combinatorial interpretation of the difference

$$\Delta_{J,J'}\Delta_{\overline{J},\overline{J'}} - \Delta_{I,I'}\Delta_{\overline{I},\overline{I'}}, \tag{3.4}$$

when this difference is nonnegative for all totally nonnegative matrices.

Corollary 3.3. *Let I, I', J and J' be subsets of $[n]$, define the sets I'' and J'' as in Theorem 3.2, and assume that the inequality*

$$\max\{|B \cap I''|, |B \smallsetminus I''|\} \geq \max\{|B \cap J''|, |B \smallsetminus J''|\}. \tag{3.5}$$

holds for all even subintervals B of $[2n]$. Then for any $n \times n$ totally nonnegative matrix A, and any planar network G whose weighted path matrix is A, the difference (3.4) is equal to the weighted sum of path families π in G which obey the (J, J') crossing rule and which cannot be covered by any path family which obeys the (I, I') crossing rule.

As a second corollary of Theorem 3.2, we obtain the following specialization to principal minors, first stated in [3, Thm 4.10].

Corollary 3.4. *Let I and J be subsets of $[n]$. Then the following statements are equivalent.*
(1) In each totally nonnegative matrix of size at least $n \times n$, the principal minors $\Delta_{I,I}$, $\Delta_{\overline{I},\overline{I}}$, $\Delta_{J,J}$, and $\Delta_{\overline{J},\overline{J}}$ satisfy

$$\Delta_{I,I}\Delta_{\overline{I},\overline{I}} \leq \Delta_{J,J}\Delta_{\overline{J},\overline{J}}. \tag{3.6}$$

(2) For each even subinterval B of $[n]$, the sets I and J satisfy

$$\max\{|B \cap I|, |B \smallsetminus I|\} \geq \max\{|B \cap J|, |B \smallsetminus J|\}. \tag{3.7}$$

Proof. $(1 \Rightarrow 2)$ follows from Theorem 3.2.

$(2 \Rightarrow 1)$ Suppose that I and J satisfy (3.7) for all even subintervals of $[n]$, and define the subsets I'' and J'' of $[2n]$ as before Theorem 3.2. If I'' and J'' fail to satisfy (3.3) for some even subinterval of $[2n]$, then this interval must be of the form $B = [b_1, 2n+1-b_2]$, where $b_1, b_2 \leq n$. We cannot have $b_1 = b_2$, for then the four sets $B \cap I''$, $B \smallsetminus I''$, $B \cap J''$, $B \smallsetminus J''$ would all have cardinality equal to $n - b_1 + 1$. On the other hand, we cannot have $b_1 \neq b_2$, for then I and J would fail to satisfy (3.7) for the interval whose endpoints are b_1 and b_2. Thus I'' and J'' must satisfy (3.3) and the principal minors $\Delta_{I,I}$, $\Delta_{\overline{I},\overline{I}}$, $\Delta_{J,J}$, and $\Delta_{\overline{J},\overline{J}}$ must satisfy (3.6). □

Note the similarity between the inequalities (3.3) and (3.7). This implies that the map $(I, I', n) \mapsto (I'', 2n)$ induces a bijective correspondence between inequalities

$$\Delta_{I,I'}\Delta_{\overline{I},\overline{I'}} \leq \Delta_{J,J'}\Delta_{\overline{J},\overline{J'}} \tag{3.8}$$

in products of nonprincipal minors of $n \times n$ matrices and inequalities

$$\Delta_{I'',I''}\Delta_{\overline{I''},\overline{I''}} \leq \Delta_{J'',J''}\Delta_{\overline{J''},\overline{J''}}$$

in products of principal $n \times n$ minors of $2n \times 2n$ matrices.

4. Generalization to noncomplementary index sets

A more general inequality than (3.8) has the form

$$\Delta_{I,I'}\Delta_{K,K'} \leq \Delta_{J,J'}\Delta_{L,L'}. \tag{4.1}$$

That is, the pairs (I, K), etc. need not be complements. However, any such inequality which holds for all totally nonnegative matrices can be deduced from Theorem 3.2 or equivalently from Corollary 3.4. We will give a combinatorial proof of this fact using the families of graphs introduced in Section 2.

Let p and p' be the cardinalities of $I \cup K$ and $I' \cup K'$, and let q and q' be the cardinalities of $I \cap K$ and $I' \cap K'$. Necessarily, $p + q = p' + q' = |I| + |K|$. Applying Observation 1.3, we may interpret the product of minors $\Delta_{I,I'}\Delta_{K,K'}$ of any totally nonnegative matrix to be the weighted sum of path families $\pi = (\pi_1, \ldots, \pi_{p+q})$ in a planar network which connect sources indexed by I (K) to sinks indexed by I' (K') and in which all S_I to $T_{I'}$ paths (S_K to $T_{K'}$ paths) are vertex-disjoint. We will say that such a path family obeys *the* (I, I', K, K') *crossing rule.* The following neccessary condition for eight sets to satisfy (4.1) for all totally nonnegative matrices was first stated in [3, Prop 2.2].

Observation 4.1. *Let $I, I', J, J', K, K', L, L'$ be subsets of $[m]$. Unless $I \cup K$ and $J \cup L$ are equal as multisets, and $I' \cup K'$ and $J' \cup L'$ are equal as multisets, the products $\Delta_{I,I'}\Delta_{K,K'}$ and $\Delta_{J,J'}\Delta_{L,L'}$ are incomparable as functions on totally nonnegative matrices.*

Proof. Suppose i is an index which appears with greater multiplicity in $I \cup K$ than in $J \cup L$, and let G be any planar network of order m in which the unique edge leaving source s_i has weight c. If c is large enough, then the weighted path matrix of G satisfies

$$\Delta_{I,I'}\Delta_{K,K'} > \Delta_{J,J'}\Delta_{L,L'}.$$

On the other hand, if c is close enough to zero, we have the opposite strict inequality. Similarly, if $I' \cup K'$ and $J' \cup L'$ are not equal as multisets, then the products are again incomparable. □

Without loss of generality, we shall assume that the sets $I \cup K$ and $I' \cup K'$ are equal to $[p]$ and $[p']$, respectively. Otherwise we can delete appropriate matrix rows and columns to make this true. Necessary and sufficient conditions for eight sets to satisfy (4.1) are analogous to the inequalities (3.3) and (3.7). The appropriate choices of I'', J'', and n are as follows. Let n be the number $\frac{1}{2}(p-q+p'-q')$, let η be the unique order preserving map

$$\eta : (I \smallsetminus K) \cup (K \smallsetminus I) \to [p-q],$$

and let η' be the unique order *reversing* map

$$\eta' : (I' \smallsetminus K') \cup (K' \smallsetminus I') \to [p-q+1, 2n].$$

Define the subsets I'' and J'' of $[2n]$ by

$$I'' = \eta(I \smallsetminus K) \cup \eta'(K' \smallsetminus I'),$$
$$J'' = \eta(J \smallsetminus L) \cup \eta'(L' \smallsetminus J').$$

Theorem 4.2. *Let $I, I', J, J', K, K', L, L'$ be subsets of $[m]$, and define p, p', q, q', n, η, η', I'', and J'', as above. Then the following statements are equivalent.*

(1) *In each totally nonnegative matrix of size at least $m \times m$, the minors $\Delta_{I,I'}$, $\Delta_{J,J'}$, $\Delta_{K,K'}$, $\Delta_{L,L'}$ satisfy*

$$\Delta_{I,I'}\Delta_{K,K'} \leq \Delta_{J,J'}\Delta_{L,L'}.$$

(2) *The multisets* $I \cup K$ *and* $J \cup L$ *are equal, the multisets* $I' \cup K'$ *and* $J' \cup L'$ *are equal, and the sets* I'', J'' *satisfy*

$$\max\{|B \cap I''|, |B \smallsetminus I''|\} \geq \max\{|B \cap J''|, |B \smallsetminus J''|\}$$

for each even subinterval B *of* $[2n]$.

Proof. ($2 \Rightarrow 1$) Suppose (1) is false. Then there exists a planar network in which more path families obey the $(I, I'K, K')$ crossing rule than obey the (J, J', L, L') crossing rule. This network contains a subnetwork G which is a union of $(p+q)$ paths from p sources to p' sinks with the property that more path families $\pi = (\pi_1, \ldots, \pi_{p+q})$ which cover G obey the (I, I', K, K') crossing rule than obey the (J, J', L, L') crossing rule.

Applying the procedure defining ϕ to G, we obtain a graph in which exactly n connected components are paths whose endpoints belong to the $2n$-element set

$$S_{I \smallsetminus K} \cup S_{K \smallsetminus I} \cup T_{I' \smallsetminus K'} \cup T_{K' \smallsetminus I'}.$$

By the discussion following Corollary 2.2, these n paths define a perfect matching of $S_{(I \smallsetminus K)} \cup T_{(K' \smallsetminus I')}$ with $S_{(K \smallsetminus I)} \cup T_{(I' \smallsetminus K')}$, which is not a perfect matching of $S_{(J \smallsetminus L)} \cup T_{(L' \smallsetminus J')}$ with $S_{(L \smallsetminus J)} \cup T_{(J' \smallsetminus L')}$. Let H be the graph in $\mathcal{G}_3(2n)$ realizing this matching, in which vertex i ($1 \leq i \leq p-q$) corresponds to the source in $S_{I \smallsetminus K} \cup S_{K \smallsetminus I}$ with the ith smallest index and vertex j ($p - q < j \leq p - q + p' - q'$) corresponds to the sink in $T_{I' \smallsetminus K'} \cup T_{K' \smallsetminus I'}$ with the jth greatest index. Let $B = [b_1, b_2]$ be a minimal interval of $[2n]$ such that (b_1, b_2) is an edge of H and b_1, b_2 both belong to J'' or both belong to $\overline{J''}$. Then we have

$$\max\{|B \cap I''|, |B \smallsetminus I''|\} < \max\{|B \cap J''|, |B \smallsetminus J''|\}. \tag{4.2}$$

($1 \Rightarrow 2$) Let $B = [b_1, b_2]$ be a minimal subinterval of $[2n]$ which satisfies (4.2), and let j_1 and j_2 be the preimages of these numbers with respect to the maps η and/or η'. Create a graph $H = (V, E)$ as follows.

(1) Place $2p + 2q$ vertices on a horizontal line.
(2) Define six sets of symbols

$$\begin{aligned}
S &= \{s_i \,|\, i \in (I \smallsetminus K) \cup (K \smallsetminus I)\},\\
U &= \{u_i \,|\, i \in (I \cap K)\},\\
U' &= \{u'_i \,|\, i \in (I \cap K)\},\\
T &= \{t_i \,|\, i \in (I' \smallsetminus K') \cup (K' \smallsetminus I')\},\\
V &= \{v_i \,|\, i \in (I' \cap K')\},\\
V' &= \{v'_i \,|\, i \in (I' \cap K')\}.
\end{aligned}$$

(3) Label the leftmost $p+q$ vertices by the $p+q$ symbols $S \cup U \cup U'$, in order of nondecreasing indices. Label the rightmost $p+q$ vertices by the $p+q$ symbols $T \cup V \cup V'$, in order of nonincreasing indices.
(4) Connect each of the $q + q'$ pairs of the form (u_i, u'_i) or (v_i, v'_i) by $q + q'$ noncrossing arcs.
(5) Connect the b_1st (from the left) singleton to the b_2nd (from the left) singleton by an arc.
(6) Draw $p + q - 1$ more arcs above the vertices to complete a noncrossing perfect matching of $S_{I \smallsetminus K} \cup T_{K' \smallsetminus I'}$ with $S_{K \smallsetminus I} \cup T_{I' \smallsetminus K'}$.

The arc drawn in step (5) prevents H from inducing a perfect matching of $S_{J \smallsetminus L} \cup T_{L' \smallsetminus J'}$ with $S_{L \smallsetminus J} \cup T_{J' \smallsetminus L'}$. Now consider the planar network G which is obtained from H by creating $\rho(H)$ and then identifying all pairs of vertices of the form (u_i, u'_i) or (v_i, v'_i). There are no path families in G which obey the (J, J', L, L') crossing rule, but there are some which obey the (I, I', K, K') crossing rule. This contradicts (1). $\square$

Example 4.1. Consider the two products of minors

$$\Delta_{\{1,2,3,6\},\{1,2,4,5\}}\Delta_{\{3,4\},\{2,5\}}, \quad \Delta_{\{1,3,6\},\{1,2,5\}}\Delta_{\{2,3,4\},\{2,4,5\}}.$$

We claim that these products are incomparable. To see this, let the map $\eta : \{1, 2, 4, 6\} \to \{1, 2, 3, 4\}$ preserve order, and let the map $\eta' : \{1, 4\} \to \{5, 6\}$ reverse order. Then define

$$I'' = \{\eta(1), \eta(2), \eta(6)\} \cup \emptyset = \{1, 2, 4\},$$
$$J'' = \{\eta(1), \eta(6)\} \cup \{\eta'(4)\} = \{1, 4, 5\}.$$

Since the sets I'', J'' satisfy

$$\max\{|[4, 5] \cup I''|, |[4, 5] \smallsetminus I''|\} < \max\{|[4, 5] \cup J''|, |[4, 5] \smallsetminus J''|\},$$
$$\max\{|[5, 6] \cup I''|, |[5, 6] \smallsetminus I''|\} > \max\{|[5, 6] \cup J''|, |[5, 6] \smallsetminus J''|\},$$

the products of minors are incomparable.

5. OPEN PROBLEMS

Note that the results of this paper reduce the problem of comparing minors in totally nonnegative matrices to the problem of counting *unweighted* path families in planar networks. This is curious, since Theorem 1.2 does not guarantee that for every totally nonnegative integer matrix A, there exists a planar network G such that A counts unweighted paths in G. This suggests the following open problem.

Problem 5.1. Characterize the totally nonnegative integer matrices $A = [a_{ij}]$ for which there exists a planar network G such that each entry a_{ij} counts *unweighted* paths in G from source s_i to sink t_j.

Other possibilities for extending the present work are the following.

Problem 5.2. Let $P(n)$ be the poset whose elements are set partitions of $[2n]$ into two blocks of size n, ordered by I-$\overline{I} \leq_{P(n)} J$-$\overline{J}$ whenever the inequality

$$\Delta_{I,I}\Delta_{\overline{I},\overline{I}} \leq \Delta_{J,J}\Delta_{\overline{J},\overline{J}}$$

holds for all totally nonnegative matrices. Find a simple description for $P(n)$.

Problem 5.3. Characterize the inequalities in products of k minors which are satisfied by all totally nonnegative matrices, for $k > 2$.

6. ACKNOWLEDGMENTS

I would like to thank Sergey Fomin, John Stembridge, Eric Babson, Misha Gekhtman, Francesco Brenti, Alex Postnikov, and Harm Derksen for many helpful discussions.

REFERENCES

[1] F. BRENTI. Combinatorics and total positivity. *J. Combin. Theory Ser. A*, **71** (1995) pp. 175–218.
[2] C. W. CRYER. Some properties of totally positive matrices. *Lin. Alg. Its Appl.*, **15** (1976) pp. 1–25.
[3] S. M. FALLAT, M. I. GEKHTMAN, AND C. R. JOHNSON. Multiplicative principal-minor inequalities for totally nonnegative matrices, 2001. Preprint.
[4] S. FOMIN AND A. ZELEVINSKY. Total positivity: Tests and parametrizations. *Math. Intelligencer*, (2001) pp. 23–33.
[5] I. GESSEL AND G. VIENNOT. Binomial determinants, paths, and hook length formulae. *Advances in Mathematics*, **58** (1985) pp. 300–321.
[6] I. GESSEL AND G. VIENNOT. Determinants and plane partitions, 1989. Preprint.

[7] S. Karlin and G. McGregor. Coincidence probabilities. *Pacific J. Math.*, **9** (1959) pp. 1141–1164.
[8] B. Lindström. On the vector representations of induced matroids. *Bull. London Math. Soc.*, **5** (1973) pp. 85–90.
[9] C. Loewner. On totally positive matrices. *Math. Z.*, **63** (1955) pp. 338–340.
[10] G. Lusztig. Total positivity in reductive groups. In *Lie Theory and Geometry: in Honor of Bertram Kostant*, vol. 123 of *Progress in Mathematics*. Birkhäuser, Boston, 1994 pp. 531–568.
[11] R. Stanley. *Enumerative Combinatorics*, vol. 2. Cambridge University Press, Cambridge, 1999.
[12] A. Whitney. A reduction theorem for totally positive matrices. *J. d'Analyse Math.*, **2** (1952) pp. 88–92.

Department of Mathematics, University of Michigan, 2074 East Hall, 525 East University, Ann Arbor, Michigan, 48109-1109
E-mail address: mskan@umich.edu

HALL-LITTLEWOOD ANALOGS IN THE Q-FUNCTION ALGEBRA

GEANINA TUDOSE AND MICHAEL ZABROCKI

Abstract. We present a family of analogs of the Hall-Littlewood symmetric functions in the Q-function algebra. The change of basis coefficients between this family and Schur's Q-functions are q-analogs of numbers of marked shifted tableaux. These coefficients exhibit many parallel properties to the Kostka-Foulkes polynomials.

Résumé. Nous présentons une famille des analogues des fonctions symétriques de Hall-Littlewood dans l'algébre des Q-fonctions. Le changement des coefficients de base sont les q-analogues des nombres de tableaux deplacés et marqués. Ces coefficients présentent beaucoup de propriétés parallèles aux polynômes de Kostka-Foulkes.

1. Introduction

The symmetric functions form a fundamental algebra Λ associated to the representation theory and the combinatorics of the symmetric group. The Q-functions form a sub-algebra $\Gamma \subset \Lambda$ of the symmetric functions and are associated to the projective representations of the symmetric group. The fundamental bases for these algebras are (respectively) the Schur functions, $s_\lambda[X]$, indexed by partitions, and Schur's Q-functions, $Q_\lambda[X]$, indexed by strict partitions.

Another important basis of Λ is the Hall-Littlewood symmetric functions. They were introduced by Hall [3] as an algebra whose structure coefficients count chains of submodules of a certain type. In addition, they have remarkable combinatorial and algebraic properties and interpolate several other bases with a parameter q. They may also be seen in other contexts such as a Demazure character formula for the Hecke algebra, or as the Frobenius series of certain symmetric group or $GL(n)$-modules.

The change of basis coefficients between the Hall-Littlewood symmetric functions and the Schur basis are known as the Kostka-Foulkes polynomials. They can be seen as a q-generating function for the set of column strict tableaux of fixed shape and content. The combinatorial tools of the RSK-algorithm, jeu de taquin and the plactic monoid [13] were developed in part to help answer the question of a combinatorial interpretation for these coefficients. Lascoux and Schützenberger [12] produced a solution by the introduction of the statistic *charge* on column strict tableaux.

The aim of this paper is to introduce the analogs of the Hall-Littlewood symmetric functions and the Kostka-Foulkes polynomials for the Q-function algebra. These analogs do not exist in the literature and are certainly missing elements in the combinatorial theory of the Q-function algebra. The combinatorics of the shifted tableaux developed in [2] [15] [19] [21] correspond to the Q-functions in the same way that the column strict tableaux correspond to the symmetric functions. The functions that we introduce here share many of the same properties as the Hall-Littlewood symmetric functions. The analogs of Kostka-Foulkes polynomials are q-generating functions of the set of marked shifted tableaux of fixed shape and content. Some of the most important properties of these functions are still conjecture and they suggest a yet undiscovered structure in the combinatorics of shifted tableaux including a poset structure associated to a charge-like statistic.

This abstract is divided into two sections. In the first section we present some well known theory related to the symmetric functions and in particular the Hall-Littlewood functions. For a more detailed account of this theory we refer the reader to [14]. The second section is a development of the functions $G_\lambda[X;q]$ (the functions that are analogs of the Hall-Littlewoods in Γ) and the change of basis coefficients $L_{\lambda\mu}(q)$ (analogs of the Kostka-Foulkes polynomials). We present several formulas for computing these coefficients and list some properties and conjectures. In an appendix we also list a table of transition coefficients $L_{\lambda\mu}(q)$ and draw as an example a conjectured poset for the marked shifted tableaux of content $(4,3,1)$.

Finally, we briefly consider the *parabolic* version of $G_\lambda[X;q]$ which are analogs of the functions introduced in [16, 17]. The definition follows the generalization of Jing's Hall-Littlewood vertex operator to a more general class of operators, as was considered in [18]. The coefficients that appear in this generalization can be viewed as q-analogs of the structure coefficients of Schur's Q-functions.

2. Notation and Definitions

2.1. **Symmetric functions, partitions, tableaux.** Define the ring of symmetric functions as the polynomial ring $\Lambda = \mathbb{C}[p_1, p_2, p_3, \ldots]$ with $deg(p_k) = k$. A typical monomial of degree n in this ring will be $p_{\lambda_1} p_{\lambda_2} \cdots p_{\lambda_\ell} := p_\lambda$, where $\sum_i \lambda_i = n$ and a basis will indexed by the sequences λ such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$.

The sequence λ is a partition of n (denoted by $\lambda \vdash n$) if the entries are non-negative integers and are is weakly decreasing. The size of λ is given by $|\lambda| := \sum_i \lambda_i = n$. The entries of λ are called the parts of the partition. The number of parts that are of size i in λ will be represented by $m_i(\lambda)$ and the total number of non-zero parts is represented by $\ell(\lambda) = \sum_i m_i(\lambda)$. A common statistic on partitions λ is $n(\lambda) := \sum_i (i-1)\lambda_i$.

The dominance order, $\lambda \leq \mu$ if and only if $\sum_{i=1}^k \lambda_i \leq \sum_{i=1}^k \mu_i$ for all $1 \leq k \leq \ell(\lambda)$, is a partial order on partitions. Using this partial order, the operators

$$R_{ij}\lambda = (\lambda_1, \ldots, \lambda_i + 1, \ldots, \lambda_j - 1, \ldots, \lambda_{\ell(\lambda)})$$

for $1 \leq i \leq j \leq \ell(\lambda)$ have the property that $R_{ij}\lambda \geq \lambda$ if $R_{ij}\lambda$ is a partition.

We will consider three fundamental bases of Λ here. Following the notation of [14], we define the homogeneous (complete) symmetric functions are $h_\lambda := h_{\lambda_1} h_{\lambda_2} \cdots h_{\lambda_{\ell(\lambda)}}$ where $h_n = \sum_{\lambda \vdash n} p_\lambda / z_\lambda$ and $z_\lambda = \prod_{i=1}^{\ell(\lambda)} i^{m_i(\lambda)} m_i(\lambda)!$. The elementary symmetric functions are $e_\lambda := e_{\lambda_1} e_{\lambda_2} \cdots e_{\lambda_{\ell(\lambda)}}$ where $e_n = \sum_{\lambda \vdash n} (-1)^{n-\ell(\lambda)} p_\lambda / z_\lambda$. By convention we set $p_0 = h_0 = e_0 = 1$ and $p_{-k} = h_{-k} = e_{-k} = 0$ for $k > 0$. The Schur functions are given by $s_\lambda = det\, |h_{\lambda_i + i - j}|_{1 \leq i,j \leq \ell(\lambda)}$. The sets $\{p_\lambda\}_{\lambda \vdash n}$, $\{h_\lambda\}_{\lambda \vdash n}$, $\{e_\lambda\}_{\lambda \vdash n}$ and $\{s_\lambda\}_{\lambda \vdash n}$ all form bases for the symmetric functions of degree n.

The fundamental theorem of symmetric functions says that the subring $\mathbb{C}[p_1, p_2, \ldots, p_n]$ is isomorphic to the ring of symmetric polynomials $\Lambda^{X_n} = \mathbb{C}[x_1, x_2, \ldots, x_n]^{S_n}$ (the polynomials in n variables which are invariant under the action $\sigma(x_i) = x_{\sigma(i)}$ for any $\sigma \in S_n$) using the map that sends $p_k \to x_1^k + x_2^k + \cdots + x_n^k$. The space Λ^X of symmetric series in an infinite number of variables $x_1, x_2, x_3, \ldots$ of finite degree is isomorphic to Λ under the map that sends $p_k \to x_1^k + x_2^k + x_3^k + \cdots$.

Much of our notation for the symmetric functions thus far has reflected that of [14], but we will concentrate on operations involving the Hopf algebra structure of the symmetric functions and specialization of variables. To this end we extend the notation for these maps in a natural manner and represent a set of variables as a sum $X = x_1 + x_2 + x_3 + \ldots$ and act on this sum with elements of Λ. We define $p_k[X] = x_1^k + x_2^k + x_3^k + \cdots$ and for any $P \in \Lambda$ we set $P[X]$ equal to P with p_k replaced by $p_k[X]$. That is for $P = \sum_\lambda c_\lambda p_\lambda$,

$$P[X] = \sum_\lambda c_\lambda p_{\lambda_1}[X] p_{\lambda_2}[X] \cdots p_{\lambda_{\ell(\lambda)}}[X]. \tag{1}$$

It is clearly true for two sets of variables X and $Y = y_1 + y_2 + y_3 + \cdots$ that $p_k[X + Y] = p_k[X] + p_k[Y]$ and to extend this linearly we set $p_k[X - Y] = p_k[X] - p_k[Y]$ and $p_k[XY] = p_k[X]p_k[Y]$. We will also consider the Cauchy element

$$\Omega = \sum_{n\geq 0} \sum_{\lambda \vdash n} p_\lambda / z_\lambda = \sum_{n \geq 0} h_n \tag{2}$$

in the completion of Λ. This special element has the property that $\Omega[X+Y] = \Omega[X]\Omega[Y]$, $\Omega[X - Y] = \Omega[X]/\Omega[Y]$ and $\Omega[X] = \prod_i (1 - x_i)^{-1}$.

Notice that for an arbitrary element $c \in \mathbb{C}$, we have $p_k[cX] = cp_k[X]$. This implies that cX does not represent $cx_1 + cx_2 + cx_3 + \cdots$, instead it represents c 'copies of' the variables X. We introduce a special parameter q or t that interacts with the variable set in that $p_k[qX] = q^k p_k[X]$. Sometimes this element will be an arbitrary parameter and other times we will specialize it to values in the base field $\mathbb{C}$. To obtain operations such as replacing x_i by cx_i in a symmetric function we use our special parameter q and at the end of our calculations we specialize this parameter to c. In particular, the operation of replacing x_i by $-x_i$ is useful and we will represent it with the notation

$$P[\epsilon X] = P[qX]\Big|_{q=-1}. \tag{3}$$

We also have the relations $p_k[\epsilon X] = (-1)^k p_k[X]$, $\Omega[\epsilon X] = \prod_i (1 + x_i)^{-1}$ and $h_n[X] = e_n[-\epsilon X]$. Of course if the symmetric function P or the set of variables X already has a parameter q, the one that is set to -1 is unique and does not interfere with parameters in P or X.

It follows from the definition of the Schur function and the expansion of the Vandermonde determinant $det|x_i^{j-1}|_{1 \leq i,j \leq n} = \prod_{1 \leq i < j \leq n} (x_i - x_j)$ that $s_\lambda[X] = \prod_{1 \leq i < j \leq n} (1 - R_{ij}) h_\lambda[X]$, where $R_{ij} h_\lambda[X] = h_{R_{ij}\lambda}[X]$. Since the coefficient of z^λ in $\Omega[Z_n X]$ is $h_\lambda[X]$ and $(z_j/z_i)^{-1} z_\lambda = z^{R_{ij}\lambda}$, then the Schur function is equal to

$$s_\lambda[X] = \Omega[Z_n X] \prod_{1 \leq i < j \leq n} (1 - z_j/z_i)\Big|_{z^\lambda}. \tag{4}$$

Remark: We follow [14] in the use of R_{ij} acing on symmetric functions, however one should note that these operators are not associative. This issue can be resolved however and is dealt with in more detail in [1] or [8].

Now for any symmetric function $P \in \Lambda$ define $\mathbf{S}(z)P[X] := P\left[X - \frac{1}{z}\right]\Omega[zX]$. Since we have that $\mathbf{S}(z_1)\mathbf{S}(z_2)\cdots\mathbf{S}(z_n)1 = \Omega[Z_n X]\prod_{1 \leq i < j \leq n}(1 - z_j/z_i)$, then the operator $\mathbf{S}_m P[X] = \mathbf{S}(z)P[X]\Big|_{z^m}$ raises the degree of a symmetric function by m and has the property that $\mathbf{S}_m(s_\lambda[X]) = s_{(m,\lambda)}[X]$ as long as $m \geq \lambda_1$. The $\mathbf{S}_m$ operators also have the commutation relations $\mathbf{S}_m\mathbf{S}_{m+1} = 0$ and $\mathbf{S}_m\mathbf{S}_n = -\mathbf{S}_{n-1}\mathbf{S}_{m+1}$.

A Young diagram for a partition will be a collection of cells of the integer grid lying in the first quadrant. For a partition λ, $Y(\lambda) = \{(i,j) : 0 \leq j < \ell(\lambda) \text{ and } 0 \leq i \leq \lambda_j\}$. The reason why we consider empty cells rather than say points is because we wish to consider fillings of these cells. A tableau is a map from the set $Y(\lambda)$ to $\mathbb{N}$, this may be represented on a Young diagram by writing integers within the cells of a graphical representation of a

Young diagram (see figure 1). The shape of the tableau is the partition λ. We say that a tableau T is column strict if $T(i,j) \leq T(i+1,j)$ and $T(i,j) < T(i,j+1)$ whenever the points $(i+1,j)$ or $(i,j+1)$ are in $Y(\lambda)$. Let $m_k(T)$ represent the number of points p in $Y(\lambda)$ such that $T(p) = k$. The vector $(m_1(T), m_2(T), \ldots)$ is the content of the tableau T.

The Pieri rule describes a combinatorial method for computing the product of $h_m[X]$ and $s_\mu[X]$ expanded in the Schur basis. We will use the notation $\lambda/\mu \in \mathcal{H}_m$ to represent that $|\lambda| - |\mu| = m$ and for $1 \leq i \leq \ell(\lambda)$, $\mu_i \leq \lambda_i$ and $\mu_i \geq \lambda_{i+1}$. It may be easily shown that

$$h_m[X]s_\mu[X] = \sum_{\lambda/\mu \in \mathcal{H}_m} s_\lambda[X]. \tag{5}$$

This gives a method for computing the expansion of the $h_\mu[X]$ basis in terms of the Schur functions. Consider the coefficients $K_{\lambda\mu}$ defined by the expression

$$h_\mu[X] = \sum_{\lambda \vdash |\mu|} K_{\lambda\mu} s_\lambda[X]. \tag{6}$$

$K_{\lambda\mu}$ are called the Kostka numbers and are equal to the number of column strict tableaux of shape λ and content μ. Now define a q analog of the $\{h_\lambda\}$ basis by setting

$$H_\lambda[X;q] = \prod_{i<j} \frac{1 - R_{ij}}{1 - qR_{ij}} h_\lambda[X] = \prod_{i<j} (1 + (q-1)R_{ij} + (q^2 - q)R_{ij}^2 + \cdots) h_\lambda[X]. \tag{7}$$

Since the coefficient of z^λ in $\Omega[Z_k X]$ is $h_\lambda[X]$, it is clear that we have the formula

$$H_\lambda[X;q] = \Omega[Z_k X] \prod_{1 \leq i < j \leq k} \frac{1 - z_j/z_i}{1 - qz_j/z_i}\Big|_{z^\lambda}. \tag{8}$$

This leads us to a 'vertex operator' definition for these functions. If we define the operation $\mathbf{H}(z)P[X] = P\left[X - \frac{1-q}{z}\right]\Omega[zX]$, then

$$\mathbf{H}(z_1)\mathbf{H}(z_2)\cdots\mathbf{H}(z_k)1 = \Omega[Z_k X] \prod_{1 \leq i < j \leq k} \frac{1 - z_j/z_i}{1 - qz_j/z_i}, \tag{9}$$

and therefore defining the operator $\mathbf{H}_m$ that raises the degree of a symmetric function by m as $\mathbf{H}_m P[X] := \mathbf{H}(z)P[X]\Big|_{z^m}$, has the property that $\mathbf{H}_m H_\lambda[X;q] = H_{(m,\lambda)}[X;q]$ as long as $m \geq \lambda_1$. The vertex operator also satisfies the relations $\mathbf{H}_{m-1}\mathbf{H}_m = q\mathbf{H}_m\mathbf{H}_{m-1}$ and $\mathbf{H}_{m-1}\mathbf{H}_n - q\mathbf{H}_m\mathbf{H}_{n-1} = q\mathbf{H}_n\mathbf{H}_{m-1} - \mathbf{H}_{n-1}\mathbf{H}_m$.

The functions $H_\lambda[X;q]$ interpolate between the functions $s_\lambda[X] = H_\lambda[X;0]$ and $h_\lambda[X] = H_\lambda[X;1]$. The Kostka-Foulkes polynomials are defined as the q-polynomial coefficient of $s_\lambda[X]$ in $H_\mu[X;q]$ and hence we have the expansion analogous to (6).

$$H_\mu[X;q] = \sum_{\lambda \vdash |\mu|} K_{\lambda\mu}(q) s_\lambda[X]. \tag{10}$$

The coefficients $K_{\lambda\mu}(q)$ are clearly polynomials in q, but it is surprising to find that the coefficients of the polynomials are non-negative integers. A defining recurrence can be derived $K_{\lambda\mu}(q)$ in terms of the Kostka-Foulkes polynomials indexed by partitions of size $|\mu| - \mu_1$ using the formula for $\mathbf{H}_m$. This recurrence is often referred to as the 'Morris recurrence' for the Kostka-Foulkes polynomials.

The Kostka-Foulkes polynomials and the generating functions $H_\mu[X;q]$ have the following important properties which we simply list here so that we may draw a connection to analogous formulae. For a more detailed reference of these sorts of properties we refer the interested reader to the excellent survey article [1].

$$\begin{array}{|c|c|c|c|c|c|}\hline 4&6&7\\\hline 3&5&5\\\hline 2&2&3&4&6\\\hline 1&1&1&1&2&3\\\hline\end{array} \qquad \begin{array}{ccccccc} &&&5'&&&\\ &&3'&3&4&5&\\ &2'&2&3'&4'&4&\\ 1'&1&2'&2&2&3&3\end{array}$$

FIGURE 1. The diagram on the left represents a column strict tableau of shape $(6,5,3,3)$ and content $(4,3,3,2,2,2,1)$. The diagram on the right represents a shifted marked tableau of shape $(7,5,4,1)$ and content $(2,5,5,3,2)$. This tableau has labels which are marked on the diagonal.

(1) $K_{\lambda\mu}(q)$ has non-negative integer coefficients.
(2) $K_{\lambda\mu}(q) = \sum_T q^{c(T)}$, where the sum is over all column strict tableaux of shape λ and content μ and $c(T)$ denotes the charge of a tableau T (see [12]). In addition there is a combinatorial interpretation for these coefficients in terms of objects called rigged configurations (see [10]).
(3) The degree in q of $K_{\lambda\mu}(q)$ is $n(\mu) - n(\lambda)$.
(4) $K_{\lambda\mu}(0) = \delta_{\lambda\mu}$ which implies $H_\mu[X;0] = s_\mu[X]$, $K_{\lambda\mu}(1) = K_{\lambda\mu}$, so that $H_\mu[X;1] = h_\mu[X]$, $K_{\lambda\lambda}(q) = 1$ and $K_{(|\mu|)\mu}(q) = q^{n(\mu)}$. We also have that $K_{\lambda\mu}(q) = 0$ if $\lambda < \mu$.
(5) $H_{(1^n)}[X;q] = e_n\left[\frac{X}{1-q}\right](q;q)_n$ where $(q;q)_n = \prod_{i=1}^n (1-q^i)$.
(6) If ζ is k^{th} root of unity, $H_\mu[X;\zeta]$ factors into a product of symmetric functions.
(7) Set $K'_{\mu\lambda}(q) := q^{n(\lambda)-n(\mu)} K_{\mu\lambda}(1/q)$, then $K'_{\mu\lambda}(q) \geq K'_{\mu\nu}(q)$ for $\lambda \leq \nu$.
(8) $K_{\lambda+(a),\mu+(a)}(q) \geq K_{\lambda,\mu}(q)$, where $\lambda + (a)$ represents the partition λ with a part of size a inserted into it.
(9) $K_{\lambda\mu}(q) = \sum_{w\in S_n} sign(w)\mathcal{P}_q(w(\lambda+\rho) - (\mu+\rho))$ where $\mathcal{P}_q(\alpha)$ is the coefficient of x^α in $\prod_{1\leq i<j\leq n}(1 - qx_i/x_j)^{-1}$, a q analog of the Kostant partition function and $\rho = (\ell(\mu)-1, \ell(\mu)-2, \ldots, 1, 0)$.
(10) $H_\mu[X;q]H_\lambda[X;q] = \sum_\gamma d^\nu_{\lambda\mu}(q) H_\nu[X;q]$, for some coefficients $d^\nu_{\lambda\mu}(q)$ with the property that if the Littlewood-Richardson coefficient $c^\nu_{\lambda\mu} = 0$ then $d^\nu_{\lambda\mu}(q) = 0$. These coefficients are a transformation of the Hall algebra structure coefficients.
(11) For the scalar product $\langle s_\lambda[X], s_\mu[X]\rangle = \delta_{\lambda\mu}$, we have that $\langle H_\lambda[X;q], H_\mu[X(1-q);q]\rangle = 0$ if $\lambda \neq \mu$.

2.2. **Schur's Q-functions, strict partitions, and marked shifted tableaux.** The Q-function algebra is a sub-algebra of the symmetric functions $\Gamma = \mathbb{C}[p_1, p_3, p_5, \ldots]$. A typical monomial in this algebra will be p_λ, where λ is a partition and λ_i is odd. A partition λ is strict if $\lambda_i > \lambda_{i+1}$ for all $1 \leq i \leq \ell(\lambda)-1$ and a partition λ is odd if λ_i is odd for $1 \leq i \leq \ell(\lambda)$. We will use the notation $\lambda \vdash_s n$ (respectively $\lambda \vdash_o n$) to denote that λ is a partition of size n that is strict (respectively odd). Note that the number of strict partitions of size n and the number of odd partitions of size n is the same (proof: write out a generating function for each sequence).

The analog of the homogeneous and elementary symmetric functions in Γ are the functions $q_\lambda := q_{\lambda_1} q_{\lambda_2} \cdots q_{\lambda_{\ell(\lambda)}}$, where $q_n = \sum_{\lambda\vdash_o n} 2^{\ell(\lambda)} p_\lambda / z_\lambda$. Define an algebra morphism $\theta : \Lambda \to \Gamma$ by the action on the p_n generators as $\theta(p_n) = (1-(-1)^n)p_n$. That is $\theta(p_n) = 2p_n$ if n is odd and $\theta(p_n) = 0$ for n even. θ has the property that $\theta(h_n) = \theta(e_n) = q_n$ and may be represented in our notation as $\theta(p_n[X]) = p_n[(1-\epsilon)X]$. Under this morphism, our Cauchy element may also be considered a generating function for the q_n elements since

$$\Omega[(1-\epsilon)X] = \sum_{n\geq 0} q_n[X] = \prod_i \frac{1+x_i}{1-x_i}. \tag{11}$$

It follows that $\{p_\lambda\}_{\lambda\vdash_o n}$, $\{q_\lambda\}_{\lambda\vdash_o n}$, $\{q_\lambda\}_{\lambda\vdash_s n}$ are all bases for the subspace of Q-functions of degree n. Another fundamental basis for this space are the Schur's Q-functions $Q_\lambda[X] = \theta(H_\lambda[X;-1])$. These functions hold a similar place in the Q-function algebra that the Schur functions hold in Λ. In particular, $\{Q_\lambda[X]\}_{\lambda\vdash_s n}$ is a basis for the Q-functions of degree n.

In analogy with the Schur functions, $Q_\lambda[X]$ may also be defined with a raising operator formula by setting $q=-1$ and applying the θ homomorphism to equation (7). We arrive at the formula:

$$Q_\lambda[X] = \prod_{i<j} \frac{1-R_{ij}}{1+R_{ij}} q_\lambda[X] = \prod_{i<j}(1-2R_{ij}+2R_{ij}^2-\cdots)q_\lambda[X], \tag{12}$$

where the operators now act as $R_{ij}q_\lambda[X] = q_{R_{ij}\lambda}[X]$. Furthermore, they have a formula as the coefficient in a generating function:

$$Q_\lambda[X] = \Omega[(1-\epsilon)Z_nX] \prod_{1\le i<j\le n} \frac{1-z_j/z_i}{1+z_j/z_i}\Big|_{z^\lambda}. \tag{13}$$

As with Schur functions and the Hall-Littlewood functions, the raising operator formula leads us to a vertex operator definition. By setting $\mathbf{Q}(z)P[X] = P\left[X-\frac{1}{z}\right]\Omega[(1-\epsilon)zX]$, it is easily shown that $\mathbf{Q}(z_1)\mathbf{Q}(z_2)\cdots\mathbf{Q}(z_n)1 = \Omega[(1-\epsilon)Z_nX]\prod_{1\le i<j\le n}\frac{1-z_j/z_i}{1+z_j/z_i}$, and hence if we set $\mathbf{Q}_mP[X] = \mathbf{Q}(z)P[X]\Big|_{z^m}$ then $\mathbf{Q}_m(Q_\lambda[X]) = Q_{(m,\lambda)}[X]$ as long as $m>\lambda_1$. The commutation relations for the $\mathbf{Q}_m$ are

$$\mathbf{Q}_m\mathbf{Q}_n = -\mathbf{Q}_n\mathbf{Q}_m \text{ for } m\neq -n, \tag{14}$$

$$\mathbf{Q}_m\mathbf{Q}_{-m} = 2(-1)^m - \mathbf{Q}_{-m}\mathbf{Q}_m \text{ if } m\neq 0, \tag{15}$$

$$\mathbf{Q}_m^2 = 0 \text{ if } m\neq 0 \text{ and } \mathbf{Q}_0^2 = 1. \tag{16}$$

These formulas allow us to straighten the $Q_\mu[X]$ functions when they are not indexed by a strict partition.

A shifted Young diagram for a partition will again be a collection of cells lying in the first quadrant. For a strict partition λ, let $YS(\lambda) = \{(i,j) : 0\le j\le \ell(\lambda) \text{ and } j-1\le i\le \lambda_j+j-1\}$. A marked shifted tableau T of shape λ is a map from $YS(\lambda)$ to the set of marked integers $\{1'<1<2'<2<\ldots\}$ that satisfy the following conditions

- $T(i,j)\le T(i+1,j)$ and $T(i,j)\le T(i,j+1)$
- If $T(i,j)=k$ for some integer k (i.e. has an unmarked label) then $T(i,j+1)\neq k$
- If $T(i,j)=k'$ for some marked label k' then $T(i+1,j)\neq k'$.

We may represent these objects graphically with a diagram representing λ and the cells filled with the marked integer alphabet. If T is a marked shifted tableau, then we will set $m_i(T)$ as the number of occurrences of i and i' in T. The sequence $(m_1(T), m_2(T), m_3(T),\ldots)$ is the content of T.

The combinatorial definition of the marked shifted tableaux is defined so that it reflects the change of basis coefficients between the q_λ and Q_μ basis. The rule for computing the product of $q_m[X]$ and $Q_\mu[X]$ when expanded in the Schur Q-functions is the analog of the Pieri rule for the Γ space. If $\lambda/\mu\in\mathcal{H}_m$ then $a(\lambda/\mu)$ will represent 1+ the number of $1<j\le\ell(\lambda)$ such that $\lambda_j>\mu_j$ and $\mu_{j-1}>\lambda_j$. We may show that

$$q_m[X]Q_\mu[X] = \sum_{\lambda/\mu\in\mathcal{H}_m} 2^{a(\lambda/\mu)-\ell(\lambda)+\ell(\mu)}Q_\lambda[X]. \tag{17}$$

Denote by $L_{\lambda\mu}$ the number of marked shifted tableaux T of shape λ and content μ (where λ is a strict partition) such that $T(i,i)$ is not a marked integer. We may expand the function $q_\mu[X]$ in terms of the Q-functions using (17) to show

$$q_\mu[X] = \sum_{\lambda \vdash |\mu|} L_{\lambda\mu} Q_\lambda[X]. \tag{18}$$

3. The Q-Hall-Littlewood basis $G_\lambda(x;q)$ for the algebra Γ

Note: From here, unless otherwise stated, all partitions are considered strict.

3.1. **Raising operator formula.** We define the following analog of the Hall-Littlewood functions in the subalgebra Γ

$$G_\lambda[X;q] := \prod_{1\le i<j\le n} \left(\frac{1+qR_{ij}}{1-qR_{ij}}\right)\left(\frac{1-R_{ij}}{1+R_{ij}}\right) q_\lambda[X] = \prod_{1\le i<j\le n} \left(\frac{1+qR_{ij}}{1-qR_{ij}}\right) Q_\lambda[X]. \tag{19}$$

We call the functions $G_\lambda \in \Gamma \otimes_{\mathbb{C}} \mathbb{C}(q)$ the *Q-Hall-Littlewood functions.*

In $\Gamma \otimes \mathbb{C}(q)$ this family can be expressed in the basis of Q-functions as

$$G_\mu[X;q] = \sum_\lambda L_{\lambda\mu}(q) Q_\lambda[X], \tag{20}$$

which can be viewed as a q-analog of (18). We call the coefficients $L_{\lambda\mu}(q)$ the *Q-Kostka polynomials.* We shall see that this family of polynomials shares many of the same properties with the classical Kostka-Foulkes polynomials. Tables of these coefficients are given in an Appendix. It follows from (19) that $L_{\lambda\mu}(q)$ have integer coefficients and $L_{\lambda\mu}(q) = 0$ if $\lambda < \mu$. This shows

Proposition 1. *The G_λ, λ strict, form a $\mathbb{Z}$-basis for $\Gamma \otimes_{\mathbb{Z}} \mathbb{Z}(q)$.*

The basis G_λ interpolates between the Schur's Q-functions and the functions q_μ because $G_\lambda[X;0] = Q_\lambda[X]$ and $G_\lambda[X;1] = q_\lambda[X]$ as is clear from (19).

Since the coefficient of z^λ in $\Omega[(1-\epsilon)Z_nX]$ is $q_\lambda[X]$ equation (19) implies

$$G_\lambda[X;q] = \prod_{1\le i<j\le n} \left(\frac{1-z_j/z_i}{1+z_j/z_i}\right)\left(\frac{1+qz_j/z_i}{1-qz_j/z_i}\right) \Omega[(1-\epsilon)Z_nX]\Big|_{z^\lambda}. \tag{21}$$

By defining $\mathbf{G}(z)P[X] = P[X - \frac{1-q}{z}]\Omega[(1-\epsilon)zX]$, we may show that

$$\mathbf{G}(z_1)\mathbf{G}(z_2)\cdots\mathbf{G}(z_n)1 = \prod_{1\le i<j\le n} \left(\frac{1-z_j/z_i}{1+z_j/z_i}\right)\left(\frac{1+qz_j/z_i}{1-qz_j/z_i}\right) \Omega[(1-\epsilon)Z_nX]. \tag{22}$$

This implies that if we define the operator

$$\mathbf{G}_m P[X] = P\left[X - \frac{1-q}{z}\right] \Omega[(1-\epsilon)zX]\Big|_{z^m}, \tag{23}$$

then

$$G_\lambda[X;q] = \mathbf{G}_{\lambda_1}\ldots\mathbf{G}_{\ell(\lambda)}(1).$$

The operator $\mathbf{G}_m$ satisfies the following commutation relation.

Proposition 2. *For all $r,s \in \mathbb{Z}$ we have*

$$(1-q^2)(\mathbf{G}_r\mathbf{G}_s+\mathbf{G}_s\mathbf{G}_r)+q(\mathbf{G}_{r-1}\mathbf{G}_{s+1}-\mathbf{G}_{s+1}\mathbf{G}_{r-1}+\mathbf{G}_{s-1}\mathbf{G}_{r+1}-\mathbf{G}_{r+1}\mathbf{G}_{s-1}) = 2(-1)^r(1-q)^2\delta_{r,-s}.$$

For $q = 0$ in the equation above we recover the commutation relations of the operator $\mathbf{Q}$ given in equations (14), (15) and (16).

We can use formula (23) to derive the action of this operator on the basis of Schur's Q-functions.

Proposition 3. *For $m > 0$,*

$$\mathbf{G}_m(Q_\lambda[X]) = \sum_{i\geq 0} q^i \sum_{\mu:\lambda/\mu\in\mathcal{H}_i} 2^{a(\lambda/\mu)}(-1)^{\epsilon(m+i,\mu)} Q_{\mu+(m+i)}[X], \tag{24}$$

where $\mu + (k)$ denotes the partition formed by adding a part of size k to the partition μ, and $\epsilon(k,\mu)+1$ represents which part k becomes in $\mu+(k)$. For $m \leq 0$ a similar statement can be made using the commutation relations (14), (15) and (16).

Proof From (23) the action of $\mathbf{G}_m$ on a function $P[X] \in \Gamma$ can be written as

$$\begin{aligned}
\mathbf{G}_m P[X] &= P[X-(1-q)/z]\Omega[(1-\epsilon)zX]\Big|_{z^m} \\
&= \sum_{i\geq 0} q^i (q_i^\perp P)[X-1/z]\Omega[(1-\epsilon)zX]\Big|_{z^m} \\
&= \sum_{i\geq 0} q^i \mathbf{Q}_{m+i} q_i^\perp P[X]
\end{aligned}$$

where $q_i^\perp$ is

$$\mathbf{Q}[X+z]\Big|_{z^i} = q_i^\perp Q_\lambda[X] = \sum_{\mu:\lambda/\mu\in\mathcal{H}_i} 2^{a(\lambda/\mu)} Q_\mu[X],$$

and thus equation (24) follows from (14) and (15). □

Example 1. *We compute $G_{(3,2,1)}[X;q]$ using the Proposition above. We have*

$$\begin{aligned}
G_{(3,2,1)}[X;q] = \mathbf{G}_3(\mathbf{G}_2(Q_{(1)}[X])) &= \mathbf{G}_3\left(\sum_{i\geq 0}\sum_{(1)/\mu\in\mathcal{H}_i} 2^{a((1)/\mu)}(-1)^{\epsilon(2+i,\mu)}Q_{\mu+(2+i)}[X]\right) \\
= \mathbf{G}_3(Q_{(2,1)}) + 2q\mathbf{G}_3(Q_{(3)}) &= \sum_{i\geq 0}\sum_{(2,1)/\mu\in\mathcal{H}_i} 2^{a((2,1)/\mu)}(-1)^{\epsilon(3+i,\mu)}Q_{\mu+(3+i)}[X] + \\
&\quad +2q\left(\sum_{i\geq 0}\sum_{(3)/\nu\in\mathcal{H}_i} 2^{a((3)/\nu)}(-1)^{\epsilon(3+i,\nu)}Q_{\nu+(3+i)}[X]\right) \\
&= (q^0 2^0 Q_{(3,2,1)} + q^1 2^1 Q_{(4,2)} + q^2 2^1 Q_{(5,1)}) + 2q(q^1 2^1 Q_{(4,2)} + q^2 2^1 Q_{(5,1)} + q^3 2^1 Q_{(6)}) \\
&= Q_{(3,2,1)} + (2q+4q^2)Q_{(4,2)} + (2q^2+4q^3)Q_{(5,1)} + 4q^4 Q_{(6)}.
\end{aligned}$$

3.2. **Properties of the polynomials $L_{\lambda\mu}(q)$.** The Q-Kostka polynomials introduced here have a number of remarkable properties that are very similar to those of Kostka Foulkes polynomials listed in the previous section. We have already seen the analog of Property 4 holds for Q-Kostka polynomials. In what follows we will consider the other remaining properties.

An important consequence of equation (24) is a Morris-like recurrence which expresses the Q-Kostka polynomials $L_{\lambda\mu}(q)$ in terms of smaller ones.

Proposition 4. *We have the following recurrence*

$$L_{\alpha,(n,\mu)}(q) = \sum_{s=1}^{t:\alpha_t \geq n} (-1)^{s-1} q^{\alpha_s - n} \sum_{\lambda:\lambda/\alpha^{(s)} \in \mathcal{H}_{(\alpha_s - n)}} 2^{a(\lambda/\alpha^{(s)})} L_{\lambda\mu}(q), \tag{25}$$

where $n > \mu_1$ and $\alpha^{(s)}$ is α with part α_s removed.

Proof If $n > \mu_1$ we have that

$$\mathbf{G}_n G_\mu[X;q] = G_{(n,\mu)}[X;q] = \sum_\alpha L_{\alpha,(n,\mu)}(q) Q_\alpha[X]. \tag{26}$$

On the other hand $G_\mu[X;q] = \sum_\lambda L_{\lambda\mu}(q) Q_\lambda[X]$ and so

$$\mathbf{G}_n \left(\sum_\lambda L_{\lambda\mu}(q) Q_\lambda[X] \right) = \sum_\mu L_{\lambda\mu}(q) \mathbf{G}_n(Q_\lambda[X]).$$

Using the action in (24) we have

$$\mathbf{G}_n G_\mu[X;q] = \sum_\lambda L_{\lambda\mu}(q) \sum_{i\geq 0} q^i \sum_{\nu:\lambda/\nu \in \mathcal{H}_i} 2^{a(\lambda/\nu)} (-1)^{\epsilon(n+i,\nu)} Q_{\nu+(n+i)}[X]. \tag{27}$$

For $\alpha = \nu + (n+i)$, equating the coefficients of Q_α in (26) and (27) we get

$$L_{\alpha,(n,\mu)}(q) = \sum_\lambda \sum_{i\geq 0} q^i 2^{a(\lambda/\alpha-(n+i))} (-1)^{\epsilon(n+i,\alpha-(n+i))} L_{\lambda\mu}(q).$$

By reindexing $i := \alpha_s - n$ for $\alpha_s - n \geq 0$ we obtain the desired recurrence (25). □

Example 2. *Let $n = 5$ and $L_{(6,2),(5,2,1)}(q) = 2q + 4q^2$. Using the recurrence we have one s such that $\alpha_s \geq 5$, i.e. $\alpha_1 = 6$. So*

$$L_{(6,2),(5,2,1)}(q) = q^{6-5} \sum_{\lambda/(2)\in\mathcal{H}_1} 2^{a(\lambda/(2))} L_{\lambda(2,1)}(q)$$

$$= q\big(2L_{(21),(21)}(q) + 2L_{(3),(21)}(q)\big) = q(2 + 2\cdot 2q) = 2q + 4q^2.$$

As a consequence of the Morris-like recurrence we have the following

Corollary 5. *Let $\mu \leq \lambda$ in dominance order.*
1. *If $n > \lambda_1$ then $L_{(n,\lambda),(n,\mu)}(q) = L_{\lambda\mu}(q)$.*
2. *$L_{\lambda\lambda}(q) = 1$ and $L_{(|\lambda|)\lambda}(q) = 2^{\ell(\lambda)-1} q^{n(\lambda)}$.*
3. *$2^{\ell(\mu)-\ell(\lambda)}$ divides $L_{\lambda\mu}(q)$.*

Proof 1. There is only one term in the recurrence (25) in this case which is exactly $L_{\lambda\mu}(q)$.
2. The first is a consequence of (1). For the second, we have that the only term on the right hand side is $q^{|\lambda|-\lambda_1} 2 L_{(|\lambda|-\lambda_1)(\lambda_2,\ldots)}(q)$ which by induction is $q^{|\lambda|-\lambda_1+n((\lambda_2,\ldots))} 2\cdot 2^{\ell(\lambda)-2} = 2^{\ell(\lambda)-1} q^{n(\lambda)}$. This is the analog of Property 4 for the Kostka-Foulkes polynomials.
3. This property can be easily derived by induction from the recurrence. □

Using the Morris-like recurrence one can obtain a formula for the degree of $L_{\lambda\mu}(q)$ similar to Property 3 for Kostka-Foulkes.

Proposition 6. *If $\mu \leq \lambda$ in dominance order, we have*

$$deg_q L_{\lambda\mu}(q) = n(\mu) - n(\lambda).$$

The property that is most suggestive that these polynomials are analogs of the Kostka-Foulkes polynomials is

Conjecture 7. *The Q-Kostka polynomials $L_{\lambda\mu}(q)$ have non-negative coefficients.*

We can prove this conjecture for some particular cases. In general we believe that there should exist a similar combinatorial interpretation as for the Kostka-Foulkes polynomials. More precisely there should exist a statistic function d on the set of marked shifted tableaux, similar to the charge function on column strict tableaux, such that

$$L_{\lambda\mu}(q) = \sum_T q^{d(T)}$$

summed over marked shifted tableaux of shifted shape λ and content μ with diagonal entries unmarked.

In addition, we conjecture that this function must have the property that if T and S are two marked shifted tableaux such that by erasing the marks the two resulting tableaux coincide, then $d(T) = d(S)$.

For some of the polynomials $L_{\lambda\mu}(q)$, this observation determines completely the statistic on the tableaux. For instance there are two marked shifted tableaux classes of shape $(5,3)$ and content $(4,3,1)$ and $L_{(5,3),(4,3,1)}(q) = 2q + 4q^2$. Clearly the tableau with a 3 in the first row must have statistic 1 and with 3 in the second row has statistic 2. On the other hand, $L_{(6,2),(4,3,1)}(q) = 4q^2 + 4q^3$. This polynomial does not uniquely determine which of the two tableaux have statistic 2 and 3. We have used the function $G_{(4,3,1)}[X;q]$ to draw a conjectured tableau poset (similar to the case of column strict tableau) for the marked shifted tableaux with unmarked diagonals of content $(4,3,1)$ in an appendix.

We also note that monotonicity properties, similar to Property 7 and 8, hold for the Q-Kostka polynomials.

Conjecture 8. *Let $L'_{\lambda\mu}(q) := q^{n(\mu)-n(\lambda)} L_{\lambda\mu}(q^{-1})$. We have*

$$L'_{\lambda\mu}(q) \geq 2^{\ell(\nu)-\ell(\mu)} L'_{\lambda\nu}(q), \qquad \textit{for } \mu \leq \nu \textit{ in dominance order.}$$

We can prove this fact by using induction and the recurrence (25) for the case $\mu_1 = \nu_1$.

Example 3. *Let $\lambda = (6,2)$, $\mu = (4,3,1)$, $\nu = (5,2,1)$. We have $n(\lambda) = 2$, $n(\mu) = 5$, and $n(\nu) = 4$. The L' polynomials are*

$$L'_{\lambda\mu} = q^{5-2}(4/q^2 + 4/q^3) = 4 + 4q, \qquad L'_{\lambda\nu} = q^{4-2}(2/q + 4/q^2) = 4 + 2q,$$

and thus $L'_{\lambda\mu}(q) \geq 2^{3-3} L'_{\lambda\nu}(q)$.

Another property of the Kostka-Foulkes polynomials case that seems to hold in our case refers to the growth of the polynomials L. For the Kostka-Foulkes polynomials the conjecture is due to Gupta (see [1] and references therein).

Conjecture 9. *If r is an integer that is not a part in either partitions λ or μ, then*

$$L_{\lambda+(r),\mu+(r)}(q) \geq L_{\lambda\mu}(q).$$

The case where $r > \lambda_1$ (which also ensures that $r > \mu_1$) is obviously true since $L_{(r,\lambda),(r,\mu)}(q) = L_{\lambda\mu}(q)$ (see Corollary 5).

Example 4. *Let $\lambda = (5,3), \mu = (4,3,1)$ and $r = 2$. We have*

$$L_{(5,3,2),(4,3,2,1)}(q) - L_{(5,3),(4,3,1)}(q) = 2q + 4q^2 + 8q^3 - (2q + 4q^2) = 8q^3.$$

The polynomials $L_{\lambda\mu}(q)$ have a similar interpretation to property 9 using an analog of the q-Kostant partition function. Using the formal inversion from [1], equation (12) may be written as

$$q_\lambda[X] = \prod_{i<j} \left(\frac{1-R_{ij}}{1+R_{ij}}\right)^{-1} Q_\lambda[X]. \tag{28}$$

In fact if we let $\zeta_n := \prod_{i<j} \left(\frac{1-x_i/x_j}{1+x_i/x_j}\right)^{-1}$, we have that $\zeta_n = \sum_{\alpha\in\mathbb{Z}^n} \mathcal{R}(\alpha)e^\alpha$ where $\mathcal{R}(\alpha) = \sum_t a_t 2^t$ and a_t counts the number of ways the vector α can be written as a sum of positive roots of type A_{n-1}, t of which are distinct. The positive roots in the root lattice of A_{n-1} are $\{e_i - e_j\}_{1\le i<j\le n}$, where $e_i = (0, \dots, 1, \dots 0)$ is the canonical basis of $\mathbb{Z}^n$.
The q-analog of ζ_n is defined to be

$$\zeta_n(q) := \prod_{i<j} \left(\frac{1-qx_i/x_j}{1+qx_i/x_j}\right)^{-1},$$

and thus $\zeta_n(q) = \sum_{\alpha\in\mathbb{Z}^n} \mathcal{R}_q(\alpha)e^\alpha$ where $\mathcal{R}_q(\alpha) = \sum_{t,k} a_{t,k}2^t q^k$ and $a_{t,k}$ counts the number of ways the vector α can be written as a sum of k positive roots, t of which are distinct.

We can express the Q-Kostka polynomials in terms of $\mathcal{R}_q(\alpha)$ as

$$L_{\lambda\mu}(q) = \sum_{\alpha\colon Q_{\alpha+\mu}=\pm 2^t Q\lambda} \pm 2^t \mathcal{R}_q(\alpha).$$

It is possible to express the equation above using the action of the symmetric group on Schur's Q-functions, yielding an alternating sum similar to Property 9. Unfortunately the action of the symmetric group on Schur's Q-functions indexed by a general integer vector is not as elegant as for Schur functions (due to relation (15)).
Remark: Most of the properties of the Q-Kostka polynomials $L_{\lambda\mu}(q)$ are analogous to the Kostka-Foulkes. A few properties for the Kostka-Foulkes polynomials do not have a corresponding property for the Q-Kostka polynomials.

(1) The analog of Property 6 does not seem to hold since computations of $G_\lambda[X;q]$ where q is set to a root of unity do not factor.
(2) There does not seem to exist an elegant relationship between $G_\lambda[X;q]$ and its dual basis (Property 11).
(3) A property similar to that of Property 10 does not seem to hold. We do not know if there is a relationship between $G_\lambda[X;q]$ and a Hall-like algebra.
(4) The symmetries of the Macdonald symmetric function in Λ cannot hold in Γ and do not suggest what a two parameter analog of what these functions must be.

3.3. **Generalized (parabolic) Q-Kostka polynomials.** Shimozono and Weyman [17], defined a generalization of the Kostka-Foulkes polynomials that are a q-analog of the Littlewood-Richardson coefficients. They were originally defined as the coefficient of a Schur function in a symmetrized rational series, however it became clear in later work [18] that they can be defined as coefficients in families of symmetric functions using formulas similar to those presented here.

This construction exists in complete analogy within the Q-function algebra. We will create a family of functions in Γ which are indexed by a sequence of strict partitions. Let $\mu^* = (\mu^{(1)}, \mu^{(2)}, \dots, \mu^{(k)})$ where $\mu^{(i)}$ is a strict partition and set $\eta = (\ell(\mu^{(1)}), \ell(\mu^{(2)}), \dots, \ell(\mu^{(k)}))$. Define $Roots_\eta = \{(i,j) : 1 \le i \le \eta_1 + \cdots + \eta_r < j \le n \text{ for some } r\}$ and then define the function

$$G_{\mu^*}[X;q] = \prod_{(i,j)\in Roots_\eta} \frac{1+qR_{ij}}{1-qR_{ij}} Q_{\bar{\mu^*}}[X] \tag{29}$$

A generating function, vertex operator, and a Morris-like recurrence analogous to equations (21), (23) and (25) may be derived from this definition.

If we set $\bar{\mu^*}$ equal to the concatenation of the partitions in μ^*, then $G_{\mu^*}[X;0] = Q_{\bar{\mu^*}}[X]$ and $G_{\mu^*}[X;1] = Q_{\mu^{(1)}}[X]Q_{\mu^{(2)}}[X]\cdots Q_{\mu^{(k)}}[X]$. Define the polynomials $L_{\lambda\mu^*}(q)$ by the expansion

$$G_{\mu^*}[X;q] = \sum_\lambda L_{\lambda\mu^*}(q)Q_\lambda[X]. \tag{30}$$

Computing these coefficients suggests the following remarkable conjecture and indicates that these coefficients are an important q-analog of the structure coefficients of the $Q_\lambda[X]$ functions in the same way that the parabolic Kostka coefficients are q-analogs of the Littlewood-Richardson coefficients.

Conjecture 10. *For a sequence of partitions μ^*, if $\bar{\mu^*}$ is a partition then $L_{\lambda\mu^*}(q)$ is a polynomial in q with non-negative integer coefficients.*

4. Appendix: Tables of $2^{\ell(\lambda)-\ell(\mu)}L_{\lambda\mu}(q)$ for $n = 4,5,6,7,8,9$

$$\begin{bmatrix} (3,1) & (4) \\ 1 & q \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} (3,2) & (4,1) & (5) \\ 1 & 2\,q & q^2 \\ 0 & 1 & q \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} (3,2,1) & (4,2) & (5,1) & (6) \\ 1 & 2\,q^2+q & 2\,q^3+q^2 & q^4 \\ 0 & 1 & 2\,q & q^2 \\ 0 & 0 & 1 & q \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} (4,2,1) & (4,3) & (5,2) & (6,1) & (7) \\ 1 & q & 2\,q^2+q & 2\,q^3+q^2 & q^4 \\ 0 & 1 & 2\,q & 2\,q^2 & q^3 \\ 0 & 0 & 1 & 2\,q & q^2 \\ 0 & 0 & 0 & 1 & q \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{array}{cccccc}
(4,3,1) & (5,2,1) & (5,3) & (6,2) & (7,1) & (8)
\end{array}$$
$$\begin{bmatrix}
1 & 2\,q & 2\,q^2+q & 2\,q^2+2\,q^3 & q^3+2\,q^4 & q^5 \\
0 & 1 & q & 2\,q^2+q & 2\,q^3+q^2 & q^4 \\
0 & 0 & 1 & 2\,q & 2\,q^2 & q^3 \\
0 & 0 & 0 & 1 & 2\,q & q^2 \\
0 & 0 & 0 & 0 & 1 & q \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

$$\begin{array}{cccccccc}
(4,3,2) & (5,3,1) & (5,4) & (6,2,1) & (6,3) & (7,2) & (8,1) & (9)
\end{array}$$
$$\begin{bmatrix}
1 & 2\,q+4\,q^2 & 2\,q^3+q^2 & 2\,q^2+4\,q^3 & q^2+2\,q^4+4\,q^3 & 4\,q^4+q^3+2\,q^5 & 2\,q^6+2\,q^5 & q^7 \\
0 & 1 & q & 2\,q & 2\,q^2+q & 2\,q^2+2\,q^3 & q^3+2\,q^4 & q^5 \\
0 & 0 & 1 & 0 & 2\,q & 2\,q^2 & 2\,q^3 & q^4 \\
0 & 0 & 0 & 1 & q & 2\,q^2+q & 2\,q^3+q^2 & q^4 \\
0 & 0 & 0 & 0 & 1 & 2\,q & 2\,q^2 & q^3 \\
0 & 0 & 0 & 0 & 0 & 1 & 2\,q & q^2 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & q \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

5. Appendix: example of conjectured tableaux poset of content $(4,3,1)$

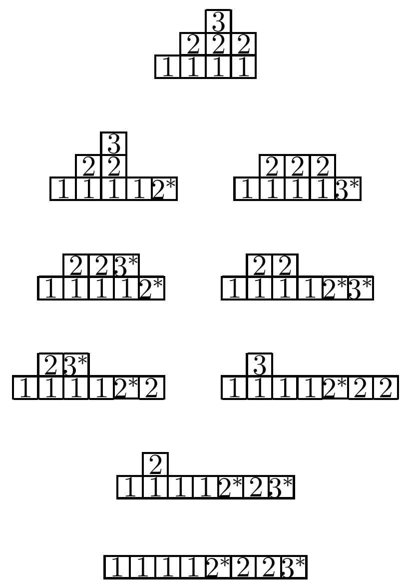


Figure 2. The cells marked with a k^* can be labeled with either k or k', we conjecture that the statistic is independent of these markings. The value of $G_{(4,3,1)}[X;q]$ determines the position of each of the shifted tableaux here except for the two of shape $(6,2)$. The covering relation is unknown, but the rank function indicates that it is not the same as the charge statistic.

Acknowledgement: Thank you to Nantel Bergeron for many helpful suggestions on this research.

References

[1] Désarménien J., Leclerc B., and Thibon J.-Y., "Hall-Littlewood Functions and Kostka-Foulkes Polynomials in Representation Theory." *Sèminaire Lotharingien de Combinatoire*, **B32c**(1994), 38 pp.

[2] Haiman M., "On Mixed Insertion, Symmetry, and Shifted Young Tableaux." *J. Comb Theory (A)*, **50**(1989) 196–225.

[3] P. Hall, "The algebra of partitions." *Proc. 4th Canadian Math. Congress, Banff*, (1959).

[4] Jing, N., "Vertex Operators, Symmetric Functions and the Spin Group Γ_n." *J. Algebra* **138**(1991), 340–398.

[5] Jing N., "Vertex Operators and Hall-Littlewood Symmetric Functions." *Advances in Math.* **87**(1991), 226–248.

[6] Foulkes H. O., "A survey of some combinatorial aspects of symmetric functions." in *Permutations.* (1974) Gauthier-Villars, Paris.

[7] Fulton W., *Young Tableaux.* Cambridge University Press, Cambridge, 1997.

[8] Garsia, A. M., "Orthogonality of Milne's polynomials and raising operators." *Discrete Math*, **99**, 247–64.

[9] Kirillov A., "Ubiquity of Kostka polynomials." *Phys. and Comb.*, *Proceedings of the Nagoya 1999 International Workshop on Physics and Combinatorics, Nagoya University*, August 23-27, (1999), 85–200.

[10] A. N. Kirillov and N. Yu. Reshetikhin, *The Bethe Ansatz and the combinatorics of Young tableaux*, J. Soviet Math. **41** (1988), 925–955.

[11] Lascoux A., Pragacz P., "Orthogonal Divided Differences and Schubert Polynomials, $\tilde{P}$-Functions and Vertex Operators." *Mich. Math. J.* **48**(2000), 417–441.

[12] Lascoux A., Schützenberger M. P., "Sur une conjecture de H. O. Foulkes." *C. R. Acad. Sci. Paris,* **286A**(1978) 323–324.

[13] Lascoux A. and Schützenberger M. P., "Le Monoid plaxique." *Noncommutative structures in Algebra and Geometric Combinatorics*, CNR, Rome. **109**(1981), 129–156.

[14] Macdonald, I. G., *Symmetric Functions and Hall Polynomials.* Cambridge University Press, Cambridge, 1995.

[15] Sagan, B. E., "Shifted Tableaux, Schur Q-functions and a Conjecture of R. Stanley." *J. Comb. Theory (A)*, **45**(1987), 62–103.

[16] Schilling A. and Warnaar O., *Inhomogeneous lattice paths, generalized Kostka polynomials and A_{n-1} supernomials*, Comm. in Math. Phys. 202, (1999), 359–401.

[17] Shimozono M., Weyman J., "Graded characters of modules supported in the closure of a nilpotent conjugacy class." *European J. Combin.* **21** no. 2(2000) 257–288.

[18] Shimozono M., Zabrocki M., "Hall-Littlewood Vertex Operators and Generalized Kostka Polynomials." *Adv. Math.* **158**(2001), 66–85.

[19] Stembridge, J. R., "Shifted Tableaux and the Projective Representations of Symmetric Groups." *Advances in Math.* **74**(1989), 87–134.

[20] Stembridge, J. R., "On Schur's Q-functions and the Primitive Idempotents of a Commutative Hecke Algebra." *J. Alg. Combinatorics* **1**(1992), 71–95.

[21] Worley, D. R., "A Theory of Shifted Young Tableaux." Thesis, MIT (1984).

School of Mathematics, University of Minnesota, Minneapolis, MN, 55455 USA
E-mail address: `tudose@math.umn.edu`

Department of Mathematics and Statistics, York University, Toronto, Ontario, M3J 1P3 CANADA
E-mail address: `zabrocki@mathstat.yorku.ca`

EXCHANGE RELATIONS, DYCK PATHS AND COPOLYMER ADSORPTION

E. J. JANSE VAN RENSBURG

ABSTRACT. We consider a lattice model of fully directed copolymer adsorption equivalent to the enumeration of vertex-coloured Dyck paths. For two infinite families of periodic colourings we are able to solve the model exactly using a type of symmetry we call an *exchange relation*. For one of these families we are able to find an asymptotic expression for the location of the critical adsorption point as a function of the period of the colouring. This expression describes the effect of a regular inhomogeneity in the polymer on the adsorption transition. We have found similar results for other directed path models.

RÉSUMÉ. Nous considérons un modèle dirigé discret d'adsorption de polymères qui est équivalent à l'énumération de chemins de Dyck colorés aux sommets. Pour deux familles infinies de colorations périodiques, nous pouvons résoudre le modèle de manière exacte en utilisant un type de symétrie que nous appelons 'relation d'échange'. Pour une de ces familles, nous donnons une expression asymptotique pour le point critique d'adsorption comme fonction de la période de la coloration. Cette expression décrit l'effet d'une inhomogénéité régulière du polymère dans la transition d'adsorption. Nous avons obtenu des résultats semblables pour d'autres modèles de chemins dirigés.

1. INTRODUCTION

One of the most active areas of research at the interface of combinatorics and statistical mechanics has been the investigation of the physical properties of polymers in solution. The canonical model of this is the self-avoiding walk [14]. A self-avoiding walk is a path on a regular lattice that does not intersect itself. By considering additional properties and restrictions the self-avoiding walk model of polymers can be used to mimic a variety of physical situations. In this paper we consider *polymer adsorption*. In particular, the adsorption of polymers whose component molecules (which are called *monomers*) have different properties; such polymers are called *copolymers*.

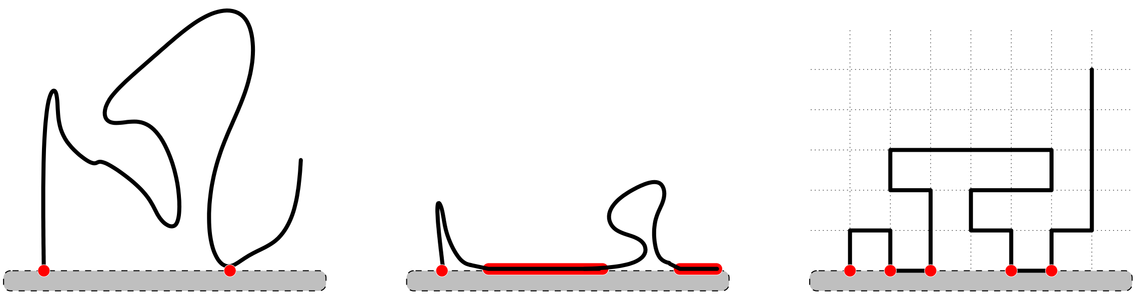


FIGURE 1. From left-to-right; a polymer in its desorbed or free phase, a polymer in the adsorbed phase, and a self-avoiding walk with vertex-visits highlighted.

1.1. **Polymers and Dyck paths.** Consider a long chain polymer in solution close to the wall of the container (see Figure 1). If there is a weak attractive force between the wall and the polymer, then the fraction of monomers in contact with the wall will be zero as the length of the polymer goes to infinity; in this case we say that the polymer is *desorbed* or *free*. If the attractive force is strong, then the limiting fraction of monomers in contact with the wall will be positive; we say that the polymer is *adsorbed*. Each of these distinct behaviours is called a *phase*, the change between the two phases is called a *phase transition*, and the point at which the transition occurs is called a *critical point*.

A variety of lattice models of polymer adsorption have received much attention in the literature over the last two decades. Perhaps the most well known model of this type is an adsorbing self-avoiding walk in a half-space first defined in [9]. In this model configurations are weighted according to the number of vertices or edges lying in the boundary (which is the X-axis in two dimensions and XY-plane in three dimensions). We refer to such vertices and edges as *vertex-visits* and *edge-visits*.

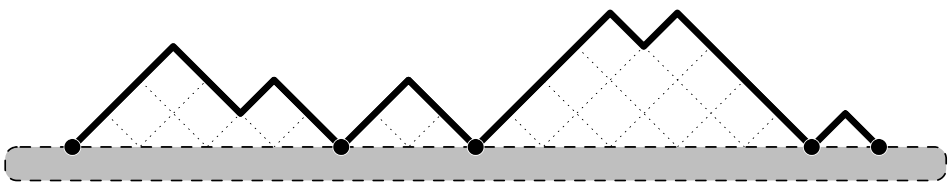


FIGURE 2. An example of a Dyck-path. The horizontal axis can be considered an adsorbing wall.

Underlying the "physical" properties of the adsorbing self-avoiding walk are its combinatorial properties. Unfortunately there are very few rigorous combinatorial results known for this model [14], and attention is instead focused on directed versions of the above problem [10], see also references [3, 5, 7, 15, 17]. The problem of adsorbing directed walks is equivalent to the problem of enumerating Dyck paths [7] according to their half-length and number of visits. The generating function of this model is given by

$$D(z,1) = \frac{1-\sqrt{1-4z}}{2z}, \tag{1}$$

$$D(z,v) = \frac{v}{1-\frac{v}{2}\left(1-\sqrt{1-4z}\right)}, \tag{2}$$

where z is conjugate to the half-length of the path, and v is conjugate to the number of vertex-visits.

This model of Dyck paths can be interpreted as a model of homopolymers since each vertex-visit is given the same weight. In nature one can find many examples of copolymers, such as DNA, whose monomers (potentially) have different physical and chemical properties. To study the effect of such inhomogeneity we consider a variation of the Dyck path model in which different vertex-visits may have different weights. In particular we consider a *fixed* colouring of the even vertices[1] of Dyck paths, and one wishes to enumerate the number of vertex-visits of each different colour. See Figure 3.

[1] If the vertices along the path are labeled sequentially from the left by $0, 1, 2, 3, \ldots$, then those with even labels are called "even vertices". The remaining vertices are "odd" and cannot visit the adsorbing diagonal — hence we can safely ignore their colour.

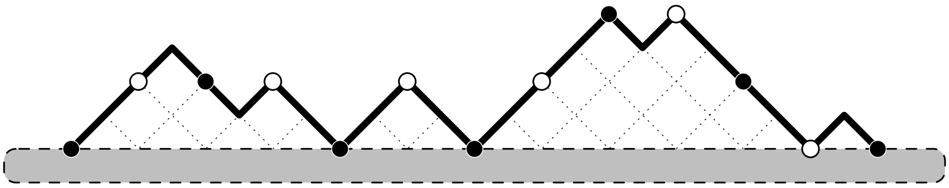


FIGURE 3. An example of a vertex-coloured Dyck-path. Every even vertex is coloured $A, B, A, B \ldots$. There are 4 A-visits and 1 B-visit.

1.2. **From generating functions to phase transitions.** There is a close relationship between the statistical mechanics and combinatorics in this model, and we describe its "physical" behaviour from the behaviour of the generating function. If the numerical value of v in $D(z, v)$ is increased, then paths with larger numbers of visits will contribute more to the generating function and, since they are in some sense "more important", determine the thermodynamic phase of the model — whether or not it is adsorbed or desorbed. Consider now the following:

$$D(z,v) = \sum_{n\geq 0} \left(\sum_{m\geq 1} c_{n,m} v^m \right) z^n = \sum_{n \geq 0} Z_n(v) z^n \tag{3}$$

where $Z_n(v)$ is the *partition function* of the model and is related to the radius of convergence of the model by

$$\log z_c(v) = \left(\lim_{n\to\infty} \frac{1}{n} \log Z_n(v) \right)^{-1} = -\mathcal{F}(v), \tag{4}$$

where $z_c(v)$ is the radius of convergence, and $\mathcal{F}(v)$ is the *limiting canonical free-energy density*[2] [11]. This relation between $z_c(v)$ and $\mathcal{F}(v)$ explicitly connects the combinatorics and thermodynamics of the model. The change of phase, from desorbed to adsorbed, is signalled by a non-analytic point in $\mathcal{F}(v)$ and hence also $z_c(v)$. Describing the location of this transition for vertex-coloured Dyck paths is one of the major goals of this paper.

From equation (2) one sees that:

$$z_c(v) = \begin{cases} 1/4, & v \leq 2 \\ (v-1)/v^2, & v > 2 \end{cases}, \tag{5}$$

from which the limiting free energy, $\mathcal{F}(v)$, can be explicitly computed:

$$\mathcal{F}(v) = -\log z_c(v) = \begin{cases} 2\log 2, & v \leq 2 \\ 2\log v - \log(v-1), & v > 2. \end{cases} \tag{6}$$

Observe that for $v > 0$, the free energy is a continuous function of v, but that $\mathcal{F}(v)$ is non-analytic at $v = v_c = 2$. The non-analytic point is interpreted a phase transition in this model; we refer to this value of v as the *critical adsorption point*, and denote it by v_c. Since the first derivative of $\mathcal{F}(v)$ to $\log v$ is also continuous, we call this phase transition a *continuous transition*. A phase transition is a *first order transition* if the first derivative of $\mathcal{F}(v)$ is discontinuous.

[2]We note that the derivative of the free energy to $\log(v)$ is the limiting density of visits, *i.e.* the limiting average number of visits per length. The second derivative of the limiting free energy to $\log(v)$ is the *specific heat* which is a measure of the fluctuations in the density of visits.

A singularity analysis of this generating function, allows one to compute the mean number of visits as a function of the length of the path:

$$\text{mean number of visits}(n) \sim \begin{cases} \frac{2}{2-v} & + o(1), & v < 2 \\ \sqrt{\pi}\,\sqrt{n} & + O(1), & v = v_c = 2 \\ \frac{v-2}{v-1}\, n & + O(\sqrt{n}), & v > 2 \end{cases} \tag{7}$$

This shows that that in the $n \to \infty$ limit, the density of visits is 0 in the desorbed phase and *at* the critical point, and is positive in the adsorbed phase. This can also be seen by an analysis of the free energy (see [11] for details). If $v < 2$, then all the derivatives of the free energy, with respect to v, are zero; the density of visits is zero, and the free energy is determined entirely by a class of Dyck paths which visits the adsorbing line with zero density. Whereas when $v > 2$, the first derivative of the free energy is positive and so the free energy is dominated by a class of Dyck paths which has a non-zero density of visits. Further analysis shows that the second derivative of $\mathcal{F}(v)$ with respect to $\log(v)$ (*i.e.* the specific heat) is:

$$\frac{\mathrm{d}^2\mathcal{F}(v)}{\mathrm{d}(\log v)^2} = \begin{cases} 0, & v \le 2; \\ \frac{v}{(v-1)^2}, & v > 2. \end{cases} \tag{8}$$

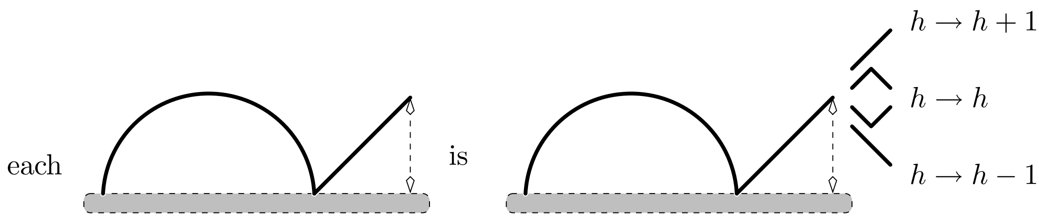
This is a measure of the fluctuations in the number of visits and it is maximised at $v = v_c = 2$. We expect this to be the case, since when the model is away from the critical point, typical configurations have roughly the same number of visits, however, when the model is close to the critical point and there is a change in the behaviour of the model, we expect that typical configurations will have widely different numbers of visits.

In the next section we describe two commonly used techniques for computing generating functions of lattice models and demonstrate why they are not practical for the enumeration of vertex-coloured Dyck paths. We introduce a type of symmetry that we call an exchange relation that allows us to find generating functions for two infinite families of vertex-coloured Dyck paths. These generating functions all have a similar form and for one of the two families we are able to use it to find an asymptotic expression for the location of the adsorption critical point. In Section 3 we find similar results for other directed models of copolymers.

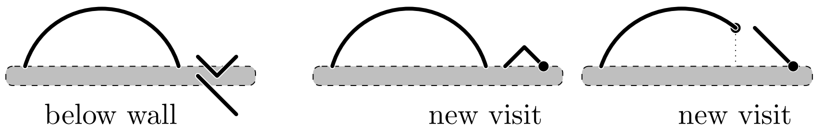
2. Exchange relations

Perhaps the two most successful techniques for the enumeration of lattice animals are the Temperley method and recursive constructions that are sometimes called “wasp-waist” factorisations. Both of these methods are readily applied to the enumeration of Dyck paths according to their length and number of visits. We shall quickly review these standard techniques in order to demonstrate that they are not well suited to the problem of vertex-coloured Dyck paths, and that one must, in fact, resort to other methods.

The Temperley method is based around the idea that many lattice-animals can be constructed column-by-column [2, 16]. One can apply this method to Dyck paths, by constructing left-factors of Dyck paths two steps at a time, while keeping track of the number of visits and the height of the last vertex:



but the following cases must be treated separately.

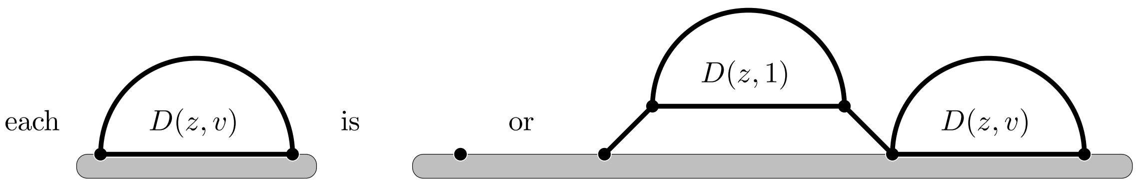


Let us write $D(s;z,v)$ as the generating function of left-factors of Dyck paths, where s, z and v are respectively conjugate to the height of the final vertex, the half-length of the path, and the number of visits. The above construction translates directly (see for example [2]) into a linear functional equation:

$$D(s;z,v) = v + z(s+2+1/s)D(s;z,v) + z(v-2-1/s)D(0;z,v) + z(v-1)\left(\frac{\partial D}{\partial s}\right)\Big|_{s=0}, \tag{9}$$

This can then be solved (see [1, 2], for example) to give $D(0;z,v)$ which is the generating function of Dyck paths given above.

The wasp-waist factorisation, on the other hand, consists of cutting a Dyck path at an especially thin point into two smaller Dyck paths. More specifically



This factorisation shows that every adsorbing Dyck path is either a single visit, or has a prefactor which is a Dyck path with exactly two visits (its first and last vertices), and then followed by an arbitrary adsorbing Dyck path (which may consist of a single vertex). This gives:

$$D(z,v) = v + zvD(z,1)D(z,v), \tag{10}$$

which can be solved to recover the Dyck path generating function.

Consider now the enumeration of vertex-coloured Dyck paths for some fixed periodic colouring with period p:

- To apply the Temperley method one must consider the colour of the vertex being added and this leads to a system of p equations involving p generating functions — one for each different half-length modulo p. These equations can be reduced to a single functional equation which may be interpreted as appending $2p$ edges to the end of a left-factor. As p increases, the equation contains higher and higher derivative order terms which make solving it increasingly difficult.

- The wasp-waist factorisation given above does not preserve the colouring and one must take into account the colour of the first and second visit-vertices in the factorisation. This leads to a system of p simultaneous equations in p generating functions — each one counts Dyck-paths whose vertices are coloured starting from a different point in the period.

For small periods (say up to 4 or 5) these equations may be solved by hand or computer (see [11, 17]), but beyond this the calculations quickly use up available human perseverance, as well as computer time and memory.

Though we have not investigated *all* the techniques that may been used to enumerate Dyck paths, we expect that similar problems will be encountered — though it would be interesting to see this problem tackled using other techniques in the literature.

In this paper we have succeeded in solving this problem for two infinite families of periodic colourings [12, 13]. The key to their solution is a type of symmetry that we call an *exchange relation.* This relation establishes a symmetry between $D(z,v)$ and $D(z,1)$ with respect to $1 \longleftrightarrow v$. While this equation cannot be used to solve for $D(z,v)$ it does have the advantage that it also holds for certain vertex-colourings and enables us to find generating functions and asymptotic expressions for the adsorption critical point.

2.1. **Exchange relation in Adsorbing Dyck paths.** Let P be a non-empty and unweighted Dyck path. Start at the left-most vertex in P at the origin, and weight visits in P by v in sequence. After a certain arbitrary number of visits have been weighted, but not all, the situation is as depicted in the top half of Figure 2.1. This configuration factors into two halves: the left part is a weighted Dyck path, which may be a single vertex and so is enumerated by $D(z,v)$. The second half is an unweighted Dyck path, which may not be a single vertex (since the last vertex of P is not weighted), and so is enumerated by $\big(D(z,1)-1\big)$.

If the next visit in the path is now weighted then the lower half of Figure 2.1 is obtained. Again the walk factors into two halves — one weighted and one unweighted. The weighted part of the walk is longer than before and so cannot be a single vertex, and consequently is counted by $\big(D(z,v)-v\big)$. The unweighted half is shorter and may now possibly be a single vertex, and is counted by $D(z,1)$. The statistics of this new configuration are the same as the starting configuration, except that there is exactly one more weighted visit, and so exactly one extra factor of v.

This construction creates a correspondence between pairs of weighted and unweighted Dyck paths, and applying it to all possible pairs of weighted and unweighted paths gives the following functional relation involving $D(z,v)$ and $D(z,1)$:

$$vD(z,v)\Big(D(z,1)-1\Big) = \Big(D(z,v)-v\Big)D(z,1) \tag{11}$$

This relation between $D(z,v)$ and $D(z,1)$ exhibits an exchange symmetry which exchanges $v \longleftrightarrow 1$ between the generating functions of absorbing and free Dyck paths; notice the role reversal of the generating functions on both sides of the equation.

Solving this equation for $D(z,v)$ gives

$$D(z,v) = \frac{vD(z,1)}{v+(1-v)D(z,1)}. \tag{12}$$

This solution is not identical to the relation in equation (10), but using the fact that $D(z,1) = 1 + zD(z,1)^2$, shows them to be equivalent. Further, it is not possible to solve for $D(z,v)$ from equation (11); a solution for $D(z,1)$ is needed as well and cannot be obtained by setting $v = 1$ in the above. This indicates that the exchange relation is not equivalent to

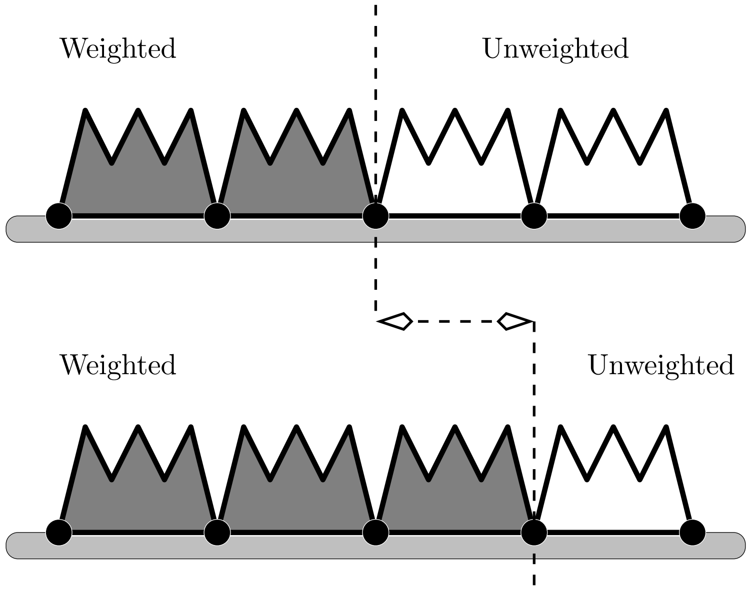


FIGURE 4. This is a schematic representation of a partially weighted Dyck path. Starting from the left and working towards the right, each visit is weighted v. At some point *before* the end of the path is reached the situation is as represented in the top half of the illustration. Proceeding on to the next visit gives the lower illustration. The correspondence between these two pictures gives equation (11).

equations (10) and (9), and at the same time, it does not appear to be particularly useful since $D(z,1)$ must be computed by some other means.

On the other hand, it establishes a relation which must be satisfied by the generating functions of certain vertex-coloured Dyck path models, where it proves to be very useful. It also appears to hold, in very similar forms, for other models (see Section 3).

2.2. **Generating Functions of paths coloured by $\{AB^{p-1}\}^*A$ and $\{BA^{p-1}\}^*B$.** The exchange relation in equation (11) can be applied to certain models of vertex-coloured Dyck paths. We consider a model in which the vertices of the Dyck path are coloured with two colours A and B, and A-visits are weighted by v_a and B-visits by v_b. In this and subsequent sections we shall put $v_b = 1$ and $v_a = a$.

As described above, applying the Temperley or factorisation methods to this model is not practical excepting for very short periodic colourings (and we believe that other techniques in the literature will encounter similar difficulties). The exchange relation we have defined above does not provide us with results for general colourings, however it proves extremely useful for two families of periodic colourings; in particular it enables us to find solutions in the case of walks coloured by the periodic sequence $\{AB^{p-1}\}^*A$ and the "complementary" colouring $\{BA^{p-1}\}^*B$ (for a given fixed period p). Other colourings seem not to be amenable to the treatment here.

We proceed in much the same manner as above. Consider a weighted Dyck path (all visits are weighted v), and start colouring its even vertices by the sequence $\chi = \{AB^{p-1}\}^*A$. Each A-vertex which is also a visit is given a weight a, while B vertices are ignored. Let us stop colouring the Dyck path at some A-visit, leaving the remaining part of the path uncoloured. This configuration factors into two Dyck paths, one part is coloured and has half-length equal to 0 (mod p), while the other is uncoloured.

If the path is then labelled up to and including the next A-visit, then the half-length of the coloured part of the path is increased by a multiple of p again, while the unweighted part of the path decreases in half-length by the same amount — and again the configuration factors into a coloured path of half-length 0 (mod p) and an uncoloured path. This construction forms the exchange relation — and to underline this symmetry we require that both the coloured and uncoloured parts of the walk have half-length 0 (mod p).

Let us now define the following generating functions, in which z is conjugate to the half-length, v is conjugate to the total number of visits (coloured or otherwise) and a is conjugate to the number of A-visits:

Definition 1. *Fix p, the period of the colouring. Then:*

- *$U(z, v|p)$ is the generating function of all* uncoloured *Dyck paths of half-length* 0 *(mod p), with visits generated by v;*
- *$L(z, v, a|p)$ is the generating function of all Dyck paths* coloured *or* labelled *by $\chi = \{AB^{p-1}\}^*A$, with half-length* 0 *(mod p), with visits generated by v and A-visits generated by a.*

The exchange relation obtained in the construction above can be written in terms of $U(z, v|p)$ and $L(z, v, a|p)$:

Theorem 1. *For any fixed p, the generating functions U and L satisfy*

$$aL(z, v, a|p)\big(U(z, v|p) - v\big) = \big(L(z, v, a|p) - va\big)U(z, v|p), \tag{13}$$

and therefore

$$L(z, v, a|p) = \frac{vaU(z, v|p)}{va + (1 - a)U(z, v|p)}. \tag{14}$$

□

Notice that $L(z, v, a|p)$ is not the generating function of all coloured Dyck paths of arbitrary half-length; only those paths of half-length 0 (mod p) are counted. This is somewhat disappointing from the combinatorial point of view, but it is nevertheless enough to describe the physics of the model. In particular the radius of convergence $z_c(v, a)$ of $L(z, v, a|p)$ is equal to the radius of convergence of the full generating function. Fortunately we are able to find the full generating function, but first we require two more definitions.

Definition 2. *Let $F(z, v, a|p)$ be the full generating function of Dyck paths coloured by $\chi = \{AB^{p-1}\}^*A$. Further, let $\overline{F}(z, v|p) = \lim_{a\to 0}(F(z, v, a|p)/a)$, which is the generating function of these coloured Dyck paths, such that all but the first A-visit are forbidden.*

Notice that $L(z, v, a|p)$ is obtained by taking every p-th coefficient of $F(z, v, a|p)$, and that $\overline{F}(z, v|p)$ is also the coefficient of a^1 in $F(z, v, a|p)$.

Theorem 2. *The generating functions $F(z, v, a|p)$ and $\overline{F}(z, v|p)$ are related by*

$$F(z, v, a|p) = L(z, v, a|p)\overline{F}(z, v|p)/v. \tag{15}$$

Consequently

$$F(z, v, a|p) = L(z, v, a|p)D(z, v)/U(z, 1|p) = \frac{vaD(z, v)}{va + (1 - a)U(z, v|p)} \tag{16}$$

Proof. Any Dyck path coloured by $\{AB^{p-1}\}^*A$ may be uniquely factored into a Dyck path of length 0 (mod p) and a Dyck path with no subsequent A-visits, by cutting it at the rightmost A-visit. This proves the first equality.

Setting $a = 1$ gives $\overline{F}(z, v|p) = F(z, v, 1|p)/L(z, v, 1|p)$, and back-substitution gives the main result once we notice that $L(z, v, 1|p) = U(z, v|p)$ and $F(z, v, 1|p) = D(z, v)$. □

By setting $v = 1$ in the above, $F(z, 1, a|p)$ is obtained. This is the generating function of Dyck paths coloured by $\chi = \{AB^{p-1}\}^*A$ where A-visits are counted by a. On the other hand, if we let $a \to 1/a$ and $v = a$ instead, then $F(z, a, 1/a|p)$ counts Dyck paths coloured with the "complementary" colouring $\chi = \{BA^{p-1}\}^*B$ with a conjugate to the number of A-visits. This gives the following corollary to Theorem 2:

Corollary 3. *The generating function of a Dyck path model of adsorbing copolymers coloured by $\{AB^{p-1}\}^*A$ is given by $F(z, 1, a|p)$, and of adsorbing copolymers coloured by $\{BA^{p-1}\}^*B$ is given by $F(z, a, 1/a|p)$, where a is conjugate to the number of A-visits.*

2.3. **The Location of the Adsorption Critical Point.** Let us first consider the homopolymer case. The generating function of adsorbing Dyck paths is given by equation (2); and its radius of convergence $z_c(w)$ is given in equation (5). $z_c(w)$ is non-analytic at $w_c = 2$. This point is the intersection of a line of branch points in $D(z, w)$ along $z = 1/4$ and a line of simple poles along $z = (w - 1)/w^2$. This may also be interpreted as the location of the adsorption transition; one can show that the density of average visit vertices in walks of half-length n is $O(1)$ below this point, $O(\sqrt{n})$ at this point and and $O(n)$ above this point.

Let us now consider Dyck paths coloured by $\chi = \{AB^{p-1}\}^*A$. The generating function is given by

$$F(z, 1, a|p) = \frac{aD(z, 1)}{a + (1 - a)U(z, 1|p)} \tag{17}$$

where $U(z, v|p)$ the generating function of Dyck paths of length 0 (mod p), and can be explicitly expressed as

$$U(z, v|p) = \frac{1}{p}\sum_{j=0}^{p-1} D(\beta^j z, v), \qquad \text{where } \beta = e^{2\pi i/p}. \tag{18}$$

This expression simplifies to give

$$U(z, 1|p) = \frac{-1}{p}\sum_{j=0}^{p-1} \frac{\sqrt{1 - 4z\beta^j}}{2z\beta^j}. \tag{19}$$

One can show that the radius of convergence of $F(z, 1, a|p)$ is very similar to that of $D(z, v)$ — for small a it is given by $z = 1/4$, while for larger a it is determined by the simple pole encountered when the denominator in equation (17) vanishes. These singularities cross at the adsorption critical point, so that the critical value of a is given by the solution of

$$\frac{a_c}{a_c - 1} = U(1/4, 1|p), \qquad \text{or} \qquad a_c = \frac{U(1/4, 1|p)}{U(1/4, 1|p) - 1}. \tag{20}$$

For short periods (small values of p) it is possible to evaluate $U(1/4, 1|p)$, and hence a_c, exactly (this was done for alternating coloured paths in reference [11]). For larger values of p this is no longer the case, and instead we explore the asymptotic behaviour of a_c as a function of p.

In order to find an asymptotic form for $U(1/4, 1|p)$ and a_c, we find a uniform asymptotic estimate of the summands of $U(z, 1|p)$ and then sum them together. The starting point for this is to note that

$$D(z,1) = 1 + \sum_{n\geq 1} \binom{2n}{n} \frac{z^n}{n+1}, \qquad \text{and so} \tag{21}$$

$$U(1/4,1|p) = 1 + \sum_{n\geq 1} \binom{2np}{np} \frac{4^{-np}}{np+1}. \tag{22}$$

The uniform asymptotics of the summands of $U(1/4,1|p)$ may be found from the asymptotics of the coefficients of $D(z,1)$, and may be calculated by evaluating the contour integral $\frac{1}{2\pi i}\oint [D(z,1)/z^{n+1}]dz$, with a contour which circles the origin (see [6] for example). This gives the following lemma:

Lemma 1. *There exists $M \in [0,\infty)$ such that for sufficiently large n,*

$$\left| \frac{1}{n+1}\binom{2n}{n}\frac{\sqrt{\pi n^3}}{4^n} - \left(1 - \frac{9}{8n} + \frac{145}{128n^2}\right)\right| < \frac{M}{n^3}. \tag{23}$$

By replacing the summands in the above expression for $U(1/4,1|p)$ with their uniform asymptotic expansions given by the above lemma, we are able to obtain the asymptotic behaviour of $U(1/4,1|p)$ for large p, and this in turn gives an asymptotic expression for the location of a_c, the adsorption critical point. This expression gives an idea of how the adsorption transition is affected by a regular inhomogeneity in the polymer.

Theorem 4. *The function $U(1/4,1|p)$ is (as a function of p) asymptotic to:*

$$U(1/4,1|p) \sim 1 + \frac{1}{\sqrt{p^3\pi}}\left(\zeta(3/2) - \frac{9\zeta(5/2)}{8p} + \frac{145\zeta(7/2)}{128p^2} + O(1/p^3)\right). \tag{24}$$

Thus the adsorption critical point $a_c(p) = \frac{U(1/4,1|p)}{U(1/4,1|p)-1}$ is asymptotic to

(25)

$$a_c(p) \sim \frac{\sqrt{\pi}}{\zeta(3/2)}p^{3/2} + \frac{9\sqrt{\pi}\zeta(5/2)}{8\zeta(3/2)^2}p^{1/2} + 1 + \frac{\sqrt{\pi}\left(162\zeta(5/2)^2 - 145\zeta(7/2)\zeta(3/2)\right)}{128\zeta(3/2)^3}p^{-1/2} + O(p^{-3/2})$$

Unfortunately when we attempt to apply the above ideas to the analysis of complementary colouring, $\chi = \{BA^{p-1}\}^*B$, we encounter some problems. We are able to show that the radius of convergence has the same general form, but we are unable to find a similar asymptotic expression for the location of critical point.

We have already shown that the generating function for this model is given by $F(z,a,1/a|p)$, which is given more explicitly by:

$$F(z,a,1/a|p) = \frac{D(z,a)}{1+(1-1/a)U(z,a|p)}, \tag{26}$$

with $D(z,a)$ and $U(z,a|p)$ defined as above. The location of the critical point is given (as was the case above) by the zero of denominator when $z = 1/4$:

$$a + (a-1)U(1/4,a|p) = 0. \tag{27}$$

We have been unable to find an asymptotic expression for the solution to this non-linear equation.

Since we were unable to proceed analytically, we hypothesised a scaling form using numerical data. Using the CLN[3] library for c++, we computed $a_c(p)$ to 300 significant digits

[3]The CLN package provides, amongst many other things, arbitrary precision complex number arithmetic functions for c++. At the time of writing, it was available from `http://clisp.cons.org/~haible/packages-cln.html`

for p from 10 to 400, and then to 1000 digits for p from 1000 to 1100. Plotting this data and using the techniques described in [4, 8], we reached the following hypothesis for the asymptotic behaviour of $a_c(p)$:

$$a_c(p) \sim 2 + 1/p + c_1/p^{3/2} + c_2/p^2 + c_3/p^{5/2} + O(p^{-3}) \tag{28}$$

where

$$\begin{aligned} c_1 &= 0.41198(2) \\ c_2 &= 0.792(2) \\ c_3 &= 0.83(2) \end{aligned}$$

The estimates of the constant in the p^{-1} term rapidly approach 1 as p is increases, and it appears not to differ from 1 by more than 10^{-4}. In these circumstances, it is not unreasonable to hypothesise that it is equal to 1. That the first coefficient in the asymptotic expansion of $a_c(p)$ is so close to 1 (if not exactly equal to 1) is quite suggestive that this leading asymptotic behaviour could perhaps be solved exactly. We also note here that a similar numerical analysis on $F(z, a, 1|p)$ agrees with the results of Theorem 4.

Period	$\{AB^{p-1}\}^*A$		$\{BA^{p-1}\}^*B$	
p	actual	asymptotic	actual	asymptotic
1	2	1.965	No transition	
2	$2+\sqrt{2}$	3.399	$2+\sqrt{2}$	5.03
3	5.152712190	5.144	2.631303464	2.99
4	7.165355763	7.159	2.403090211	2.55
5	9.419630950	9.415	2.295052084	2.38
10	23.66348531	23.6618	2.124630022	2.1236
20	63.41544315	63.4148	2.057152564	2.0571
30	114.6142480	114.614	2.036915291	2.0369
40	175.1068722	175.107	2.027215547	2.0272
50	243.6370630	243.637	2.021533738	2.0215
1000	21468.92712	21468.927	2.001013847	2.00101
∞	No transition		2	2

TABLE 1. A table of the critical adsorption points for vertex-coloured Dyck paths. For $\chi = \{AB^{p-1}\}^*A$ we have computed $a_c(p)$ using equation (20), while for the complementary colouring, $\chi = \{BA^{p-1}\}^*B$, we have computed $a_c(p)$ by solving equation (27) numerically using the CLN high-precision numerics library for c++. For the sake of comparison, we have also included estimates using the asymptotic expressions in Theorem 4 and equation (28).

3. Extensions to Motzkin paths and bargraphs

A natural extension of this work is to search for similar relations in other coloured path models. We have found such relations in a number of different models based on Motzkin paths and bargraphs (see Figure 5). Motzkin paths are a generalisation of Dyck paths, in which the path is also allowed to step east. Bargraphs are partially directed self-avoiding walks that lie on or above the horizontal axis. In these two models both edges and vertices may lie on the axis, and so one may also consider edge-coloured paths as well; i.e. in which one considers the number of edge-visits rather than vertex-visits.

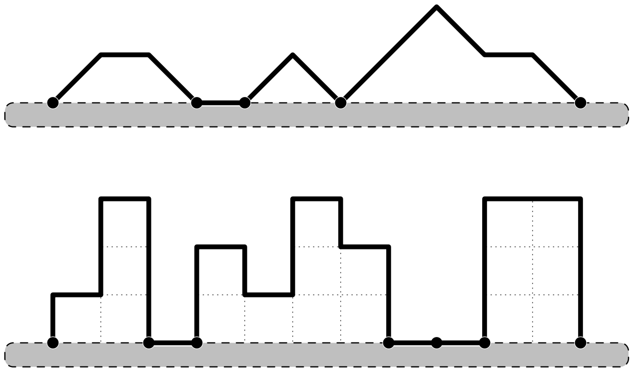


FIGURE 5. (top): a Motzkin path with 5 vertex-visits and 1 edge-visit. (bottom): a bargraph with 7 vertex-visits and 3 edge-visits.

In the following theorems we have fixed the colouring $\chi = \{AB^{p-1}\}^* A$, and all the generating functions are for paths of length 0 (mod p). We have found the following exchange relations:

Vertex-coloured Motzkin paths behave extremely similarly to vertex-coloured Dyck paths:

Theorem 5. *The generating function of vertex-coloured Motzkin paths, $M(z,v,a|p)$, satisfies the following exchange relation:*

$$\alpha M(z,v,a|p)(M(z,v,1|p)-v) = (M(z,v,a|p)-va)M(z,v,1|p). \tag{29}$$

We find that the critical point is given by

$$a_c(p) = \frac{2\sqrt{3\pi}}{9\zeta(3/2)}p^{3/2} + \frac{13\sqrt{3\pi}\zeta(5/2)}{24\zeta(3/2)^2}p^{1/2} + 1 + O(p^{-1/2}) \tag{30}$$

On the other hand, the edge-colouring problem is more complicated:

Theorem 6. *The generating function of edge-coloured Motzkin paths, $M(z,w,\alpha|p)$, satisfies the the exchange relation*

$$\begin{aligned} \alpha M(z,w,\alpha|p)H(z,w,1|p) - &\Big(M(z,w,\alpha|p) - M(z,w,0|p)\Big) \\ &= \Big(M(z,w,\alpha|p) - M(z,w,0|p)\Big)H(z,w,1|p) \end{aligned} \tag{31}$$

where α is conjugate to the number of A-edge-visits and $H(z,w,1|p)$ is the generating function of Motzkin paths that end in a horizontal step.

The asymptotic position of the adsorption critical point is given by:

$$\alpha_c(p) \sim \frac{2\sqrt{3\pi}}{3\zeta(3/2)}p^{3/2} + \frac{5\zeta(5/2)\sqrt{3\pi}}{8\zeta(3/2)^2}p^{1/2} + 1 + O(p^{-1/2}). \tag{32}$$

Notice that $\lim_{p\to\infty} \alpha_c(p)/a_c(p) = 3$. We do not, as yet, have a simple explanation of why this is so.

The edge-coloured bargraph and Motzkin path models behave very similarly and we find that the exchange relations are the same:

Theorem 7. *The generating function of edge-coloured bargraphs, $B(z, w, \alpha|p)$, satisfies the following exchange relation:*

$$\begin{aligned}\alpha B(z,w,\alpha|p)C(z,w,1|p) - &\Big(B(z,w,\alpha|p) - B(z,w,0|p)\Big) \\ &= \Big(B(z,w,\alpha|p) - B(z,w,0|p)\Big)C(z,w,1|p)\end{aligned} \tag{33}$$

where $C(z, w, 1|p)$ is the generating function of edge-coloured bargraphs that end in a horizontal step.

The asymptotic position of the critical point is given by

$$\alpha_c(p) \sim \frac{\sqrt{\varphi\pi}}{\zeta(3/2)}p^{3/2} + \frac{3\zeta(5/2)(71\sqrt{2}-100)\sqrt{\pi}}{16\sqrt{\varphi^7}\zeta(3/2)^2}p^{1/2} + 1 + O(p^{-1/2}) \tag{34}$$

where $\varphi = \sqrt{2} - 1$.

In the case of vertex-coloured bargraphs the exchange relation is complicated by the presence of two unknown generating functions — one for bargraphs that end in a horizontal step, $X(z, v, a|p)$, and one for bargraphs that end in a vertical step, $Y(z, v, a|p)$. This complication does not effect the edge-visit problem since if a bargraph ends in an edge-visit, then it must end in a horizontal step.

Theorem 8. *The generating functions $X(z, v, a|p)$ and $Y(z, v, a|p)$ satisfy the following exchange relation:*

$$\begin{aligned}\alpha\Big(X(a)\big(X(1)-v\big) + X(a)Y(1) + &Y(a)\big(X(1)-v\big)\Big) = \\ &\big(X(a)-va\big)X(1) + \big(X(a)-va\big)Y(1) + Y(a)X(1),\end{aligned} \tag{35}$$

where we have used $X(a)$ and $Y(a)$ as shorthand for $X(z, v, a|p)$ and $Y(z, v, a|p)$.

Since we have been unable to find an additional relation between these two unknown generating functions we are unable to proceed on to find an expression for the location of the critical point.

4. Conclusions

We have found a type of symmetry relation that we call an *exchange relation*. This relation allows us to solve two infinite families of vertex-coloured Dyck paths. These can be interpreted as models of fully directed copolymer adsorption. If the period of the colouring is short, then we are able to find exact expressions for the location of the adsorption transition, while for moderately long periods we are able to compute it numerically (see Table 1). For one of the two families we are also able to find an exact asymptotic expression for the critical point in terms of the period of the colouring. This expression gives an idea of the effect of a regular inhomogeneity on the adsorption transition. Unfortunately for the other family, we are only able to hypothesise an asymptotic form based on numerical estimates.

Exchange relations have also been found for infinite families of vertex-coloured and edge-coloured Motzkin paths and bargraphs. Unfortunately, it is not clear that the exchange-relation technique may be applied to more general colourings, however these models are certainly worthy of further investigation.

Acknowledgements

The authors would like to thank M. Zabrocki, M. Bousquet-Mélou, C. Chauve and S. G. Whittington for their helpful discussions. E. J. Janse van Rensburg is supported by an operating grant from NSERC (Canada).

References

[1] C. Banderier and P. Flajolet, 2001, Basic Analytic Combinatorics of Directed Lattice Paths. To appear in *Theo. Comp. Sci.*

[2] M. Bousquet-Mélou, 1996, A Method for the Enumeration of Various Classes of Column-Convex Polygons, *Discrete Math.*, **154**, 1–25.

[3] R. Brak, J.M. Essam and A.L. Owczarek, 1998, New Results for Directed Vesicles and Chains near an Attractive Wall. *J. Stat. Phys.* **93**, 155–192.

[4] A. R. Conway and A. J. Guttmann, 1996, Square Lattice Self-Avoiding Walks and Corrections to Scaling. *Phys. Rev. Lett.*, **77**, 5284–5287.

[5] K. De'Bell and T. Lookman, 1993, Surface Phase Transitions in Polymer Systems. *Rev. Mod. Phys.* **65**, 87–114.

[6] P. Flajolet and R. Sedgewick, 2001, Analytic Combinatorics: Functional Equations, Rational and Algebraic Functions. *INRIA Rapport de Recherche No. 4103*, January 2001, chapter 5.

[7] I.M. Gessel, 1986, A Probabilistic Method for Lattice Path Enumeration. *J. Stat. Plan. and Infer.* **14**, 49–58.

[8] A. J. Guttmann, I. Jensen, L. H. Wong and I. G. Enting, 2000, Punctured Polygons and Polyominoes on the Square Lattice. *J. Phys. A: Math. Gen.* **33**, 1735–1764.

[9] J.M. Hammersley, G.M. Torrie and S.G. Whittington, 1982, Self-Avoiding Walks Interacting with a Surface. *J. Phys. A: Math. Gen.* **15**, 539–571.

[10] E.J. Janse van Rensburg, 1999, Adsorbing Staircase Walks and Staircase Polygons. *Ann. Comb.* **3**, 451–473.

[11] E.J. Janse van Rensburg, 2000, *The Statistical Mechanics of Interacting Walks, Polygons, Animals and Vesicles.* Oxford Lecture Series in Mathematics and its Applications **18** (OUP Inc., New York).

[12] E.J. Janse van Rensburg and A. Rechnitzer, 2001, Exchange Relations, Dyck Paths and Copolymers. Submitted to . *Applied Discrete Math.*

[13] E.J. Janse van Rensburg and A. Rechnitzer, 2001, Exchange Symmetries in Partially and Fully Directed Models of Copolymer Adsorption. In preparation.

[14] N. Madras and G. Slade, 1992, *The Self-Avoiding Walk.* Probability and its Applications (Birkhäuser, Boston).

[15] V. Privman, G. Forgacs and H.L. Frisch, 1988, New Solvable Models of Polymer Chain Adsorption near a Surface. *Phys. Rev.* **B37**, 9897–9900.

[16] H.N.V. Temperley, 1956, Combinatorial Problems suggested by the Statistical Mechanics of Domains and of Rubber-like Molecules. *Phys. Rev.* **103**, 1–16.

[17] S.G. Whittington, 1998, A Directed Walk Model of Copolymer Adsorption. *J. Phys. A: Math. Gen.* **31**, 8797–8803.

Department of Mathematics and Statistics, York University, North York, Ontario, Canada, M3J-1P3.

E-mail address: rensburg@mathstat.yorku.ca

THE LOST NOTEBOOK AND PARTIAL THETA FUNCTIONS

S. OLE WARNAAR

In the Lost Notebook, Ramanujan wrote many amazing identities for partial theta functions, such as

$$
\begin{aligned}
1 &+ \frac{x^2(1-x)}{(1+ax^2)(1+\frac{x^2}{a})} + \frac{x^4(1-x)(1-x^3)}{(1+ax^2)(1+ax^4)(1+\frac{x^2}{a})(1+\frac{x^4}{a})} + \cdots \\
&= (1+a)(1-ax+a^2x^3-a^3x^6+\cdots) \\
&\quad - a(1-x)(1-x^3)(1-x^5)\cdots\frac{1-ax+a^3x^5-a^6x^8+\cdots}{(1+ax^2)(1+ax^4)\cdots(1+\frac{x^2}{a})(1+\frac{x^4}{a})\cdots}
\end{aligned}
$$

and

$$
\begin{aligned}
1 &+ \frac{q(1-q)}{(1+aq)(1+\frac{q}{a})} + \frac{q^2(1-q)(1-q^3)}{(1+aq)(1+aq^2)(1+\frac{q}{a})(1+\frac{q^2}{a})} \\
&= (1+a)(1-aq+a^2q^3-a^3q^6+\cdots) \\
&\quad - a\,\frac{(1-q)(1-q^3)\cdots(1-a^2q^2+a^5q^6-\cdots)}{(1+aq)(1+aq^2)\cdots(1+\frac{q}{a})(1+\frac{q^2}{a})\cdots}
\end{aligned}
$$

In this talk I will try to explain the origin of these identities, and will show that most partial-theta formulae from the Lost Notebook can be embedded in infinite hierarchies of such identities. This will reveal an unexpected connection between partial theta functions and Rogers–Ramanujan-type identities.

Department of Mathematics and Statistics, The University of Melbourne, Vic 3010, Australia

E-mail address: warnaar@ms.unimelb.edu.au

CRITICAL RESONANCE IN THE NON-INTERSECTING LATTICE PATH MODEL

RICHARD W. KENYON AND DAVID B. WILSON

ABSTRACT. We study the phase transition in the honeycomb dimer model (equivalently, monotone non-intersecting lattice path model). At the critical point the system has a strong long-range dependence; in particular, periodic boundary conditions give rise to a "resonance" phenomenon, where the partition function and other properties of the system depend sensitively on the shape of the domain.

1. INTRODUCTION

We study a model of monotone non-intersecting lattice paths in $\mathbb{Z}^2$. Applications of this model include random surfaces [2, 23, 15], magnetic flux lines in superconductors [32, 3], and a number of other physical phenomena [20, 11, 8], including spin-domain boundaries of the three-dimensional Ising model at zero temperature [4]. (See also [22, 13].)

Let $R_{m,n}$ be a domain consisting of the $m \times n$ rectangle in $\mathbb{Z}^2$ with periodic boundary conditions ($R_{m,n}$ is a graph on a torus). On $R_{m,n}$ consider configurations consisting of collections of vertex-disjoint, monotone northeast-going lattice paths (or rather loops, since the paths are required to eventually close up, possibly after winding several times around the torus: there are no "free ends"). See Figure 1. We do not restrict the number of disjoint loops but rather give a configuration an energy $E_bN_b + E_cN_c$ where N_b is the total number of "east" steps of the paths and N_c is the total number of "north" steps of the paths. The Boltzmann measure μ at temperature T is the probability measure assigning a configuration a probability proportional to $e^{-(E_bN_b+E_cN_c)/T}$. We study these Boltzmann measures near the critical temperature T, which is the temperature at which $e^{-E_b/T} + e^{-E_c/T} = 1$ [14]. Letting $b = e^{-E_b/T}$ and $c = e^{-E_c/T}$, a configuration has a probability proportional to $b^{N_b}c^{N_c}$.

This process is equivalent to another well-known model, the model of dimers (weighted perfect matchings) on the honeycomb lattice H [30, 31]. Weight the edges of the honeycomb lattice $a = 1$, b, or c according to their direction as in Figure 1. See Figure 1 for an illustration of the weight-preserving bijection between dimer configurations and lattice paths.

Let $H_{m,n}$ be a finite graph which is a quotient of H by the horizontal and vertical translations of length m, n respectively as in Figure 1. Let $\mu_{m,n} = \mu_{m,n}(a, b, c)$ be the Boltzmann measure on dimer configurations on the toroidal graph $H_{m,n}$. This measure assigns a dimer configuration a probability proportional to $a^{N_a}b^{N_b}c^{N_c}$, where there are N_a edges of weight a, N_b edges of weight b, and N_c edges of weight c. The exact probability is $a^{N_a}b^{N_b}c^{N_c}/Z$, where the normalizing constant Z is called the partition function of the system.

As $m, n \to \infty$ the measures $\mu_{m,n}(a, b, c)$ have a unique limiting Gibbs measure $\mu(a, b, c)$ [25, 6]. The measure $\mu(a, b, c)$ is a measure on dimer configurations on H, and is well-understood for all a, b, c [6]. As a, b, c vary the measure $\mu(a, b, c)$ undergoes a phase transition –the "solid-liquid" transition– when one of a, b, or c is equal to the sum of the other two, for example $a = b + c$. When $a \geq b + c$ the system is frozen: with probability 1

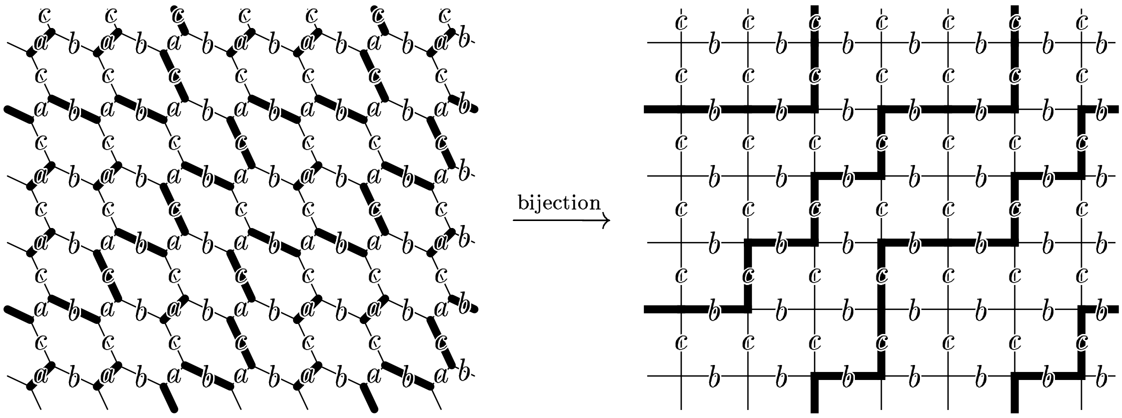


FIGURE 1. The classical bijection [30] between perfect matchings of the hexagonal lattice and non-intersecting north-east lattice paths, shown for a $m \times n = 7 \times 6$ region. Edges of type 'a' in the hexagonal lattice $H_{m,n}$ are contracted to obtain $R_{m,n}$. The dimer configuration has weight $a^{16}b^{14}c^{12}$, as does the lattice path configuration (when $a = 1$ or else one introduces a factor of a for each vertex not in a lattice path).

only edges of weight a are present. See Figure 2. When a, b, c satisfy the strict triangle inequality (each is less than the sum of the others) dimers of all types are present in a typical configuration of $\mu(a, b, c)$. Likewise in the lattice path model, when $a \geq b + c$ there are no lattice paths. When a, b, c satisfy the strict triangle inequality, the lattice paths in a configuration are dense, that is, on average they lie within a constant distance of one another.

The finite-volume measures $\mu_{m,n}(a, b, c)$ are less well understood near $a = b + c$. When $a = b+c$ the lattice paths exist but are spread out; the average distance between strands is on the order of the square root of the system size. The ratio n/m imposes fairly rigid entropic constraints on the way these loops can join up. Surprisingly these constraints become stronger with increasing system size, so that the scaling limit has nontrivial structure.

Our main result is a computation of the partition function as a function of a, b, c, m, n, for parameters near $a = b+c$. The partition function Z (normalized by $(\text{area})^{1/4}$) is largest when $nb/(mc)$ is rational with small numerator and denominator.

For example we have

Theorem 1. *When $a = 1, b = c = 1/2$ and $n/m = p/q$ in lowest terms we have*

$$\log Z = \frac{(mn)^{1/4}\zeta(3/2)(1 - 2^{-1/2})}{2\sqrt{\pi}(pq)^{3/4}}(1 + o(1)).$$

as m and n tend to ∞ while p and q remain fixed (here ζ is the Riemann zeta function). Furthermore the number N_c of 'c'-type edges tends to a Gaussian with expectation

$$\langle N_c \rangle = \frac{(mn)^{3/4}\zeta(1/2)(1 - 2^{1/2})}{2\sqrt{\pi}(pq)^{1/4}}(1 + o(1))$$

and variance

$$\sigma^2(N_c) = \frac{(mn)^{5/4}(pq)^{1/4}\zeta(-1/2)(1 - 2^{3/2})}{2\sqrt{\pi}}(1 + o(1)).$$

(Both $\zeta(1/2)$ and $\zeta(-1/2)$ are negative.)

We see that the system greatly prefers domains of *rational* modulus over those with irrational modulus. Here by rational we mean, a lattice path with average slope c/b will

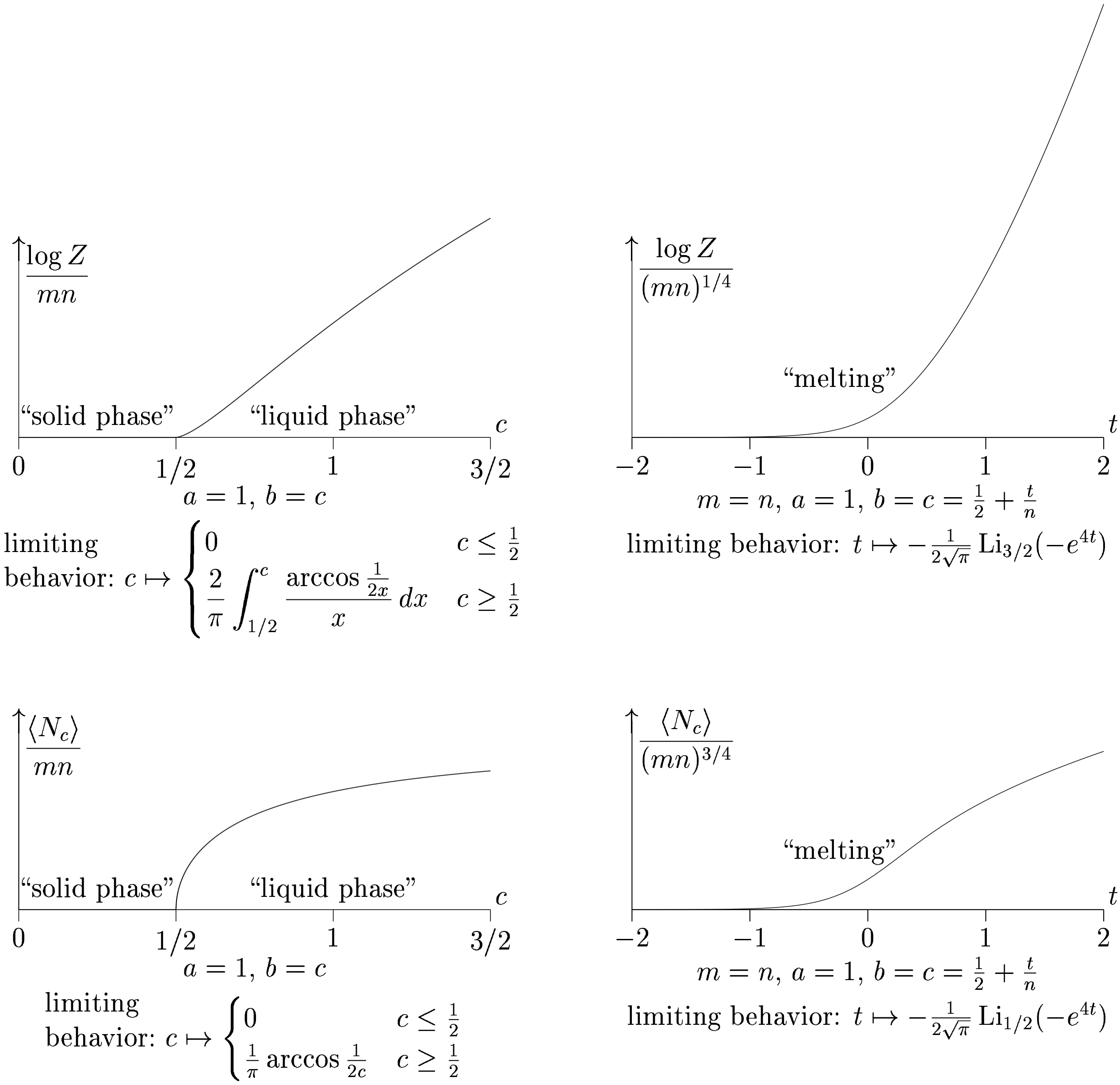


FIGURE 2. The partition function is shown in the upper panels, and the expected number of edges of type 'c' is shown in the lower panels. The solid and liquid phases are shown in the left panels, and the melting transition between between these phases is shown in the right panels. The formulas on the left are derived in [14, 29, 6] and are not used here. The formulas on the right follow from Theorem 2. The melting transition depends quite sensitively on the aspect ratio of the region (see Figure 3 and Theorem 1).

close up after winding a small number of times around the torus. In such a case the number of loops can be large, whereas in an irrational case each loop must wind many times around before closing up (unless it pays a large entropic cost).

One can think of the partition function Z, taken as a function of m and n, as an indicator of rationality of n/m. See Figure 3 which plots $\log Z/(\text{area})^{1/4}$ and $\langle N_c\rangle/(\text{area})^{3/4}$ as a function of $\log(n/m)$ near area $= 10^7$ for $a = 1$ and $b = c = \frac{1}{2}$.

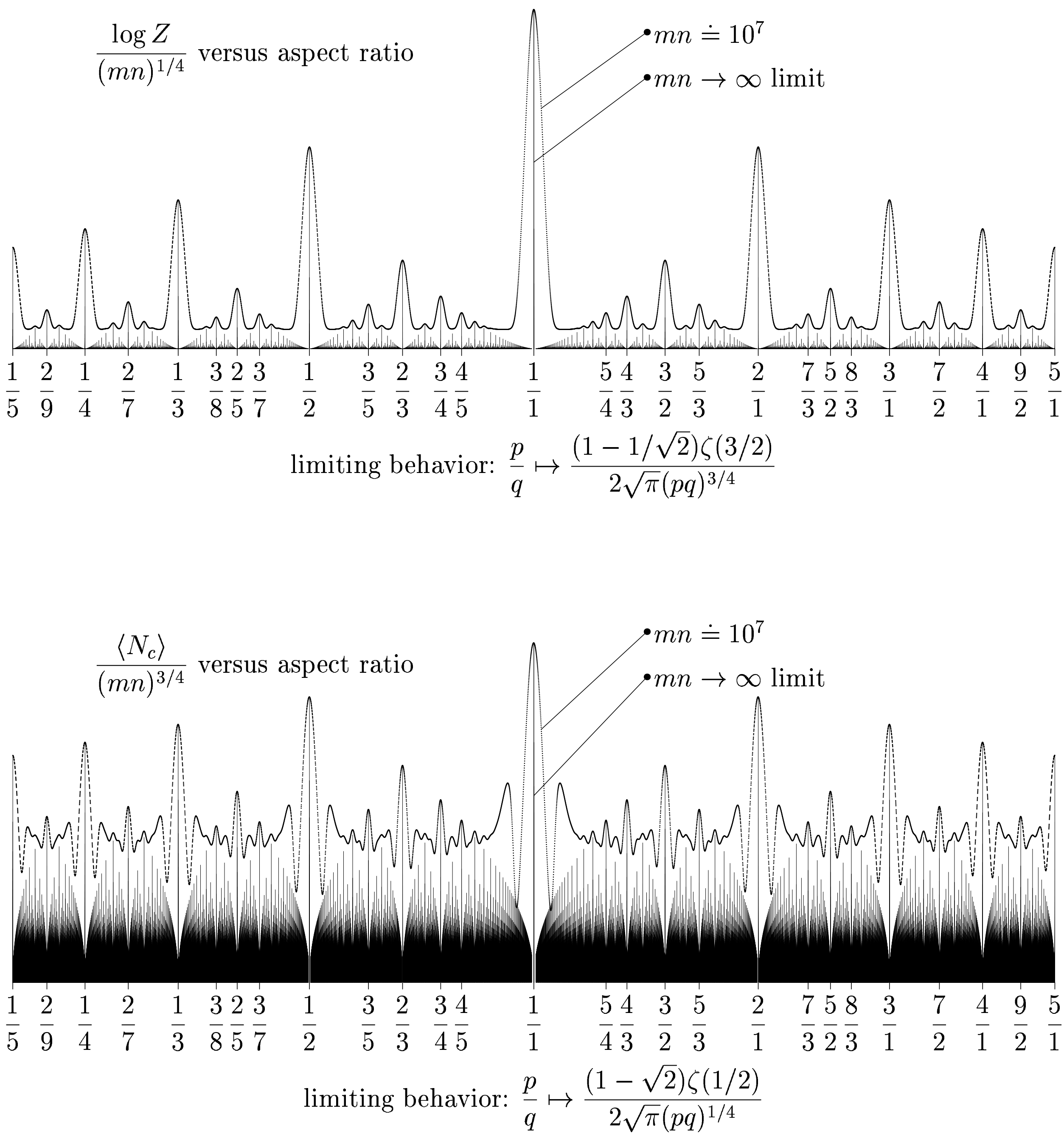


FIGURE 3. Resonant spikes in $\log Z$ (upper panel) and in the edge density (lower panel) for very large regions when $a = 1$ and $b = c = 1/2$. The aspect ratio is plotted on a log scale.

This is not the only interesting phenomenon in this model. For a rational domain, there is a non-trivial behavior as we vary a, b, c away from the critical point. Letting $A = (c/(a-b))^n$, the partition function as a function of A has an infinite number of nonanalyticities (in the large-n limit) which correspond to abrupt changes in the winding number of curves in a typical configuration. That is, the curves "ratchet" at well-defined values as A increases: see Figure 5 and § 5.4. We did not prove (although we believe) that at a typical value of A the curves are in a well-defined integer homology class (i.e., have well-defined winding number around the torus), and this homology class changes at discrete values of A. We prove only that the $\mathbb{Z}_2$ homology class (winding number modulo 2) changes at these well-defined points. We also show that, away from these transition points, the number of such curves is a Gaussian; at the transition it is a mixture of two Gaussians, coming from a Gaussian for each homology class.

It would be very interesting to study this same model on higher-genus surfaces. On higher-genus surfaces (with translation structures having conical singularities) it would be very useful to be able to detect "rationality", in the form of bands of parallel closed geodesics: this is an important problem in billiards [26]. Moreover the ratcheting phenomenon must be significantly richer in the presence of a non-abelian fundamental group.

2. Review of $\log Z$ and the distribution of N_c

The partition function Z is

$$Z = \sum_{\text{configurations}} a^{N_a} b^{N_b} c^{N_c}.$$

It contains all the information about the distribution of the total number of edges of each type. We can view Z as a polynomial in c, where the coefficient of c^{N_c} is the weighted sum of configurations which contain N_c edges of type 'c'. Thus the expected number of edges of type 'c' is $\langle N_c \rangle = \frac{c}{Z}\frac{\partial Z}{\partial c}$. Similarly, the ℓth moment of the number of edges of type 'c' is given by

$$\langle N_c^\ell \rangle = \frac{1}{Z} \overbrace{c\frac{\partial}{\partial c} \cdots c\frac{\partial}{\partial c}}^{\ell} Z.$$

The ℓth *cumulant* K_ℓ is defined by

$$K_\ell = \overbrace{c\frac{\partial}{\partial c} \cdots c\frac{\partial}{\partial c}}^{\ell} \log Z,$$

which we will in effect estimate later. Here we recall several properties of cumulants [7]. Of course $\langle N_c^1 \rangle = K_1$. The variance in the number of edges of type 'c' is K_2:

$$\begin{aligned} c\frac{\partial}{\partial c} c\frac{\partial \log Z}{\partial c} &= c\frac{\partial}{\partial c}\frac{c}{Z}\frac{\partial Z}{\partial c} \\ &= \frac{c}{Z}\frac{\partial}{\partial c} c\frac{\partial Z}{\partial c} - \frac{c}{Z^2}\frac{\partial Z}{\partial c} c\frac{\partial Z}{\partial c} \\ &= \langle N_c^2 \rangle - \langle N_c \rangle^2 = \sigma^2(N_c). \end{aligned}$$

Later we will use higher moments to show that the number of 'c'-type edges tends to a Gaussian, and to this end we express these moments in terms of the K_ℓ's. Since

$$\langle N_c^{\ell+1} \rangle = \frac{1}{Z}\frac{c\partial}{\partial c}(Z\langle N_c^\ell \rangle),$$

when $\langle N_c^\ell \rangle$ is expressed as a polynomial in the variables $K_1, \ldots, K_\ell$, we may calculate $\langle N_c^{\ell+1} \rangle$ from $\langle N_c^\ell \rangle$ by replacing each monomial $K_{i_1} K_{i_2} \cdots K_{i_k}$ of $\langle N_c^\ell \rangle$ with

$$K_1 K_{i_1} K_{i_2} \cdots K_{i_k} + K_{i_1+1} K_{i_2} \cdots K_{i_k} + K_{i_1} K_{i_2+1} \cdots K_{i_k} + \cdots + K_{i_1} K_{i_2} \cdots K_{i_k+1}.$$

Thus for example we have

$$\begin{aligned}\langle N_c^1\rangle =&K_1\\ \langle N_c^2\rangle =&K_2+K_1^2\\ \langle N_c^3\rangle =&K_3+3K_2K_1+K_1^3\\ \langle N_c^4\rangle =&K_4+4K_3K_1+3K_2^2+6K_2K_1^2+K_1^4\\ \langle N_c^5\rangle =&K_5+5K_4K_1+10K_3K_2+10K_3K_1^2+15K_2^2K_1+10K_2K_1^3+K_1^5.\end{aligned}\tag{1}$$

These polynomials $\langle N_c^j\rangle = Y_j(K_1, K_2, \dots)$ are the *complete Bell polynomials* [7]. We see that $\langle N_c^\ell\rangle$ contains a monomial for each partition of ℓ, and the coefficient associated with partition with distinct part sizes $s_1 > s_2 > \cdots > s_k$ and r_i parts of size s_i is

$$\frac{\ell!}{s_1!^{r_1}r_1!\cdots s_k!^{r_k}r_k!}.\tag{2}$$

It will be more useful to work with moments about the mean rather than moments about the origin. Note that if we replace Z with $Z^* = Zc^{-\mu}$, then the above derivation shows us how to express $\langle (N_c-\mu)^\ell\rangle$ in terms of the K_ℓ^*'s defined by

$$K_\ell^* = \overbrace{c\frac{\partial}{\partial c}\cdots c\frac{\partial}{\partial c}}^{\ell}\log(Zc^{-\mu}) = K_\ell + \overbrace{c\frac{\partial}{\partial c}\cdots c\frac{\partial}{\partial c}}^{\ell}\log(c^{-\mu}).$$

As $K_1^* = K_1 - \mu$ and $K_\ell^* = K_\ell$ for $\ell > 1$, upon substituting $\mu = \langle N_c\rangle = K_1$ we see that the ℓth moment of N_c about the mean may be obtained from the above expressions for $\langle N_c^\ell\rangle$ by deleting all monomials that contain the variable K_1.

3. Product form of the partition function

We compute an expression for the partition function as a function of a, b, and c. We are interested in approximating Z to within $1 + o(1)$ multiplicative errors when $a = 1$, $b, c \in (0, 1)$ and $b + c$ is close to a. The interesting range is when $(c/(a-b))^n$ is of constant order, that is, $a - b - c = O(1/n)$. In what follows we always set $a = 1$, although we keep using a for notational convenience.

Recall $H_{m,n}$, the $m \times n$ hexagonal toroidal graph shown in Figure 1. By Kasteleyn [14] (see also [27, 12, 24, 18, 19] for extensions and further developments), the partition function $Z = Z(a, b, c)$ for dimer coverings of $H_{m,n}$ is a sum of four expressions,

$$Z = \frac{1}{2}(-Z_{00} + Z_{01} + Z_{10} + Z_{11}),\tag{3}$$

where $Z_{\sigma\tau}$ is the determinant of a signed version of the adjacency matrix of $H_{m,n}$, and counts dimer coverings with a sign according to the homology class (in $H_1(\text{torus}, \mathbb{Z}_2) \cong \mathbb{Z}_2^2$) of the corresponding system of loops, as follows. Let $N(\varepsilon_{\hat{x}}, \varepsilon_{\hat{y}})$ denote the total weight of dimer coverings whose corresponding loops have $\varepsilon_{\hat{x}}$ mod 2 crossings of the line $x = 0$ and $\varepsilon_{\hat{y}}$ mod 2 crossings of the line $y = 0$. Each $Z_{\sigma\tau}$ is a linear combination of the $N(\varepsilon_{\hat{x}}, \varepsilon_{\hat{y}})$ with coefficients ±1 as follows:

(4)

	$N(0,0)$	$N(1,0)$	$N(0,1)$	$N(1,1)$
Z_{00}	$+1$	-1	-1	-1
Z_{10}	$+1$	$+1$	-1	$+1$
Z_{01}	$+1$	-1	$+1$	$+1$
Z_{11}	$+1$	$+1$	$+1$	$-1.$

Note three important facts, which follow from this table:

Proposition 1. *The sum of any two of $-Z_{00}, Z_{01}, Z_{10}, Z_{11}$ has only nonnegative coefficients. The difference between any two of $-Z_{00}, Z_{01}, Z_{10}, Z_{11}$ is bounded by the sum of the other two. The difference between the coefficients of $a^\alpha b^\beta c^\gamma$ in any two of $-Z_{00}, Z_{01}, Z_{10}, Z_{11}$ is bounded by the sum of the coefficients of $a^\alpha b^\beta c^\gamma$ in the other two.*

Kasteleyn [14] evaluated the determinants $Z_{\sigma\tau}$ by multiplying eigenvalues obtained through Fourier analysis, giving

$$\begin{aligned}
Z_{\sigma\tau} &= \prod_{\substack{(-z)^m=(-1)^\sigma \\ (-w)^n=(-1)^\tau}} [a+bz+cw] \\
&= \prod_{(-z)^m=(-1)^\sigma} [(a+bz)^n - (-1)^\tau c^n] \\
&= \prod_{(-z)^m=(-1)^\sigma} (a+bz)^n \prod_{(-z)^m=(-1)^\sigma} \left[1-(-1)^\tau \left(\frac{c}{a+bz}\right)^n\right] \\
&= (a^m - (-1)^\sigma b^m)^n \prod_{(-z)^m=(-1)^\sigma} \left[1-(-1)^\tau \left(\frac{c}{a+bz}\right)^n\right]. \qquad (5)
\end{aligned}$$

We can ignore the b^m term which is exponentially smaller than a^m. When $b+c$ is close to a, unless z is close to -1, $|a+bz|$ will be greater than c; in particular $|a+bz|^n$ is exponentially larger than c^n. So we can ignore the factors in the products for which z is not close to -1. We will expand the remaining factors near $z=-1$. As z is a root of unity, let $z = z_k = -e^{i\theta_k}$, where $\theta_k = 2\pi k/m$ for $k \in \mathbb{Z}_m + \sigma/2$. Of course $k \equiv k+m$, so when doing series expansions we can take $-m/2 < k \le m/2$.

Define r_k, ϕ_k by

$$1 \pm \left(\frac{c}{a+bz_k}\right)^n = 1 \pm \left(\frac{c}{a-b}\right)^n r_k e^{i\phi_k},$$

so that

$$r_k = (a-b)^n |a+bz_k|^{-n} \quad \text{and} \quad \phi_k = \arg(a+bz_k)^{-n}.$$

We make the simplifying assumptions $b/a = \Theta(1)$, $1-b/a = \Theta(1)$, and $n = \Theta(m)$. Then

$$r_k = \exp\left(-n\theta_k^2 \frac{ab}{2(a-b)^2} + O(n\theta_k^4)\right) = \exp(-\epsilon k^2 + O(k^4/m^3))$$

where we have defined

$$\epsilon = \frac{2\pi^2 nab}{m^2(a-b)^2} = O(1/m),$$

and

$$\phi_k = \frac{n\theta_k b}{a-b} + nO(\theta_k^3) = \phi k + O(k^3/m^2)$$

where we define

$$\phi = \frac{2\pi nb}{m(a-b)}.$$

Letting $A = \left(\frac{c}{a-b}\right)^n$, we have

$$Z_{\sigma\tau} = (a^m - (-1)^\sigma b^m)^n \prod_{k \in \mathbb{Z}_m + \sigma/2} (1 - (-1)^\tau A r_k e^{i\phi_k}). \qquad (6)$$

In logarithmic form we can write

$$\log Z_{\sigma\tau} = -\sum_{k\in\mathbb{Z}_m+\sigma/2} \mathrm{Li}_1\left((-1)^\tau A r_k e^{i\phi_k}\right) + \underbrace{n\log(a^m-(-1)^\sigma b^m)}_{\text{negligible}} \tag{7}$$

where the polylogarithm function Li_ν is defined by $\mathrm{Li}_\nu(z) = \sum_{n=1}^\infty \frac{z^n}{n^\nu}$ for $|z|<1$ and by analytic continuation elsewhere (see Appendix A for background on polylogarithms). Here the second term is essentially zero, and the terms in the summation become negligible when $|k|$ is larger than $\Theta(\sqrt{m})$.

In this expression for $\log Z_{\sigma\tau}$ we need only keep track of its real part: from (6) we see that (since $b<a$) $Z_{\sigma\tau}$ is real and nonnegative if $A\le 1$, and real and nonnegative if $A>1$ except for Z_{00} which is strictly negative. In particular when $A>1$ the expression (3) is a sum of nonnegative terms. When $A\le 1$ we shall see that Z_{00} is negligible compared to Z so its sign is irrelevant.

More generally we find, using (7), $z\frac{d}{dz}\mathrm{Li}_\nu(z) = \mathrm{Li}_{\nu-1}(z)$, and $(c\partial/\partial c)A = nA$, that

$$\overbrace{c\frac{\partial}{\partial c}\cdots c\frac{\partial}{\partial c}}^{\ell}\log(Z_{\sigma\tau}) = -n^\ell \sum_{k\in\mathbb{Z}_m+\sigma/2} \mathrm{Li}_{1-\ell}\left((-1)^\tau A r_k e^{i\phi_k}\right) \\ (+\underbrace{n\log(a^m-(-1)^\sigma b^m)}_{\text{negligible}} \text{ if } \ell=0). \tag{8}$$

4. Rational tori

The expressions (8) are non-trivial to evaluate, mostly because they are describing behavior which depends sensitively on the parameters defining the system. In this section we compute the asymptotics of (8) for "nearly rational" domains.

We consider the toroidal hexagonal graph $H_{m,n}$ to be "nearly rational" when $\phi/(2\pi) = \frac{nb}{m(a-b)}$ is close to a simple rational p/q, where p and q are relatively prime integers. (Note that this depends not only on m,n but also on a,b,c.) We keep p and q fixed as m and n tend to infinity, and by "close" we mean that $nb/(mc)-p/q$ is not too large compared to $1/\sqrt{qn}$. We will determine the asymptotic shape of the resonant peaks in Figure 3.

4.1. Spokes, spirals, and clouds.

For fixed p and q, as the area mn gets large, the terms $1\pm Ar_ke^{i\phi_k} \approx 1\pm Ar_k e^{2\pi i(p/q)k}$ accumulate on q different spokes (or radii) of a circle with radius A centered at 1. If $\phi/(2\pi)$ is only approximately p/q, then each of these spokes becomes a spiral, which spirals out from 1 when k is negative and increasing, and then spirals back in towards 1 when k is positive and increasing. If $\phi/(2\pi)$ is far from a simple rational, then the terms $1\pm Ar_ke^{i\phi_k}$ form a cloud within the disk of radius A centered at 1. In our analysis for nearly rational tori, we will assume that $\phi/(2\pi)$ is sufficiently close to a simple enough rational p/q that the terms $1\pm Ar_ke^{i\phi_k}$ form what appear to be q continuous spokes or spirals (in a sense we define more precisely below). See Figure 4.

It is useful to re-express (8) to reflect the presence of the q spirals. Since $\phi\approx 2\pi\frac{p}{q}$, every qth term lies on a given spiral, so we break the sum apart into q different sums, one for each spiral. Of course q may not evenly divide m. The most convenient way to re-express the summation is

$$\sum_{k\in\mathbb{Z}_m+\sigma/2} f(k) = \frac{1}{q}\sum_{j\in\mathbb{Z}_q+\sigma/2}\ \sum_{u\in\mathbb{Z}_m} f(j+qu).$$

Since spirals are continuous objects rather than discrete sets of points, we wish to approximate $\sum_{u\in\mathbb{Z}_m}$ with $\int_0^m du$. To this end we subtract $2\pi pu$ from ϕ_{j+qu}; then the angle

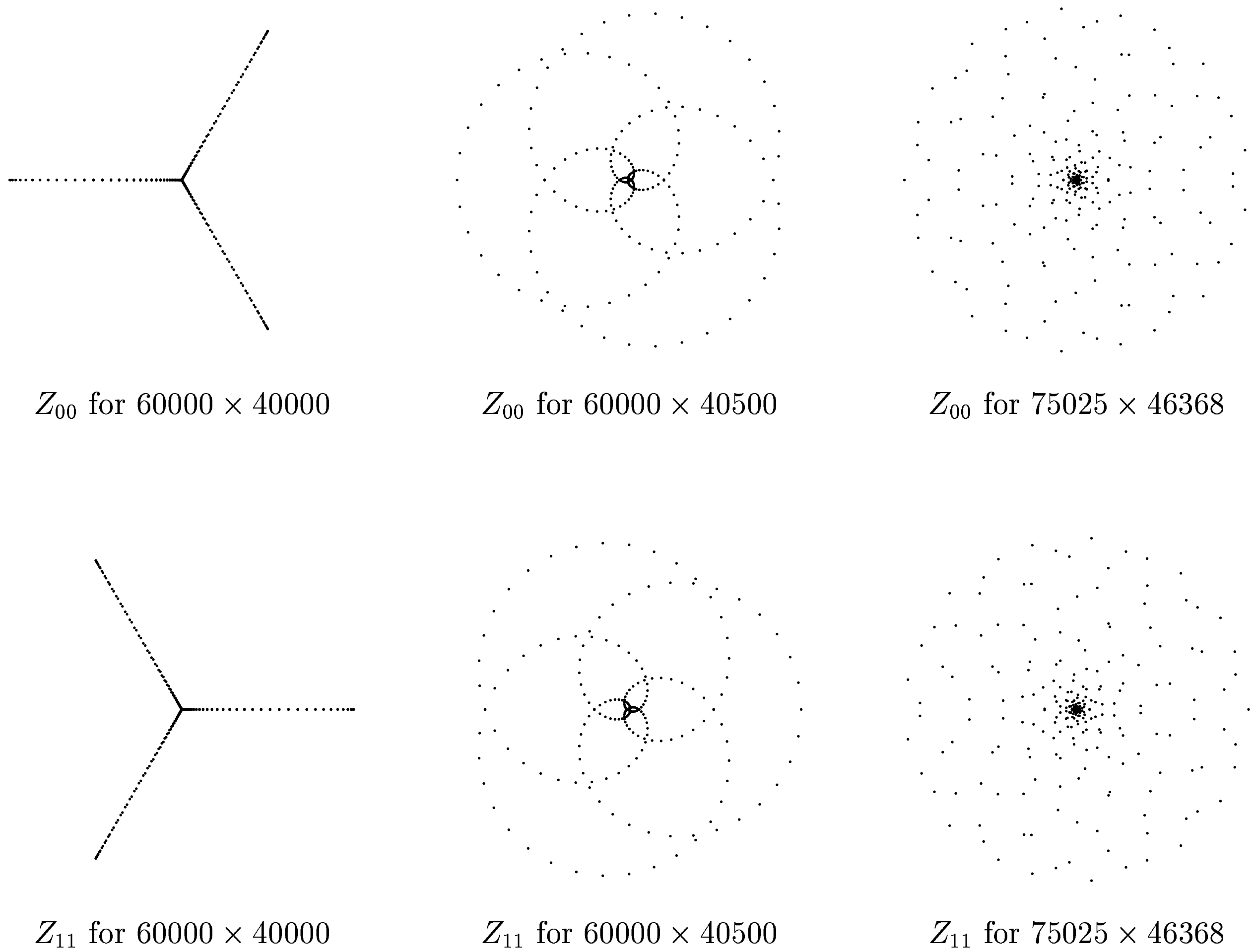


FIGURE 4. The multiplicands for Z_{00} and Z_{11} for a domain whose aspect ratio is (a) simple rational (b) nearly simple rational (c) far from simple rational. The multiplicands are complex and accumulate towards 1.

(mod 2π) is unchanged for integer u, while for continuous u it is slowly varying so we can hope that the integral approximates the sum. Thus we re-express (8) as (ignoring the negligible $O(nb^m)$ error term when $\ell = 0$)

$$\overbrace{c\frac{\partial}{\partial c}\cdots c\frac{\partial}{\partial c}}^{\ell}\log(Z_{\sigma\tau}) \doteq -n^{\ell}\frac{1}{q}\sum_{j\in\mathbb{Z}_q+\sigma/2}\ \sum_{u\in\mathbb{Z}_m}\mathrm{Li}_{1-\ell}\left((-1)^{\tau}Ar_{j+qu}e^{i(\phi_{j+qu}-2\pi pu)}\right)$$

Later we will quantify the error introduced by approximating these sums with integrals, and show it to be insignificant so long as the points appear to line up on q spirals which miss the singularity. For now we proceed with the integral approximation and simplify it:

$$\approx -n^{\ell}\frac{1}{q}\sum_{j\in\mathbb{Z}_q+\sigma/2}\int_0^m \mathrm{Li}_{1-\ell}\left((-1)^{\tau}Ar_{j+qu}e^{i(\phi_{j+qu}-2\pi pu)}\right)du \tag{9}$$

changing variables to $k = j + qu$ and using the fact that the integrand is periodic,

$$= -n^\ell \frac{1}{q^2} \int_0^{qm} \sum_{j \in \mathbb{Z}_q + \sigma/2} \mathrm{Li}_{1-\ell}\left((-1)^\tau A r_k e^{i(\phi_k - 2\pi k p/q)} e^{2\pi i j p/q)}\right) dk$$

using the replication formula $\frac{1}{q}\sum_{\omega^q=1} \mathrm{Li}_\nu(\omega z) = \frac{1}{q^\nu}\mathrm{Li}_\nu(z^q)$ and the fact that $\gcd(p,q) = 1$,

$$= -(qn)^\ell \frac{1}{q^2} \int_0^{qm} \mathrm{Li}_{1-\ell}\left((-1)^{\tau q} A^q r_k^q e^{i(\phi_k - 2\pi k p/q)q} e^{2\pi i (\sigma/2) p}\right) dk$$

and again using the periodicity of ϕ_k and r_k,

$$(10) \qquad = -\frac{(qn)^\ell}{q} \int_{-m/2}^{m/2} \mathrm{Li}_{1-\ell}\left((-1)^{\tau q + \sigma p} A^q r_k^q e^{i(\phi_k - 2\pi k p/q)q}\right) dk,$$

which, as a function of σ and τ, only depends upon the parity of $\tau q + \sigma p$. To obtain the next formula we substitute the estimate $r_k e^{i\phi_k} \approx e^{-\epsilon(k \bmod m)^2} e^{i\phi(k \bmod m)}$ (where we take $k \bmod m$ to lie between $-m/2$ and $m/2$). Doing this substitution introduces an error, but we postpone the error analysis until later. Note that we can substitute this approximation for $r_k e^{i\phi_k}$ either in (10) or just prior to the integral approximation (9), since the only property of $r_k e^{i\phi_k}$ that we used in the intervening steps is that it is periodic in k with period m.

$$(11) \qquad \approx -\frac{(qn)^\ell}{q} \int_{-m/2}^{m/2} \mathrm{Li}_{1-\ell}\left((-1)^{\tau q + \sigma p} A^q e^{-q\epsilon k^2} e^{i(\phi - 2\pi p/q)qk}\right) dk.$$

We extend the range of integration to all of $\mathbb{R}$, introducing a negligible error,

$$(12) \qquad \approx -\frac{(qn)^\ell}{q} \int_{-\infty}^{\infty} \mathrm{Li}_{1-\ell}\left((-1)^{\tau q + \sigma p} A^q e^{-q\epsilon k^2} e^{i(\phi - 2\pi p/q)qk}\right) dk.$$

To measure the closeness of $\phi/(2\pi)$ to p/q, we define α so that $\phi/(2\pi) = p/q(1 + \alpha W)$, where we define $W = \sqrt{q\epsilon}/(\pi p)$. Then $(\phi - 2\pi p/q)qk = 2\pi p \alpha W k = 2\alpha\sqrt{q\epsilon}k$. We change variables to $x = \sqrt{q\epsilon}k$ to obtain

$$(13) \qquad = -\frac{(qn)^\ell}{\sqrt{q^3\epsilon}} \int_{-\infty}^{\infty} \mathrm{Li}_{1-\ell}\left((-1)^{\tau q + \sigma p} A^q e^{2i\alpha x - x^2}\right) dx.$$

The reader may wonder about the apparent asymmetry in equation (13) (when $\ell = 0$), e.g. why does q appear but not p, while Z is symmetrical with respect to width and height? But when $b + c \approx a \approx 1$ and $(nb)/(mc) \approx p/q$ we have

$$W = \frac{\sqrt{q\epsilon}}{\pi p} \approx \sqrt{\frac{2}{pmc}} \approx \sqrt{\frac{2}{qnb}} \approx \frac{\sqrt{2}}{(pqmnbc)^{1/4}},$$

$$(14) \qquad \frac{1}{\sqrt{q^3\epsilon}} = \frac{1}{\pi p q W} \approx \frac{(mnbc)^{1/4}}{\pi\sqrt{2}(pq)^{3/4}},$$

$$\log A^q = nq \log \frac{c}{a-b} \approx (b + c - a)\frac{nq}{c} \approx (b + c - a)\frac{mp}{b},$$

$$(qn)^\ell \approx (pqmnc/b)^{\ell/2}.$$

The asymmetry between b and c when $\ell > 0$ should not be unexpected since we differentiated with respect to c rather than b.

Referring back to the spirals in Figure 4, q counts the number of spirals, the parameter A measures the radius of the spirals, $\sqrt{\epsilon}$ is a measure of how far apart the points are on the spiral, and α is a measure of the "spirality". α is the right parameter against which to plot the shape of the spikes, making W a measure if their width. When $\alpha = 0$ the spirals are spokes, when α gets too large (for a given $\sqrt{\epsilon}$) the spirals break up into a cloud, by which time the integral approximation (9) breaks down.

4.2. **Error analysis.** In the interest of simplicity, we only consider the case when p, q, A, and α are held fixed while $m \to \infty$ and $n \to \infty$. Most of the interesting behavior already shows up in this case. It is also quite interesting to ask how much these parameters can vary (e.g. can p and q be as large as $n^{1/3}$?), but we do not pursue that in this article.

Lemma 1. *When we fix p, q, A, and α while $m \to \infty$ and $n \to \infty$, we have*

$$\frac{\sqrt{q^3\epsilon}}{(qn)^\ell}\overbrace{c\frac{\partial}{\partial c}\cdots c\frac{\partial}{\partial c}}^{\ell}\log(Z_{\sigma\tau}) = -\int_{-\infty}^{\infty}\operatorname{Li}_{1-\ell}\left((-1)^{\tau q+\sigma p}A^q e^{2i\alpha x - x^2}\right)dx + o(1),$$

provided that the curve $(-1)^{\tau q+\sigma p}A^q e^{2i\alpha x-x^2}$ does not contain the point 1. If the curve does contain 1, then the convergence for $\ell = 0$ is still valid provided that no multiplicand of $Z_{\sigma\tau}$ is closer than $e^{-o(\sqrt{n})}$ to 0, in which case the right-hand side is merely an upper bound.

These integrals are explicitly evaluated in § 5.1.

Proof. Much of the proof has already been given in § 4.1, what we have left to do is justify the approximations that we made in (9) and (11). For this error analysis we do the $r_k e^{i\phi_k} \approx e^{-\epsilon k^2 + i\phi k}$ substitution before the integral approximation. With $k = j + qu$ and $j \in \mathbb{Z}_q + \sigma/2$,

$$\begin{aligned}\operatorname{Li}_{1-\ell}\left((-1)^\tau A r_{j+qu} e^{i(\phi_{j+qu}-2\pi p u)}\right) &= \operatorname{Li}_{1-\ell}\left((-1)^\tau e^{2\pi i j p/q} A e^{i(\phi-2\pi p/q)k - \epsilon k^2 + O(k^3/m^2)}\right)\\ &= \operatorname{Li}_{1-\ell}\left((-1)^\tau e^{2\pi i j p/q} A e^{2i\alpha\sqrt{\epsilon/q}\,k - \epsilon k^2 + O(k^3/m^2)}\right)\end{aligned}$$

using $\phi - 2\pi p/q = 2\pi(p/q)\alpha W = 2\alpha\sqrt{\epsilon/q}$. Next we use the fact that for integer $\ell \geq 0$, $\frac{d}{dz}\operatorname{Li}_{1-\ell}(e^z) = \operatorname{Li}_{-\ell}(e^z)$ is a rational function of e^z with a pole at $e^z = 1$ but which is bounded outside a neighborhood of this pole,

$$= \operatorname{Li}_{1-\ell}\left((-1)^\tau e^{2\pi i j p/q} A e^{2i\alpha\sqrt{\epsilon/q}\,k - \epsilon k^2}\right) + O(k^3/m^2)$$

which is valid as long as $(-1)^\tau e^{2\pi i j p/q} A e^{2i\alpha\sqrt{\epsilon/q}\,k-\epsilon k^2}$ lies outside a neighborhood of 1, and the $O(k^3/m^2)$ error term is much smaller than the radius of this neighborhood.

It is not hard to see that the q curves $(-1)^\tau e^{2\pi i j p/q} A e^{2i\alpha\sqrt{\epsilon/q}\,k-\epsilon k^2}$ are bounded away from 1 if and only if the curve $(-1)^{\tau q+\sigma p}A^q e^{2i\alpha x - x^2}$ avoids 1.

Adding up the errors $O(k^3/m^2)$ over the range $|k| < \Theta(\sqrt{n})$ gives $O(1)$. When $|k| \gg \Theta(\sqrt{n})$, both r_k and its approximation $e^{-\epsilon k^2}$ are exponentially decreasing in $|k|$, so using $\operatorname{Li}_{1-\ell}(z) \approx z$ for small z, and $r_k = e^{-\epsilon k^2}(1 + O(k^3/m^2))$ when $k \leq n^{2/3}$, we see that doing the substitution for $n^{1/2} \leq k \leq n^{2/3}$ also introduces $O(1)$ error, and that the substitution for $k \geq n^{2/3}$ gives $o(1)$ error. Upon multiplying by $\sqrt{q^3\epsilon}/(qn)^\ell$, all these errors become $o(1)$.

For the integral approximation (9), the integrands in (9) are continuous (except at the branch cut) as long as the curves $(-1)^\tau e^{2\pi i j p/q} A e^{2i\alpha\sqrt{\epsilon/q}\,k-\epsilon k^2}$ avoid the point 1. Moreover they converge exponentially fast to 0 when $|k| \to \infty$. Therefore they are Riemann summable and the error in converting the sums to integrals tends to zero. The error introduced by extending the range of integration to the reals is exponentially small.

What happens when a curve passes through the singularity? When $\ell = 0$, the integral in expression (9) for $\log Z_{\sigma\tau}$ converges and is an upper bound for $\log Z_{\sigma\tau}$: the Riemann sum for $\log Z_{\sigma\tau}$ converges to its integral on the complement of a small neighborhood of the singularity, and the Riemann sum near the logarithmic singularity has a negligible contribution except possibly for the point which is closest to the singularity. If the distance of the closest point to the singularity is no smaller than $e^{-o(\sqrt{n})}$, then the contribution of this point is $o(\sqrt{n})$ and so can be ignored. $\square$

4.3. **The distribution of the number of edges of type 'c'.** By Lemma 1, to first order $\log Z_{\sigma\tau}$ only depends on the parity of $\tau q + \sigma p$, so we define

$$Z_- = \frac{1}{2} \sum_{\substack{\sigma,\tau \\ \tau q + \sigma p \text{ odd}}} Z_{\sigma\tau} \qquad \text{and} \qquad Z_+ = \frac{1}{2} \sum_{\substack{\sigma,\tau \\ \tau q + \sigma p \text{ even}}} \varepsilon_{\sigma\tau} Z_{\sigma\tau}$$

where $\varepsilon_{\sigma\tau} = -1$ if $\sigma = 0$ and $\tau = 0$, and $\varepsilon_{\sigma\tau} = 1$ otherwise. We have $Z = Z_- + Z_+$, and from Proposition 1, both Z_- and Z_+ have only nonnegative coefficients, so they can be interpreted as distributions. Lemma 1 shows that typically one of Z_- or Z_+ is exponentially larger than the other one, so the distribution of the number of c edges is governed by whichever of Z_- or Z_+ is dominant.

We use the method of moments to determine the distribution of the number N_c of type-'c' edges. We saw in § 2 how to express the ℓth moment N_c about its mean (call it C_ℓ) in terms of $K_\ell = \overbrace{c\frac{\partial}{\partial c} \cdots c\frac{\partial}{\partial c}}^{\ell} \log Z$, but in Lemma 1 we evaluated $K_{\ell,\sigma\tau} = \overbrace{c\frac{\partial}{\partial c} \cdots c\frac{\partial}{\partial c}}^{\ell} \log Z_{\sigma\tau}$. Define $C_{\ell,\sigma\tau}$ to be the same expression, except with the $K_{\ell,\sigma\tau}$'s replacing the K_ℓ's (see also (15) below), and similarly define the $K_{\ell,\pm}$'s and the $C_{\ell,\pm}$'s.

Lemma 2. *Under the assumptions of Lemma 1, if the curve $(-1)^{\tau q+\sigma p} A^q e^{2i\alpha x - x^2}$ does not contain the point 1 and $\int_{-\infty}^{\infty} \mathrm{Li}_{-1}\left((-1)^{\tau q+\sigma p} A^q e^{2i\alpha x - x^2}\right) dx \neq 0$, then $C_{\ell,\sigma\tau}/C_{2,\sigma\tau}^{\ell/2} \to (\ell-1)!!$. (Here as usual $\ell!! = \ell(\ell-2)\cdots(3)(1)$ when ℓ is odd and $\ell!! = 0$ when ℓ is even.)*

Proof. From Lemma 1 we have $K_{j,\sigma\tau} = O((qn)^j/\sqrt{q^3\epsilon})$, and since the above integral is nonzero, $C_{2,\sigma\tau} = K_{2,\sigma\tau} = \Theta((qn)^2/\sqrt{q^3\epsilon})$. Thus each monomial in the polynomial (1) for $C_{\ell,\sigma\tau}$ has magnitude $O((qn)^\ell/(\sqrt{q^3\epsilon})^{\text{degree}})$. Recall that $q^3\epsilon \ll 1$. The monomial degree is uniquely maximized by the $K_{2,\sigma\tau}^{\ell/2}$ term (ℓ even) or the $K_{3,\sigma\tau}K_{2,\sigma\tau}^{(\ell-3)/2}$ term (ℓ odd). Thus when ℓ is odd, $C_{\ell,\sigma\tau}/C_{2,\sigma\tau}^{\ell/2} = O((q^3\epsilon)^{1/4}) \to 0$, and when ℓ is even, $C_{\ell,\sigma\tau}/C_{2,\sigma\tau}^{\ell/2}$ tends to the coefficient of the monomial $K_{2,\sigma\tau}^{\ell/2}$ in $C_{\ell,\sigma\tau}$, which by (2) is $(\ell-1)(\ell-3)\cdots(3)(1)$. $\square$

Thus the $C_{\ell,\sigma\tau}$'s converge to the moments of a Gaussian, but recall that we cannot view $Z_{\sigma\tau}$ as a distribution since it may have some negative coefficients. Next we show that the moments $C_{\ell,\pm}$ of $Z_\pm$ (which are genuine distributions) are close to the corresponding $C_{\ell,\sigma\tau}$'s.

Lemma 3. *Suppose $\ell \in \mathbb{N}$, $\varsigma = (-1)^{\tau_1 q+\sigma_1 p} = (-1)^{\tau_2 q+\sigma_2 p}$, $(\sigma_1,\tau_1) \neq (\sigma_2,\tau_2)$, and either $\varsigma = -1$, or else $\varsigma = +1$ but $A \geq 1$. Under the assumptions of Lemmas 1 and 2, if $|K_{1,\sigma_1\tau_1} - K_{1,\sigma_2\tau_2}| \ll n^{5/4}$ then $C_{\ell,\varsigma}/C_{2,\sigma_1\tau_1}^{\ell/2} = (\ell-1)!! + o(1)$.*

Proof. Viewing $Z_{\sigma\tau}$ as a polynomial in c, $Z_{\sigma\tau} = \sum_\imath \gamma_{\imath,\sigma\tau}\, c^\imath$, we have

$$C_{\ell,\sigma\tau} Z_{\sigma\tau} = \sum_\imath \gamma_{\imath,\sigma\tau}\, (\imath - K_{1,\sigma\tau})^\ell c^\imath. \tag{15}$$

As Z_ς is the average of $\varepsilon_{\sigma_1\tau_1} Z_{\sigma_1\tau_1}$ and $\varepsilon_{\sigma_2\tau_2} Z_{\sigma_2\tau_2}$, we have

$$\begin{aligned} C_{\ell,\varsigma} Z_\varsigma &= \frac{1}{2} \sum_{\sigma,\tau} \varepsilon_{\sigma\tau} \sum_\imath \gamma_{\imath,\sigma\tau}\, c^\imath (\imath - K_{1,\varsigma})^\ell \\ &= \frac{1}{2} \sum_{\sigma,\tau} \varepsilon_{\sigma\tau} \sum_\imath \gamma_{\imath,\sigma\tau}\, c^\imath \sum_{j=0}^{\ell} \binom{\ell}{j} (\imath - K_{1,\sigma\tau})^j (K_{1,\sigma\tau} - K_{1,\varsigma})^{\ell-j} \\ &= \frac{1}{2} \sum_{\sigma,\tau} \varepsilon_{\sigma\tau} \sum_{j=0}^{\ell} \binom{\ell}{j} C_{j,\sigma\tau} Z_{\sigma\tau} (K_{1,\sigma\tau} - K_{1,\varsigma})^{\ell-j}. \end{aligned}$$

Since $\varsigma = -1$ or else $\varsigma = +1$ but $A \geq 1$, $\varepsilon_{\sigma_1\tau_1} Z_{\sigma_1\tau_1}$ and $\varepsilon_{\sigma_2\tau_2} Z_{\sigma_2\tau_2}$ have the same sign, so $K_{1,\varsigma}$ is a convex combination of $K_{1,\sigma_1\tau_1}$ and $K_{1,\sigma_2\tau_2}$, and it too can differ by at most $\ll n^{5/4} = O(C_{2,\sigma\tau}^{1/2})$ from them. Substituting $C_{j,\sigma\tau} = ((j-1)!! + o(1)) C_{2,\sigma\tau}^{j/2}$ we get

$$C_{\ell,\varsigma} Z_\varsigma = \frac{1}{2} \sum_{\sigma,\tau} \varepsilon_{\sigma\tau} ((\ell-1)!! + o(1)) C_{2,\sigma\tau}^{\ell/2} Z_{\sigma\tau},$$

so $C_{\ell,\varsigma}$ is a convex combination of $((\ell-1)!! + o(1)) C_{2,\sigma_1\tau_1}^{\ell/2}$ and $((\ell-1)!! + o(1)) C_{2,\sigma_2\tau_2}^{\ell/2}$. $\square$

We have not computed $K_{1,\sigma\tau}$ to the precision that Lemma 3 would appear to suggest that we need, but all we really need is that the two relevant $K_{1,\sigma\tau}$'s are quite close.

Lemma 4. *Suppose $\varsigma = (-1)^{\tau_1 q + \sigma_1 p} = (-1)^{\tau_2 q + \sigma_2 p}$. Under the assumptions of Lemma 1, if*

$$-\int_{-\infty}^{\infty} \mathrm{Li}_1(-\varsigma A^q e^{2i\alpha x - x^2})\, dx < -\int_{-\infty}^{\infty} \mathrm{Li}_1(\varsigma A^q e^{2i\alpha x - x^2})\, dx,$$

then $|K_{1,\sigma_1\tau_1} - K_{1,\sigma_2\tau_2}| \leq \exp(-\Theta(\sqrt{n}))$.

Proof. For expository convenience say that the two $\sigma\tau$'s for which $(-1)^{\tau q+\sigma p} = \varsigma$ are 01 and 10. From Lemma 1, Z_{01} and Z_{10} dominate $-Z_{00}$ and Z_{11} by a factor of $\exp(\Theta(\sqrt{n}))$, and then from Proposition 1, $|Z_{01}/Z_{10} - 1| \leq \exp(-\Theta(\sqrt{n}))$. Writing $Z_{\sigma\tau}$ as a polynomial in c, $Z_{\sigma\tau} = \sum_\imath \gamma_{\imath,\sigma\tau}\, c^\imath$, again from Proposition 1 we have

$$\begin{aligned}
|\gamma_{\imath,01} - \gamma_{\imath,10}| &\leq -\gamma_{\imath,00} + \gamma_{\imath,11} \\
|\gamma_{\imath,01}\, \imath c^\imath - \gamma_{\imath,10}\, \imath c^\imath| &\leq \imath c^\imath(-\gamma_{\imath,00} + \gamma_{\imath,11}) \leq mn(-\gamma_{\imath,00}\, c^\imath + \gamma_{\imath,11}\, c^\imath) \\
\left|\sum_\imath \gamma_{\imath,01}\, \imath c^\imath - \sum_\imath \gamma_{\imath,10}\, \imath c^\imath\right| &\leq mn\left[-\sum_\imath \gamma_{\imath,00}\, c^\imath + \sum_\imath \gamma_{\imath,11}\, c^\imath\right] \\
|Z_{01}K_{1,01} - Z_{10}K_{1,10}| &\leq mn(-Z_{00} + Z_{11}).
\end{aligned}$$

As Z_{01} and Z_{10} are exponentially close to each other and exponentially dominate $-Z_{00}$ and Z_{11}, it must be that $K_{1,01}$ and $K_{1,10}$ are exponentially close. □

Theorem 2. *When we fix p, q, A, and α while $m \to \infty$ and $n \to \infty$, we have*

$$\log Z = \frac{(mnbc)^{1/4}}{\pi\sqrt{2}(pq)^{3/4}}\left[\max_\pm - \int_{-\infty}^{\infty} \operatorname{Li}_1\left(\pm A^q e^{2i\alpha x - x^2}\right) dx + o(1)\right], \tag{16}$$

and with the exceptions noted below, the number N_c of edges of type 'c' converges in distribution to a Gaussian, with mean

$$\langle N_c\rangle = \frac{(mnc)^{3/4}}{\pi\sqrt{2}(pqb)^{1/4}}\left[-\int_{-\infty}^{\infty} \operatorname{Li}_0\left(\pm A^q e^{2i\alpha x - x^2}\right) dx + o(1)\right], \tag{17}$$

and variance

$$\sigma^2(N_c) = \frac{(mnc)^{5/4}(pq)^{1/4}}{\pi\sqrt{2}b^{3/4}}\left[-\int_{-\infty}^{\infty} \operatorname{Li}_{-1}\left(\pm A^q e^{2i\alpha x - x^2}\right) dx + o(1)\right], \tag{18}$$

where the choice of $\pm$ in (17) *and* (18) *is the value that maximizes* (16)*. The exceptions are*

(1) *When both $+$ and $-$ maximize* (16)*, the distribution of type-'c' edges is a mixture of the two Gaussians defined above, provided that exceptions 2 and 3 do not also occur. (In the interest of space we omit the proof about the mixture of Gaussians, but we can supply it to the interested reader upon request.)*
(2) *If for the dominant choice of $\pm$ in* (16)*, the curve $\pm A^q e^{2i\alpha x - x^2}$ passes through* 1*, we do not say anything about the distribution of type-'c' edges. (We have reason to believe, but have not proved, that this scenario never occurs.)*
(3) *In the event that the integral in* (18) *evaluates to* 0*, the formulas are still valid, but we no longer claim that the distribution is a Gaussian. (We believe that this scenario never occurs, Lemma 6 in § 5.2 rules it out when $A \leq 1$.)*

Proof. Immediate from Lemma 1, Lemma 2, Lemma 3 (with the fact that Lemma 5 shows the dominant choice of $\pm$ to be $-$ when $A \leq 1$), Lemma 4, the fact that the moments are those of a Gaussian random variable, the method of moments, approximation (14) for $1/\sqrt{q^3\epsilon}$, and $qn \approx \sqrt{pqmnc/b}$. □

Theorem 1 follows by plugging into Theorem 2 $A = 1$ and $\alpha = 0$, using the explicit evaluation of the integrals in § 5.1, and using $-\operatorname{Li}_{3/2}(-1) > -\operatorname{Li}_{3/2}(1)$, $-\operatorname{Li}_{-1/2}(-1) > 0$, and $\operatorname{Li}_\nu(-1) = (2^{1-\nu} - 1)\zeta(\nu)$.

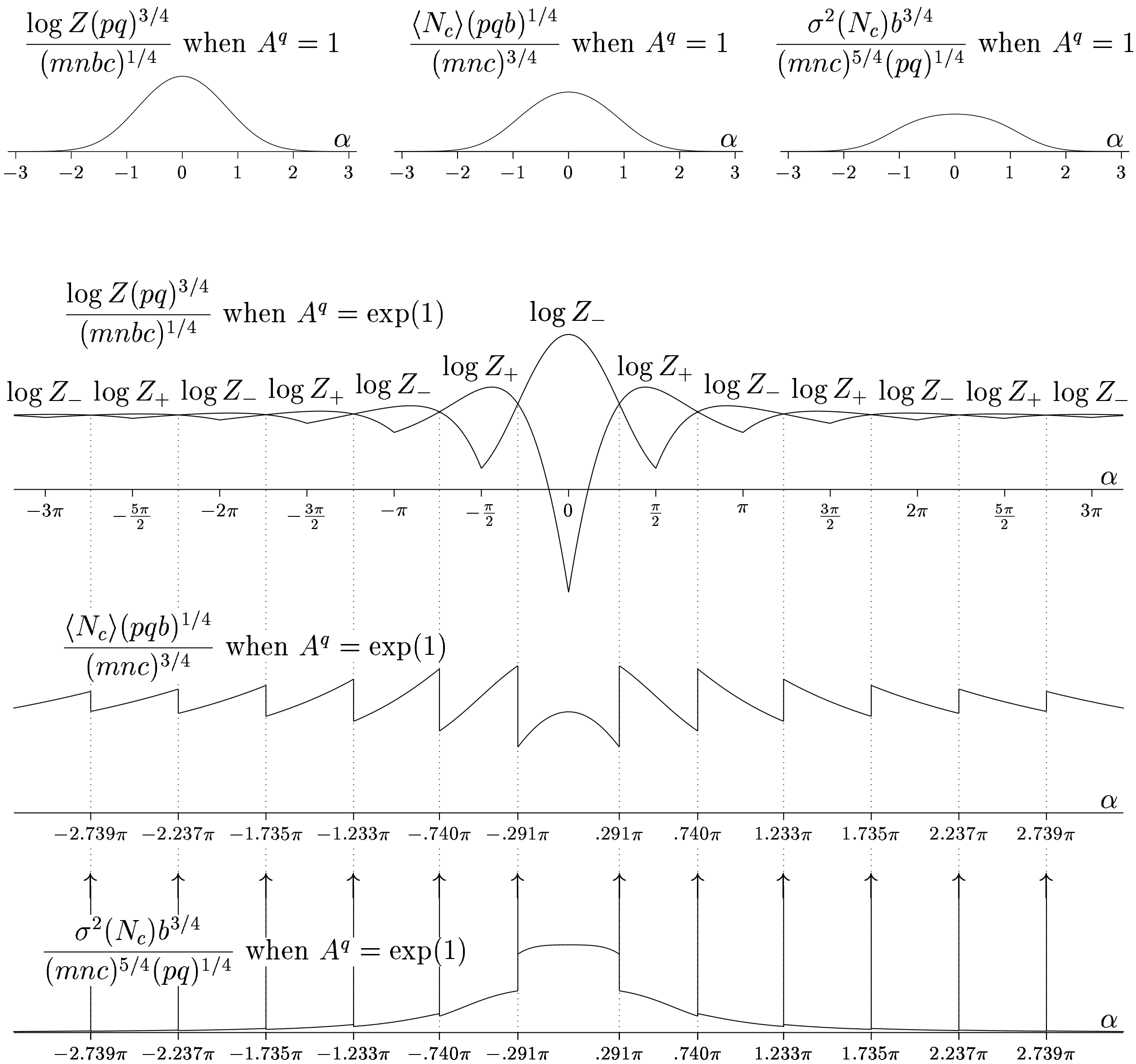


FIGURE 5. Anatomy of the resonant spikes. In the upper panel we show the curves for $\log Z$, $\langle N_c\rangle$, and $\sigma^2(N_c)$ as a function of α when $A = 1$. When $A \leq 1$, Z_- is dominant, and all three curves are analytic and unimodal. The situation is quite different when $A > 1$. In the next panel we show the curves for $\log Z_+$ and $\log Z_-$ as a function of α when $A^q = \exp(1)$ ($\log Z$ is the max of these two curves). When $A > 1$ there are singularities in $\log Z_+$ and $\log Z_-$ that occur when their corresponding spirals cross the singularity, that is, when $\alpha = (\pi/\sqrt{\log A^q})\mathbb{Z}$ for $\log Z_+$ and when $\alpha = (\pi/\sqrt{\log A^q})(\mathbb{Z} + 1/2)$ for $\log Z_-$. In the lower two panels we show the curves for $\langle N_c\rangle$ and $\sigma^2(N_c)$. These curves were computed using (17) and (18) as explicitly evaluated in (21) with whichever of Z_+ or Z_- is significant. The "crossover points", where $\log Z_-$ and $\log Z_+$ alternate in significance, are in a sense a phase transition within a phase transition. At each crossover point, the curve for $\log Z$ is continuous but nonanalytic, the curve for $\langle N_c\rangle$ is discontinuous, and the curve for $\sigma^2(N_c)$ has a delta function. When a spiral hits the singularity, the corresponding $Z_\pm$ appears to be the insignificant one. The large-α asymptotics of the curves for $\log Z$, $\langle N_c\rangle$, and $\sigma^2(N_c)$ are given in Theorem 3.

5. Understanding the resonant spikes

The resonant spikes, an example of which is shown in Figure 5, exhibit nontrivial behavior. We saw already that this behavior is determined by the integrals

$$\int_{-\infty}^{\infty} \operatorname{Li}_\nu\left(\beta e^{2i\alpha x - x^2}\right) dx,$$

with $\nu = 1$ governing $\log Z$, $\nu = 0$ governing $\langle N_c \rangle$, and $\nu = -1$ governing $\sigma^2(N_c)$. We start by explicitly evaluating these integrals in § 5.1, and then investigate some of their properties in § 5.2, § 5.3, and § 5.4. These subsections may be read in any order.

5.1. **The integral** $\int_{-\infty}^{\infty} \operatorname{Li}_\nu\left(\beta e^{2i\alpha x - x^2}\right) dx$**.** We assume that $\beta \in \mathbb{C}$, $\alpha \in \mathbb{R}$, and $\nu \in \mathbb{C}$, though later we restrict ν to integers ≤ 1. For convenience we assume $\alpha \ge 0$ since the integral is an even function of α. When $|\beta| > 1$ there is a branch cut ($\operatorname{Li}_\nu(z)$ has a branch cut $\{z \in \mathbb{R} : z \ge 1\}$) that may be encountered when we vary x, so we need to specify which branch of the polylogarithm we are integrating over. Since the principal branch is the one for which $\operatorname{Li}_\nu\left(\beta e^{2i\alpha x - x^2}\right) \to 0$ as $x \to \pm\infty$, we specify that the integrand is the principal branch of Li_ν, even though this may make the integrand only piecewise analytic as a function of x as it ranges from $-\infty$ to ∞. In the event that $\alpha = 0$ and β is real and ≥ 1, so as to ensure continuity in α, we specify that $\beta e^{2i\alpha x - x^2}$ lies above the branch cut for positive x and below the cut for negative x.

We set $z = x + iy$ ($x, y \in \mathbb{R}$) and integrate instead within the complex plane. Rather than integrate $\operatorname{Li}_\nu\left(\beta e^{2i\alpha z - z^2}\right)$ along the real axis $\Im z = 0$, it is more convenient to the integrate along the line $\Im z = \alpha$. When we deform the contour of integration and set $z = x + i\alpha$ we get

$$\int_{-\infty}^{\infty} \operatorname{Li}_\nu\left(\beta e^{2i\alpha x - x^2}\right) dx = \int_{-\infty}^{\infty} \operatorname{Li}_\nu\left(\beta e^{-\alpha^2} e^{-x^2}\right) dx + \begin{array}{l}\text{terms from the singularities and}\\ \text{branch cuts of } \operatorname{Li}_\nu\left(\beta e^{2i\alpha z - z^2}\right) \text{ in}\\ \text{the complex } z\text{-plane.}\end{array}$$

In particular if $|\beta| \le 1$ there are no singularities or branch cuts encountered when deforming the contour of integration, so there are no additional terms. We shall evaluate the main term in first. For now assume that $|\beta| e^{-\alpha^2} < 1$ so that the series expansion of the polylogarithm is absolutely convergent. This enables us to write

$$\begin{aligned}
\int_{-\infty}^{\infty} \operatorname{Li}_\nu\left(\beta e^{-\alpha^2} e^{-x^2}\right) dx &= \int_{-\infty}^{\infty} \sum_{n=1}^{\infty} \frac{(\beta e^{-\alpha^2})^n}{n^\nu} e^{-nx^2}\, dx \\
&= \sum_{n=1}^{\infty} \frac{(\beta e^{-\alpha^2})^n}{n^\nu} \int_{-\infty}^{\infty} e^{-nx^2}\, dx \\
&= \sum_{n=1}^{\infty} \frac{(\beta e^{-\alpha^2})^n}{n^\nu} \frac{\sqrt{\pi}}{\sqrt{n}} \\
&= \sqrt{\pi} \operatorname{Li}_{\nu+1/2}\left(\beta e^{-\alpha^2}\right).
\end{aligned}$$

The singularities of $\operatorname{Li}_\nu\left(\beta e^{2i\alpha z - z^2}\right)$ in the complex z-plane occur when

$$\begin{aligned}
\beta e^{2i\alpha z - z^2} &= 1 \\
2i\alpha z - z^2 &= -\log\beta - 2\pi i k \\
z &= i\alpha \pm i\sqrt{\alpha^2 - \log\beta - 2\pi i k}
\end{aligned}$$

where k is an integer. Since we are moving the contour between $\Im z = 0$ and $\Im z = \alpha$ (see Figure 6), the relevant singularities are of the form $i\alpha - i\sqrt{\alpha^2 - \log\beta - 2\pi i k}$ where the

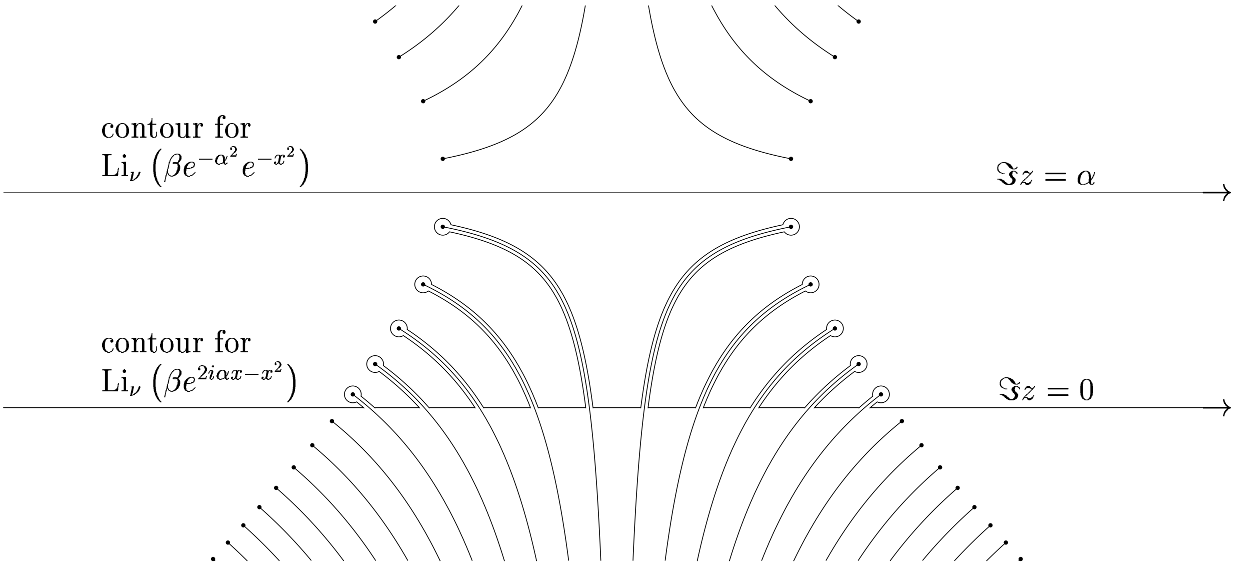


FIGURE 6. Singularities, branch cuts, and contours of integration for $\mathrm{Li}_\nu\left(\beta e^{2i\alpha z - z^2}\right)$. When pushing the contour from $\Im z = \alpha$ to $\Im z = 0$, the contour must deform to travel around the singularities and branch cuts, so these contribute to the integral.

principal square root is taken. (If $\arg\beta = 0$, $\alpha^2 < \log|\beta|$, and $k = 0$ then both roots are relevant.) The branch cuts of the integrand occur where $\beta e^{2i\alpha z - z^2}$ is real and ≥ 1. With $z = x + iy$ and $x, y \in \mathbb{R}$ these are

$$2\alpha x - 2xy = -\arg\beta - 2\pi k \qquad -2\alpha y - x^2 + y^2 \geq -\log|\beta|$$

$$x = \frac{1}{2}\frac{2\pi k + \arg\beta}{y - \alpha} \qquad (y-\alpha)^2 - x^2 \geq \alpha^2 - \log|\beta|$$

To identify if and where the kth cut intersects the line $\Im z = 0$ we set $y = 0$ and find that there is an intersection at

$$-\frac{2\pi k + \arg\beta}{2\alpha}$$

provided

$$(2\pi k + \arg\beta)^2 \leq 4\alpha^2 \log|\beta|.$$

For the moment suppose that $|\beta| e^{-\alpha^2} < 1$ so that there are no singularities or branch cuts of the integrand on the line $\Im z = \alpha$. Push this contour downwards toward the line $\Im z = 0$, except that when a branch cut is encountered, the contour must go around the branch cut, and travels along the cut upwards on its left side in downwards on its right side (Figure 6). The integrand increases by $(2\pi i/\Gamma(\nu))(\log(\beta e^{2i\alpha z - z^2}))^{\nu-1}$ when going from the left side of the cut to the right side, so the net contribution of this cut to the integral is

$$\frac{2\pi i}{\Gamma(\nu)} \int_C (\log(\beta e^{2i\alpha z - z^2}))^{\nu-1}\, dz$$

(plus another term from the singularity) where the contour C is a branch cut travelling from the branch point to the point where the branch cut intersects the line $\Im z = 0$. For the kth branch cut the endpoints of integration are $i\alpha - i\sqrt{\alpha^2 - \log\beta - 2\pi i k}$ and $-(2\pi k +$

$\arg\beta)/(2\alpha)$. When $\nu = 1$ the singularity does not contribute and this integral is easy to evaluate and we get

$$2\pi i\left[-i\alpha + i\sqrt{\alpha^2 - \log\beta - 2\pi i k} - \frac{2\pi k + \arg\beta}{2\alpha}\right].$$

Since this term gets added to the integral over $\Im z = 0$, when we evaluate the integral over $\Im z = 0$ we subtract this term from the integral over $\Im z = \alpha$. Thus

(19)
$$\int_{-\infty}^{\infty} \mathrm{Li}_1\left(\beta e^{2i\alpha x - x^2}\right) dx = \sqrt{\pi}\,\mathrm{Li}_{3/2}\left(\beta e^{-\alpha^2}\right) - 2\pi i \sum_{\substack{k\in\mathbb{Z}\\ (2\pi k+\arg\beta)^2 \le 4\alpha^2\log|\beta|}} \left[-i\alpha + i\sqrt{\alpha^2 - \log\beta - 2\pi i k} - \frac{2\pi k + \arg\beta}{2\alpha}\right].$$

There is a term in the summation for each time the spiral $\beta e^{2i\alpha x - x^2}$ encloses the singularity at 1. This formula is valid when $|\beta| e^{-\alpha^2} < 1$, and we would like to show that it is valid (for real β and α) without this restriction. We can do this via analytic continuation in (complex) β and α. While both sides of equation (19) are analytic in α (except where the spiral $\beta e^{2i\alpha x - x^2}$ crosses the singularity), neither side of (19) is analytic in β. The $\arg\beta$ terms on the right-hand side are nonanalytic, and the left-hand side is nonanalytic due to the branch cut in $\mathrm{Li}_1(z)$ — as $\arg\beta$ is varied, portions of the integrand cross the branch cut. So long as the spiral avoids the singularity, we can rotate the branch cut as $\arg\beta$ is changed so as to ensure that the integrand does not cross the moving branch cut for any x. This re-interpreted integral is then analytic except where the spiral $\beta e^{2i\alpha x - x^2}$ crosses the singularity, and the $\arg\beta$ terms on the right-hand side become constant. For a given β and α, we may continuously decrease $|\beta|$ and increase α until $|\beta| e^{-\alpha^2} < 1$, where we know that (19) is valid, while keeping the spiral from crossing the singularity, so that both sides of the re-interpreted (19) remain analytic. (The nonanalyticity in $\sqrt{\pi}\,\mathrm{Li}_{3/2}\left(\beta e^{-\alpha^2}\right)$ where $\beta e^{-\alpha^2} = 1$ is cancelled by the nonanalyticity in the $k = 0$ term in the summation.) Thus (19) is valid for all complex β and α for which the spiral avoids the singularity. As both sides of (19) are continuous, (19) is also valid when the spiral contains the singularity. We revert to the principal branch cut interpretation of the integral, restoring the $\arg\beta$ terms on the right-hand side of (19).

When β is real, upon summing over k the terms $(2\pi k + \arg\beta)/(2\alpha)$ cancel: when $\beta < 0$ the kth term cancels the $(-1-k)$th term, and when $\beta > 0$ the kth term cancels the $-k$th term, while the 0th term does not contribute. For real β we get

$$\int_{-\infty}^{\infty} \mathrm{Li}_1\left(\beta e^{2i\alpha x - x^2}\right) dx = \sqrt{\pi}\,\mathrm{Li}_{3/2}\left(\beta e^{-\alpha^2}\right) - 2\pi \sum_{\substack{k\in\mathbb{Z}\\ (2\pi k+\arg\beta)^2 \le 4\alpha^2\log|\beta|}} \left(\alpha - \sqrt{\alpha^2 - \log\beta - 2\pi i k}\right). \tag{20}$$

When $\beta e^{-\alpha^2} > 1$ the $k = 0$ "branch-cut term" exactly cancels the imaginary part of the $\mathrm{Li}_{3/2}$ term. Applying $\beta\frac{\partial}{\partial\beta}$ multiple times to (20) we find for real β and nonpositive integer ν

(21)
$$\int_{-\infty}^{\infty} \mathrm{Li}_\nu\left(\beta e^{2i\alpha x - x^2}\right) dx = \sqrt{\pi}\,\mathrm{Li}_{\nu+1/2}\left(\beta e^{-\alpha^2}\right) - \sqrt{\pi}\,\Gamma\left(\tfrac{1}{2} - \nu\right) \sum_{\substack{k\in\mathbb{Z}\\ (2\pi k+\arg\beta)^2 \le 4\alpha^2\log|\beta|}} [\alpha^2 - \log\beta + 2\pi i k]^{-1/2+\nu}.$$

5.2. **Inequalities.** To use Theorem 2 we need to know which of $Z_\pm$ is significant. We already saw that they can alternate in significance (as α is varied) when $A > 1$. We claim that when $A \leq 1$, or else $A > 1$ but $\alpha = 0$, that Z_- is the significant one.

Lemma 5. *If* $0 < \beta \leq 1$, *or else* $\beta > 1$ *but* $\alpha = 0$, *then*

$$-\int_{-\infty}^{\infty} \mathrm{Li}_1\left(\beta e^{2i\alpha x - x^2}\right) dx < -\int_{-\infty}^{\infty} \mathrm{Li}_1\left(-\beta e^{2i\alpha x - x^2}\right) dx. \tag{22}$$

Proof: Under either of the hypotheses, equation (20) simplifies to

$$\int_{-\infty}^{\infty} \mathrm{Li}_1\left(\pm\beta e^{2i\alpha x - x^2}\right) dx = \sqrt{\pi}\,\mathrm{Li}_{3/2}\left(\pm\beta e^{-\alpha^2}\right) = \int_{-\infty}^{\infty} \mathrm{Li}_1\left(\pm\beta e^{-\alpha^2 - x^2}\right) dx,$$

but
□

$$-\Re\,\mathrm{Li}_1\left(\beta e^{-\alpha^2-x^2}\right) = \Re\log\left(1-\beta e^{-\alpha^2-x^2}\right) < \Re\log\left(1+\beta e^{-\alpha^2-x^2}\right) = -\Re\,\mathrm{Li}_1\left(-\beta e^{-\alpha^2-x^2}\right).$$

To show that the distribution of N_c is a Gaussian, we need to know that the integral expression in (18) (for the dominant choice of $\pm$) is nonzero. The integral clearly cannot be negative, because it has an interpretation in terms of variance, but *a priori* it could be zero, in which case the $o(1)$ error term would control the variance, and we would be unable to characterize the distribution of N_c. Ideally we would like to show that it always positive, but we only do this for $A \leq 1$, or else $A > 1$ but $\alpha = 0$. Recall that under these conditions Z_- is dominant.

Lemma 6. *If* $0 < \beta \leq 1$, *or else* $\beta > 1$ *but* $\alpha = 0$, *then*

$$-\int_{-\infty}^{\infty} \mathrm{Li}_{-1}\left(-\beta e^{2i\alpha x - x^2}\right) dx > 0.$$

Proof. As above, under either of the hypotheses, equation (21) simplifies to

$$\int_{-\infty}^{\infty} \mathrm{Li}_{-1}\left(-\beta e^{2i\alpha x - x^2}\right) dx = \sqrt{\pi}\,\mathrm{Li}_{-1/2}\left(-\beta e^{-\alpha^2}\right) = \int_{-\infty}^{\infty} \mathrm{Li}_{-1}\left(-\beta e^{-\alpha^2 - x^2}\right) dx,$$

but $-\mathrm{Li}_{-1}\left(-\beta e^{-\alpha^2-x^2}\right) = \beta e^{-\alpha^2-x^2}/\left(1+\beta e^{-\alpha^2-x^2}\right)^2 > 0$. □

Lemma 7. *The curve for* $\log Z$ *attains its (unique) maximum value when* $\alpha = 0$.

Proof. When $0 < \beta$ we have $\log(1+\beta e^{-x^2}) \geq \Re\log(1\pm\beta e^{2i\alpha x - x^2})$, with strict inequality when $\pm e^{2i\alpha x} \neq 1$, so $-\int_{-\infty}^{\infty} \mathrm{Li}_1\left(\mp\beta e^{2i\alpha x - x^2}\right) dx \leq -\int_{-\infty}^{\infty} \mathrm{Li}_1\left(-\beta e^{-x^2}\right) dx$, with strict inequality unless both $\alpha = 0$ and $\mp = -$. □

In contrast, the curve for $\langle N_c\rangle$ does not always attain its maximum value when $\alpha = 0$.

5.3. **Asymptotics for large** α**.** In Figure 3 the curves for $\log Z$ and $\langle N_c\rangle$ appear to asymptote out to a positive constant, while the curve for $\sigma^2(N_c)$ decays to 0. The following theorem gives the large-α asymptotics of these curves which are given by the integrals in (16), (17), and (18) in Theorem 2.

Theorem 3. *For positive β and real α,*

$$\lim_{|\alpha|\to\infty} -\int_{-\infty}^{\infty} \mathrm{Li}_1\left(\pm\beta e^{2i\alpha x - x^2}\right) dx = \begin{cases} \frac{4}{3}(\log\beta)^{3/2} & \beta\geq 1 \\ 0 & \beta\leq 1\end{cases}$$

$$\lim_{|\alpha|\to\infty} -\int_{-\infty}^{\infty} \mathrm{Li}_0\left(\pm\beta e^{2i\alpha x - x^2}\right) dx = \begin{cases} 2(\log\beta)^{1/2} & \beta\geq 1 \\ 0 & \beta\leq 1\end{cases}$$

$$\lim_{|\alpha|\to\infty} -\alpha^2\int_{-\infty}^{\infty} \mathrm{Li}_{-1}\left(\pm\beta e^{2i\alpha x - x^2}\right) dx = \begin{cases} (\log\beta)^{1/2} & \beta\geq 1 \\ 0 & \beta\leq 1\end{cases}$$

Proof. These integrals are evaluated in (20) and (21). When $|\alpha|\to\infty$, the main terms in (20) and (21), $-\sqrt{\pi}\,\mathrm{Li}_{\nu+1/2}(\pm\beta e^{-\alpha^2})$ for $\nu=1,0,-1$, become negligible. Thus we only have to compute the "branch-cut terms" when $\beta>1$. For (20) these terms can be approximated for large α as follows. Let $\gamma=\log\beta-\alpha^2$. The identity

$$\sqrt{x+iy}+\sqrt{x-iy}=\sqrt{2\big(x+\sqrt{x^2+y^2}\big)}$$

is easily verified by squaring both sides, and allows us to rewrite the sum in (20) as

$$\sum_{|k|<|\alpha|\sqrt{\log\beta}/\pi} 2\pi\left(|\alpha|-\sqrt{\big(-\gamma+\sqrt{\gamma^2+4\pi^2k^2}\big)/2}\right)$$

where k is integer for positive β (for Z_+) or half-integer for negative β (for Z_-).

Suppose $|\alpha|\gg\sqrt{\log\beta}$. Then k, which ranges up to $(|\alpha|/\pi)\sqrt{\log\beta}$, is much smaller than $|\gamma|=(1+o(1))\alpha^2$, and we have

$$\begin{aligned}\sqrt{\big(-\gamma+\sqrt{\gamma^2+4\pi^2k^2}\big)/2} &= \sqrt{-\gamma}\sqrt{\big(1+\sqrt{1+4\pi^2k^2/\gamma^2}\big)/2}\\ &= \sqrt{-\gamma}\sqrt{1+(\pi^2+o(1))k^2/\gamma^2}\\ &= \sqrt{\alpha^2-\log\beta}(1+(\pi^2/2+o(1))k^2/\gamma^2)\\ &= |\alpha|(1-(1/2+o(1))\log\beta/\alpha^2)(1+(\pi^2/2+o(1))k^2/\alpha^4)\\ &= |\alpha|-\frac{1+o(1)}{2|\alpha|}(\log\beta-\pi^2k^2/\alpha^2).\end{aligned}$$

In particular

$$2\pi\left(|\alpha|-\sqrt{\big(-\gamma+\sqrt{\gamma^2+4\pi^2k^2}\big)/2}\right)=\frac{\pi+o(1)}{|\alpha|}(\log\beta-\pi^2k^2/\alpha^2)$$

and we sum this over k's for which the quantity is positive. If $|\alpha|\gg 1/\sqrt{\log\beta}$ then we can approximate this sum with an integral, and the integral is

$$\frac{\pi+o(1)}{|\alpha|}\frac{2}{3}\log\beta\times\frac{2|\alpha|}{\pi}\sqrt{\log\beta}=(1+o(1))\frac{4}{3}(\log\beta)^{3/2}.$$

These are the asymptotics for $\nu=1$. For integer $\nu\leq 0$ the range of k is the same as it is for $\nu=1$, i.e. up to about $|\alpha|\sqrt{\log\beta}/\pi$. Thus we need to evaluate

$$\sqrt{\pi}\Gamma(1/2-\nu)\sum_{|k|<|\alpha|\sqrt{\log\beta}/\pi}[\alpha^2-\log\beta+2\pi ik]^{-1/2+\nu}.$$

For $\alpha\gg\sqrt{\log\beta}$, $k\ll\alpha^2$ so that

$$[\alpha^2-\log\beta+2\pi ik]^{-1/2+\nu}+[\alpha^2-\log\beta-2\pi ik]^{-1/2+\nu}\approx 2|\alpha|^{2\nu-1}$$

and if $\alpha \gg 1/\sqrt{\log \beta}$ the summation is approximately an integral which is asymptotically

$$\frac{2}{\sqrt{\pi}}\Gamma(1/2-\nu)\alpha^{2\nu}\sqrt{\log \beta}.$$

Taking $\nu = 0$ and $\nu = -1$ give the desired results for the $\langle N_c \rangle$ and $\sigma^2(N_c)$ integrals. □

5.4. **Crossover locations.** If $A \leq 1$ then Z_- always exceeds Z_+. But for larger A there are crossover values for α at which Z_- and Z_+ alternate in significance. Since Z_- and Z_+ count configurations in different $\mathbb{Z}_2$-homology classes (see (4)), we conclude that the crossover values are the places where the typical homology class of a lattice path changes. So there is a phase transition at these points: the topology of a typical configuration changes.

The crossover values α for a fixed $\beta = A^q$ satisfy an implicit (and transcendental) equation. Instead of solving these equations directly we can derive an analytic expression for the crossover α's as a function of $\gamma = \log \beta - \alpha^2$, and then given γ and the crossover α we can calculate the corresponding β. In this way we can parametrically plot these critical pairs (β, α) as a function of γ. For example, for the 0th crossover we have

$$\begin{aligned}
\epsilon^{1/2}q^{3/2}Z_- + o(1) &= -\sqrt{\pi}\,\mathrm{Li}_{3/2}(-\beta e^{-\alpha^2}) \\
\epsilon^{1/2}q^{3/2}Z_+ + o(1) &= -\sqrt{\pi}\,\mathrm{Li}_{3/2}(\beta e^{-\alpha^2}) + 2\pi(\alpha - \sqrt{\alpha^2 - \log \beta})
\end{aligned}$$

from which we can solve

$$\alpha = \frac{\mathrm{Li}_{3/2}(e^\gamma) - \mathrm{Li}_{3/2}(-e^\gamma)}{2\sqrt{\pi}} + \sqrt{-\gamma}.$$

Similarly, for the rth crossover we have

$$(-1)^r\alpha = \frac{\mathrm{Li}_{3/2}(e^\gamma) - \mathrm{Li}_{3/2}(-e^\gamma)}{2\sqrt{\pi}} + \sum_{k=-r}^{r}(-1)^k\sqrt{-\gamma + k\pi i}.$$

Next let us approximate these crossovers for r fixed and γ large. For large γ we can substitute $\nu = 3/2$ into the asymptotic series expansions (25) and (26) for $\mathrm{Li}_\nu(\pm e^\gamma)$ to write

$$-\mathrm{Li}_{3/2}(-e^\gamma) \approx \frac{4}{3}/\pi^{1/2}\gamma^{3/2} + \frac{1}{6}\pi^{3/2}/\gamma^{1/2} + \frac{7}{480}\pi^{7/2}/\gamma^{5/2} + \cdots$$

$$\mathrm{Li}_{3/2}(e^\gamma) \approx -\frac{4}{3}/\pi^{1/2}\gamma^{3/2} - 2\sqrt{\pi}i\sqrt{\gamma} + \frac{1}{3}\pi^{3/2}/\gamma^{1/2} + \frac{1}{60}\pi^{7/2}/\gamma^{5/2} + \cdots$$

so that

$$\begin{aligned}
(-1)^r\alpha &\approx \frac{\pi}{4}\gamma^{-1/2} + \sum_{k=1}^{r}(-1)^k\left[\sqrt{-\gamma+k\pi i}+\sqrt{-\gamma-k\pi i}\right] \\
&= \frac{\pi}{4}\gamma^{-1/2} + \sum_{k=1}^{r}(-1)^k\sqrt{2\big(-\gamma+\sqrt{\gamma^2+k^2\pi^2}\big)} \\
&\approx \frac{\pi}{4}\gamma^{-1/2} + \sum_{k=1}^{r}(-1)^k\sqrt{2(-\gamma+\gamma(1+k^2\pi^2/(2\gamma^2)))} \\
&= \frac{\pi}{4}\gamma^{-1/2} + \sum_{k=1}^{r}(-1)^k\sqrt{k^2\pi^2/\gamma} \\
&= \frac{\pi}{4}\gamma^{-1/2} + \sum_{k=1}^{r}(-1)^k k\pi/\gamma^{1/2} \\
&= (-1)^r(r/2+1/4)\pi\gamma^{-1/2}.
\end{aligned}$$

Since α is small, $\gamma \approx \log\beta$, and so the rth crossover occurs at $\alpha \approx (r/2+1/4)\pi/\sqrt{\log\beta}$.

By comparison the nonanalyticities in the curve for Z_- occur exactly at $\alpha = \pi/\sqrt{\log\beta}(\mathbb{Z}+1/2)$, and the nonanalyticities in the curve for Z_+ occur exactly at $\alpha = \pi/\sqrt{\log\beta}\mathbb{Z}$.

6. Open problems

Our analysis for of the lattice paths at the critical point is geared to regions whose aspect ratio is close to a simple rational number, but we do not know how to treat "irrational domains" which do not have a simple rational approximation. Consider for instance the case $a = 1$, $b = c = 1/2$ when the side lengths m and n are successive Fibonacci numbers. Then the aspect ratio of the region is very close to the Golden ratio, which is not well approximated by simple rationals. In this case we believe that the partition function Z is $\Theta(1)$ (about 2.1), so that with $\Theta(1)$ probability there are no lattice paths. In the event that there are lattice paths, we believe that they connect up into $\Theta(1)$ loops each of length $\Theta(n^{4/3})$. We have also found empirically that as one varies the aspect ratio of large regions, the smallest value that the partition function Z takes on is very close to 2. We do not know how to prove any of these conjectures.

In Figure 3 we plotted $\log Z$ and $\langle N_c\rangle$ versus n/m when $a = 1$, $b = c = 1/2$. As predicted there are spikes in $\log Z$ and $\langle N_c\rangle$ when the aspect ratio n/m is near a simple rational p/q. But there also appear to be flanking secondary spikes in $\langle N_c\rangle$ when $n/m \approx p/q$ but $|n/m - p/q| \gg 1/\sqrt{n}$. We do not understand this phenomenon.

When $a = 1$, $b = c = 1/2 + \Theta(1/n)$ and the aspect ratio is nearly a simple rational p/q, the partition function and edge density as a function of α are nonanalytic at certain points. We know that the homology type of the strands changes at these nonanalyticities. We conjecture that when $\alpha = 0$ each loop winds around exactly p times horizontally and q times vertically. For large enough values of α, when one follows a loop for $mp+nq$ steps, one does not return to the starting point. We conjecture that the number of nonanalyticities between α and 0 determines how many strands away one ends up after following a loop for $mp + nq$ steps.

Despite our explicit formulas for $\log Z_\pm$ and for the expected value and variance of the number of edges, there are some basic properties about these functions that we have not been able to derive. For example, we conjecture that when our formula for $\log Z_-$ is nonanalytic, our formula for $\log Z_+$ gives a *strictly* larger value, and vice versa, and that the formula for $\sigma^2(N_c)$ is strictly positive. We also conjecture that for $A > 1$ our formula for $\log Z$ as a function of α always exceeds its limiting value as $\alpha \to \infty$. These conjectures seem fairly evident from the graphs in Figure 5, but it is not obvious how to prove them. One would also like to determine for fixed $A > 1$ the maximum and minimum values given by our formula for $\langle N_c \rangle$, since both of these can be different than the limiting $\alpha \to \infty$ value.

References

[1] Harry Bateman and A. Erdélyi *et al.* *Higher Transcendental Functions*, volume 1. McGraw-Hill, 1953. Based in part on notes left by Bateman, edited by Erdélyi *et al.*

[2] H. W. J. Blöte and H. J. Hilhorst. Roughening transitions and the zero-temperature triangular Ising antiferromagnet. *Journal of Physics A*, 15(11):L631–L637, 1982.

[3] C. Borgs, J. T. Chayes, and C. King. Meissner phase for a model of oriented flux lines. *Journal of Physics A*, 28(23):6483–6499, 1995.

[4] Raphaël Cerf and Richard Kenyon. The low-temperature expansion of the Wulff crystal in the 3D Ising model. *Communications in Mathematical Physics*, 222:147–179, 2001.

[5] Joseph T. Chang and Yuval Peres. Ladder heights, Gaussian random walks and the Riemann zeta function. *Annals of Probability*, 25(2):787–802, 1997.

[6] Henry Cohn, Richard Kenyon, and James Propp. A variational principle for domino tilings. *Journal of the American Mathematical Society*, 14(2):297–346, 2001. arXiv:math.CO/0008220.

[7] Louis Comtet. *Advanced Combinatorics*. D. Reidel Publishing Co., 1974.

[8] Marcel den Nijs. The domain wall theory of two-dimensional commensurate-incommensurate phase transitions. In C. Domb and J. L. Lebowitz, editors, *Phase Transitions and Critical Phenomena*, volume 12, pages 219–333. Academic Press, 1988.

[9] R. B. Dingle. On the Bose-Einstein integrals $\mathcal{B}_p(\eta) = (p!)^{-1} \int_0^\infty \varepsilon^p (e^{\varepsilon-\eta} - 1)^{-1} d\varepsilon$. *Applied Scientific Research*, 6(4):240–244, 1957.

[10] R. B. Dingle. On the Fermi-Dirac integrals $\mathcal{F}_p(\eta) = (p!)^{-1} \int_0^\infty \varepsilon^p (e^{\varepsilon-\eta} + 1)^{-1} d\varepsilon$. *Applied Scientific Research*, 6(4):225–239, 1957.

[11] Michael E. Fisher. Walks, walls, wetting and melting. *Journal of Statistical Physics*, 34(5–6):667–729, 1984.

[12] Anna Galluccio and Martin Loebl. On the theory of Pfaffian orientations. I. Perfect matchings and permanents. *Electronic Journal of Combinatorics*, 6(1):#R6, 1999.

[13] H. Y. Huang, F. Y. Wu, H. Kunz, and D. Kim. Interacting dimers on the honeycomb lattice: An exact solution of the five-vertex model. *Physica A*, 228(1–4):1–32, 1996. arXiv:cond-mat/9510161.

[14] P. W. Kasteleyn. Graph theory and crystal physics. In Frank Harary, editor, *Graph Theory and Theoretical Physics*. Academic Press, 1967.

[15] Richard Kenyon. Local statistics of lattice dimers. *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques*, 33(5):591–618, 1997. arXiv:math.CO/0105054.

[16] Leonard Lewin. *Polylogarithms and Associated Functions*. North Holland, 1981.

[17] Ernst Lindelöf. *Le Calcul des Résidus et ses Applications a la Théorie des Fonctions*. Gauthier-Villars, 1905.

[18] W. T. Lu and F. Y. Wu. Dimer statistics on the Möbius strip and the Klein bottle. *Physics Letters A*, 259(2):108–114, 1999. arXiv:cond-mat/9906154.

[19] Wentau T. Lu and F. Y. Wu. Close-packed dimers on nonorientable surfaces, 2001. arXiv:cond-mat/0110035.

[20] J. F. Nagle. Theory of biomembrane phase transitions. *Journal of Chemical Physics*, 58:252–264, 1973.

[21] W. F. Pickard. On polylogarithms. *Publicationes Mathematicae*, 15(1–4):33–43, 1968.

[22] V. Popkov, Doochul Kim, H. Y. Huang, and F. Y. Wu. Lattice statistics in three dimensions: solution of layered dimer and layered domain wall models. *Physical Review E*, 56(4):3999–4008, 1997. arXiv:cond-mat/9703065.

[23] M. Prähofer and H. Spohn. An exactly solved model of three-dimensional surface growth in the anisotropic KPZ regime. *Journal of Statistical Physics*, 88(5–6):999–1012, 1997. arXiv:cond-mat/9612209.

[24] Tullio Regge and Riccardo Zecchina. Combinatorial and topological approach to the 3D Ising model. *Journal of Physics A*, 33:741–761, 2000. arXiv:cond-mat/9909168.

[25] Scott Sheffield. Gibbs measure uniqueness results for Lipschitz random surfaces, 2001. In preparation.

[26] Serge Tabachnikov. *Billiards*. Panoramas et Synthèses. Société Mathématique de France, 1995.

[27] Glenn Tesler. Matchings in graphs on non-orientable surfaces. *Journal of Combinatorial Theory, Series B*, 78:198–231, 2000.
[28] C. Truesdell. On a function which occurs in the theory of the structure of polymers. *The Annals of Mathematics*, 46(1):144–157, 1945.
[29] F. Y. Wu. Exactly soluable model of the ferroelectric phase transition in two dimensions. *Physical Review Letters*, 18(15):605–607, 1967.
[30] F. Y. Wu. Remarks on the modified potassium dihydrogen phosphate model of a ferroelectric. *Physical Review*, 168(2):539–543, 1968.
[31] F. Y. Wu and H. Y. Huang. Exact solution of a vertex model in d dimensions. *Letters in Mathematical Physics*, 29(3):205–213, 1993.
[32] F. Y. Wu and H. Y. Huang. Exact solution of a lattice model of flux lines in superconductors. *Physika A*, 205:31–40, 1994.

Appendix A. Polylogarithms

The polylogarithm function $\mathrm{Li}_\nu(z)$ is defined by

$$\mathrm{Li}_\nu(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^\nu}$$

for $|z| < 1$ and by analytic continuation elsewhere. We may for instance write $\mathrm{Li}_1(z) = -\log(1-z)$, $\mathrm{Li}_0(z) = z/(1-z)$, and $\mathrm{Li}_{-1}(z) = z/(1-z)^2$. The polylogarithm function has the convenient property that

$$z\frac{\partial\,\mathrm{Li}_\nu(z)}{\partial z} = \mathrm{Li}_{\nu-1}(z).$$

Appell's integral expression

$$\mathrm{Li}_\nu(z) = \frac{1}{\Gamma(\nu)} \int_0^\infty \frac{z s^{\nu-1}\,ds}{e^s - z} \tag{23}$$

is valid for $\Re(\nu) > 0$ and $z \notin [1,\infty)$, and defines the principal branch of the polylogarithm. The polylogarithm has an interesting Riemann surface. When one crosses the branch cut $[1,\infty)$ in the positive direction, the polylogarithm increases by $\frac{2\pi i}{\Gamma(\nu)}(\log z)^{\nu-1}$. (Both this and the defining series expansion for Li_ν are readily derived from Appell's integral expression.) For nonpositive integer ν this quantity is 0, consistent with the fact that Li_ν is a rational function for these values of ν. Unless ν is an integer ≤ 1, the $\frac{2\pi i}{\Gamma(\nu)}(\log z)^{\nu-1}$ term creates a second branch point at $z = 0$ off the principal branch. The function $\mathrm{Li}_\nu(z)$ is analytic in both z and ν, except for a singularity at $z = 1$ (and $z = 0$ off the principal branch). For further background see Bateman and Erdélyi *et al.* [1, Chapt. 1 §11], Truesdell [28], Dingle [10, 9], and Lewin [16].

When z is on the principal branch and near 1, the series expansion

$$\mathrm{Li}_\nu(z) = \Gamma(1-\nu)(-\log z)^{\nu-1} + \sum_{n=0}^{\infty} \zeta(\nu-n)\frac{(\log z)^n}{n!} \tag{24}$$

was given by Lindelöf [17, pp. 138–141] (derivations are also given in [28] and [5]), and is absolutely convergent when $|\log z| < 2\pi$.

We also use the asymptotic series expansions for $\mathrm{Li}_\nu(z)$ when $|z| \to \infty$. For large positive x these are

$$\mathrm{Li}_\nu(-x) = -\cos(\pi\nu)\,\mathrm{Li}_\nu(-1/x) + 2\sum_{k=0}^{\infty} \frac{\left(1/2^{2k-1} - 1\right)\zeta(2k)}{\Gamma(\nu+1-2k)}(\log x)^{\nu-2k} \tag{25}$$

and

$$\mathrm{Li}_\nu(x) = -\cos(\pi\nu)\,\mathrm{Li}_\nu(1/x) \pm \pi i \frac{(\log x)^{\nu-1}}{\Gamma(\nu)} + 2\sum_{k=0}^{\infty} \frac{\zeta(2k)}{\Gamma(\nu+1-2k)}(\log x)^{\nu-2k}. \tag{26}$$

Aside from the $-\cos(\pi\nu)\,\mathrm{Li}_\nu(\pm 1/x)$ terms, the asymptotic series expansions (25) and (26) were derived by Sommerfeld and Clunie respectively. Rhodes first derived (25) for the case of integer ν, and for these ν the expansions (25) and (26) have only finitely many nonzero terms. For noninteger ν the expansions (25) and (26) diverge, and the $-\cos(\pi\nu)\,\mathrm{Li}_\nu(\pm 1/x)$ term is dominated by each term in the divergent series expansion, so one may wonder what role it plays. Dingle [10, eqn 17] [9, eqn 11] showed how to make practical computational use of these series by truncating the series after finitely many terms and providing convergent series expansions for the remainder when the $-\cos(\pi\nu)\,\mathrm{Li}_\nu(\pm 1/x)$ term is present. For positive $x > 1$ we are evaluating $\mathrm{Li}_\nu(x)$ on the branch cut of the principal branch; the $\pm$ sign is positive when the branch cut is just above the real axis, which is the usual convention. See also Pickard [21, eqn 3.5] for the asymptotics for large complex values of z.

Appendix B. Notation

a, b, c: weights of edges in the three different directions. a is weight for vertex not being in a loop, b is weight for horizontal step, and c is weight for vertical step.
m: horizontal length of torus
n: vertical length of torus
Simplifying assumptions: $b < a$, $b/a = \Theta(1)$, $1 - b/a = \Theta(1)$, $n = \Theta(m)$
p/q: rational approximation of $nb/(mc)$, $\gcd(p,q) = 1$. Intuitively p is the number of horizontal windings of loops, and q is the number of vertical windings of loops, but this intuition is not quite accurate if "ratcheting" takes place.
$W = \sqrt{q\epsilon}/(\pi p) = \sqrt{2qnab}/(pm(a-b)) \approx \sqrt{2/(pmc)} \approx \sqrt{2/(qnb)}$
$\dfrac{nbq}{mcp} \approx \dfrac{nbq}{m(a-b)p} = 1 + \alpha W.$
$\phi = \frac{2\pi nb}{m(a-b)}$
$\epsilon = \frac{2\pi^2 nab}{m^2(a-b)^2}$
$A = \left(\frac{c}{a-b}\right)^n$
$z_k = -e^{i\theta_k}$, $\theta_k = 2\pi k/m$; $k \in \mathbb{Z}_m + \sigma/2$ for $Z_{\sigma\tau}$
$r_k = (a-b)^n |a + bz_k|^{-n} = e^{-\epsilon k^2 + O(k^4/m^3)}$
$\phi_k = \arg(a + bz_k)^{-n} = \phi k + O(k^3/m^2)$

CNRS UMR 8628, Laboratoire de Mathématiques, Université Paris-Sud, France
E-mail address: kenyon@topo.math.u-psud.fr

Microsoft Research, One Microsoft Way, Redmond, Washington, U.S.A.
E-mail address: dbwilson@microsoft.com

ON THE AUTOMORPHISM GROUP OF THE SUBMODULE LATTICE OF A MODULE OVER COMPLETE DISCRETE VALUATION RING

KAN YASUDA

Abstract. The automorphism group $\operatorname{Aut}\mathcal{L}(M)$ of the submodule lattice $\mathcal{L}(M)$ of a finite-length module M over complete discrete valuation ring $\mathfrak{o}$ is studied. Let $\lambda = (\lambda_1, \cdots, \lambda_l)$ be the type of M. We show that for those M with $\lambda_1 = \lambda_2$, $\operatorname{Aut}\mathcal{L}(M)$ can be analyzed by computing a certain subgroup of the bijections on a quotient of the scalar ring $\mathfrak{o}$. In particular, when $\mathfrak{o}$ is $\mathbb{Z}_p$ (hence M is a finite abelian p-group) or the ring of Witt vectors over the finite field $\mathbb{F}_q$, we can compute $\operatorname{Aut}\mathcal{L}(M)$ more explicitly.

Résumé. Étudié est le groupe $\operatorname{Aut}\mathcal{L}(M)$ des automorphismes du réseau $\mathcal{L}(M)$ des sous-modules d'un module M de type fini sur un anneau $\mathfrak{o}$ de valuation discrète complet. Soit $\lambda = (\lambda_1, \cdots, \lambda_l)$ le type de M. Nous montrons que, pour tels M que $\lambda_1 = \lambda_2$, le groupe $\operatorname{Aut}\mathcal{L}(M)$ peut être analysé par un calcul d'un certain sous-groupe du groupe des permutations dans un quotient de l'anneau $\mathfrak{o}$ scalaire. En particulier, si $\mathfrak{o}$ est $\mathbb{Z}_p$ (et, par conséquent, M est un p-groupe abelien fini), ou l'anneau de vecteurs de Witt sur un corps $\mathbb{F}_q$ fini, nous pouvons calculer le groupe $\operatorname{Aut}\mathcal{L}(M)$ plus explicitement.

1. Introduction

Let $\mathfrak{o}$ be a complete discrete valuation ring with the maximal ideal $\mathfrak{p}$, a prime element π and the valuation v. Let M be an $\mathfrak{o}$-module of finite length. Then, since $\mathfrak{o}$ is a principal ideal domain, M can be written as a sum of cyclic $\mathfrak{o}$-submodules:

$$M \cong \mathfrak{o}/\mathfrak{p}^{\lambda_1} \oplus \cdots \oplus \mathfrak{o}/\mathfrak{p}^{\lambda_l},$$

with $\lambda = (\lambda_1, \cdots, \lambda_l)$ being some partition. λ is called the *type* of M. Let $\mathcal{L}(M)$ denote the set of $\mathfrak{o}$-submodules of M. $\mathcal{L}(M)$ inherits a lattice structure by inclusion relation. Our main objective is to compute $\operatorname{Aut}\mathcal{L}(M)$, the automorphism group of the lattice $\mathcal{L}(M)$, for λ with $\lambda_1 = \lambda_2 \geq \lambda_3$.

When $\mathfrak{o} = \mathbb{Z}_p$, the ring of p-adic integers, M becomes nothing but a finite abelian p-group and $\mathcal{L}(M)$ the subgroup lattice of M. This can be generalized by considering the case $\mathfrak{o} = W[\mathbb{F}_q]$, the ring of Witt vectors over the finite field $\mathbb{F}_q$, for $W[\mathbb{F}_p] \cong \mathbb{Z}_p$.

We call $e = (e_1, \cdots, e_l) \in M^l$ an *ordered basis* for M if $M = \bigoplus_{i=1}^{l} \mathfrak{o}e_i$ and $\mathfrak{o}e_i \cong \mathfrak{o}/\mathfrak{p}^{\lambda_i}$. Let e be fixed, and let us define a subset $\mathcal{F}(e)$ of $\mathcal{L}(M)$ by

$$\mathcal{F}(e) = \{\, \mathfrak{o}e_1, \ldots, \mathfrak{o}e_l \,\} \cup \{\, \mathfrak{o}(e_1 + e_2), \ldots, \mathfrak{o}(e_1 + e_l) \,\}.$$

Let $R(e)$ denote the element-wise stabilizer of $\mathcal{F}(e)$ in $\operatorname{Aut}\mathcal{L}(M)$. In most cases it boils down to computing $R(e)$ in order to analyze $\operatorname{Aut}\mathcal{L}(M)$, in the sense of the following.

Since an autormophism of $\mathfrak{o}$-module M induces an automorphism of the lattice $\mathcal{L}(M)$, we have the natural group homomorphism

$$\xi : \operatorname{Aut} M \to \operatorname{Aut}\mathcal{L}(M).$$

It can be shown that $\operatorname{Ker}\xi \cong (\mathfrak{o}/\mathfrak{p}^{\lambda_1})^{\times}$ and that $\operatorname{Aut} M$ can be expressed in matrix form, as mentioned at the end of this section. Naturally $\operatorname{Aut}\mathcal{L}(M)$ contains a subgroup isomorphic to $\operatorname{Aut} M/\operatorname{Ker}\xi$, and we let $\operatorname{PAut} M$ denote this subgroup.

These two subgroups $R(e)$ and $\mathrm{PAut}\,M$ are closely related; the subgroup $R(e)$ is a permutable complement of the subgroup $\mathrm{PAut}\,M$ in $\mathrm{Aut}\,\mathcal{L}(M)$. Namely, we have

$$R(e)\cdot \mathrm{PAut}\,M = \mathrm{Aut}\,\mathcal{L}(M)$$
$$R(e)\cap \mathrm{PAut}\,M = 1.$$

Also, we remark that if e and e' are ordered base for M, then it is easily checked that $\varphi R(e)\varphi^{-1} = R(e')$, where $\varphi \in \mathrm{PAut}\,M$ is the lattice automorphism induced by the module automorphism of M defined by $e_i \mapsto e'_i$ $(1 \le i \le l)$. Hence the isomorphism type of $R(e)$ does not depend on the choice of e. We content ourselves with computing $R(e)$ instead of computing $\mathrm{Aut}\,\mathcal{L}(M)$ for our purpose.

Let us mention the relation with earlier results. We consider the case when the residue field of $\mathfrak{o}$ is the finite field $\mathbb{F}_p$. Let $M = \mathfrak{o}/\mathfrak{p}\oplus\mathfrak{o}/\mathfrak{p} \cong \mathbb{F}_p\oplus\mathbb{F}_p$. Then $\mathrm{Aut}\,\mathcal{L}(M)$ is isomorphic to the symmetric group $\mathfrak{S}_{p+1}$ and $\mathrm{PAut}\,M$ isomorphic to the projective general linear group $PGL(2,p)$ (Note that $|PGL(2,p)| = (p+1)p(p-1)$). In this case, $R(e)$ is a subgroup that fixes three points and isomorphic to $\mathfrak{S}_{p-2}$.

More generally, for $M = \mathbb{Z}_p/p^{\lambda_2}\mathbb{Z}_p \oplus \mathbb{Z}_p/p^{\lambda_2}\mathbb{Z}_p$ ($\mathfrak{o} = \mathbb{Z}_p$ is the ring of p-adic integers), Holmes' result [5] states that $\mathrm{Aut}\,\mathcal{L}(M)$ is isormorphic to $\mathfrak{S}_p^{\wr(\lambda_2-1)} \wr \mathfrak{S}_{p+1}$, where $\mathfrak{S}_p^{\wr n}$ means $\mathfrak{S}_p \wr\cdots\wr \mathfrak{S}_p$ (n times). In this case, $\mathrm{PAut}\,M$ is nothing but $PGL_2(\mathbb{Z}_p/p^{\lambda_2}\mathbb{Z}_p)$, and we note that $|PGL_2(\mathbb{Z}_p/p^{\lambda_2}\mathbb{Z}_p)| = (p+1)p(p-1)\cdot(p^{\lambda_2-1})^3$. $R(e)$ is the subgroup that fixes three points $\mathbb{Z}_p(1,0)$, $\mathbb{Z}_p(0,1)$ and $\mathbb{Z}_p(1,1)$; it is indeed isomorphic to

$$(\mathfrak{S}_p^{\wr(\lambda_2-1)} \wr \mathfrak{S}_{p-2}) \times \left\{\prod_{i=0}^{\lambda_2-2} (\mathfrak{S}_p^{\wr i} \wr \mathfrak{S}_{p-1})\right\}^3 .$$

Another example is the case when $\lambda_1 = \lambda_2 = \lambda_3$, which is essentially the well-known result of Baer [2]. In this case, $R(e)$ becomes isomorphic to the group $\mathrm{Aut}\,\overline{\mathfrak{o}}$ of ring automorphisms, where $\overline{\mathfrak{o}} = \mathfrak{o}/\mathfrak{p}^{\lambda_2}$. More specifically, we have

$$\mathrm{Aut}\,\overline{\mathfrak{o}} \ltimes \mathrm{PAut}\,M \cong \mathrm{Aut}\,\mathcal{L}(M).$$

In particular, when $\lambda_1 = \cdots = \lambda_l = 1$ $(l \ge 3)$, M becomes a vector space over the residue filed k of $\mathfrak{o}$, and $\mathrm{Aut}\,\mathcal{L}(M)$ is isomorphic to $P\varGamma L(l,k)$, the group of projective semi-linear automorphisms. This result is a variation of so called *the Fundamental Theorem of Finite Projective Geometry.*

There have been works to bridge the gap between Holmes' result and Baer's. Costantini-Holmes-Zacher[3] and Costantini-Zacher[4] treated the case of abelian groups in a rather general framework. Yasuda[11] studied the case of abelian groups for $\lambda_1 > \lambda_2 = \lambda_3$ with explicit computation of $R(e)$ and $\mathrm{Aut}\,\mathcal{L}(M)$. In this work, we shall treat the case $\lambda_1 = \lambda_2 \ge \lambda_3$, in the general setting of finite-length modules over complete valuation ring.

We end this section with the description of the automorphism group $\mathrm{Aut}\,M$ of an $\mathfrak{o}$-module M. Let e be fixed. The action of $f \in \mathrm{Aut}\,M$ is then determined by its action on $e = (e_1,\cdots,e_l)$. Write

$$f(e_j) = \sum_{i=1}^{l} a_{ij}e_i$$

and express f as the matrix $(a_{ij})_{i,j=1}^{l}$. Rewriting $\lambda = (\lambda_1,\ldots,\lambda_l) = \langle {d_1}^{m_1},\ldots,{d_r}^{m_r}\rangle$ $(d_1 > \cdots > d_r)$, $\mathrm{Aut}\,M$ can be expressed in matrix form as

$$\begin{pmatrix} GL_{m_1}(\mathfrak{o}/\mathfrak{p}^{d_1}) & \cdots & \mathrm{Hom}((\mathfrak{o}/\mathfrak{p}^{d_r})^{\oplus m_r}, (\mathfrak{o}/\mathfrak{p}^{d_1})^{\oplus m_1}) \\ \vdots & \ddots & \vdots \\ \mathrm{Hom}((\mathfrak{o}/\mathfrak{p}^{d_1})^{\oplus m_1}, (\mathfrak{o}/\mathfrak{p}^{d_r})^{\oplus m_r}) & \cdots & GL_{m_r}(\mathfrak{o}/\mathfrak{p}^{d_r}) \end{pmatrix},$$

with respect to the ordered basis e. Here, the block matrix in the diagonal

$$A \in GL_{m_i}(\mathfrak{o}/\mathfrak{p}^{d_i})$$

is of size $m_i \times m_i$ and has elements of $\mathfrak{o}/\mathfrak{p}^{d_i}$ in its components, satisfying $\pi \nmid \det A$. Also, the block matrix at (i,j)-position $(i \neq j)$

$$A \in \mathrm{Hom}((\mathfrak{o}/\mathfrak{p}^{d_j})^{\oplus m_j}, (\mathfrak{o}/\mathfrak{p}^{d_i})^{\oplus m_i})$$

is of size $m_i \times m_j$ and in its components has elements of $\mathfrak{p}^{d_i - \min(d_j, d_i)}(\mathfrak{o}/\mathfrak{p}^{d_i})$, that is, for $i < j$ ($\Longrightarrow$ $d_i > d_j$) elements of $\mathfrak{p}^{d_i - d_j}(\mathfrak{o}/\mathfrak{p}^{d_i})$, and for $i > j$ ($\Longrightarrow$ $d_i < d_j$) elements of $\mathfrak{o}/\mathfrak{p}^{d_i}$.

2. Results

2.1. $\mathfrak{u}$-crossed automorphisms.

Lemma 1. *For $\varphi \in R(e)$, there exists a unique bijective map $\sigma : \overline{\mathfrak{o}} \to \overline{\mathfrak{o}}$ satisfying*

$$\varphi(\mathfrak{o}(e_1 + ae_2)) = \mathfrak{o}(e_1 + \sigma(a)e_2).$$

Note that since $\lambda_1 = \lambda_2$, $(e_2, e_1, e_3, \cdots, e_l)$ is also an ordered basis for M. Then with this lemma we also see that there exists a bijection $\tau \in \mathfrak{S}(\overline{\mathfrak{o}})$ such that $\varphi(\mathfrak{o}(ae_1 + e_2)) = \mathfrak{o}(\tau(a)e_1 + e_2)$ for all $a \in \overline{\mathfrak{o}}$. Thus we have obtained the map $R(e) \to \mathfrak{S}(\overline{\mathfrak{o}}) \times \mathfrak{S}(\overline{\mathfrak{o}})$ ($\varphi \mapsto (\tau, \sigma)$).

Put $\overline{\mathfrak{p}} = \mathfrak{p}\overline{\mathfrak{o}} \subset \overline{\mathfrak{o}}$. Let $\mathfrak{u}$ denote the kernel of the natural group homomorphism $\overline{\mathfrak{o}}^\times \to (\mathfrak{o}/\mathfrak{p}^{\lambda_3})^\times$. In other words,

$$\mathfrak{u} = 1 + \overline{\mathfrak{p}}^{\lambda_3} \subset \overline{\mathfrak{o}}^\times.$$

For $a, b \in \overline{\mathfrak{o}}$, we say that a and b are $\mathfrak{u}$-*similar* and write

$$a \overset{\mathfrak{u}}{\sim} b$$

if $a \in \mathfrak{u}b$. It is easily seen that $\overset{\mathfrak{u}}{\sim}$ defines an equivalence relation on $\overline{\mathfrak{o}}$. We call the bijection $\sigma \in \mathfrak{S}(\overline{\mathfrak{o}})$ a $\mathfrak{u}$-*crossed automorphism* of $\overline{\mathfrak{o}}$ if σ satisfies the following three conditions:

(1) $\sigma(\overline{\mathfrak{p}}) \subset \overline{\mathfrak{p}}$,
(2) $\sigma(a - b) \overset{\mathfrak{u}}{\sim} \sigma(a) - \sigma(b)$ for all $a, b \in \overline{\mathfrak{o}}$, and
(3) $\sigma(ab) \overset{\mathfrak{u}}{\sim} \sigma(a)\sigma(b)$ for all $a, b \in \overline{\mathfrak{o}}$.

Clearly a ring automorphism $\sigma : \overline{\mathfrak{o}} \to \overline{\mathfrak{o}}$ is a $\mathfrak{u}$-crossed automorphism.

Lemma 2. *Let $\varphi \in R(e)$ and $\sigma \in \mathfrak{S}(\overline{\mathfrak{o}})$ the derived map from φ as in Lemma 1. Then σ is a $\mathfrak{u}$-crossed automorphism.*

We list some basic properties of $\mathfrak{u}$-crossed automorphisms.

Lemma 3. *Let $\sigma : \overline{\mathfrak{o}} \to \overline{\mathfrak{o}}$ be a $\mathfrak{u}$-crossed automorphism. Then*

(1) *$\sigma(\overline{\mathfrak{p}}^i) = \overline{\mathfrak{p}}^i$ for all $i \geq 0$. That is, we have $v(\sigma(a)) = v(a)$ for all $a \in \overline{\mathfrak{o}}$.*
(2) *For all $a, b \in \overline{\mathfrak{o}}$ and $i \geq 0$, we have $a \equiv b \mod \overline{\mathfrak{p}}^i$ if and only if $\sigma(a) \equiv \sigma(b) \mod \overline{\mathfrak{p}}^i$. That is, $v(a - b) = v(\sigma(a) - \sigma(b))$ for all $a, b \in \overline{\mathfrak{o}}$.*
(3) *$\sigma(1) \in \mathfrak{u}$. That is, $\sigma(1) \overset{\mathfrak{u}}{\sim} 1$.*
(4) *For all $a, b \in \overline{\mathfrak{o}}$, we have $a \overset{\mathfrak{u}}{\sim} b$ if and only if $\sigma(a) \overset{\mathfrak{u}}{\sim} \sigma(b)$. That is, $\sigma(\mathfrak{u}a) = \mathfrak{u}\sigma(a)$ for all $a \in \overline{\mathfrak{o}}$.*

With this lemma, we see that modulo $\mathfrak{p}^i$ ($i \leq \lambda_2$) reduction of σ induces a $\mathfrak{u}$-crossed automorphism $\mathfrak{o}/\mathfrak{p}^i \to \mathfrak{o}/\mathfrak{p}^i$. In particular, modulo $\mathfrak{p}^{\lambda_3}$ reduction induces a ring automorphism $\mathfrak{o}/\mathfrak{p}^{\lambda_3} \to \mathfrak{o}/\mathfrak{p}^{\lambda_3}$.

Let $\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}$ denote the set of $\mathfrak{u}$-crossed automorphisms of $\overline{\mathfrak{o}}$. Then

Lemma 4. $\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}$ *forms a subgroup of* $\mathfrak{S}(\overline{\mathfrak{o}})$.

Since a ring automorphism is a $\mathfrak{u}$-crossed automorphism, $\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}$ contains the subgroup $\mathrm{Aut}\,\overline{\mathfrak{o}}$ of ring automorphisms. Also, modulo $\mathfrak{p}^{\lambda_3}$ reduction gives us the group homomorphism $\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}} \to \mathrm{Aut}\,\mathfrak{o}/\mathfrak{p}^{\lambda_3}$.

Thus far we have obtained the group homomorphism $R(e) \to \mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}} \times \mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}$ ($\varphi \mapsto (\tau,\sigma)$). Next we see that the $\mathfrak{u}$-crossed automorphisms τ and σ derived from φ are related as in:

Lemma 5. *Let* $\varphi \in R(e)$ *and* $(\tau,\sigma) \in (\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}})^2$ *be derived from* φ. *Then*

(1) $\tau(a)^{-1} = \sigma(a^{-1})$ *for all* $a \in \overline{\mathfrak{o}} \setminus \overline{\mathfrak{p}}$,

(2) $\tau(a) \overset{\mathfrak{u}}{\sim} \sigma(a)$ *for all* $a \in \overline{\mathfrak{p}}$.

Now let $\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}_1$ denote the stabilizer of 1, i.e., $\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}_1 = \{\sigma \in \mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}} \mid \sigma(1) = 1\}$. Also, define $\tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}_1)^2$ to be the set of $(\tau,\sigma) \in (\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}_1)^2$ with the properties of Lemma 5. Then $\tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}_1)^2$ forms a subgroup of $(\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}_1)^2$. So we have the group homomorphism $R(e) \to \tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}_1)^2$.

Let $\sigma \in \mathrm{Aut}\,\overline{\mathfrak{o}}$. Then $\varphi : \mathcal{L}(M) \to \mathcal{L}(M)$ defined by $\varphi(\mathfrak{o}(\sum_{i=1}^{l} a_i e_i)) = \mathfrak{o}(\sum_{i=1}^{l} \sigma(a_i) e_i)$ is an element of $R(e)$. Thus $R(e)$ contains a subgroup isomorphic to $\mathrm{Aut}\,\overline{\mathfrak{o}}$. Also, $\tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}_1)^2$ contains the subgroup $\Delta(\mathrm{Aut}\,\overline{\mathfrak{o}})^2 = \{(\sigma,\sigma) \in \mathrm{Aut}\,\overline{\mathfrak{o}} \times \mathrm{Aut}\,\overline{\mathfrak{o}}\}$. The diagonal map $\Delta : \mathrm{Aut}\,\overline{\mathfrak{o}} \to \Delta(\mathrm{Aut}\,\overline{\mathfrak{o}})^2$ is compatible with the homomorphism $R(e) \to \tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}_1)^2$.

2.2. **Main isomorphism theorem.** We now state our main result:

Theorem 1 (Main isomorphism theorem)**.** *We have the isomorphism of groups*

$$R(e) \cong \tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}_1)^2.$$

One direction of the isomorphism

$$\eta : R(e) \to \tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}_1)^2$$

is already described in the previous subsection. To construct the isomorphism in the other direction

$$\zeta : \tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}_1)^2 \to R(e),$$

we need to divide the map in two stages. Since every bijective map $f \in \mathfrak{S}(M)$ induces a bijective map $\varphi : 2^M \to 2^M$, we can define $\mathfrak{S}_{R(e)}(M)$ to be the set of bijections $f \in \mathfrak{S}(M)$ such that the induced map φ satisfies $\varphi(\mathcal{L}(M)) \subset \mathcal{L}(M)$ and the lattice automorphism $\varphi\,|_{\mathcal{L}(M)} : \mathcal{L}(M) \to \mathcal{L}(M)$ is an element of $R(e)$. Then $\mathfrak{S}_{R(e)}(M)$ is a subgroup of $\mathfrak{S}(M)$, and we obtain the group homomorphism

$$\iota : \mathfrak{S}_{R(e)}(M) \to R(e).$$

The other half of the map

$$\zeta' : \tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}}\,\overline{\mathfrak{o}}_1)^2 \to \mathfrak{S}_{R(e)}(M)$$

is constructed as follows. We partition M into two disjoint sets, by defining

$$M_e^{\leq} = \left\{ \sum_{i=1}^{l} a_i e_i \in M \;;\; v(a_1) \leq v(a_2) \right\}, \text{ and}$$

$$M_e^{>} = \left\{ \sum_{i=1}^{l} a_i e_i \in M \;;\; v(a_1) > v(a_2) \right\},$$

so that $M = M_e^{\leq} \sqcup M_e^{>}$. Given $(\tau, \sigma) \in \tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}} \bar{\mathfrak{o}}_1)^2$, we shall define the map $f \in \mathfrak{S}_{R(e)}(M)$ on $M_e^{\leq}$ and on $M_e^{>}$ separately. For $a \in M_e^{\leq}$, write $a = a_1 e_1 + a_1 \tilde{a}_2 e_2 + \sum_{i=3}^{l} a_i e_i$ and define

$$f(a) = \tau(a_1)e_1 + \tau(a_1)\sigma(\tilde{a}_2)e_2 + \sum_{i=3}^{l} \tau(a_i)e_i.$$

Similarly, for $a \in M_e^{>}$, write $a = \tilde{a}_1 a_2 e_1 + a_2 e_2 + \sum_{i=3}^{l} a_i e_i$ and define

$$f(a) = \tau(\tilde{a}_1)\sigma(a_2)e_1 + \sigma(a_2)e_2 + \sum_{i=3}^{l} \sigma(a_i)e_i.$$

Lemma 6. *The map $f : M \to M$ above is well-defined and bijective.*

Thus we have the group homomorphism $\tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}} \bar{\mathfrak{o}}_1)^2 \to \mathfrak{S}(M)$. To prove that the map $\varphi : 2^M \to 2^M$ induced by f satisfies $\varphi(\mathcal{L}(M)) \subset \mathcal{L}(M)$ and that $\varphi |_{\mathcal{L}(M)}: \mathcal{L}(M) \to \mathcal{L}(M)$ is a lattice automorphism, we make a use of the following lemma, which is a slightly modified variation of a well-known principle in the theory of subgroup lattices (cf. [7]).

Lemma 7. *Let R be a unitary commutative ring, M an R-module, and G a subgroup of the group $\mathfrak{S}(M)$ of all bijections on M. If every $f \in G$ satisfies*

$$f(Ra) \subset Rf(a)$$
$$f(a+b) \in Rf(a) + Rf(b)$$

for all $a, b \in M$, then every $f \in G$ induces a lattice automorphism on $\mathcal{L}(M)$.

Also, we have

Lemma 8. *The map $f : M \to M$ defined by $(\tau, \sigma) \in \tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}} \bar{\mathfrak{o}}_1)^2$ satisfies the conditions of Lemma 7 and hence induces a lattice automorphism $\varphi : \mathcal{L}(M) \to \mathcal{L}(M)$. Moreover, thus induced φ is indeed an element of $R(e)$.*

Hence we have obtained the group homomorphism $\zeta' : \tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}} \bar{\mathfrak{o}}_1)^2 \to \mathfrak{S}_{R(e)}(M)$ and the series of group homomorphisms

$$\iota \circ \zeta' : \tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}} \bar{\mathfrak{o}}_1)^2 \to \mathfrak{S}_{R(e)}(M) \to R(e),$$

which by letting $\zeta = \iota \circ \zeta'$ gives us the desired isomorphism $\zeta : \tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}} \bar{\mathfrak{o}}_1)^2 \to R(e)$.

To show that η and ζ are indeed isomorphisms, we prove that $\eta \circ \zeta = 1$ and $\zeta \circ \eta = 1$. The fact $\eta \circ \zeta = 1$ is almost clear from the constructions. To see $\zeta \circ \eta = 1$, we use the

Lemma 9. *Let $\varphi \in R(e)$ and $(\tau, \sigma) \in \tilde{\Delta}(\mathrm{Aut}_{\mathfrak{u}} \bar{\mathfrak{o}}_1)^2$ be derived from φ. Then we have*

(1) $\varphi(\mathfrak{o}(e_1 + a_2 e_2 + a_3 e_3 + \cdots + a_l e_l)) = \mathfrak{o}(e_1 + \sigma(a_2)e_2 + \sigma(a_3)e_3 + \cdots + \sigma(a_l)e_l)$.
(2) $\varphi(\mathfrak{o}(a_2 e_2 + a_3 e_3 + \cdots + a_l e_l)) = \mathfrak{o}(\sigma(a_2)e_2 + \sigma(a_3)e_3 + \cdots + \sigma(a_l)e_l)$.

Using this lemma, we can show that the action of $\zeta \circ \eta(\varphi)$ on $\mathcal{L}(M)$ is actually the same as that of φ, thus showing $\zeta \circ \eta = 1$.

2.3. **The case of the ring of Witt vectors over $\mathbb{F}_q$.** In this section we apply our main result to the case $\mathfrak{o} = W[\mathbb{F}_q]$, the ring of Witt vectors over the finite field $\mathbb{F}_q$. Let $q = p^r$. Recall that we have the group homomorphism

$$\rho : \mathrm{Aut}_{\mathfrak{u}} W[\mathbb{F}_q]/(p)^{\lambda_2}{}_1 \to \mathrm{Aut}\, W[\mathbb{F}_q]/(p)^{\lambda_3}.$$

Since we have $\mathrm{Aut}\, W[\mathbb{F}_q]/(p)^{\lambda_3} \cong \mathbb{Z}/r\mathbb{Z} \cong \mathrm{Aut}\, W[\mathbb{F}_q]/(p)^{\lambda_2}$, the exact sequence

$$1 \to \mathrm{Ker}\, \rho \to \mathrm{Aut}_{\mathfrak{u}} W[\mathbb{F}_q]/(p)^{\lambda_2}{}_1 \to \mathrm{Aut}\, W[\mathbb{F}_q]/(p)^{\lambda_3} \to 1$$

splits, and we have

$$\mathrm{Aut_u}\, W[\mathbb{F}_q]/(p)^{\lambda_2}{}_1 \cong \mathbb{Z}/r\mathbb{Z} \ltimes \operatorname{Ker}\rho.$$

So it remains to compute $\operatorname{Ker}\rho$. When $\lambda_1(=\lambda_2) \le 2\lambda_3 + 1 \ge 3$, $\operatorname{Ker}\rho$ becomes a (finite) abelian p-group and indeed can be calculated explicitly. When $\lambda_1 \ge 2\lambda_3$, $\operatorname{Ker}\rho$ is not necessarily abelian, and explicit computation of $\operatorname{Ker}\rho$ (and of $R(e)$) becomes more complicated. The author has obtained two different results in this case, using different approach, and is currently in the stage of determining which one should be the correct result.

References

[1] Baer, R. "The significance of the system of subgroups for the structure of the group." *Amer. J. Math.* **61** (1939), 1-44.

[2] Baer, R. "A unified theory of projective spaces and finite Abelian groups." *Trans. Amer. Math. Soc.* **52** (1942), 282-343.

[3] Costantini, M., Holmes, C. S., and Zacher, G. "A representation theorem for the group of autoprojectivities of an abelian p-Group of finite exponent." *Ann. Mat. Pura. Appl. Ser. 4.* **175** (1998), 119-140.

[4] Costantini, M., and Zacher, G. "On the group of autoprojectivities of periodic modular groups." *J. Group Theory.* no. 4 **1** (1998), 369-394.

[5] Holmes, C. "Automorphisms of the lattice of subgroups of $\mathbb{Z}_{p^m} \times \mathbb{Z}_{p^n}$." *Arch. Math.* **51** (1988), 491-495.

[6] Macdonald, I.G. *Symmetric Functions and Hall polynomials, 2nd. ed..* Oxford: Clarendon Press, 1995.

[7] Schmidt, R. *Subgroup Lattices of Groups.* Berlin: Walter de Gruyter, 1994.

[8] Serre, J.-P. *Local Fields.* New York - Berlin: Springer-Verlag, 1979.

[9] Suzuki, M. *Structure of a Group and the Structure of its Lattice of Subgroups.* Berlin: Springer-Verlag, 1967.

[10] Vogt, F. "Subgroup lattices of finite Abelian groups: structure and cardinality." K. A. Baker, R. Wille eds.. *Lattice Theory and its Applications.* Heldermann Verlag, 1995; 241-259.

[11] Yasuda, K. *On the Automorphism Group of the Subgroup Lattice of a Finite Abelian p-Group.* Master's thesis, Univeristy of Tokyo, 2000.

Graduate School of Mathematical Sciences, University of Tokyo, Tokyo, Japan

E-mail address: `kan@msf.biglobe.ne.jp`

DEGREE BOUNDS IN QUANTUM SCHUBERT CALCULUS

ALEXANDER YONG

Abstract. Fulton and Woodward [8] have recently identified the smallest degree of q that appears in the expansion of the product of two Schubert classes in the (small) quantum cohomology ring of a Grassmannian. We present a combinatorial proof of this result, and provide an alternative characterization of this smallest degree in terms of the rim hook formula [2] for the quantum product.

Résumé. Fulton et Woodward ont récenemment identifié le degré minimum de q dans l'expansion du produit de deux classes de Schubert dans le (petit) anneau de cohomologie quantique d'une grassmannienne. Nous présentons une démonstration combinatoire de ce résultat et une characterization du dégrée minimal, en terme de rubans, dans la formule Bertram, Ciocan-Fontanine et Fulton pour le produit quantique.

1. Introduction and Main Results

Let $X = Gr(l, \mathbb{C}^n)$ be the Grassmannian of l-dimensional subspaces in $\mathbb{C}^n$. The classical cohomology ring $\mathrm{H}^*(X, \mathbb{Z})$ has an additive basis of *Schubert classes* $\{\sigma_\lambda\}$, indexed by the Young diagrams λ (identified with the corresponding partitions) contained in the $l \times k$ rectangle, where $k = n - l$ (we denote this by $\lambda \subseteq l \times k$). The product of two Schubert classes in $\mathrm{H}^*(X, \mathbb{Z})$ is given by

$$\sigma_\lambda \cdot \sigma_\mu = \sum_{\nu \subseteq l \times k} c^{\nu}_{\lambda,\mu} \sigma_\nu, \tag{1}$$

where $c^{\nu}_{\lambda,\mu}$ is the Littlewood-Richardson coefficient (see, e.g., [6, 14]).

The (small) quantum cohomology ring $\mathrm{QH}^*(X)$ is a certain deformation of $\mathrm{H}^*(X, \mathbb{Z})$ that has been extensively studied in recent years; see, e.g., [1, 7, 13] and references therein. This ring is canonically isomorphic to the Verlinde algebra of $\mathfrak{sl}_l$ at level k (see, e.g., [2, 15]); consequently, the results presented here can be reformulated in representation theoretic language.

The additive structure of $\mathrm{QH}^*(X)$ is essentially the same as that of $\mathrm{H}^*(X, \mathbb{Z})$: the Schubert classes form a basis of $\mathrm{QH}^*(X)$ as a free module over $\mathbb{Z}[q]$, where q is an indeterminate. The multiplicative structure of $\mathrm{QH}^*(X)$ is defined by

$$\sigma_\lambda * \sigma_\mu = \sum_{\nu \subseteq l \times k} \sum_{d \geq 0} q^d \langle \lambda, \mu, \nu^\vee \rangle_d \sigma_\nu, \tag{2}$$

where $\nu^\vee = (k-\nu_l, \ldots, k-\nu_1)$ is the complement of ν in the $l \times k$ rectangle, and $\langle \lambda, \mu, \nu^\vee \rangle_d$ is a three-point, genus-zero Gromov-Witten invariant of X. Note that we use "$*$" to distinguish the quantum multiplication from the product in the classical cohomology ring. Setting $q = 0$ recovers (1) because $c^{\nu}_{\lambda,\mu} = \langle \lambda, \mu, \nu^\vee \rangle_0$.

1991 *Mathematics Subject Classification.* Primary 14M15; Secondary 05E05, 14N10.

Key words and phrases. Gromov-Witten invariants, quantum cohomology, Grassmannian, Schubert calculus.

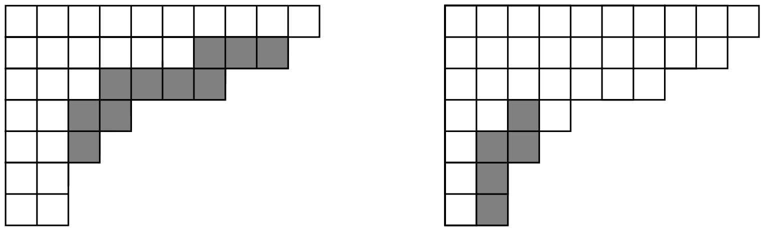


FIGURE 1. Legal and illegal rim hooks for $\lambda = (10, 9, 7, 4, 3, 2, 2)$

Bertram, Ciocan-Fontanine and Fulton [2] have given a combinatorial rule to compute $\sigma_\lambda * \sigma_\mu$ and thus the Gromov-Witten invariants $\langle \lambda, \mu, \nu^\vee \rangle_d$. To describe this rule, we need some terminology and notation. An *n-rim hook* of a Young diagram λ is a connected subset of n boxes of λ that does not contain a 2×2 square. The *width* of an n-rim hook is the number of columns it occupies. An n-rim hook is *legal* if removing it from a Young diagram gives a valid Young diagram (see Figure 1). For a partition ρ, we define its n-core, denoted $\mathrm{core}_n(\rho)$, to be the partition corresponding to the Young diagram obtained by repeatedly removing legal n-rim hooks from ρ until further removals are not possible. It is well known (see, e.g., [9]) that this procedure defines $\mathrm{core}_n(\rho)$ uniquely. Let $r_n(\rho) = \frac{|\rho| - |\mathrm{core}_n(\rho)|}{n}$ be the number of n-rim hooks removed in this process, and set $\epsilon(\rho) = (-1)^{\sum (k - width(R_i))}$, where $R_1, \ldots, R_{r_n(\rho)}$ are these n-rim hooks. With this notation, the rule obtained in [2] is as follows:

$$\sigma_\lambda * \sigma_\mu = \sum_{\nu \subseteq l \times k} \sum_{\substack{c^\rho_{\lambda,\mu} \neq 0 \\ \rho_1 \leq k \\ \mathrm{core}_n(\rho) = \nu}} q^{r_n(\rho)} \epsilon(\rho) c^\rho_{\lambda,\mu} \sigma_\nu. \tag{3}$$

The Gromov-Witten invariants $\langle \lambda, \mu, \nu^\vee \rangle$ are known to be nonnegative; thus the coefficient of each term $q^d \sigma_\nu$ in the right-hand side of (3) is nonnegative. It remains an open problem to find a direct combinatorial proof of this statement; such a proof would provide a generalization of the Littlewood-Richardson rule to the quantum Schubert calculus of the Grassmannian.

Let $\mathrm{d}_{\min}$ denote the smallest degree of q such that $q^{\mathrm{d}_{\min}}$ appears in (2) with nonzero coefficient. The following theorem provides an affirmative answer to a question posed by A. Ram (private communication).

Theorem 1.1. *Let $\lambda, \mu \subseteq l \times k$. Among all ρ with $\rho_1 \leq k$ and $c^{\rho}_{\lambda,\mu} \neq 0$, pick those with the largest n-core (equivalently, the smallest value of $r_n(\rho)$). Then their contributions to* (3) *do not cancel each other out. In other words,* $\mathrm{d}_{\min}$ *equals the smallest degree of q that appears in the right-hand side of* (3) *(before cancellations).*

Our proof of Theorem 1.1 will also prove a recent result of Fulton and Woodward (see Theorem 1.2 below). For partitions $\lambda, \mu \subseteq l \times k$, place λ against the upper left corner of the rectangle. Then rotate μ 180 degrees and place it in the lower right corner (see Figure 2). We will refer to rotate(μ) as the resulting subshape of $l \times k$. Let $\mathfrak{d}$ be the side length of the largest square that fits inside $\lambda \cap \text{rotate}(\mu)$. The following theorem was conjectured by Fulton, and later proved by Fulton and Woodward [8], using moduli spaces.[1]

Theorem 1.2. [8] $\mathrm{d}_{\min} = \mathfrak{d}$.

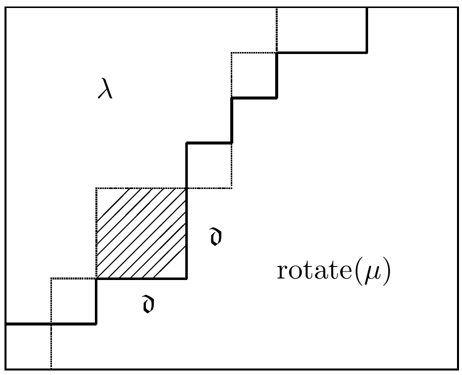


FIGURE 2. λ, rotate(μ) and $\mathfrak{d} \times \mathfrak{d}$

An alternative proof of Theorem 1.2 was later given by A. Buch [3], using an elegant geometric argument combined with combinatorics. We will utilize the combinatorial part of Buch's proof below.

Let us now describe our proof of Theorems 1.1 and 1.2. The proof is entirely combinatorial once the nonnegativity of the Gromov-Witten invariants and the fact that $\mathrm{QH}^*(X)$ is an associative ring [10, 12] are granted. Our argument is based on the following result.

Theorem 1.3. *Let $a \times A$, $b \times B$ and $c \times C$ be rectangular Young diagrams contained in $l \times k$. Then the following are equivalent:*

(I) *there exist Young diagrams $\lambda, \mu, \nu \subseteq l \times k$ containing $a \times A$, $b \times B$ and $c \times C$, respectively, such that $\sigma_\lambda \cdot \sigma_\mu \cdot \sigma_\nu \neq 0$;*
(II) $\sigma_{a\times A} \cdot \sigma_{b\times B} \cdot \sigma_{c\times C} \neq 0$*;*
(III) *all of the five conditions below hold:*
  (i) $a + b \leq l$ *or* $A + B \leq k$*;*
  (ii) $a + c \leq l$ *or* $A + C \leq k$*;*
  (iii) $b + c \leq l$ *or* $B + C \leq k$*;*
  (iv) $a + b + c \leq l$ *or* $A + B + C \leq 2k$*;*
  (v) $a + b + c \leq 2l$ *or* $A + B + C \leq k$.

[1]Actually, Fulton and Woodward proved a more general result that applies to any homogeneous space $X = G/P$, where G is a simply connected complex semisimple Lie group and P its parabolic subgroup.

A combinatorial proof of Theorem 1.3 is given in Section 2. S. Fomin (private communication) has observed that the necessity of the conditions (i)-(v) can be derived from the Horn inequalities describing the "Klyachko cone" $\{(\lambda, \mu, \nu) \mid c_{\lambda,\mu}^{\nu} \neq 0\}$; specifically see [5, (11)].

Corollary 1.4. *Let λ, μ, ρ be partitions such that $\lambda, \mu \subseteq l \times k$ and $\rho_1 \leq k$. If $c_{\lambda,\mu}^{\rho} \neq 0$, then $(l + \mathfrak{d}) \times \mathfrak{d} \subseteq \rho$.*

Proof. Suppose $c_{\lambda,\mu}^{\rho} \neq 0$ but $(l + \mathfrak{d}) \times \mathfrak{d} \not\subseteq \rho$. Choose any positive integer $L \geq l$ such that $\rho \subseteq L \times k$. Then we have $\sigma_\lambda \cdot \sigma_\mu \cdot \sigma_{\rho^\vee} \neq 0$. Notice that $(L - (l + \mathfrak{d}) + 1) \times (k - \mathfrak{d} + 1) \subseteq \rho^\vee$. By the definition of $\mathfrak{d}$, there exist rectangles $a \times A \subseteq \lambda$ and $b \times B \subseteq \mu$ such that $a + b \geq l + \mathfrak{d}$ and $A + B \geq k + \mathfrak{d}$. Set $\nu = \rho^\vee, c = L - (l + \mathfrak{d}) + 1$ and $C = k - \mathfrak{d} + 1$, then it is easy to check that both corresponding inequalities (iv) are violated, a contradiction of Theorem 1.3 (see Figure 3 below). $\square$

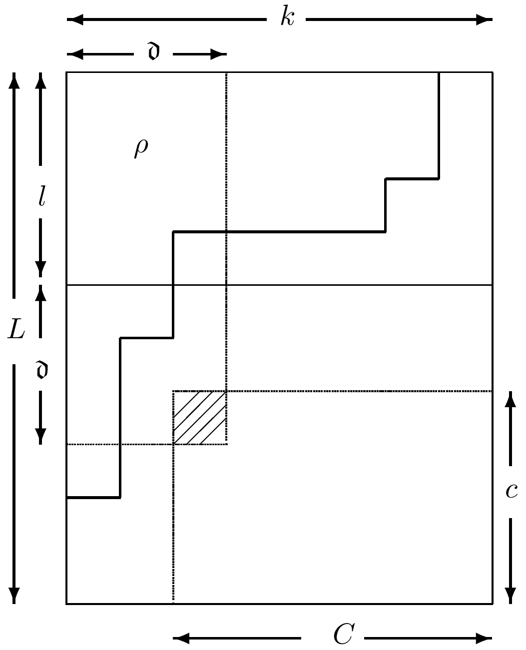


FIGURE 3. Proof of Corollary 1.4.

Proof of Theorems 1.1 and 1.2. Let ρ be a partition whose first part is at most k and whose n-core is contained in $l \times k$. If $c_{\lambda,\mu}^{\rho} \neq 0$, then Corollary 1.4 gives $(l + \mathfrak{d}) \times \mathfrak{d} \subseteq \rho$, implying that at least $\mathfrak{d}$ rim hook removals are needed to obtain $\mathrm{core}_n(\rho)$. Hence the smallest degree of q that occurs in (3), and therefore $\mathrm{d}_{\min}$, is at least $\mathfrak{d}$.

To complete the proofs, it remains to show that $\mathrm{d}_{\min}$ and thus the smallest degree of q that appears in (3) is at most $\mathfrak{d}$. To this end, we borrow an argument of A. Buch [3], used in his own proof of Theorem 1.2. We reproduce his argument below:

Let $\widetilde{\lambda}$ and $\overline{\lambda}$ be the partitions obtained by removing the leftmost $\mathfrak{d}$ columns and the top $\mathfrak{d}$ rows of λ, respectively. Also, let $\widetilde{\overline{\lambda}}$ be the partition obtained by removing both the leftmost $\mathfrak{d}$ columns and the top $\mathfrak{d}$ rows of λ. Set $\alpha = (k + \mathfrak{d} - \lambda_{\mathfrak{d}}, \ldots, k + \mathfrak{d} - \lambda_1)$ and let $\beta_i = \max(\mathfrak{d} - \lambda_{l+1-i}, 0)$ for $1 \leq i \leq l$. In other words, α is the complement of the bottom $\mathfrak{d}$ rows of $\mathrm{rotate}(\lambda)$ in $l \times (k + \mathfrak{d})$, and β is the complement of $\mathrm{rotate}(\lambda)$ in the rightmost $\mathfrak{d}$ columns (see Figure 4 below).

It follows from the Littlewood-Richardson rule that the expansion of $\sigma_\lambda * \sigma_\beta$ contains the class $\sigma_{(\mathfrak{d}^l)+\widetilde{\lambda}} = \sigma_{(\mathfrak{d}^l)} * \sigma_{\widetilde{\lambda}}$. It also follows that $\sigma_{\widetilde{\lambda}} * \sigma_\alpha$ contains $\sigma_{(k^{\mathfrak{d}}),\widetilde{\widetilde{\lambda}}} = \sigma_{(k^{\mathfrak{d}})} * \sigma_{\widetilde{\widetilde{\lambda}}}$. It is not hard to check directly from (3) that $\sigma_{(\mathfrak{d}^l)} * \sigma_{(k^{\mathfrak{d}})} = q^{\mathfrak{d}}$. By the nonnegativity of the Gromov-Witten invariants and the associativity of $\mathrm{QH}^*(X)$, $\sigma_\lambda * \sigma_\beta * \sigma_\alpha$ contains the product $\sigma_{(\mathfrak{d}^l)} * \sigma_{(k^{\mathfrak{d}})} * \sigma_{\widetilde{\widetilde{\lambda}}} = q^{\mathfrak{d}} \sigma_{\widetilde{\widetilde{\lambda}}}$. Note that $\widetilde{\widetilde{\lambda}} \cap \mathrm{rotate}(\mu) = \emptyset$, which is well known to be equivalent to $\sigma_{\widetilde{\widetilde{\lambda}}} \cdot \sigma_\mu \neq 0$. Thus we conclude that $\sigma_\lambda * \sigma_\mu * \sigma_\alpha * \sigma_\beta$ contains $q^{\mathfrak{d}}$ times some Schubert class. Therefore, the product $\sigma_\lambda * \sigma_\mu$ must have a term involving a degree of q less than or equal to $\mathfrak{d}$, and we are done. $\square$

$A \cup B \cup C \cup D = \mathrm{rotate}(\lambda)$
$A \cup B = \mathrm{rotate}(\overline{\lambda})$
$A \cup C = \mathrm{rotate}(\widetilde{\lambda})$
$A = \mathrm{rotate}(\widetilde{\widetilde{\lambda}})$
$D = \mathrm{rotate}(\mathfrak{d} \times \mathfrak{d})$

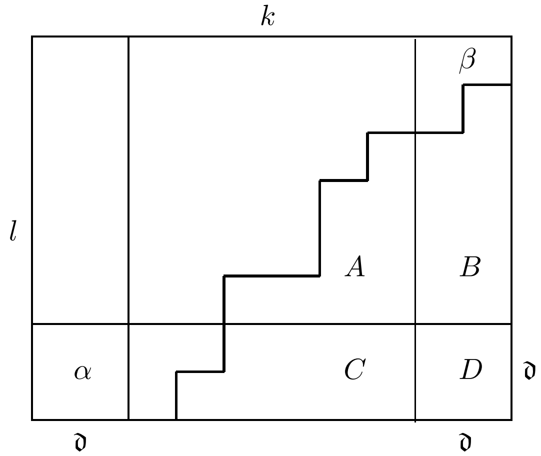


FIGURE 4. Proof of Theorems 1.1 and 1.2

2. Proof of Theorem 1.3

For $\alpha, \beta \subseteq l \times k$, define $\varkappa(\alpha, \beta)$ to be the intersection of all Young diagrams $\gamma \subseteq l \times k$ such that $c^{\gamma}_{\alpha,\beta} \neq 0$. In the event the set of such γ is empty, we set $\varkappa(\alpha, \beta) = l \times k$ by convention. We will make use of the following observation, which follows immediately from this definition:

Lemma 2.1. *Let $\alpha, \beta, m \times M \subseteq l \times k$. Then $\sigma_\alpha \cdot \sigma_\beta \cdot \sigma_{m \times M} \neq 0$ if and only if $\varkappa(\alpha, \beta) \cap \mathrm{rotate}(m \times M) = \emptyset$.*

Our proof of Theorem 1.3 also uses the following two lemmas.[2]

Lemma 2.2. *Let $\alpha, \beta, \lambda, \mu \subseteq l \times k$ be such that $\alpha \subseteq \lambda$ and $\beta \subseteq \mu$. Then $\varkappa(\alpha, \beta) \subseteq \varkappa(\lambda, \mu)$.*

Proof. It suffices to prove the case when $\lambda = \alpha$ and $\mu \setminus \beta$ is a single box. Since every term of $(\sigma_\alpha \cdot \sigma_\beta) \cdot \sigma_1$ is indexed by a partition containing $\varkappa(\alpha, \beta)$, the same is true for $\sigma_\alpha \cdot (\sigma_\beta \cdot \sigma_1) = \sigma_\lambda \cdot (\sigma_\mu + \text{nonnegative terms})$. The claim follows. $\square$

The next lemma is proved by a straightforward application of the Littlewood-Richardson Rule. Details are left to the reader.

[2]Although we will not need it here, this Lemma 2.2 generalizes to a statement about the cohomology of any homogeneous space G/P. The proof is analogous to the one given above.

Lemma 2.3. *Let $m\times M, n\times N \subseteq l\times k$. If $\sigma_{m\times M}\cdot\sigma_{n\times N}=0$, then $\varkappa(m\times M, n\times N) = l\times k$. Otherwise,*

$$\begin{aligned}\varkappa(m\times M, n\times N) &= (m\times M)\cup(n\times N)\\ &\cup((m+n)\times(M+N-k))\\ &\cup((m+n-l)\times(M+N)),\end{aligned}$$

where each of the last two rectangles in the right-hand side is understood to be empty if one of its dimensions is negative, or if it does not fit inside $l\times k$.

Proof of Theorem 1.3. Suppose (I) holds. Then there exists $\rho\subseteq l\times k$ with $c_{\lambda,\mu}^{\rho}\neq 0$ and $\rho\cap\mathrm{rotate}(\nu)=\emptyset$. Hence $\varkappa(\lambda,\mu)\cap\mathrm{rotate}(\nu)=\emptyset$ which by Lemma 2.2 implies

$$\varkappa(a\times A, b\times B)\cap\mathrm{rotate}(c\times C)=\emptyset. \tag{4}$$

By Lemma 2.1, (4) is equivalent to (II), which itself trivially implies (I). Lastly, it follows from Lemma 2.3 that (4) is equivalent to (III). $\square$

3. An Upper Bound, Open Problems and Conjectures

It is an open problem to determine $\mathrm{d_{max}}$, the largest degree of q such that $q^{\mathrm{d_{max}}}$ appears in (2) with nonzero coefficient. In particular, $\mathrm{d_{max}}$ is not equal to the largest degree of q appearing in (3). Next, we present a simple upper bound for $\mathrm{d_{max}}$. For a Young diagram α, let $\mathrm{diag}(\alpha)$ be the size of the largest square contained in α.

Proposition 3.1. $\mathrm{d_{max}}\leq\min(\mathrm{diag}(\lambda),\mathrm{diag}(\mu))$.

We note that for a large class of pairs (λ,μ), this inequality is sharper than the obvious upper bound $\mathrm{d_{max}}\leq\frac{|\lambda|+|\mu|}{n}$.

Proof of Proposition 3.1 Let λ be a Young diagram and $\lambda=(\alpha_1,\ldots\alpha_t\mid\beta_1,\ldots,\beta_t)$ its *Frobenius notation*; the following is a well known identity for Schur functions (see, e.g. [11]):

$$s_\lambda=\det(s_{(\alpha_i|\beta_j)})_{1\leq i,j\leq t}. \tag{5}$$

From (5) and our description of $\mathrm{QH}^*(X)$ it follows that

$$\sigma_\lambda=\det(\sigma_{(\alpha_i|\beta_j)})_{1\leq i,j\leq t}. \tag{6}$$

It is easy to check from the Littlewood-Richardson rule that for any $(\alpha\mid\beta),\lambda\subseteq l\times k$, the largest degree of q that appears in $\sigma_{(\alpha|\beta)}\cdot\sigma_\lambda$ is 1. By (6) we have

$$\sigma_\lambda*\sigma_\mu=\det(\sigma_{(\alpha_i|\beta_j)})_{1\leq i,j\leq t}*\sigma_\mu=\sum_{\tau\in\mathcal{S}_n}\mathrm{sign}(\tau)(\prod_{i=1}^{t}\sigma_{(\alpha_i|\beta_{\tau(i)})})*\sigma_\mu.$$

Thus $M\leq\mathrm{diag}(\lambda)$. Switching the roles of λ and μ gives $M\leq\mathrm{diag}(\mu)$. $\square$

It is an interesting problem to determine precisely which degrees of q appear in the product $\sigma_\lambda*\sigma_\mu$. Based on extensive computational evidence, we conjecture:

Conjecture 3.2. *The product $\sigma_\lambda*\sigma_\mu$ involves q^d for any $d\in[\mathrm{d_{min}},\mathrm{d_{max}}]$.*

More seems to be true.

Conjecture 3.3. *Let $d\geq 1$ be an integer. If $\langle\lambda,\mu,\nu\rangle_d\neq 0$, then either there exists $\alpha\supseteq\nu$ such that $\langle\lambda,\mu,\alpha\rangle_{d-1}\neq 0$, or else $\langle\lambda,\mu,\alpha\rangle_j=0$ for all $\alpha\subseteq l\times k$ and all $0\leq j\leq d-1$.*

Acknowledgments

This work was partially carried out while the author was visiting the Fields Institute in Toronto, and later at the Isaac Newton Institute in Cambridge. We thank both institutes for their hospitality. We are most deeply indebted to Sergey Fomin, whose guidance greatly improved the content and form of this paper. We benefited from conversations with Anders Buch, Bill Fulton, Arun Ram and Chris Woodward. Most of our computational investigations were done using Anders Buch's package [4].

References

[1] A. Bertram, *Quantum Schubert Calculus*, Adv. Math. **128** (1997) 289-305.

[2] A. Bertram, I. Ciocan-Fontanine and W. Fulton, *Quantum Multiplication of Schur polynomials*, Journal of Algebra **219** (1999) 728-746.

[3] A. Buch, *Quantum cohomology of Grassmannians*, e-print `math.AG/0106268`.

[4] A. Buch, *Littlewood-Richardson Calculator*, software available at `http://www-math.mit.edu/~abuch`.

[5] W. Fulton, *Eigenvalues, invariant factors, highest weights, and Schubert calculus*, Bull. Amer. Math. Soc. (N.S.) **37** (2000), no. 3, 209-249.

[6] W. Fulton, *Young tableaux*, Cambridge University Press, 1997.

[7] W. Fulton and R. Pandharipande, *Notes on stable maps and quantum cohomology*, Algebraic geometry—Santa Cruz 1995, Amer. Math. Soc., Providence, RI, 1997, 45-96.

[8] W. Fulton and C. Woodward, preprint, 2001.

[9] G. James and A. Kerber, *The representation theory of the symmetric group*, Encyclopedia of Mathematics and its Applications, 16. Addison-Wesley Co. Reading, Mass., 1981.

[10] M. Kontsevich and Y. Manin, *Gromov-Witten classes, quantum cohomology, and enumerative geometry*, Mirror Symmetry II, Amer. Math. Soc., Providence, RI, 1997, 607-653.

[11] I. Macdonald, *Symmetric functions and Hall polynomials, Second edition,* Oxford University Press, 1995.

[12] Y. Ruan and G. Tian, *A mathematical theory of quantum cohomology,* Math. Res. Lett. **1** (1994), 269-278.

[13] F. Sottile, *Rational Curves on Grassmannians: systems theory, reality, and transverality*, to appear in Contemporary Mathematics, 2001.

[14] R. Stanley, *Enumerative Combinatorics, Volume 2*, Cambridge University Press, 1999.

[15] E. Witten, *The Verlinde algebra and the cohomology of the Grassmannian*, Geometry, topology & physics, Conf. Proc. Lecture Notes Geom. Topology, IV, Internat. Press, Cambridge, MA, 1995, 357-422.

Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109
E-mail address: `ayong@umich.edu`

q AND q,t-ANALOGS OF NON-COMMUTATIVE SYMMETRIC FUNCTIONS

MICHAEL ZABROCKI

Abstract. We introduce two families of non-commutative symmetric functions that have analogous properties to the Hall-Littlewood and Macdonald symmetric functions.

Résumé. Nous présentons deux familles des fonctions symétriques non-commutatives qui ont les propriétés analogues aux fonctions symétriques de Hall-Littlewood et de Macdonald.

1. Introduction

It was noticed in [16] that many q and q,t analogs that are commonly studied in the space of symmetric functions arise from an unusual q-twisting of the symmetric function found by setting $q=0$. In particular we see that the Hall-Littlewood and Macdonald [10] symmetric functions arise from this construction by taking a q-analog of operators that add a row on the Schur basis or a column on the Hall-Littlewood symmetric functions.

The q-twisting that was given in that article may easily be expressed in terms of the notation of Hopf algebras (since the symmetric functions form a commutative and co-commutative Hopf algebra structure). Within this context it can be seen that this may be generalized to any graded Hopf algebra. It is not clear when or if this q-analog will be interesting in another setting, but it creates a context for looking for 'natural' examples of q-analogs within other spaces.

The non-commutative symmetric functions [5] form such a graded Hopf algebra and are an obvious place to begin searching for such examples. The natural analog of the Schur functions within this space are the ribbon Schur functions.

In a manner analogous to that in the symmetric functions, it is possible to define operators that add a row to the ribbon Schur functions. The q-analog of these operators are a natural analog of the operators that add a row to the Hall-Littlewood symmetric functions. Remarkably, we see that this action gives rise to a family of non-commutative symmetric functions that when expanded in terms of ribbon Schur functions have coefficients that are a power of q.

These NC-symmetric functions have properties that suggest they are an analog of the Hall-Littlewood symmetric functions, but they are not equivalent to the non-commutative Hall-Littlewood symmetric functions of Hivert [7]. They do however share some of the same properties of Hivert's NC-Hall-Littlewoods including a factorization property by setting q equal to a root of unity (Proposition 13). It should be remarked that the differences between the factorization properties of these and the Hivert functions suggest that the Hivert functions are associated to the length statistic on compositions in the same way that the Hall-Littlewood functions presented here are associated to the size statistic on compositions.

This family of NC-symmetric functions $\{\mathbf{H}^q_\alpha\}_\alpha$ have the following distinguishing properties.

1. They are triangularly related to the ribbon Schur basis. Namely, we have

$$\mathbf{H}_\alpha^q = \mathbf{s}_\alpha + \sum_{\beta > \alpha} c_{\alpha\beta}^q \mathbf{s}_\beta.$$

2. The coefficient of a single ribbon function in $\mathbf{H}_\alpha^q$ is either 0 or a power of q. The coefficient of the ribbon function indexed by a single part in $\mathbf{H}_\alpha^q$ is $q^{n(\alpha)}$ where $n(\alpha) = \sum_i i$ where the sum is over all i which are descents of α.

3. When $q = 1$, $\mathbf{H}_\alpha^q$ becomes $\mathbf{h}_\alpha$, the non-commutative analogs of the homogeneous symmetric functions. When $q = 0$, $\mathbf{H}_\alpha^q$ specializes to the ribbon Schur function $\mathbf{s}_\alpha$. When q is a root of unity then $\mathbf{H}_\alpha^q$ specializes to a product of non-commutative symmetric functions.

Considering the morphism χ that sends the non-commutative symmetric function $\mathbf{h}_\alpha$ to the commutative version h_α, the image of the functions $\mathbf{H}_\alpha^q$ are the commutative Hall-Littlewood functions whenever the composition α represents a partition (i.e. when α is a hook). That is,

4. (Proposition 9) $\chi(\mathbf{H}_{(1^a,b)}^q) = H_{(b,1^a)}^q$ where $H_\lambda^q = \sum_\mu K_{\mu\lambda}(q) s_\mu$.

We introduce an inner product on the space of non-commutative symmetric functions by setting the ribbon Schur functions as 'semi-self' dual in the following manner

$$\langle \mathbf{s}_\alpha, \mathbf{s}_\beta \rangle = (-1)^{|\alpha| + \ell(\alpha)} \delta_{\alpha\beta^c}.$$

This inner product does not seem to appear elsewhere in the literature, but does share similar properties with the inner product of the symmetric functions and can be a useful tool for calculation within this space. The surprising property that we observe is that the non-commutative analogs of the elementary, homogeneous and Hall-Littlewood bases also share this 'semi-self' duality property. That is, we have in addition to the properties mentioned above,

5. (Proposition 8)

$$\left\langle \mathbf{H}_\alpha^q, \mathbf{H}_\beta^q \right\rangle = (-1)^{|\alpha| + \ell(\alpha)} \delta_{\alpha\beta^c}.$$

Most of these properties are analogous to ones that exist for the non-commutative Hall-Littlewood analogues of Hivert [7], however this last property is not shared by Hivert's noncommutative Hall-Littlewood functions.

Next, we consider a q,t-analog of the non-commutative symmetric functions where we look for properties that are analogous to the Macdonald symmetric functions. Of course, many q,t-analogues are possible and we consider one that that has properties which generalize those for our version of the non-commutative Hall-Littlewood and seem to be analogous to the Macdonald symmetric functions. The family $\{\mathbf{H}_\alpha^{qt}\}_\alpha$ has the following important properties:

1. There is a triangular relation between the family $\{\mathbf{H}_\alpha^t\}_\alpha$ and $\{\mathbf{H}_\alpha^{qt}\}_\alpha$.

$$\mathbf{H}_\alpha^{qt} = \sum_{\beta \le \alpha} c_{\alpha\beta}^{qt} \mathbf{H}_\beta^t.$$

2. The coefficient of a single ribbon function in $\mathbf{H}_\alpha^{qt}$ is of the form $q^a t^b$ (with $a, b \ge 0$). The coefficient of a ribbon Schur function indexed by a single part in $\mathbf{H}_\alpha^{qt}$ is $t^{n(\alpha)}$, the coefficient of a ribbon Schur function indexed by a composition of 1s is $q^{n(\alpha')}$.

3. We have the specialization $\mathbf{H}_\alpha^{0t} = \mathbf{H}_\alpha^t$, and $\mathbf{H}_\alpha^{1t}$ is a product of non-commutative symmetric functions (Proposition 16).

4. The $\mathbf{H}^{qt}_{\alpha}$ satisfy the following two relations

$$\mathbf{H}^{tq}_{\alpha} = \omega' \mathbf{H}^{qt}_{\alpha'},$$

$$q^{n(\alpha')} t^{n(\alpha)} \mathbf{H}^{\frac{1}{q}\frac{1}{t}}_{\alpha} = \omega^c \mathbf{H}^{qt}_{\alpha}.$$

5. (Proposition 20) $\chi(\mathbf{H}^{qt}_{(1^a,b)}) = H^{qt}_{(b,1^a)}$, where $H^{qt}_{\lambda} = \sum_{\mu} K_{\mu\lambda}(q,t) s_{\mu}$.
6. (Proposition 17)

$$\left\langle \mathbf{H}^{qt}_{\alpha}, \mathbf{H}^{qt}_{\beta} \right\rangle = (-1)^{|\alpha|+\ell(\alpha)} \delta_{\alpha\beta^c} \prod_{i=1}^{n-1} (1 - q^i t^{n-i}).$$

The most remarkable property that arises from these functions is the existence of an operator ∇ that has the NC-Macdonald functions as eigenfunctions. That is, if we set

$$\nabla \widetilde{\mathbf{H}}^{qt}_{\alpha} = q^{n(\alpha')} t^{n(\alpha)} \widetilde{\mathbf{H}}^{qt}_{\alpha},$$

where $\widetilde{\mathbf{H}}^{qt}_{\alpha} = t^{n(\alpha)} \mathbf{H}^{q\frac{1}{t}}_{\alpha}$, then this operator has an elegant action on the ribbon Schur functions and shares many of the same properties that exist in the commutative case [3]. Unlike in the commutative case however, formulas for this operator are immediately solvable. In a beautiful analogy, where the commutative version of the operator ∇ produced a q,t grading of the space of parking functions, the non-commutative ∇ produces a grading of the space of preferential arrangements.

In searching for an interesting non-commutative analog of the Macdonald symmetric functions, we considered many possibilities (including the analog considered in [8]). None of the functions except for the one we discuss here had an equivalent ∇ operator and it was this property that indicated to us that these functions are indeed remarkable.

It is not known yet if this family has a representation theoretical model analogous to the $n!$-conjecture or the diagonal harmonics that motivate the existence of these functions. We do see however that the non-commutative versions of these functions and operators share many of the same properties with the commutative case. Independent of their own interest, it is at least hopeful that they will give some insight into why the some of the conjectures for the commutative case are true.

2. Notation for compositions, partitions, Hopf algebras, symmetric, NC-symmetric, and Quasi-symmetric functions

2.1. **Compositions.** We will say that α is a composition of n and write $\alpha \models n$ if α is a sequence of positive integers such that $\alpha_1 + \alpha_2 + \cdots + \alpha_{\ell(\alpha)} = |\alpha| = n$. The length of the sequence is denoted by the symbol $\ell(\alpha)$.

For any two compositions α and β, define the *concatenate* and the *attach* of α and β to be the compositions (respectively)

$$\alpha \cdot \beta = (\alpha_1, \alpha_2, \ldots, \alpha_{\ell(\alpha)}, \beta_1, \beta_2, \ldots, \beta_{\ell(\beta)}) \tag{1}$$

and

$$\alpha | \beta = (\alpha_1, \alpha_2, \ldots, \alpha_{\ell(\alpha)} + \beta_1, \beta_2, \ldots, \beta_{\ell(\beta)}). \tag{2}$$

For a composition α of n define the descent set of α to be the subset of $\{1, 2, 3, \ldots, n-1\}$ as $D(\alpha) = \{\alpha_1, \alpha_1 + \alpha_2, \ldots, \alpha_1 + \alpha_2 + \cdots + \alpha_{\ell(\alpha)-1}\}$. The size of the descent set of α is one less than it's length and it is easily seen that the compositions of n are in one-to-one correspondence with the subsets of $\{1, 2, \ldots, n-1\}$.

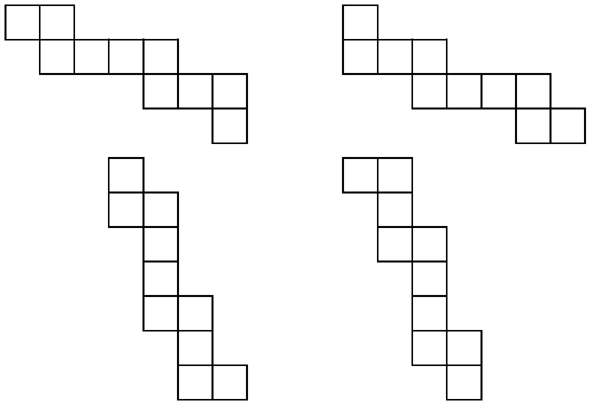


FIGURE 1. Images of the three involutions. If $\alpha = (2,4,3,1)$, then $\overline{\alpha} = (1,3,4,2)$, $\alpha^c = (1,2,1,1,2,1,2)$ and $\alpha' = (2,1,2,1,1,2,1)$.

There is a natural partial order on the compositions of n. Say that a composition α is finer than a composition β (or β is coarser than α) and write $\alpha \leq \beta$ if there exists compositions $\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(k)}$ such that $\alpha = \gamma^{(1)} \cdot \gamma^{(2)} \cdot \ldots \cdot \gamma^{(k)}$ and $\beta = \gamma^{(1)} | \gamma^{(2)} | \ldots | \gamma^{(k)}$. Alternatively, in terms of descent sets we say that $\alpha \leq \beta$ if and only if $D(\beta) \subseteq D(\alpha)$.

There are three standard involutions on the set of compositions. The first involution reverses the order of the sequence. We set $\overleftarrow{\alpha} = (\alpha_{\ell(\alpha)}, \alpha_{\ell(\alpha)-1}, \ldots, \alpha_1)$. If the descent set of α is $D(\alpha) = \{i_1, i_2, \ldots, i_k\}$ then $D(\overleftarrow{\alpha}) = \{|\alpha| - i_1, |\alpha| - i_2, \ldots, |\alpha| - i_k\}$.

The second involution corresponds to taking the complement of the descent set. Define α^c to be the composition with $D(\alpha^c) = \{1, 2, \ldots, |\alpha| - 1\} - D(\alpha)$. Notice that if α is a composition of length k then α^c is a composition of length $n + 1 - k$.

Finally, the third involution is the composition of the previously two defined. Let $\alpha' = \overleftarrow{\alpha^c} = \overleftarrow{\alpha}^c$. The composition α' also has length $n + 1 - k$.

For some formulas we will need a total order on the compositions of size n. We set $\phi(\alpha) = \sum_{i \in D(\alpha)} 2^{i-1}$. This map associates each composition with an integer between 0 and $2^{n-1} - 1$. This map induces a total order from the integers on this set which is a refinement of the partial order defined above.

For the composition α, there is a standard statistic we will use frequently given by $n(\alpha) = \sum_{i \in D(\alpha)} i$.

2.2. **Partitions.** A partition λ of n is a composition of n with the property that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{\ell(\lambda)}$. We will indicate that λ is a partition by the notation that $\lambda \vdash n$. We say that a partition μ is contained in a partition λ and write that $\mu \subseteq \lambda$ if $\ell(\mu) \leq \ell(\lambda)$ and $\mu_i \leq \lambda_i$ for all $1 \leq i \leq \ell(\mu)$. Define then a skew partition to be represented by λ/μ where λ and μ are partitions such that $\mu \subseteq \lambda$.

There is a partial order on the set of partitions. We will say that the partition $\lambda \geq \mu$ if $\lambda_1 + \lambda_2 + \cdots + \lambda_i \geq \mu_1 + \mu_2 + \cdots + \mu_i$ for all $i \geq 1$.

Partitions will sometimes be represented by Ferrers diagrams, a graphical representation of a partition formed by placing rows of square cells aligned on the left hand edge with λ_i cells in the i^{th} row. We will use the cartesian convention where we place the 1^{st} row of cells on the bottom of the diagram (the matrix convention is to place the 1^{st} row of cells on the top). A skew Ferrers diagram for a skew partition λ/μ is Ferrers diagram for the partition λ where the cells that correspond to the partition μ are not drawn.

Define the conjugate partition to λ to be the partition λ' such that λ'_i is the number of parts of λ that have size greater than or equal to i. This corresponds to the partition formed by flipping λ across the line $x = y$.

Any composition of n may be associated with a 'ribbon,' a skew partition with no 2×2 sub-diagrams. This ribbon is usually represented by a skew-Ferrers diagram. The composition $\alpha \vdash n$ is mapped to the skew partition $(\alpha_1 + \alpha_2 + \cdots + \alpha_{\ell(\alpha)} - \ell(\alpha) + 1, \alpha_1 + \alpha_2 + \cdots + \alpha_{\ell(\alpha)-1} - \ell(\alpha) + 2, \ldots, \alpha_1)/(\alpha_1 + \alpha_2 + \cdots + \alpha_{\ell(\alpha)-1} - \ell(\alpha) + 1, \alpha_1 + \alpha_2 + \cdots + \alpha_{\ell(\alpha)-2} - \ell(\alpha) + 2, \ldots, \alpha_1 + \alpha_2 - 2, \alpha_1 - 1)$. For the example in Figure 1, $\alpha = (2, 4, 3, 1)$ is associated to the skew partition $(7, 7, 5, 2)/(6, 4, 1)$.

We will label the cells of the x, y-coordinate lattice and say that a point (i, j) is in the diagram of a partition μ if $1 \leq i \leq \mu_j$. The arm of the cell $s = (i, j)$ in a partition μ is denoted by $a_\mu(s) := \mu_j - i$. The leg will be denoted by the value $l_\mu(s) := a_{\mu'}(j, i)$.

2.3. **Hopf algebras.** For general facts about Hopf algebras, we refer the reader to [1] or [12].

Let R be a commutative ring and H an R module. We say that H is an algebra if there are maps $\mu : H \otimes H \to H$ (multiplication) and $\eta : R \to H$ (unit) that satisfy the following two conditions:
1) $\mu \circ (id \otimes \mu) = \mu \circ (\mu \otimes id)$
2) $\mu \circ (\eta \otimes id) = id = \mu \circ (id \otimes \eta)$

We say that H has a co-algebra structure if there are maps $\Delta : H \to H \otimes H$ (comultiplication) and $\varepsilon : H \to R$ (counit) that satisfy the following two conditions:
1) $(\Delta \otimes id) \circ \Delta = (id \otimes \Delta) \circ \Delta$
2) $(\varepsilon \otimes id) \circ \Delta = id = (id \otimes \varepsilon) \circ \Delta$

If H has at the same time an algebra and a co-algebra structure (H, μ, η) and (H, Δ, ε) and Δ is a homomorphism of algebras then we call H together with these corresponding operations, $(H, \mu, \eta, \Delta, \varepsilon)$, a bialgebra. H is called a Hopf algebra if it is a bialgebra with a map $S : H \to H$ that satisfies the following identity:
1) $\mu \circ S \otimes id \circ \Delta = \mu \circ id \otimes S \circ \Delta = \eta \circ \varepsilon$

If we define the map $\tau : H \otimes H \to H \otimes H$ by $\tau(a \otimes b) = b \otimes a$ then we say that an algebra H is commutative if $\mu \circ \tau = \mu$ and we say that a co-algebra H is co-commutative if $\tau \circ \Delta = \Delta$. It may be shown that for any Hopf algebra $\Delta \circ S = S \otimes S \circ \tau \circ \Delta$.

If the Hopf algebra is either commutative or co-commutative, then it follows that S is an involution.

An important operation that arises in this setting is the convolution of two operators $f, g \in Hom(H, H)$. We set $f * g = \mu \circ f \otimes g \circ \Delta$. Convolution is an associative binary operation and the element $\eta\varepsilon$ serves as the identity. That is, we have for $V \in Hom(H, H)$

$$\eta\varepsilon * V = V * \eta\varepsilon = V. \tag{3}$$

In addition, it follows from the defining property of the antipode that $id * S = S * id = \eta\varepsilon$.

2.4. **Symmetric functions.** We refer the reader to [10] for basic facts about the symmetric functions.

Consider the space of symmetric functions as the polynomial ring $\Lambda = \mathbb{Q}[p_1, p_2, \ldots]$ in the commuting set of variables $\{p_1, p_2, p_3, \ldots\}$. The p_i are the simple power symmetric functions and represent the symmetric formal series $p_k = x_1^k + x_2^k + x_3^k + \cdots$ (although in this context we need not consider the variables $\{x_1, x_2, x_3, \ldots\}$). Define the degree of the power symmetric function p_k within this space to have degree k. Let Λ^n represent the subspace of polynomials of degree n. Since the p_k commute, we see that Λ^n is spanned by the set of monomials $p_\lambda := p_{\lambda_1} p_{\lambda_2} \cdots p_{\lambda_{\ell(\lambda)}}$ where λ is a partition of n.

Let $n_i(\lambda)$ represent the number of parts of size i in the partition λ, then define $z_\lambda = \prod_{i\geq 1} i^{n_i(\lambda)} n_i(\lambda)!$. The simple elementary symmetric functions are defined to be $e_k = \sum_{\lambda\vdash k}(-1)^{k-\ell(\lambda)}p_\lambda/z_\lambda$ and the simple homogeneous symmetric functions are defined to be $h_k = \sum_{\lambda\vdash k} p_\lambda/z_\lambda$. Set $e_\lambda = e_{\lambda_1}e_{\lambda_2}\cdots e_{\lambda_{\ell(\lambda)}}$ and $h_\lambda = h_{\lambda_1}h_{\lambda_2}\cdots h_{\lambda_{\ell(\lambda)}}$. The Schur symmetric functions are defined to be $s_\lambda = det|h_{\lambda_i+i-j}| = det|e_{\lambda'_i+i-j}|$.

There is a natural scalar product defined on this space that is defined on the power symmetric functions by $\langle p_\lambda, p_\mu\rangle = z_\lambda \delta_{\lambda\mu}$. It may be shown that the Schur functions are self-dual with respect to the scalar product, that is, $\langle s_\lambda, s_\mu\rangle = \delta_{\lambda\mu}$.

We are interested in finding analogs for the two families of symmetric functions $H^q_\mu := \sum_{\lambda\vdash|\mu|} K_{\lambda\mu}(q)s_\lambda$ and $H^{qt}_\mu := \sum_{\lambda\vdash|\mu|} K_{\lambda\mu}(q,t)s_\lambda$ (the Hall-Littlewood and Macdonald symmetric functions). For a definition of $K_{\lambda\mu}(q)$ and $K_{\lambda\mu}(q,t)$ and their associated properties, we refer the interested reader to [10].

2.5. **Non-commutative symmetric functions.** For a more detailed reference about the non-commutative symmetric functions, we refer the reader to [5].

Consider the space of non-commutative symmetric functions as the polynomial ring $NC\Lambda = \mathbb{Q} < \mathbf{h}_1, \mathbf{h}_2, \ldots >$ in the non commuting set of variables $\{\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \ldots\}$. The degree of a monomial $\mathbf{h}_{i_1}\mathbf{h}_{i_2}\cdots\mathbf{h}_{i_\ell}$ will be the sum of the indices $i_1+i_2+\cdots+i_\ell$. The span of the monomials of $NC\Lambda$ of degree n will be denoted by $NC\Lambda^n$ so that $NC\Lambda = \bigoplus_{n\geq 0} NC\Lambda^n$ is a graded ring.

We will define $\mathbf{e}_k = \sum_{\alpha\models k}(-1)^{k-\ell(\alpha)}\mathbf{h}_\alpha$. These are the analogs of the basic homogeneous and elementary symmetric functions. For any composition we define $\mathbf{h}_\alpha = \mathbf{h}_{\alpha_1}\mathbf{h}_{\alpha_2}\cdots\mathbf{h}_{\alpha_{\ell(\alpha)}}$ and $\mathbf{e}_\alpha = \mathbf{e}_{\alpha_1}\mathbf{e}_{\alpha_2}\cdots\mathbf{e}_{\alpha_{\ell(\alpha)}}$. The sets $\{\mathbf{h}_\alpha\}_{\alpha\models n}$ and $\{\mathbf{e}_\alpha\}_{\alpha\models n}$ are all bases for the space of non-commutative symmetric functions of degree n.

The ribbon Schur functions are defined to be $\mathbf{s}_\alpha = \sum_{\beta\geq\alpha}(-1)^{\ell(\alpha)-\ell(\beta)}\mathbf{h}_\beta$. It is well known that the set $\{\mathbf{s}_\alpha\}_{\alpha\models n}$ also defines a basis for $NC\Lambda^n$. It is normal to define a pairing of the NC-symmetric functions with the Quasi-symmetric functions. Instead here we define a scalar product on this space such that the ribbon Schur functions are self-dual, that is,

$$\langle \mathbf{s}_\alpha, \mathbf{s}_\beta\rangle = (-1)^{|\alpha|+\ell(\alpha)}\delta_{\alpha\beta^c}. \tag{4}$$

We remark that this product has the property that $\langle f, g\rangle = (-1)^{deg(f)+1}\langle g, f\rangle$. This follows from the relation $\ell(\alpha)+\ell(\alpha^c) = |\alpha|+1$. We will prove in full generality in later sections (although it is not difficult to show) that we have the following relations

Proposition 1.

$$\langle \mathbf{s}_\alpha, \mathbf{s}_\beta\rangle = \langle \mathbf{h}_\alpha, \mathbf{h}_\beta\rangle = \langle \mathbf{e}_\alpha, \mathbf{e}_\beta\rangle = (-1)^{|\alpha|+\ell(\alpha)}\delta_{\alpha\beta^c}. \tag{5}$$

Corollary 2. *For any $f \in NC\Lambda^n$, we have*

$$f = \sum_{\alpha\models n}(-1)^{\ell(\alpha)+1}\langle \mathbf{s}_{\alpha^c}, f\rangle\, \mathbf{s}_\alpha = \sum_{\alpha\models n}(-1)^{|\alpha|+\ell(\alpha)}\langle f, \mathbf{s}_{\alpha^c}\rangle\, \mathbf{s}_\alpha. \tag{6}$$

A similar statement can be made for the $\mathbf{h}_\alpha$ and $\mathbf{e}_\alpha$ bases.

There is a co-commutative Hopf algebra structure on the space of non-commutative symmetric functions. Multiplication μ, the unit η and the counit ε as defined in the usual manner. The comultiplication is defined as the algebra homomorphism that sends $\Delta(\mathbf{h}_k) = \sum_{i=0}^k \mathbf{h}_i\otimes\mathbf{h}_{k-i}$ and $\Delta(\mathbf{e}_k) = \sum_{i=0}^k \mathbf{e}_i\otimes\mathbf{e}_{k-i}$. The antipode is defined by $S(\mathbf{s}_\alpha) = (-1)^{|\alpha|}\mathbf{s}_{\alpha'}$, $S(\mathbf{h}_\alpha) = (-1)^{|\alpha|}\mathbf{e}_{\overleftarrow{\alpha}}$ and $S(\mathbf{e}_\alpha) = (-1)^{|\alpha|}\mathbf{h}_{\overleftarrow{\alpha}}$. Elementary properties of the antipode and the scalar product show that $\langle S(f), S(g)\rangle = \langle g, f\rangle$.

There are three standard involutions that correspond to those that exist for the compositions. Set $\omega'(\mathbf{s}_\alpha) = \mathbf{s}_{\alpha'}$, $\overleftarrow{\omega}(\mathbf{s}_\alpha) = \mathbf{s}_{\overleftarrow{\alpha}}$ and $\omega^c(\mathbf{s}_\alpha) = \mathbf{s}_{\alpha^c}$. It is important to note that we have the relations $\omega'\overleftarrow{\omega} = \overleftarrow{\omega}\omega' = \omega^c$. These operations may be expressed on other bases as well yielding the following expressions:

$$\begin{array}{cc} \omega'(\mathbf{h}_\alpha) = \mathbf{e}_{\overleftarrow{\alpha}} & \omega'(\mathbf{e}_\alpha) = \mathbf{h}_{\overleftarrow{\alpha}} \\ \overleftarrow{\omega}(\mathbf{h}_\alpha) = \mathbf{h}_{\overleftarrow{\alpha}} & \overleftarrow{\omega}(\mathbf{e}_\alpha) = \mathbf{e}_{\overleftarrow{\alpha}} \\ \omega^c(\mathbf{h}_\alpha) = \mathbf{e}_\alpha & \omega^c(\mathbf{e}_\alpha) = \mathbf{h}_\alpha \end{array} \tag{7}$$

Of course we also see that $\langle \overleftarrow{\omega} f, \overleftarrow{\omega} g\rangle = \langle f, g\rangle$ and $\langle \omega' f, \omega' g\rangle = \langle \omega^c f, \omega^c g\rangle = \langle g, f\rangle$.

We will sometimes wish to look at the commutative versions of the non-commutative symmetric functions. To this end, we introduce the surjection $\chi : NC\Lambda \to \Lambda$ which sends $\mathbf{h}_\alpha$ to the symmetric function $h_{\alpha_1} h_{\alpha_2} \dots h_{\alpha_k}$.

2.6. **The quasi-symmetric functions.** Consider the space of polynomials in the commuting set of variables $x_1, x_2, x_3, \dots, x_n$. The quasi-symmetric functions will be denoted by $Qsym$ which will be the subspace of polynomials spanned by the functions

$$M_\alpha = \sum_f x_{f(1)}^{\alpha_1} x_{f(2)}^{\alpha_2} \cdots x_{f(\ell(\alpha))}^{\alpha_{\ell(\alpha)}}, \tag{8}$$

where the sum is over all functions $f : [\ell(\alpha)] \to [n]$ such that $f(i) < f(i+1)$.

These functions are the analogs of the monomial symmetric functions within the space of symmetric functions. There is a standard pairing between the quasi-symmetric functions and space of non-commutative symmetric functions. This pairing is defined by setting non-commutative homogeneous symmetric functions as dual to the M_α basis, that is $[M_\alpha, \mathbf{h}_\beta] = \delta_{\alpha\beta}$. This is the pairing that makes $Qsym$ and $NC\Lambda$ graded dual Hopf algebras [11].

The ribbon quasi-symmetric functions, F_α are then defined as the elements of $Qsym$ such that $[F_\alpha, \mathbf{s}_\beta] = \delta_{\alpha\beta}$. Clearly, any $A \in Qsym$ may be expanded in these bases by using the formula

$$A = \sum_{\beta \models deg(A)} [A, \mathbf{s}_\beta]\, F_\beta = \sum_{\beta \models deg(A)} [A, \mathbf{h}_\beta]\, M_\beta. \tag{9}$$

Similarly, for any element $\mathbf{A} \in NC\Lambda$, $\mathbf{A}$ may be expanded in terms of the $\mathbf{s}_\beta$ and $\mathbf{h}_\beta$ bases if we know the pairing between $\mathbf{A}$ and F_β or M_β.

$$\mathbf{A} = \sum_{\beta \models deg(\mathbf{A})} [F_\beta, \mathbf{A}]\, \mathbf{s}_\beta = \sum_{\beta \models deg(\mathbf{A})} [M_\beta, \mathbf{A}]\, \mathbf{h}_\beta \tag{10}$$

There is a simple relation between the $Qsym/NC\Lambda$ pairing and the scalar product on $NC\Lambda$. This is expressed with the following proposition.

Proposition 3. *$A \in Qsym$ and $\mathbf{A} \in NC\Lambda$ are such that for all $\mathbf{B} \in NC\Lambda$*

$$[A, \mathbf{B}] = \langle \mathbf{A}, \mathbf{B}\rangle\,,$$

if and only if A and $\mathbf{A}$ have the following relationship

$$\mathbf{A} = \sum_{\beta \models n} (-1)^{\ell(\beta)+1} [A, \mathbf{s}_\beta]\, \mathbf{s}_{\beta^c} = \sum_{\beta \models n} (-1)^{\ell(\beta)+1} [A, \mathbf{h}_\beta]\, \mathbf{h}_{\beta^c},$$

$$A = \sum_{\beta \models n} \langle \mathbf{A}, \mathbf{s}_\beta\rangle\, F_\beta = \sum_{\beta \models n} \langle \mathbf{A}, \mathbf{h}_\beta\rangle\, M_\beta. \tag{11}$$

Proof. By applying Corollary 2,

$$\langle \mathbf{A}, \mathbf{B} \rangle = \left\langle \sum_{\beta \models n} (-1)^{\ell(\beta)+1} [A, \mathbf{s}_\beta] \, \mathbf{s}_{\beta^c}, \mathbf{B} \right\rangle \tag{12}$$

$$= \left[A, \sum_{\beta \models n} (-1)^{\ell(\beta)+1} \langle \mathbf{s}_{\beta^c}, \mathbf{B} \rangle \, \mathbf{s}_\beta \right] = [A, \mathbf{B}] \, .$$

□

This last proposition implies that for a basis $\mathbf{A}_\alpha$ such that $\langle \mathbf{A}_\alpha, \mathbf{A}_\beta \rangle = (-1)^{n+\ell(\alpha)} \delta_{\alpha\beta^c}$, then to compute its dual basis in $Qsym$ it is sufficient to calculate the values of the scalar products $\langle \mathbf{A}_\alpha, \mathbf{s}_\beta \rangle$, since by equation (11), we have that $\left[(-1)^{n+\ell(\alpha^c)} \sum_{\beta \models n} \langle \mathbf{A}_{\alpha^c}, \mathbf{s}_\beta \rangle F_\beta, \mathbf{A}_\beta \right] = (-1)^{n+\ell(\alpha^c)} \langle \mathbf{A}_{\alpha^c}, \mathbf{A}_\beta \rangle = \delta_{\alpha\beta}$.

3. q-Analog bases

3.1. **Scrambled Hopf algebra operators.** Consider the following transformation on $Hom(A, A)$ that seems to arise in a natural way when considered as an operation on symmetric functions [17]. If A is a Hopf algebra, then for $V \in Hom(A, A)$ we define $\overline{V} = \mu \circ id \otimes (VS) \circ \Delta = id * (VS)$. We may show that in any co-commutative Hopf algebra A, the bar operation on $V \in Hom(A, A)$ is an involution. That is, we have

Proposition 4. *For any $V \in Hom(A, A)$ with $(A, \mu, \eta, \Delta, \varepsilon, S)$ a co-com-mutative Hopf algebra, we have that $\overline{\overline{V}} = V$.*

Our q-analog arises by starting with a graded co-commutative Hopf algebra with a function R^q such that for any element $f \in A$ that is of homogeneous degree $R^q(f) = q^{deg(f)} f$. The important properties of this function are that $R^q\big|_{q=0} = \eta\varepsilon$ and $R^q\big|_{q=1} = id$. Now for any $V \in Hom(A, A)$, we set

$$\widetilde{V}^q = \overline{\overline{V} \; \overline{R^q}}. \tag{13}$$

It is not at all obvious that the effect on V will be necessarily interesting. We do however have many examples of where this q-analog arises within the theory of symmetric functions [16]. One may only hope that within other co-commutative Hopf algebras that this operation is also interesting. In this exposition we will show that there is an application of this q-analog within the non-commutative symmetric functions.

Proposition 5. *Let A be a Hopf algebra, with the property that $\overline{V} := id{*}VS$ is an involution for all $V \in Hom(A, A)$. Also assume that there is an operator R^q such that $R^q\big|_{q=0} = \eta\varepsilon$ and $R^q\big|_{q=1} = id$. If V does not depend on q, then the operator $\widetilde{V}^q = \overline{\overline{V} \; \overline{R^q}}$ is a q-analog of V (recovered by setting $q = 0$) and $\overline{\overline{V} \; \eta\varepsilon}$ (by setting $q = 1$).*

Proof. Since $id * S = \eta\varepsilon$, we have that $\overline{id} = \eta\epsilon$. It follows then that

$$\widetilde{V}^q\big|_{q=0} = \overline{\overline{V} \; \overline{\eta\varepsilon}} = \overline{\overline{V}} = V, \tag{14}$$

$$\widetilde{V}^q\big|_{q=1} = \overline{\overline{V} \; \overline{id}} = \overline{\overline{V} \; \eta\varepsilon}. \tag{15}$$

□

3.2. **A q-twisting of $NC\Lambda$ operators.** Define an operator A_α that sends the ribbon Schur function $\mathbf{s}_\beta$ to the ribbon Schur function $\mathbf{s}_{\beta\cdot\alpha}$ (the concatenate operation) and then extend this operator linearly to the space $NC\Lambda$. Similarly define an operator B_α that sends the ribbon Schur function $\mathbf{s}_\beta$ to the ribbon Schur function $\mathbf{s}_{\beta|\alpha}$ (the attach operation) and extend the function linearly.

Define $\widetilde{A_\alpha}^q$ as the q-analog of the operator A_α using equation (13). From Proposition 5, $\widetilde{A_\alpha}^q$ has the property that when $q=0$ it is A_α. When $q=1$ we see that

$$\widetilde{A_\alpha}^q(\mathbf{s}_\beta)\big|_{q=1} = \overline{\overline{A_\alpha}\eta\varepsilon}(\mathbf{s}_\beta) = id * (\overline{A_\alpha}\eta\varepsilon S)(\mathbf{s}_\beta) = \mathbf{s}_\beta\overline{A_\alpha}(1). \tag{16}$$

It is easily shown that $\overline{A_\alpha}(1) = \mathbf{s}_\alpha$, hence we see that $\widetilde{A_\alpha}^q(\mathbf{s}_\beta)\big|_{q=1} = \mathbf{s}_\beta\mathbf{s}_\alpha = A_\alpha(\mathbf{s}_\beta) + B_\alpha(\mathbf{s}_\beta)$. In fact we may derive that there is an simple formula for the action of $\widetilde{A_\alpha}^q$.

Proposition 6. *Let β be a composition of $n > 0$, then*

$$\widetilde{A_\alpha}^q(\mathbf{s}_\beta) = A_\alpha(\mathbf{s}_\beta) + q^n B_\alpha(\mathbf{s}_\beta).$$

Also $\widetilde{A_\alpha}^q(1) = \mathbf{s}_\alpha$.

Proof. Define the numbers $\mathcal{C}^\alpha_{\beta\gamma}$ as the coefficients that arise in the coproduct $\Delta(\mathbf{s}_\alpha) = \sum_{\beta,\gamma}\mathcal{C}^\alpha_{\beta\gamma}\mathbf{s}_\beta\otimes\mathbf{s}_\gamma$. From the defining property of the antipode we have the relation

$$\sum_{\beta\gamma}(-1)^{|\beta|}\mathcal{C}^\alpha_{\beta\gamma}\mathbf{s}_{\beta'}\otimes\mathbf{s}_\gamma = \sum_{\beta\gamma}(-1)^{|\gamma|}\mathcal{C}^\alpha_{\beta\gamma}\mathbf{s}_\beta\otimes\mathbf{s}_{\gamma'} = 0 \tag{17}$$

as long as α is not the empty composition.

The notation for the q analog of any operator is given in equation (13). To show that the q-analog of A_α satisfies the proposition we must give a definition in terms of Hopf algebra operations and then demonstrate that the action reduces dramatically. For any operator, (13) may be restated as

$$\widetilde{V}^q = \mu(1\otimes V)(\mu\otimes S)(1\otimes\Delta)(1\otimes\mu)(1\otimes S\otimes R^q)(1\otimes\Delta)\Delta. \tag{18}$$

At this point it is a direct computation within the Hopf algebra of the non-commutative symmetric functions using relation (17) and the definition of A_α to arrive at the formula stated in the proposition. Since the computation is detailed and not necessary for the remainder of this exposition, we leave it to the reader as an exercise. □

3.3. **An example of q-non-commutative symmetric functions.** Define for a pair of compositions $\alpha,\beta\models n$ the statistic $c(\alpha,\beta) = \sum_{i\in D(\alpha)\cap D(\beta)} i$.

Proposition 7. *Let $\mathbf{H}^q_\alpha = \sum_\beta q^{c(\alpha,\beta^c)}\mathbf{s}_\beta$ where the sum is over all compositions β of $|\alpha|$ such that $\alpha\preceq\beta$ then*

$$\mathbf{H}^q_\alpha = \widetilde{A_{\alpha_{\ell(\alpha)}}}^q\widetilde{A_{\alpha_{\ell(\alpha)-1}}}^q\cdots\widetilde{A_{\alpha_1}}^q 1.$$

Notice that when $q=0$, then $\mathbf{H}^0_\alpha = \mathbf{s}_\alpha$. We also have that when $q=1$, then $\mathbf{H}^1_\alpha = \mathbf{h}_\alpha$. We will think of this family as a non-commutative analog of the Hall-Littlewood symmetric functions because of the following two properties. The first says that this family shares a sort of self-duality property similar to the $\mathbf{s}$, $\mathbf{h}$ and $\mathbf{e}$ bases of $NC\Lambda$, the second says that the commutative versions of these symmetric functions agree with the Hall-Littlewood symmetric functions when indexed by a composition that is equivalent to a partition.

Proposition 8.

$$\left\langle\mathbf{H}^q_\alpha,\mathbf{H}^q_\beta\right\rangle = (-1)^{|\alpha|+\ell(\alpha)}\delta_{\alpha\beta^c}. \tag{19}$$

Proof. Relation (19) will follow from Proposition 17 by setting $t = q$ and $q = 0$. $\square$

We also have the following remarkable connection with the Hall-Littlewood basis, H^q_λ.

Proposition 9. *If* $\alpha = (1^a, b)$, *then*

$$\chi(\mathbf{H}^q_\alpha) = H^q_{(b,1^a)}.$$

Proof. Due to the rule given in ([10], (5.7), p. 228) we have the following important recurrence for the Hall-Littlewood symmetric functions indexed by a hook.

$$h_b H^q_{(1^a)} = H^q_{(b,1^a)} + (1 - q^a) H^q_{(b+1,1^{a-1})} \tag{20}$$

Also consider the recurrence that we have developed for these non-commutative symmetric functions.

$$\begin{aligned} \mathbf{H}^q_{(1^a)}\mathbf{h}_b &= (A_{(b)} + q^a B_{(b)} + (1-q^a)B_{(b)})(\mathbf{H}^q_{(1^a)}) \\ &= \mathbf{H}^q_{(1^a,b)} + (1-q^a)B_{(b)}(\mathbf{H}^q_{(1^a)}) \end{aligned} \tag{21}$$

We also have that $\mathbf{H}^q_{(1^a)} = (A_{(1)} + q^{a-1}B_{(1)})(\mathbf{H}^q_{(1^{a-1})})$. Since we have that $B_{(r)}B_{(s)} = B_{(r+s)}$ and $B_{(r)}A_{(s)} = A_{(r+s)}$, then we see $B_{(b)}(\mathbf{H}^q_{(1^a)}) = \mathbf{H}^q_{(1^{a-1},b+1)}$.

This implies

$$\mathbf{H}^q_{(1^a)}\mathbf{h}_b = \mathbf{H}^q_{(1^a,b)} + (1-q^a)\mathbf{H}^q_{(1^{a-1},b+1)}. \tag{22}$$

By induction on the length of the hook we see that $\chi(\mathbf{H}^q_\alpha) = H^q_{(b,1^a)}$ (which is obviously true when either $a = 0$ or $b = 1$). $\square$

Using Proposition 8 we are able to derive an equation for the basis in $Qsym$ dual to the $\mathbf{H}^q_\alpha$. We notice that $\delta_{\alpha\beta} = (-1)^{\ell(\alpha)+1}\left\langle \mathbf{H}^q_{\alpha^c}, \mathbf{H}^q_\beta \right\rangle$. This implies that if

$$\left[P^q_\alpha, \mathbf{H}^q_\beta \right] = \delta_{\alpha\beta}, \tag{23}$$

then $P^q_\alpha = \sum_{\beta\models|\alpha|} \left\langle (-1)^{\ell(\alpha)+1}\mathbf{H}^q_{\alpha^c}, \mathbf{s}_\beta \right\rangle F_\beta$. A simple calculation using Proposition 3 implies that

Corollary 10.

$$P^q_\alpha = \sum_{\beta\le\alpha} (-1)^{\ell(\beta)-\ell(\alpha)} q^{c(\alpha^c,\beta)} F_\beta \tag{24}$$

is the basis of $Qsym$ *which is dual to the family* $\mathbf{H}^q_\alpha$ *with respect to the* $Qsym/NC\Lambda$ *pairing.*

We remark that these non-commutative symmetric functions are not equivalent to those defined in [7]. They are however remarkably similar and do agree for $\alpha = (1^a, b)$. We show this in the following proposition.

Say that $D(\alpha) = \{a_1 < a_2 < \ldots < a_{\ell(\alpha)-1}\}$ and $D(\beta) = \{b_1 < b_2 < \ldots < b_{\ell(\beta)-1}\}$. Let $Bre(\alpha,\beta)$ be the composition of $\ell(\alpha)$ with the descent set equal to $D(Bre(\alpha,\beta)) = \{\#\{a_j : a_j \le b_i\} : 1 \le i \le \ell(\beta)-1\}$. Let

$$\mathbf{W}^q_\alpha = \sum_{\beta\ge\alpha} q^{n(Bre(\alpha,\beta)^c)} \mathbf{s}_\beta.$$

This is the definition of the non-commutative analogs of the Hall-Littlewood symmetric functions given in Theorem 6.13 of [7]. We may easily see that the family $\mathbf{W}^q_\alpha$ satisfies the following recurrence

Proposition 11.

$$\mathbf{W}^q_{\alpha\cdot(m)} = A_{(m)}(\mathbf{W}^q_\alpha) + q^{\ell(\alpha)}B_{(m)}(\mathbf{W}^q_\alpha) \tag{25}$$

Proof. $D(Bre(\alpha\cdot(m),\beta\cdot(m))) = D(Bre(\alpha,\beta)) \cup \{|\alpha|\}$. At the same time we have $D(Bre(\alpha\cdot(m),\beta|(m))) = D(Bre(\alpha,\beta))$. Both $Bre(\alpha\cdot(m),\beta\cdot(m))$ and $Bre(\alpha\cdot(m),\beta|(m))$ are compositions of $\ell(\alpha)+1$, hence the proposition follows. □

This proposition should be compared to Proposition 6. This also shows the following corollary.

Corollary 12. *For $\alpha = (1^a, b)$ we have*

$$\mathbf{W}^q_\alpha = \mathbf{H}^q_\alpha.$$

One open question that arises from this definition is: Is there a Hecke algebra action on $Qsym$ such that the functions P^q_α are invariants under its action? This is the case of the functions of Hivert that are dual to the non-commutative symmetric functions $\mathbf{W}^q_\alpha$ and are given by the formula:

$$G^q_\alpha = \sum_{\beta\leq\alpha} (-1)^{\ell(\beta)-\ell(\alpha)} q^{n(\overleftarrow{Bre(\beta,\alpha)})} F_\beta. \tag{26}$$

The functions $\mathbf{H}^q_\alpha$ have a factorization property that is very similar to that held by the functions of Hivert $\mathbf{W}^q_\alpha$ and the commutative Hall-Littlewood symmetric functions [7].

Theorem 13. *Let ζ be an r^{th} root of unity. Then*

$$\mathbf{H}^\zeta_\alpha = \mathbf{H}^\zeta_{\alpha^{(1)}}\mathbf{H}^\zeta_{\alpha^{(2)}}\cdots\mathbf{H}^\zeta_{\alpha^{(k)}}$$

for any decomposition $\alpha = \alpha^{(1)}\cdot\alpha^{(2)}\cdot\dots\cdot\alpha^{(k)}$, where for $1 \leq i \leq k-1$, $\alpha^{(i)}$ is a composition of a multiple of r.

This theorem follows from the following derivation of the product rule for these functions and then evaluating q at a root of unity.

Proposition 14.

$$\mathbf{H}^q_\alpha\mathbf{H}^q_\beta = \sum_{\gamma\geq\beta} f^\gamma_{\alpha\beta}(q)(\mathbf{H}^q_{\alpha\cdot\gamma} + (1-q^{|\alpha|})\mathbf{H}^q_{\alpha|\gamma}), \tag{27}$$

where

$$f^\gamma_{\alpha\beta}(q) = q^{c(\beta,\gamma^c)}(1-q^{|\alpha|})^{\ell(\beta)-\ell(\gamma)}. \tag{28}$$

Before proceeding with the proof we introduce the following lemma.

Lemma 15.

$$\mathbf{H}^q_\alpha\mathbf{s}_\beta = \sum_{\gamma\geq\beta} g^\gamma_{\alpha\beta}(q)(\mathbf{H}^q_{\alpha\cdot\gamma} + (1-q^{|\alpha|})\mathbf{H}^q_{\alpha|\gamma}), \tag{29}$$

where

$$g^\gamma_{\alpha\beta}(q) = (-1)^{\ell(\beta)-\ell(\gamma)}q^{c(\alpha\cdot\beta,(\alpha\cdot\gamma)^c)}. \tag{30}$$

Proof. Let $n = |\alpha|+|\beta|$. We will take the scalar product defined in equation (4) of $\mathbf{H}^q_\alpha \mathbf{s}_\beta$ and $\mathbf{H}^q_{\theta^c}$. This will give the coefficient of $(-1)^{\ell(\theta)+n}\mathbf{H}^q_\theta$ in the expression $\mathbf{H}^q_\alpha \mathbf{s}_\beta$. By expanding $\mathbf{H}^q_\alpha$ in terms of $\mathbf{s}_\gamma$ and using that $\left\langle \mathbf{s}_\alpha, \mathbf{H}^q_\beta \right\rangle = (-1)^{|\alpha|+\ell(\alpha)} q^{c(\alpha,\beta)}\chi(\alpha \leq \beta^c)$, we see

$$\begin{aligned}
(-1)^{\ell(\theta)+n}\left\langle \mathbf{H}^q_\alpha \mathbf{s}_\beta, \mathbf{H}^q_{\theta^c}\right\rangle &= (-1)^{\ell(\theta)+n}\sum_{\gamma\geq\alpha} q^{c(\alpha,\gamma^c)}\left\langle \mathbf{s}_{\gamma\cdot\beta} + \mathbf{s}_{\gamma|\beta}, \mathbf{H}^q_{\theta^c}\right\rangle \\
&= \sum_{\gamma\geq\alpha} q^{c(\alpha,\gamma^c)}(-1)^{\ell(\theta)+\ell(\gamma\cdot\beta)} q^{c(\gamma\cdot\beta,\theta^c)}\chi(\gamma\cdot\beta\leq\theta) \\
&\quad + \sum_{\gamma\geq\alpha} q^{c(\alpha,\gamma^c)}(-1)^{\ell(\theta)+\ell(\gamma|\beta)} q^{c(\gamma|\beta,\theta^c)}\chi(\gamma|\beta\leq\theta).
\end{aligned} \tag{31}$$

Break the composition θ into the composition consisting of the first $|\alpha|$ 'cells' of θ: $\mu = \theta\big|_{1\ldots|\alpha|}$, and the last $|\beta|$ 'cells:' $\nu = \theta\big|_{|\alpha|+1\ldots n}$, so that either $\theta = \mu\cdot\nu$ or $\theta = \mu|\nu$.

If $\nu<\beta$ or $\mu<\alpha$ then clearly both sums will be 0.

If $\mu>\alpha$ then $D(\mu^c)\cap D(\alpha)$ is non empty and contains at least one element a. For each γ in the sums, either $D(\gamma)$ contains a or it doesn't. The terms with γ such that $a \in D(\gamma)$ are of opposite sign but the same q coefficient as those γ such that $a \notin D(\gamma)$. Therefore the two sums will again be 0.

We need only consider the θ where $\mu = \alpha$ and $\nu\geq\beta$. If $\theta = \alpha\cdot\nu$, then the second sum is clearly 0 and first sum contains only 1 term, $(-1)^{\ell(\nu)+\ell(\beta)}q^{c(\alpha\cdot\beta,\theta^c)}$ (which agrees with the statement of the lemma). If $\theta = \alpha|\nu$, then both sums have exactly one non-zero term, and the scalar product is

$$\begin{aligned}
&(-1)^{\ell(\nu)+\ell(\beta)-1}q^{c(\alpha\cdot\beta,(\alpha|\nu)^c)} + (-1)^{\ell(\nu)+\ell(\beta)}q^{c(\alpha|\beta,(\alpha|\nu)^c)} \\
&\quad = (-1)^{\ell(\nu)+\ell(\beta)}q^{c(\alpha|\beta,(\alpha|\nu)^c)}(1-q^{|\alpha|}).
\end{aligned} \tag{32}$$

$\square$

Proof of Proposition 14. Expanding $\mathbf{H}^q_\beta$ in terms of $\mathbf{s}_\gamma$ and using Lemma 15, yields

$$\begin{aligned}
&(-1)^{\ell(\theta)+n}\left\langle \mathbf{H}^q_\alpha \mathbf{H}^q_\beta, \mathbf{H}^q_{\theta^c}\right\rangle \\
&= (-1)^{\ell(\theta)+n}\sum_{\gamma\geq\beta}\sum_{\mu\geq\gamma} q^{c(\beta,\gamma^c)}\left\langle g^\mu_{\alpha\gamma}(q)(\mathbf{H}^q_{\alpha\cdot\mu} + (1-q^{|\alpha|})\mathbf{H}^q_{\alpha|\mu}), \mathbf{H}^q_{\theta^c}\right\rangle \\
&= \sum_{\gamma\geq\beta}\sum_{\mu\geq\gamma} q^{c(\beta,\gamma^c)}g^\mu_{\alpha\gamma}(q)\delta_{\alpha\cdot\mu,\theta} + q^{c(\beta,\gamma^c)}g^\mu_{\alpha\gamma}(q)(1-q^{|\alpha|})\delta_{\alpha|\mu,\theta}.
\end{aligned} \tag{33}$$

Now if $|\alpha|$ is in $D(\theta)$ then the inner product is

$$\begin{aligned}
&= \sum_{\mu\geq\gamma\geq\beta} q^{c(\beta,\gamma^c)}(-1)^{\ell(\gamma)-\ell(\mu)}q^{c((\alpha\cdot\gamma),(\alpha\cdot\mu)^c)} \\
&= \sum_{\mu\geq\gamma\geq\beta} q^{c(\beta,\gamma^c)+c(\gamma,\mu^c)}(-q^{|\alpha|})^{\ell(\gamma)-\ell(\mu)} \\
&= q^{c(\beta,\mu^c)}(1-q^{|\alpha|})^{\ell(\beta)-\ell(\mu)}.
\end{aligned} \tag{34}$$

This agrees with the formula given for $f^\mu_{\alpha\beta}(q)$. If $|\alpha|$ is not in $D(\theta)$, the result is $(1-q^{|\alpha|})$ times this result. $\square$

4. q,t-Analogs of non-commutative symmetric functions

Define the following q,t-non commutative symmetric function.

$$\mathbf{H}^{qt}_{\alpha} = \sum_{\beta \models |\alpha|} t^{c(\alpha,\beta^c)} q^{c(\alpha',\overleftarrow{\beta})} \mathbf{s}_{\beta} \tag{35}$$

Clearly from this definition, if $q=0$ and $t=q$, then all terms such that $D(\alpha') \cap D(\overleftarrow{\beta})$ is non empty vanish and we have $\mathbf{H}^{0q}_{\alpha} = \mathbf{H}^{q}_{\alpha}$. Therefore we also have the specializations, $\mathbf{H}^{00}_{\alpha} = \mathbf{s}_{\alpha}$ and $\mathbf{H}^{01}_{\alpha} = \mathbf{h}_{\alpha}$, and $\mathbf{H}^{10}_{\alpha} = \mathbf{e}_{\alpha^c}$.

Moreover, $\mathbf{H}^{qt}_{\alpha}$ satisfies the following relations which are similar to those held by the Macdonald symmetric functions in the commutative case:

$$\mathbf{H}^{tq}_{\alpha} = \omega' \mathbf{H}^{qt}_{\alpha'}. \tag{36}$$

$$q^{n(\alpha')} t^{n(\alpha)} \mathbf{H}^{\frac{1}{q}\frac{1}{t}}_{\alpha} = \omega^c \mathbf{H}^{qt}_{\alpha}. \tag{37}$$

When we set $q=1$, the $\mathbf{H}^{1t}_{\alpha}$ become products of some non-standard non-commutative symmetric functions, as seen in the following proposition.

Proposition 16. *Define the non-commutative symmetric functions* $\mathbf{H}^{q(i)}_{(m)} = \sum_{\beta \models m} q^{(\ell(\beta)-1)i + n(\overleftarrow{\beta})} \mathbf{s}_{\beta}$. *For a composition α such that $k = \ell(\alpha)$, we have*

$$\mathbf{H}^{q1}_{\alpha} = \mathbf{H}^{q(\sum_{i>1}\alpha_i)}_{(\alpha_1)} \mathbf{H}^{q(\sum_{i>2}\alpha_i)}_{(\alpha_2)} \cdots \mathbf{H}^{q(0)}_{(\alpha_k)}. \tag{38}$$

Proof. Fix α and for $1 \le i \le \ell(\alpha)$ let $\gamma^{(i)}$ be a composition of α_i. The coefficient of $\mathbf{s}_{\gamma^{(1)}}\mathbf{s}_{\gamma^{(2)}} \cdots \mathbf{s}_{\gamma^{(\ell(\alpha))}}$ in the right hand side of equation (38) is q raised to the power of

$$\sum_i n(\overleftarrow{\gamma^{(i)}}) + \sum_i (\ell(\gamma^{(i)}) - 1) \sum_{j>i} \alpha_i = \sum_i \left(\sum_{d \in D(\overleftarrow{\gamma^{(i)}})} i + \sum_{j>i} \alpha_i \right). \tag{39}$$

This agrees with $c(\alpha', \overleftarrow{\beta})$ where β is attach and concatenate of the $\gamma^{(i)}$ and hence agrees with the q coefficient on the left hand side of equation (38). □

We also have the following two additional Propositions that lead us to believe that it is an interesting generalization of the family $\mathbf{H}^{q}_{\alpha}$.

Proposition 17. *Let $\alpha \models n$, then*

$$\left\langle \mathbf{H}^{qt}_{\alpha}, \mathbf{H}^{qt}_{\beta} \right\rangle = (-1)^{|\alpha|+\ell(\alpha)} \delta_{\alpha\beta^c} \prod_{i=1}^{n-1} (1 - q^i t^{n-i}). \tag{40}$$

Proof.

$$\begin{aligned} \left\langle \mathbf{H}^{qt}_{\alpha}, \mathbf{H}^{qt}_{\beta} \right\rangle &= \sum_{\gamma} \sum_{\theta} t^{c(\alpha,\gamma^c)+c(\beta,\theta^c)} q^{c(\alpha',\overleftarrow{\gamma})+c(\beta',\overleftarrow{\theta})} \langle \mathbf{s}_{\gamma}, \mathbf{s}_{\theta} \rangle \\ &= \sum_{\gamma} (-1)^{n+\ell(\gamma)} t^{c(\alpha,\gamma^c)+c(\beta,\gamma)} q^{c(\alpha',\overleftarrow{\gamma})+c(\beta',\gamma')}. \end{aligned} \tag{41}$$

If $\alpha \neq \beta^c$, then $(D(\alpha) \cap D(\beta)) \cup (D(\alpha^c) \cap D(\beta^c))$ is non empty. Take the smallest element i of this set (although any will do) and consider the involution ϕ on the set of compositions such that the compositions that contain i in the descent set are sent to the compositions that do not contain the element i (in the most natural manner). For each $\gamma \models n$, the terms corresponding to γ and $\phi(\gamma)$ have the same weight but opposite sign, hence the sum is 0.

If $\alpha = \beta^c$, then the sum reduces to

$$= \sum_{\gamma} (-1)^{n+\ell(\gamma)} t^{c(\alpha,\gamma^c)+c(\alpha^c,\gamma)} q^{c(\alpha',\overleftarrow{\gamma})+c(\overleftarrow{\alpha},\gamma')} \tag{42}$$

$$= \sum_{S\subseteq\{1,\ldots,n\}} (-1)^{n+|S\cap D(\alpha^c)|+|S^c\cap D(\alpha)|+1} t^{\sum_{i\in S} i} q^{\sum_{i\in S} n+1-i},$$

where the subsets S represent the sets $(D(\alpha)\cap D(\gamma^c))\cup(D(\alpha^c)\cap D(\gamma))$. This is clearly equal to the product stated in the proposition. ☐

Corollary 18. *The family*

$$P^{qt}_{\alpha} = \prod_{i=1}^{n-1} \frac{1}{1-q^i t^{n-i}} \sum_{\beta} (-1)^{\ell(\beta)-\ell(\alpha)} t^{c(\alpha^c,\beta)} q^{c(\overleftarrow{\alpha},\beta')} F_{\beta}, \tag{43}$$

has the property that $\left[P^{qt}_{\alpha}, \mathbf{H}^{qt}_{\beta}\right] = \delta_{\alpha\beta}$.

Proof. If we wish that $\left[P^{qt}_{\alpha}, \mathbf{H}^{qt}_{\beta}\right] = \delta_{\alpha\beta}$, then using equation (11),

$$P^{qt}_{\alpha} = \sum_{\beta\models|\alpha|} \langle \mathbf{A}, \mathbf{s}_{\beta}\rangle F_{\beta},$$

where $\mathbf{A} = (-1)^{\ell(\alpha)+1}\prod_{i=1}^{n-1}\frac{1}{1-q^i t^{n-i}}\mathbf{H}^{qt}_{\alpha^c}$, since

$$\delta_{\alpha\beta} = (-1)^{\ell(\alpha)+1}\prod_{i=1}^{n-1}\frac{1}{1-q^i t^{n-i}}\left\langle \mathbf{H}^{qt}_{\alpha^c}, \mathbf{H}^{qt}_{\beta}\right\rangle. \tag{44}$$

A simple calculation yields the equation stated in the corollary. ☐

There is a characterization of the non-commutative q,t analogs $\mathbf{H}^{qt}_{\alpha}$ in terms of properties that are similar to those shared by the Macdonald symmetric functions. This characterization is not particularly important for our treatment, but it should not be ignored because of the similarities that it shares with the commutative case.

Define a family of non-commutative symmetric functions $\mathbf{P}^{qt}_{\alpha}$ by the following three conditions.

1. $\mathbf{P}^{qt}_{\alpha} = \mathbf{H}^{t}_{\alpha} + \sum_{\beta<\alpha} c_{\alpha\beta}(q,t)\mathbf{H}^{t}_{\beta}$ for some coefficients $c_{\alpha\beta}(q,t)$ that are rational functions in the parameters q and t.

2. $\omega'\mathbf{P}^{qt}_{\alpha} = a_{\alpha}(q,t)\mathbf{P}^{tq}_{\alpha'}$ for some coefficients $a_{\alpha}(q,t)$.

3. $\left\langle \mathbf{P}^{qt}_{\alpha}, \mathbf{P}^{qt}_{\beta}\right\rangle = 0$ if $\alpha \neq \beta^c$.

Theorem 19. *The family $\mathbf{P}^{qt}_{\alpha}$ are defined by the three conditions listed above and, moreover, $\mathbf{P}^{qt}_{\alpha} = r_{\alpha}\mathbf{H}^{qt}_{\alpha}$ where $r_{\alpha} = 1/\prod_{i\in D(\alpha^c)}(1-q^{n-i}t^i)$. The coefficients $c_{\alpha\beta}(q,t)$ are given by the formula*

$$c_{\alpha\beta}(q,t) = \prod_{i\in D(\alpha^c)\cap D(\beta)} q^{n-i}/(1-t^i q^{n-i}).$$

The coefficients $a_{\alpha}(q,t)$ mentioned in the second condition are given by the formula $a_{\alpha}(q,t) = \prod_{i\in D(\alpha)}(1-q^{n-i}t^i)/\prod_{i\in D(\alpha^c)}(1-q^{n-i}t^i)$.

Proof. The proof proceeds by induction, for there is a method of calculating the coefficients $c_{\alpha\beta}(q,t)$ from ones preceding it in some order. Say that $\mathbf{P}^{qt}_{\alpha} = \sum_{\beta \le \alpha} c_{\alpha\beta}(q,t)\mathbf{H}^{t}_{\beta}$ where we assume that $c_{\alpha\alpha}(q,t) = 1$ and that this family satisfies the three conditions given above.

Assume that the coefficients $c_{\gamma\delta}(q,t)$ are known and given by the formula stated in the theorem for all γ such that $|D(\gamma)| > |D(\alpha)|$ or for $\gamma = \alpha$ and $\delta > \beta$. To determine $c_{\alpha\beta}(q,t)$ we take the scalar product of $\mathbf{P}^{qt}_{\alpha}$ and $\omega'\mathbf{P}^{tq}_{\overleftarrow{\beta}}$ since $\beta < \alpha$, $|D(\overleftarrow{\beta})| > |D(\alpha)|$ and all coefficients in $\mathbf{P}^{qt}_{\overleftarrow{\beta}}$ have been calculated already. Since $\omega'\mathbf{P}^{tq}_{\overleftarrow{\beta}} = \mathbf{P}^{qt}_{\beta^c} = \sum_{\theta \le \beta^c} c_{\beta^c\theta}(q,t)\mathbf{H}^{t}_{\theta}$, hence we have the expression

$$\begin{aligned}
\left\langle \mathbf{P}^{qt}_{\alpha}, \omega'\mathbf{P}^{tq}_{\overleftarrow{\beta}} \right\rangle &= \left\langle \mathbf{P}^{qt}_{\alpha}, a_{\overleftarrow{\beta}}(t,q)\mathbf{P}^{qt}_{\beta^c} \right\rangle = 0 \\
&= a_{\overleftarrow{\beta}}(t,q) \sum_{\alpha \ge \theta \ge \beta} c_{\alpha\theta}(q,t)c_{\beta^c\theta^c}(q,t)(-1)^{n+\ell(\theta)} \\
&= a_{\overleftarrow{\beta}}(t,q)c_{\alpha\beta}(q,t)(-1)^{n+\ell(\beta)} + \\
&\qquad a_{\overleftarrow{\beta}}(t,q) \sum_{\alpha \ge \theta > \beta} c_{\alpha\theta}(q,t)c_{\beta^c\theta^c}(q,t)(-1)^{n+\ell(\theta)}.
\end{aligned}$$

Those values of $c_{\beta^c\theta}(q,t)$ may be calculated from what we have already determined since

$$c_{\beta^c\theta^c}(q,t) = \left\langle \mathbf{H}^{t}_{\theta}, \mathbf{P}^{qt}_{\beta^c} \right\rangle (-1)^{n+\ell(\theta)} = \left\langle \mathbf{H}^{t}_{\theta}, \omega'\mathbf{P}^{tq}_{\overleftarrow{\beta}} \right\rangle / a_{\overleftarrow{\beta}}(t,q)(-1)^{n+\ell(\theta)}.$$

Hence we see that

$$c_{\alpha\beta}(q,t) = \frac{(-1)^{n+1+\ell(\beta)}}{a_{\overleftarrow{\beta}}(t,q)} \sum_{\alpha \ge \theta > \beta} c_{\alpha\theta}(q,t) \left\langle \mathbf{H}^{t}_{\theta}, \omega'\mathbf{P}^{tq}_{\overleftarrow{\beta}} \right\rangle.$$

In addition we may calculate $a_{\overleftarrow{\beta}}(t,q)$ since

$$\left\langle \mathbf{H}^{t}_{\beta}, \omega\mathbf{P}^{tq}_{\overleftarrow{\beta}} \right\rangle = \left\langle \mathbf{H}^{t}_{\beta}, a_{\overleftarrow{\beta}}(t,q)\mathbf{P}^{qt}_{\beta^c} \right\rangle = a_{\overleftarrow{\beta}}(t,q)(-1)^{n+\ell(\beta)}.$$

Although we may use these formulas to calculate the coefficients, the only conclusion that we are going to draw from them is that the coefficients $c_{\alpha\beta}(q,t)$ are determined by assuming that the defining conditions are true, hence the family $\mathbf{P}^{qt}_{\alpha}$ which satisfies these conditions is unique.

It remains to show then that $\mathbf{P}^{qt}_{\alpha} = \mathbf{H}^{qt}_{\alpha} / \prod_{i \in D(\alpha^c)}(1 - q^{n-i}t^{i})$ satisfies the conditions listed above. Clearly they satisfy conditions 2 and 3. It remains to show that the $\mathbf{H}^{qt}_{\alpha}$ have the correct expansion in terms of $\mathbf{H}^{t}_{\alpha}$.

$$\begin{aligned}\left\langle \mathbf{H}^t_{\alpha^c}, \mathbf{H}^{qt}_\alpha \right\rangle &= \sum_{\beta \geq \alpha^c} t^{c(\alpha^c,\beta^c)} \left\langle \mathbf{s}_\beta, \mathbf{H}^{qt}_\alpha \right\rangle \\ &= \sum_{\beta \geq \alpha^c} (-1)^{n+\ell(\beta)} t^{c(\alpha,\beta)+c(\alpha^c,\beta^c)} q^{c(\alpha',\beta')} \\ &= \sum_{\beta \geq \alpha^c} (-1)^{n+\ell(\beta)} t^{c(\alpha^c,\beta^c)} q^{c(\alpha',\beta')} \\ &= \sum_{S \subseteq D(\alpha^c)} (-1)^{n+1+|S|} t^{\sum_{i\in S} i} q^{\sum_{i \in S} n-i} \\ &= (-1)^{n+\ell(\alpha^c)} \prod_{i \in D(\alpha^c)} (1 - q^{n-i} t^i)\end{aligned}$$

We also see for β is not strictly smaller than α then $D(\alpha) \cap D(\beta^c)$ is non-empty and

$$\begin{aligned}\left\langle \mathbf{H}^t_{\beta^c}, \mathbf{H}^{qt}_\alpha \right\rangle &= \sum_{\gamma \geq \beta^c} t^{c(\beta^c,\gamma^c)} \left\langle \mathbf{s}_\gamma, \mathbf{H}^{qt}_\alpha \right\rangle \\ &= \sum_{\gamma \geq \beta^c} t^{c(\beta^c,\gamma^c)} t^{c(\alpha,\gamma)} q^{c(\alpha',\gamma')} (-1)^{n+\ell(\gamma)}.\end{aligned}$$

Since there is some element a in $D(\alpha) \cap D(\beta^c)$, every composition $\gamma \geq \beta^c$ either has $a \in D(\gamma)$ or $a \in D(\gamma^c)$. There is an obvious involution between these two sets of compositions and they have the same weight but opposite sign, hence the sum is 0 in this case. □

As in the case of the family with one parameter, when the functions are indexed by composition representing a partition (i.e. a hook), then they are a generalization of the Macdonald symmetric function.

Proposition 20. *If $\alpha = (1^a, b)$, then*

$$\chi(\mathbf{H}^{qt}_\alpha) = H^{qt}_{(b,1^a)}. \tag{45}$$

Proof. Idea: same as in q case. Show that

$$\mathbf{H}^{qt}_{(1^a)} \mathbf{H}^{qt}_{(b)} = \frac{1-t^a}{1-q^b t^a} \mathbf{H}^{qt}_{(1^{a-1},b+1)} + \frac{1-q^b}{1-q^b t^a} \mathbf{H}^{qt}_{(1^a,b)} \tag{46}$$

and by a formula ([10] eq. (6.24) p. 340) we have the same recurrence in the commutative case. That is,

$$H^{qt}_{(1^a)} H^{qt}_{(b)} = \frac{1-t^a}{1-q^b t^a} H^{qt}_{(b+1,1^{a-1})} + \frac{1-q^b}{1-q^b t^a} H^{qt}_{(b,1^a)}. \tag{47}$$

By induction this implies that the commutative versions agree on hooks.

Consider the product $\mathbf{H}^{qt}_{(1^a)} \mathbf{H}^{qt}_{(b)}$. This is equal to

$$\begin{aligned}\mathbf{H}^{qt}_{(1^a)} \mathbf{H}^{qt}_{(b)} &= \sum_{\gamma \models a} \sum_{\theta \models b} t^{n(\gamma^c)} q^{n(\overleftarrow{\theta})} \mathbf{s}_\gamma \mathbf{s}_\theta \\ &= \sum_{\gamma \models a} \sum_{\theta \models b} t^{n(\gamma^c)} q^{n(\overleftarrow{\theta})} (A_\theta(\mathbf{s}_\gamma) + B_\theta(\mathbf{s}_\gamma)).\end{aligned} \tag{48}$$

We also have since $(1^a, b)' = (1^{b-1}, a+1)$.

$$\mathbf{H}^{qt}_{(1^a,b)} = \sum_{\beta \models a+b} t^{c((1^a,b),\beta^c)} q^{c((1^{b-1},a+1),\overleftarrow{\beta})} \mathbf{s}_\beta = \sum_{\gamma \models a} \sum_{\theta \models b} t^{n(\gamma^c)} q^{n(\overleftarrow{\theta})} (A_\theta(\mathbf{s}_\gamma) + t^a B_\theta(\mathbf{s}_\gamma)). \tag{49}$$

While at the same time

$$\mathbf{H}^{qt}_{(1^{a-1},b+1)} = \sum_{\beta \models a+b} t^{c((1^{a-1},b+1),\beta^c)} q^{c((1^b,a),\overleftarrow{\beta})} \mathbf{s}_\beta = \sum_{\gamma \models a} \sum_{\theta \models b} t^{n(\gamma^c)} q^{n(\overleftarrow{\theta})} (q^b A_\theta(\mathbf{s}_\gamma) + B_\theta(\mathbf{s}_\gamma)). \tag{50}$$

From here is easily shown that

$$(1 - q^b t^a)\mathbf{H}^{qt}_{(1^a)} \mathbf{H}^{qt}_{(b)} = (1 - q^b)\mathbf{H}^{qt}_{(1^a,b)} + (1 - t^a)\mathbf{H}^{qt}_{(1^{a-1},b+1)}. \tag{51}$$

□

These two properties are only an indication that $\mathbf{H}^{qt}_\alpha$ are an important generalization of the Macdonald symmetric functions. The first property does not occur in many families of non-commutative symmetric functions, the second, however, could appear for many different families (since the functions of Hivert also have the same property that they have the 'correct' statistic on hooks).

The most important indication that the family $\mathbf{H}^{qt}_\alpha$ is indeed an important analog to the Macdonald symmetric functions is the appearance of an operator 'nabla' that is analogous to the operator introduced in [2] for the symmetric functions.

First we define the analog $\widetilde{\mathbf{H}}^{qt}_\alpha = t^{n(\alpha)} \mathbf{H}^{q\frac{1}{t}}_\alpha = \sum_{\beta \models |\alpha|} t^{c(\alpha,\beta)} q^{c(\alpha',\overleftarrow{\beta})} \mathbf{s}_\beta$. Next, define ∇ to be a linear operator with the property that $\nabla(\widetilde{\mathbf{H}}^{qt}_\alpha) = t^{n(\alpha)} q^{n(\alpha')} \widetilde{\mathbf{H}}^{qt}_\alpha$. For our scalar product, we have that

$$\langle \nabla(f), \nabla(g) \rangle = q^{\binom{n}{2}} t^{\binom{n}{2}} \langle f, g \rangle. \tag{52}$$

As we will see, this operator has many properties that are analogous to those seen in the commutative case. In the non-commutative case the situation is somewhat simpler and we are able to state precisely the action of the operator on the ribbon basis.

Proposition 21. *If $\alpha \models n$, then*

$$\nabla(\mathbf{s}_\alpha) = (-1)^{n+\ell(\alpha)} q^{n(\alpha')} t^{n(\alpha^c)} \sum_{\beta \le \alpha^c} \prod_{i \in D(\alpha) \cap D(\beta)} (t^i + q^{n-i}) \mathbf{s}_\beta. \tag{53}$$

To prove this formula we will need several lemmas for the action of these operators on various bases. By choosing good notation for these operators, the proofs become almost transparent. We will order the ribbons by their descent sets using the total order described section 2.1.

For a tensor product of two matrices with $B = [b_{ij}]_{1 \le i,j \le n}$ we will use the convention that

$$A \otimes B = [b_{ij} A]_{1 \le i,j \le n}. \tag{54}$$

That is, the (r,s) entry in this matrix is

$$b_{(r \ div \ n)+1,(s \ div \ n)+1} a_{(r \ mod \ n)+1,(s \ mod \ n)+1}.$$

Lemma 22. *Let* $\mathbf{s}$ *be a column vector of* $\mathbf{s}_\alpha$ *and* $\widetilde{\mathbf{H}}$ *is a column vector of* $\widetilde{\mathbf{H}}^{qt}_\alpha$ *(both using the total order of section 2.1), then*

$$\widetilde{\mathbf{H}} = \begin{bmatrix} 1 & q^{n-1} \\ 1 & t \end{bmatrix} \otimes \begin{bmatrix} 1 & q^{n-2} \\ 1 & t^2 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & q \\ 1 & t^{n-1} \end{bmatrix} \mathbf{s}. \tag{55}$$

This lemma follows by realizing that if $\phi(\alpha) = k$, then the entries in the k^{th} row of the tensor product matrix agrees with the formula for the coefficients of $\mathbf{s}_\beta$ in $\widetilde{\mathbf{H}}^{qt}_\alpha$. By taking the inverse of this tensor product matrix we derive the inverse relation.

Corollary 23. *Let* $\mathbf{s}$ *be a column vector of* $\mathbf{s}_\alpha$ *and* $\widetilde{\mathbf{H}}$ *is a column vector of* $\widetilde{\mathbf{H}}^{qt}_\alpha$*, then*

$$\mathbf{s} = \left(\prod_{i=1}^{n-1} \frac{1}{t^i - q^{n-i}} \right) \begin{bmatrix} t & -q^{n-1} \\ -1 & 1 \end{bmatrix} \otimes \begin{bmatrix} t^2 & -q^{n-2} \\ -1 & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} t^{n-1} & -q \\ -1 & 1 \end{bmatrix} \widetilde{\mathbf{H}}. \tag{56}$$

Lemma 24. *Let* $\widetilde{\mathbf{H}}$ *be a column vector of* $\widetilde{\mathbf{H}}^{qt}_\alpha$*, then*

$$\nabla \widetilde{\mathbf{H}} = \begin{bmatrix} q^{n-1} & 0 \\ 0 & t \end{bmatrix} \otimes \begin{bmatrix} q^{n-2} & 0 \\ 0 & t^2 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} q & 0 \\ 0 & t^{n-1} \end{bmatrix} \widetilde{\mathbf{H}}. \tag{57}$$

The proof of this lemma again follows by calculating the entry in the row indexed by $\phi(\alpha)$.

Proof. (of Proposition 21) We calculate the action of ∇ on the column vector $\mathbf{s}$. This follows by first expressing $\mathbf{s}$ in terms of $\widetilde{\mathbf{H}}$ using equation (56), then using the action of ∇ on $\widetilde{\mathbf{H}}$, then reexpressing the answer in terms of $\mathbf{s}$ using (55). We calculate that

$$\frac{1}{t^i - q^{n-i}} \begin{bmatrix} t^i & -q^{n-i} \\ -1 & 1 \end{bmatrix} \begin{bmatrix} q^{n-i} & 0 \\ 0 & t^i \end{bmatrix} \begin{bmatrix} 1 & q^{n-i} \\ 1 & t^i \end{bmatrix} = \begin{bmatrix} 0 & -q^{n-i} t^i \\ 1 & (t^i + q^{n-i}) \end{bmatrix}. \tag{58}$$

Therefore we see that the action of ∇ on $\mathbf{s}$ is given by the equation

$$\nabla(\mathbf{s}) = \begin{bmatrix} 0 & -q^{n-1} t \\ 1 & (t + q^{n-1}) \end{bmatrix} \otimes \begin{bmatrix} 0 & -q^{n-2} t^2 \\ 1 & (t^2 + q^{n-2}) \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 0 & -q t^{n-1} \\ 1 & (t^{n-1} + q) \end{bmatrix} \mathbf{s}. \tag{59}$$

Now translate this tensor product directly to the action on the $\mathbf{s}_\alpha$ basis to arrive at the formula stated in equation (53). □

Interesting connections arise with combinatorics and representation theory that are analogous to the commutative case. Recall that for the standard symmetric functions we have that $\langle \nabla(e_n), h_{1^n} \rangle$ is a q, t analog of the number $(n+1)^{n-1}$, which is the number of parking functions (a function $f : [n] \to [n]$ is a parking function if the sequence $(f(1), f(2), \ldots, f(n))$ when sorted in increasing order $(a_1, a_2, \ldots, a_n)$ satisfies $a_i \leq i$). We also know that $\nabla(e_n)\big|_{t=1}$ is a graded Frobenius series for the parking function module [6]. Moreover, the top component $\nabla(e_n)\big|_{e_n}$ is known to be a q, t-analog of the number of increasing parking functions which is given by the Catalan numbers [4].

In the non-commutative case, these statements occur with exact analogy. We will see that analogs of the parking functions are the preferential arrangements ([13] p. 146). An exponential generating function for the number of preferential arrangements is given by $(2 - e^x)^{-1}$ and the number of increasing preferential arrangements is 2^{n-1}.

Proposition 25. *The quantity* $\langle \chi(\nabla(\mathbf{e}_n)), h_{1^n} \rangle$ *is a* q, t *analog for the number of preferential arrangements (the maps* $f : [n] \to [k]$ *where* $1 \leq k \leq n$ *which are onto for some* k*). Moreover, the quantity* $\langle \chi(\nabla(\mathbf{e}_n)), e_n \rangle = \prod_{i=1}^{n-1} (q^{n-i} + t^i)$ *is a* q, t *analog of* 2^{n-1}.

This proposition is a consequence of the statement that appears in full generality just below. For the moment we will provide the following example:

Example 26. *At $n = 4$, there are $125 = (4+1)^{4-1}$ parking functions, and $14 = C_4$ weakly increasing parking functions represented by the following list. The first number is the number of parking functions such that $(f(1), f(2), f(3), f(4))$ when sorted is the adjacent sequence. The sum of these numbers is 125.*

$$1 \times 1111 \quad 4 \times 1112 \quad 4 \times 1113 \quad 12 \times 1223 \quad 12 \times 1134 \quad 12 \times 1123 \quad 12 \times 1124$$

$$4 \times 1222 \quad 4 \times 1114 \quad 6 \times 1133 \quad 6 \times 1122 \quad 12 \times 1224 \quad 12 \times 1233 \quad 24 \times 1234$$

$75 = (2 - e^x)^{-1}\big|_{x^4} 4!$ *of the parking functions do not 'skip' an integer, these are the preferential arrangements. Exactly $8 = 2^3$ of the preferential arrangements are weakly increasing, those given by the following list:*

$$1111 \quad 1112 \quad 1122 \quad 1123 \quad 1222 \quad 1223 \quad 1233 \quad 1234$$

We remark that every preferential arrangement is also a parking function. This is a natural subset of the parking functions which we will denote by $Pref_n$. Just as in the case of the parking functions, there is a natural S_n action on this set formed by permuting the values of the function (that is $(\sigma f)(i) = f(\sigma_i)$) and hence $Pref_n$ forms an S_n module by defining a vector space with $Pref_n$ as the basis.

Let f be a preferential arrangement and define the content of a preferential arrangement to be the composition $\alpha^{(f)}$ such that the i^{th} component is $|f^{-1}(i)|$. We remark that two preferential arrangements are in the same S_n-orbit if $\alpha^{(f)} = \alpha^{(g)}$. The Frobenius series of the S_n module generated by the preferential arrangements of content α is given by the homogeneous symmetric function h_α.

It follows that the preferential arrangement module may be graded by a statistic on the content of the preferential arrangements. If we choose our grading to be $q^{n(\alpha')}$ (this agrees with the 'area' statistic on Dyck paths), then clearly the Frobenius series for the module of preferential arrangements is given by

$$\mathcal{F}_{Pref_n}(q) = \sum_{\alpha \models n} q^{n(\alpha')} h_\alpha. \tag{60}$$

In the commutative case it is known that

$$\nabla(e_n)\big|_{t=1} = \sum_{\mu \subseteq \delta_n} q^{\binom{n}{2} - |\mu|} e_{\lambda(\mu)}, \tag{61}$$

where $\delta_n = (n-1, n-2, \ldots, 1, 0)$ and $\lambda(\mu)$ is the sequence $(m_1(\mu), \ldots, m_{n-1}(\mu), n - \sum_{i=1}^{n-1} m_i(\mu))$ and $m_i(\mu)$ is the number of parts of size i in the partition μ. This is related to the Frobenius series for the module of parking functions by an application of the involution ω. In exact analogy with the commutative case we have the following proposition.

Proposition 27. *Set $t = 1$ in the equation for the action of ∇ on $\mathbf{e}_n$, then*

$$\nabla(\mathbf{e}_n)\big|_{t=1} = \sum_{\alpha \models n} q^{n(\alpha')} \mathbf{e}_\alpha. \tag{62}$$

Proof. With $t = 1$ the action of ∇ on $\mathbf{s}_{(1^n)}$ is given from equation (53)

$$\begin{aligned}
\nabla(\mathbf{e}_n) &= \sum_{\beta} \prod_{i \in D(\overleftarrow{\beta})} (1+q^i)\mathbf{s}_\beta \\
&= \sum_{\beta} \sum_{\gamma \geq \overleftarrow{\beta}} q^{n(\gamma)} \mathbf{s}_\beta \\
&= \sum_{\gamma} \sum_{\overleftarrow{\beta} \leq \gamma} q^{n(\gamma)} \mathbf{s}_\beta \\
&= \sum_{\gamma} q^{n(\gamma)} \mathbf{e}_{\gamma'} = \sum_{\gamma} q^{n(\gamma')} \mathbf{e}_\gamma.
\end{aligned} \tag{63}$$

□

This may be used to show that the quantity $\nabla(e_n)\big|_{t=1} - \chi(\nabla(\mathbf{e}_n))\big|_{t=1}$ is e-positive (the coefficients are polynomials in q with non-negative integer coefficients when the expression is expressed in the elementary basis.

We may use property (52) and (53) to calculate the inverse of this function as well.

Proposition 28.

$$\nabla^{-1}(\mathbf{s}_\alpha) = (-1)^{\ell(\alpha)+1} \sum_{\alpha^c \leq \beta} q^{-n(\beta')} t^{-n(\beta^c)} \prod_{i \in D(\beta^c) \cap D(\alpha^c)} (t^i + q^{n-i}) \mathbf{s}_\beta. \tag{64}$$

Proof.

$$\begin{aligned}
\nabla^{-1}(\mathbf{s}_\alpha) &= \sum_{\beta \models n} (-1)^{n+\ell(\beta)} \left\langle \nabla^{-1}(\mathbf{s}_\alpha), \mathbf{s}_{\beta^c} \right\rangle \mathbf{s}_\beta \\
&= \sum_{\beta \models n} (-1)^{n+\ell(\beta)} q^{-\binom{n}{2}} t^{-\binom{n}{2}} \left\langle \mathbf{s}_\alpha, \nabla(\mathbf{s}_{\beta^c}) \right\rangle \mathbf{s}_\beta \\
&= \sum_{\beta \models n} \sum_{\gamma \leq \beta} (-1)^{\ell(\beta)+\ell(\beta^c)} q^{n(\overleftarrow{\beta}) - \binom{n}{2}} \\
&\quad t^{n(\beta)-\binom{n}{2}} \prod_{i \in D(\beta^c) \cap D(\gamma)} (t^i + q^{n-i}) \left\langle \mathbf{s}_\alpha, \mathbf{s}_\gamma \right\rangle \mathbf{s}_\beta \\
&= \sum_{\alpha^c \leq \beta} (-1)^{\ell(\alpha)+1} q^{-n(\beta')} t^{-n(\beta^c)} \prod_{i \in D(\beta^c) \cap D(\alpha^c)} (t^i + q^{n-i}) \mathbf{s}_\beta.
\end{aligned} \tag{65}$$

□

5. Appendix: Transition matrices between $\mathbf{H}^q_\alpha$ and $\mathbf{s}_\beta$

$$\left[\begin{array}{c|cc} 2 & 1 & 0 \\ 11 & q & 1 \end{array}\right]$$

$$\left[\begin{array}{c|cccc} 3 & 1 & 0 & 0 & 0 \\ 12 & q & 1 & 0 & 0 \\ 21 & q^2 & 0 & 1 & 0 \\ 111 & q^3 & q^2 & q & 1 \end{array}\right]$$

$$\begin{array}{l} 4 \\ 13 \\ 22 \\ 112 \\ 31 \\ 121 \\ 211 \\ 1111 \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ q & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ q^2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ q^3 & q^2 & q & 1 & 0 & 0 & 0 & 0 \\ q^3 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ q^4 & q^3 & 0 & 0 & q & 1 & 0 & 0 \\ q^5 & 0 & q^3 & 0 & q^2 & 0 & 1 & 0 \\ q^6 & q^5 & q^4 & q^3 & q^3 & q^2 & q & 1 \end{bmatrix}$$

6. Appendix: Transition matrices between $\widetilde{\mathbf{H}}^{qt}_{\alpha}$ and $\mathbf{s}_{\beta}$

$$\begin{array}{l} 2 \\ 11 \end{array} \begin{bmatrix} 1 & q \\ 1 & t \end{bmatrix}$$

$$\begin{array}{l} 3 \\ 12 \\ 21 \\ 111 \end{array} \begin{bmatrix} 1 & q^2 & q & q^3 \\ 1 & t & q & tq \\ 1 & q^2 & t^2 & t^2q^2 \\ 1 & t & t^2 & t^3 \end{bmatrix}$$

$$\begin{array}{l} 4 \\ 13 \\ 22 \\ 112 \\ 31 \\ 121 \\ 211 \\ 1111 \end{array} \begin{bmatrix} 1 & q^3 & q^2 & q^5 & q & q^4 & q^3 & q^6 \\ 1 & t & q^2 & tq^2 & q & tq & q^3 & tq^3 \\ 1 & q^3 & t^2 & t^2q^3 & q & q^4 & t^2q & t^2q^4 \\ 1 & t & t^2 & t^3 & q & tq & t^2q & t^3q \\ 1 & q^3 & q^2 & q^5 & t^3 & t^3q^3 & t^3q^2 & t^3q^5 \\ 1 & t & q^2 & tq^2 & t^3 & t^4 & t^3q^2 & t^4q^2 \\ 1 & q^3 & t^2 & t^2q^3 & t^3 & t^3q^3 & t^5 & t^5q^3 \\ 1 & t & t^2 & t^3 & t^3 & t^4 & t^5 & t^6 \end{bmatrix}$$

7. Appendix: Transition matrices between $\nabla(\mathbf{s}_{\alpha})$ and $\mathbf{s}_{\beta}$

$$\begin{array}{l} 2 \\ 11 \end{array} \begin{bmatrix} 0 & -qt \\ 1 & t+q \end{bmatrix}$$

$$\begin{array}{l} 3 \\ 12 \\ 21 \\ 111 \end{array} \begin{bmatrix} 0 & 0 & 0 & q^3t^3 \\ 0 & 0 & -qt^2 & -\left(t+q^2\right)qt^2 \\ 0 & -q^2t & 0 & -\left(t^2+q\right)q^2t \\ 1 & t+q^2 & t^2+q & \left(t+q^2\right)\left(t^2+q\right) \end{bmatrix}$$

Acknowledgement: Thank you to Andrew Rechnitzer and Geanina Tudose for their valuable comments. The authors would also like to thank François Bergeron for discussions about the operator ∇ in the symmetric functions.

References

[1] E. Abe, Hopf Algebras, New York, Cambridge University Press, 1980.
[2] F. Bergeron, A. Garsia, Science Fiction and Macdonald's Polynomials, Algebraic Methods and q-Special Functions, R. Floreanini, L. Vinet (eds.), CRM Proceedings & Lecture Notes, American Mathematical Society, Volume 22, (1999), 1-52.
[3] F. Bergeron, A. Garsia, M. Haiman, G. Tesler, Identities and Positivity Conjectures for Some Remarkable Operators in the Theory of Symmetric Functions, Methods and Applications of Analysis, Volume 6, No. 3, (1999), 363-420.
[4] A. Garsia, J. Haglund, A Proof of the q, t-Catalan Positivity Conjecture, preprint.
[5] I. M. Gelfand, D. Krob, B. Leclerc, A. Lascoux, V. S. Retakh, and J.-Y. Thibon, Noncommutative symmetric functions, Adv. Math. **112** (1995), 218-348.
[6] M. Haiman, Conjectures on the quotient ring by diagonal invariants, Journal of Algebraic Combinatorics #3 (1994) 17-76.
[7] F. Hivert, Hecke Algebras, Difference Operators, and Quasi-Symmetric Functions, Adv. Math. **155** (2000), 181-238.
[8] F. Hivert, A. Lascoux, and J.-Y. Thibon, Noncommutative symmetric functions and quasi-symmetric functions with two and more parameters, arXiv: `math.CO/0106191`.
[9] A. Lascoux, B. Leclerc, and J.-Y. Thibon, Hecke algebras at roots of unity and crystal bases of quantum affine algebras, Comm. Math. Phys. 181 (1997), 205-263.
[10] I. G. Macdonald, Symmetric Functions and Hall Polynomials, Oxford Mathematical Monographs, Oxford Univ. Press, second edition, 1995.
[11] C. Malvenuto and C. Reutenauer, Duality between quasi-symmetric functions and the Solomon descent algebra, *Journal of Algebra*, **177** (1995) 967–982.
[12] M. Sweedler, Hopf Algebras, New York, W.A. Benjamin, 1969.
[13] R. Stanley, Enumerative Combinatorics, Vol 1, Wadsworth & Brooks/Cole Advanced Books & Software, 1986.
[14] B. C. V. Ung, NCSF, a Maple package for Noncommutative Symmetric Functions, Maple Tech. News. 3, No. 3 (1996), 24-29.
[15] B. C. V. Ung and S. Veigneau, ACE une environnement en combinatoire algébrique, in "Proc. of the 7th Conf. Formal Power Series and Algebraic Combinatorics, 1995."
[16] M. Zabrocki, q-Analogs of Symmetric Function Operators, Discrete Mathematics (to appear).
[17] M. Zabrocki, Ribbon Operators and Hall-Littlewood Symmetric Functions, Adv. in Math., **156** (2000), pp. 33-43.

Mathematics and Statistics, York University, Toronto, ON M3J 1P3, Canada
E-mail address: `zabrocki@mathstat.yorku.ca`

THE NUMBER OF SPANNING TREES IN GRAPHS RELATED TO CIRCULANT GRAPHS (EXTENDED ABSTRACT)*

YUANPING ZHANG† AND MORDECAI J. GOLIN

Abstract. In this paper we consider the number of spanning trees in the complete graph with circulant graphs deleted (added) from (to) it. By using the properties of *Chebyshev* polynomials, we derive closed formulae for the number of spanning trees in many classes of such graphs.

Résumé. Dans cet article, nous considérons le nombre d'arbres couvrants d'un graphe complet auquel on ajoute ou on retranche des graphes circulants. En utilisant des propriétés des polynômes de *Tchebicheff,* nous obtenons des formules closes pour le nombre d'arbres couvrants d'un grand nombre de classes de tels graphes.

1. Introduction

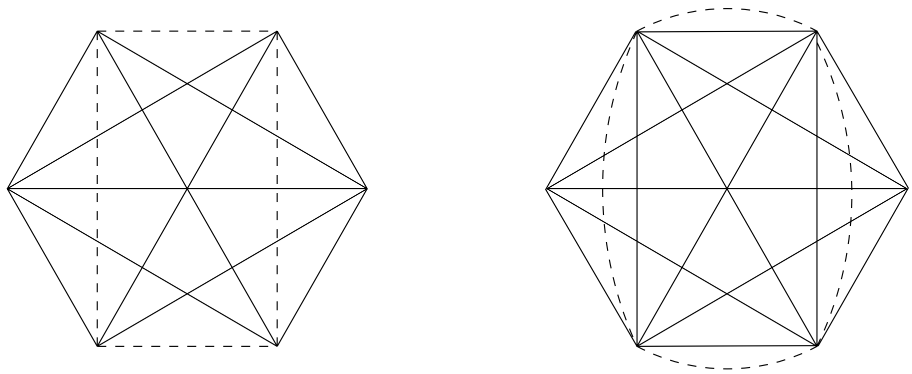
An undirected graph G is a pair (V, E), in which V is the vertex set and $E \subseteq V \times V$ is the edge set. In a graph, a *(self-)loop* is an edge joining a vertex to itself and *multiple edges* are several edges joining the same two vertices. All graphs considered in this paper are finite, and undirected with self-loops and multiple edges permitted.

The complete graph on n vertices, denoted by K_n, has one edge between each pair of distinct vertices. Let S be a subset of the edge set of K_n (or S be a subgraph of K_n). We denote by $K_n - S$ the graph remaining when all edges in S are removed from K_n. If S is a subgraph of K_n, $K_n - S$ is called the *complement* graph of S in K_n, and also denoted as $\overline{S}$. For an edge set S, we denote by $K_n + S$ the graph K_n with all edges in S added to it; if S is nonempty then $K_n + S$ is a graph containing some multiple edges. In Figure 1 we give examples of two graphs, one is the graph $K_6 - C_4$ which is the complete graph K_6 with a cycle of 4 edges C_4 deleted from it; another is the graph $K_6 + C_4$ which is K_6 with a cycle C_4 added to it. In the graph $K_6 - C_4$, the dashed lines are edges deleted; in the graph $K_6 + C_4$, the dashed lines are edges added.

A spanning tree in a graph G is a tree which has the same vertex set with G. The number of spanning trees in a graph (network) G is an important quantity to measure the reliability of G [8]. For graph G, the number of spanning trees in G is denoted as $T(G)$. The problem of calculating $T(K_n - S)$ has already been studied for many different types of S. The initial work seems to have been by *Weinberg* [18] who gave formulae for $T(K_n - S)$ when all edges in S are not adjacent or are adjacent at one vertex. Subsequently, in a series of papers [1, 2, 3, 4], *Bedrosian* extended this to show how to calculate $T(K_n - S)$ when all edges in S are not adjacent or adjacent at one vertex, or form a path, a cycle, a complete graph, or are some combination of these configurations. *Weinberg*'s results have also been generalized in [15]. Closed formulae also exist for the cases where S is a star [14], a complete K-partite graph [16], a multi-star [13, 19], and so on. The number of spanning trees in

* This work was supported by Hong Kong CERG grants HKUST6082/97E, HKUST6137/98E, HKUST6162/00E and HKUST DAG98/99.EG23.

† Also at the Department of Mathematics, Hunan Normal University, Changsha, 410081, PRC.



$K_6 - C_4^1$ $K_6 + C_4^1$

FIGURE 1. Two examples of graphs.

the complement graph is investigated in [9, 11] when the graph with maximal number of spanning trees is studied. The formulae for the number of spanning trees in the complement graphs of a disjoint union of cycles or paths are given in generic forms in [9]. Not as much is known about $T(K_n + S)$; *Bedrosian* [2] considered it for some simple configurations S, i.e., all edges in S form a cycle, complete graph, or $|S|$ is quite small but much more does not seem to be known.

In this paper we describe how to calculate $T(K_n \pm S)$ where S is a circulant graph. Let $1 \le s_1 < s_2 < \cdots < s_k$. The *undirected circulant graph*, $C_n^{s_1,s_2,\cdots,s_k}$, has n vertices labeled $0,\ 1,\ 2,\ \cdots,\ n-1$, with each vertex i $(0 \le i \le n-1)$ adjacent to $2k$ vertices $i \pm s_1,\ i \pm s_2,\ \cdots,\ i \pm s_k \bmod n$. Note that as a special case, C_n^1 is simply the n node cycle. More generally, if $(n,\ s) = 1$ then C_n^s is the n node cycle while if $(n,\ s) = d > 1$ then C_n^s is the disjoint union of d cycles $C_{n/d}^1$.

The number of spanning trees in circulant graphs has been well-studied (see [20] for exposition). But, other than for the graphs $K_n - S$ as mentioned above where S is a cycle or a disjoint union of cycles, there does not seem to have been any other work about the number of spanning trees in a complete graph with a circulant graph deleted (added) from (to) it, especially, no closed formulae for the number of spanning trees in a complete graph with a arbitrary circulant graphs deleted (added) from (to) it. This is the problem to be considered in this paper.

In Section 2 we state some lemmas about $T(G)$ and review some properties of *Chebyshev* polynomials. In Section 3 we use these lemmas and properties to derive a series of formulae for the numbers of spanning trees in the complete graphs with circulant graphs deleted (added) from (to) them. Our approach is to first start by developing a new approach to deriving a closed form for $T(K_n - C_m^s)$, where C_m^s is a cycle or union of cycles (a closed form for this was previously derived using different techniques in [9]). We then continue by showing that it is easy to generalize this approach to getting a formula for $T(K_n \pm C_m^{s_1,s_2,\cdots,s_k})$. In the case that all of the $s_i \le 4$ we will actually be able to derive a simple closed form function $g(n, m; s_1, s_2, \cdots, s_k) = T(K_n \pm C_m^{s_1,s_2,\cdots,s_k})$ of $n,\ m$. We conclude in Section 4 by describing extensions and limitations of our technique.

2. Basic concepts and lemmas

For graph G, we denote by $A(G)$ or A the adjacency matrix of $G = (V, E)$. If $V = \{v_1,\ v_2,\ \cdots,\ v_n\}$, A is the $n \times n$ matrix with $a_{i,j}$ being the number of edges connecting

v_i and v_j. The degree, d_i, of vertex v_i is $\sum_{j=1}^{n} a_{i,j}$, the number of edges adjacent to v_i. Let B denote the diagonal matrix with $\{d_1,\ d_2,\ \cdots,\ d_n\}$ as diagonal entries. The classic result known as the *Matrix Tree Theorem* [12] states that, the *Kirchhoff* matrix H defined as $H = B - A$ has all its co-factors[1] equal to $T(G)$. For example, the *Kirchhoff* matrix H of the graph $K_6 - C_4$ shown in Figure 1 is

$$H = \begin{pmatrix} 3 & 0 & -1 & -1 & 0 & -1 \\ 0 & 3 & -1 & 0 & -1 & -1 \\ -1 & -1 & 5 & -1 & -1 & -1 \\ -1 & 0 & -1 & 3 & 0 & -1 \\ 0 & -1 & -1 & 0 & 3 & -1 \\ -1 & -1 & -1 & -1 & -1 & 5 \end{pmatrix},$$

all its co-factors are 192 which is the number of spanning trees in $K_6 - C_4$.

The number of spanning trees in graph G also can be calculated from the eigenvalues of the *Kirchhoff* matrix H. All eigenvalues of a real symmetric matrix are real. Even more, by basic knowledge of linear algebra, all eigenvalues of the *Kirchhoff* matrix H of a graph with n nodes are non-negative, and 0 is one of its eigenvalues because all its row vectors add up to 0 vector. So we can let $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n (= 0)$ denote all eigenvalues of H. *Kel'mans* and *Chelnokov* [11] have shown that

$$T(G) = \frac{1}{n} \prod_{j=1}^{n-1} \mu_j. \tag{1}$$

In order to use this equation we will need to know the eigenvalues of the *Kirchhoff* matrices of the appropriate graphs. This will require the following lemmas:

Lemma 1. (Biggs [5], Page 16) *The* Kirchhoff *matrix of the circulant graph* $C_n^{s_1,s_2,\cdots,s_k}$ *has* n *eigenvalues as* $2k - \varepsilon^{-s_1 j} - \cdots - \varepsilon^{-s_k j} - \varepsilon^{s_1 j} - \cdots - \varepsilon^{s_k j}$, $1 \leq j \leq n-1$ *and* 0, *where* ε^{-j} *is the conjugate of* ε^j, $\varepsilon = e^{\frac{2\pi i}{n}}$.

Lemma 2. ([11]) *Let* G *be a graph with* n *vertices and* $\overline{G}$ *the complement graph of* G *in* K_n. *If the* Kirchhoff *matrix of* G *has eigenvalues* $\mu_1,\ \mu_2,\ \cdots,\ \mu_{n-1}$ *and* 0, *then the* Kirchhoff *matrix of* $\overline{G}$ *has eigenvalues* $n - \mu_1,\ n - \mu_2,\ \cdots,\ n - \mu_{n-1}$ *and* 0.

Using basic multilinear algebra the following lemma can be proven in a way similar to that of the proof of Lemma 2.

Lemma 3. G *is a graph with same vertex set as* K_n. *If the* Kirchhoff *matrix of* G *has eigenvalues* $\mu_1,\ \mu_2,\ \cdots,\ \mu_{n-1}$ *and* 0, *then the* Kirchhoff *matrix of* $K_n + G$ *has eigenvalues* $n + \mu_1,\ n + \mu_2,\ \cdots,\ n + \mu_{n-1}$ *and* 0.

Let $G_1 = (V_1,\ E_1)$ and $G_2 = (V_2,\ E_2)$ be two graphs with disjoint vertex sets. The join $G = G_1 \bigoplus G_2$ is defined as the graph with vertex set $V = V_1 \cup V_2$ and edge set $E = E_1 \cup E_2 \cup \{uv | u \in V_1,\ v \in V_2\}$ [7]. (Please note that in this paper we use "$\bigoplus$" to denote the join graph instead of "+" as used in some other references. This is because we are already using "+" to denote the graph that results by adding edges to some other graph.) The following lemma describes the relation of the eigenvalues of the *Kirchhoff* matrix of the join graph and the eigenvalues of *Kirchhoff* matrices of the original graphs.

Lemma 4. ([10, 11]) *If the* Kirchhoff *matrix of graph* G_1 *with* n *vertices has eigenvalues* $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n (= 0)$ *and that of graph* G_2 *with* m *vertices has eigenvalues* $\mu_1,\ \mu_2,\ \cdots,$

[1] The (i, j)th cofactor of A is the determinant of the $(n-1) \times (n-1)$ matrix that results from deleting the ith row and jth column from A.

$\mu_m(=0)$, *then the* Kirchhoff *matrix of the join* $G_1 \bigoplus G_2$ *has eigenvalues* $m+n$, λ_1+m, $\cdots$, $\lambda_{n-1}+m$ *and* μ_1+n, $\cdots$, $\mu_{m-1}+n$, 0.

Let $C_{m_1}^{s_{1_1},s_{2_1},\cdots,s_{k_1}}$, $C_{m_2}^{s_{1_2},s_{2_2},\cdots,s_{k_2}}$, $\cdots$, $C_{m_l}^{s_{1_l},s_{2_l},\cdots,s_{k_l}}$ be a collection of circulant graphs, and $\bigcup_{u=1}^{l} C_{m_u}^{s_{1_u},s_{2_u},\cdots,s_{k_u}}$ be their disjoint union. For each u, $1 \le u \le l$, suppose $m_u > 2s_{k_u}$ and let $\overline{C_{m_u}^{s_{1_u},s_{2_u},\cdots,s_{k_u}}}$ be the complement graph of $C_{m_u}^{s_{1_u},s_{2_u},\cdots,s_{k_u}}$ in K_{m_u}. Note that, for any n, $n \ge \sum_{u=1}^{l} m_u$,

$$\begin{aligned}
&K_n - \bigcup_{u=1}^{l} C_{m_u}^{s_{1_u},s_{2_u},\cdots,s_{k_u}} \\
&= \left(K_{n-\sum_{u=1}^{l} m_u}\right) \bigoplus \left(K_{m_1} - C_{m_1}^{s_{1_1},s_{2_1},\cdots,s_{k_1}}\right) \bigoplus \cdots \bigoplus \left(K_{m_l} - C_{m_l}^{s_{1_l},s_{2_l},\cdots,s_{k_l}}\right) \\
&= \left(K_{n-\sum_{u=1}^{l} m_u}\right) \bigoplus \overline{C_{m_1}^{s_{1_1},s_{2_1},\cdots,s_{k_1}}} \bigoplus \cdots \bigoplus \overline{C_{m_l}^{s_{1_l},s_{2_l},\cdots,s_{k_l}}}.
\end{aligned}$$

So, by Lemma 1, Lemma 2, Lemma 4 and (1), we have the following result.

Corollary 1. *For* $n \ge \sum_{u=1}^{l} m_u$ *and for each* u, $1 \le u \le l$, $m_u > 2s_{k_u}$,

$$\begin{aligned}
&T(K_n - \bigcup_{u=1}^{l} C_{m_u}^{s_{1u},s_{2u},\cdots,s_{ku}}) \\
&= n^{n-\sum_{u=1}^{l} m_u + l - 2} \prod_{u=1}^{l} \prod_{j=1}^{m_u-1} (n - 2k_u + \varepsilon_u^{-s_{1u}j} + \cdots + \varepsilon_u^{-s_{ku}j} + \varepsilon_u^{s_{1u}j} + \cdots + \varepsilon_u^{s_{ku}j}),
\end{aligned}$$

where $\varepsilon_u = e^{\frac{2\pi i}{m_u}}$, *for each* u, $1 \le u \le l$.

In a similar fashion, the following corollary can be derived from Lemma 1, Lemma 3, Lemma 4 and (1):

Corollary 2. *For* $n \ge \sum_{u=1}^{l} m_u$,

$$\begin{aligned}
&T(K_n + \bigcup_{u=1}^{l} C_{m_u}^{s_{1u},s_{2u},\cdots,s_{ku}}) \\
&= n^{n-\sum_{u=1}^{l} m_u + l - 2} \prod_{u=1}^{l} \prod_{j=1}^{m_u-1} (n + 2k_u - \varepsilon_u^{-s_{1u}j} - \cdots - \varepsilon_u^{-s_{ku}j} - \varepsilon_u^{s_{1u}j} - \cdots - \varepsilon_u^{s_{ku}j}),
\end{aligned}$$

where $\varepsilon_u = e^{\frac{2\pi i}{m_u}}$, *for each* u, $1 \le u \le l$.

In order to evaluate these products we will show how to transform them into functions of *Chebyshev* polynomials; to do this we will need some special properties of these polynomials [6, 17]. The properties, listed below, are taken from [6] which used them to derive the number of spanning trees of some special classes of graphs, e.g., wheels, fans, ladders, *Moebius* ladders, squares of cycles and complete prisms.

The properties of the *Chebyshev* polynomials [6, 17] are used also in this paper. The following definitions and derivations are from [6]. For positive integer m, the *Chebyshev* polynomials of the first kind are defined by

$$T_m(x) = \cos(m \arccos x). \tag{2}$$

The *Chebyshev* polynomials of the second kind are defined by

$$U_{m-1}(x) = \frac{1}{m}\frac{d}{dx}T_m(x) = \frac{\sin(m \arccos x)}{\sin(\arccos x)}. \tag{3}$$

It's easily verified that

$$U_m(x) - 2xU_{m-1}(x) + U_{m-2}(x) = 0. \tag{4}$$

Solving this recursion by using standard methods, the following explicit formula is obtained

$$U_m(x) = \frac{1}{2\sqrt{x^2-1}}[(x+\sqrt{x^2-1})^{m+1} - (x-\sqrt{x^2-1})^{m+1}]. \tag{5}$$

The definition of $U_m(x)$ easily yields its zeros and it's verified that

$$U_{m-1}(x) = 2^{m-1}\prod_{j=1}^{m-1}[x - \cos(j\pi/m)]. \tag{6}$$

Further one notes that

$$U_{m-1}(-x) = (-1)^{m-1}U_{m-1}(x). \tag{7}$$

These two results yield another formula for $U_m(x)$

$$U_{m-1}^2(x) = 4^{m-1}\prod_{j=1}^{m-1}[x^2 - \cos^2(j\pi/m)]. \tag{8}$$

3. Results

We are now ready to derive the main result of this paper, a way to calculate $T(K_n \pm S)$ when S is a circulant graph. We start by assuming that $S = C_m^s$. As previously noted, if $(m,s) = 1$ this is just the m-cycle and if $(m,s) = d > 1$ this is the disjoint union of d cycles, each of length m/d.

Before proceeding we note that *Gilbert* and *Myrvold* [9] already gave a formula for the number of spanning trees in the graph $K_n - S$ where S is the disjoint union of cycles. The following theorem can actually be derived from *Gilbert* and *Myrvold*'s formula. The proof here is new, though; we derive it since it provides a 'pure' way of illustrating the techniques we will use later.

Theorem 5. *For $n \geq m > 2s$, if $(m,\ s) = d$, then*

$$T(K_n - C_m^s) = n^{n-m-2}\left[\left(\sqrt{\frac{n}{4}} + \sqrt{\frac{n-4}{4}}\right)^{m/d} - \left(-\sqrt{\frac{n}{4}} + \sqrt{\frac{n-4}{4}}\right)^{m/d}\right]^{2d}.$$

Proof. Let $\varepsilon_1 = e^{\frac{2d\pi i}{m}}$. If $(m,\ s) = d$, then C_m^s is the disjoint union of d cycles $C_{m/d}^1$. So, by Corollary 1, we have

$$
\begin{aligned}
T(K_n - C_m^s) &= T(K_n - \bigcup_{u=1}^{d} C_{m/d}^1) \\
&= n^{n-m+d-2} \prod_{u=1}^{d} \prod_{j=1}^{\frac{m}{d}-1} (n - 2 + \varepsilon_1^{-j} + \varepsilon_1^{j}) \\
&= n^{n-m+d-2} \prod_{u=1}^{d} \prod_{j=1}^{\frac{m}{d}-1} (n - 2 + 2\cos(2dj\pi/m)) \\
&= n^{n-m+d-2} \prod_{u=1}^{d} \left[(-4)^{\frac{m}{d}-1} \prod_{j=1}^{\frac{m}{d}-1} \left(\frac{-n+4}{4} - \cos^2(dj\pi/m) \right) \right],
\end{aligned}
$$

where we are using the fact that $1 + \cos(2x) = 2\cos^2 x$.

Applying the formulas (8) and then (5) yields the required

$$
\begin{aligned}
T(K_n - C_m^s) &= n^{n-m+d-2} \prod_{u=1}^{d} \left[(-1)^{\frac{m}{d}-1} U^2_{\frac{m}{d}-1} \left(\sqrt{\frac{-n+4}{4}} \right) \right] \\
&= n^{n-m-2} \left[\left(\sqrt{\frac{n}{4}} + \sqrt{\frac{n-4}{4}} \right)^{m/d} - \left(-\sqrt{\frac{n}{4}} + \sqrt{\frac{n-4}{4}} \right)^{m/d} \right]^{2d}.
\end{aligned}
$$

□

As a first consequence of Theorem 5 we can easily derive:

Corollary 3.

$$
\begin{aligned}
T(K_n - C_3^1) &= n^{n-4}(n-3)^2,\ n \geq 3; \\
T(K_n - C_4^1) &= n^{n-5}(n-2)^2(n-4),\ n \geq 4; \\
T(K_n - C_5^1) &= n^{n-6}(n^2 - 5n + 5)^2,\ n \geq 5; \\
T(K_n - C_6^1) &= n^{n-7}(n-1)^2(n-3)^2(n-4),\ n \geq 6; \\
T(K_n - C_6^2) &= n^{n-6}(n-3)^4,\ n \geq 6.
\end{aligned}
$$

The first four formulae of the above corollary already appear in [4] where they are given in generic forms and derived from *Kel'mans* and *Chelnokov*'s result (1) by direct computation.

The proof above illustrates our general tools. We now see how to apply them when looking at the complement of a more complicated circulant graph.

Theorem 6. *For $n \geq m > 4$,*

$$
\begin{aligned}
T(K_n - C_m^{1,2}) = n^{n-m-2} &\left[\left(x_1 + \sqrt{x_1^2 - 1} \right)^m - \left(x_1 - \sqrt{x_1^2 - 1} \right)^m \right]^2 \cdot \\
&\left[\left(x_2 + \sqrt{x_2^2 - 1} \right)^m - \left(x_2 - \sqrt{x_2^2 - 1} \right)^m \right]^2,
\end{aligned}
$$

where $x_1 = \sqrt{\frac{3}{8} + \frac{1}{8}\sqrt{25 - 4n}}$, $x_2 = \sqrt{\frac{3}{8} - \frac{1}{8}\sqrt{25 - 4n}}$.

Proof. We use a very similar technique to the proof of Theorem 5. In this proof let $\varepsilon_1 = e^{\frac{2\pi i}{m}}$, and x_1, x_2 be defined as above. Then

$$\begin{aligned}
T(K_n - C_m^{1,2}) &= n^{n-m-1}\prod_{j=1}^{m-1}(n-4+\varepsilon_1^{-j}+\varepsilon_1^{-2j}+\varepsilon_1^{j}+\varepsilon_1^{2j})\\
&= n^{n-m-1}\prod_{j=1}^{m-1}(n-4-12\cos^2(j\pi/m)+16\cos^4(j\pi/m))\\
&= n^{n-m-1}16^{m-1}\prod_{j=1}^{m-1}(x_1^2-\cos^2(j\pi/m))((x_2^2-\cos^2(j\pi/m))\\
&= n^{n-m-1}U_{m-1}^2(x_1)U_{m-1}^2(x_2).
\end{aligned}$$

The closed formula in the theorem statement follows from (8) and then (5). □

As an application, Theorem 6 can directly implies the following formulae.

Corollary 4.

$$\begin{aligned}
T(K_n - C_5^{1,2}) &= n^{n-6}(n-5)^4,\ n \ge 5;\\
T(K_n - C_6^{1,2}) &= n^{n-7}(n-6)^2(n-4)^3,\ n \ge 6;\\
T(K_n - C_7^{1,2}) &= n^{n-8}(n^3-14n^2+63n-91)^2,\ n \ge 7.
\end{aligned}$$

We now examine the complement of a slightly more complicated circulant graph.

Theorem 7. *For $n \ge m > 8$, if m is odd, then*

$$T(K_n - C_m^{2,4}) = T(K_n - C_m^{1,2});$$

Otherwise, if m is even, then

$$\begin{aligned}
T(K_n - C_m^{2,4}) &= n^{n-m-2}\left[\left(x_1+\sqrt{x_1^2-1}\right)^{m/2}-\left(x_1-\sqrt{x_1^2-1}\right)^{m/2}\right]^4 \cdot\\
&\quad\left[\left(x_2+\sqrt{x_2^2-1}\right)^{m/2}-\left(x_2-\sqrt{x_2^2-1}\right)^{m/2}\right]^4,
\end{aligned}$$

where x_1 and x_2 are as defined in Theorem 6.

Proof. If m is odd then $C_m^{2,4}$ is isomorphic to $C_m^{1,2}$ so the result of Theorem 6 applies. If m is even $C_m^{2,4}$ is the disjoint union of 2 circulant graphs $C_{m/2}^{1,2}$. The proof in this case is just to combine Corollary 1 and the proof of Theorem 6. When m is even then let $\varepsilon_2 = e^{\frac{4\pi i}{m}}$, we have

$$\begin{aligned}
T(K_n - C_m^{2,4}) &= T(K_n - C_{m/2}^{1,2} \bigcup C_{m/2}^{1,2}) \\
&= n^{n-m} \left(\prod_{j=1}^{\frac{m}{2}-1} (n - 4 + \varepsilon_2^{-j} + \varepsilon_2^{-2j} + \varepsilon_2^{j} + \varepsilon_2^{2j}) \right)^2 \\
&= n^{n-m} \left(\prod_{j=1}^{\frac{m}{2}-1} (n - 4 - 12\cos^2(2j\pi/m) + 16\cos^4(2j\pi/m)) \right)^2 \\
&= n^{n-m} \left(16^{\frac{m}{2}-1} \prod_{j=1}^{\frac{m}{2}-1} (x_1^2 - \cos^2(2j\pi/m))((x_2^2 - \cos^2(2j\pi/m)) \right)^2 \\
&= n^{n-m} U^4_{\frac{m}{2}-1}(x_1) U^4_{\frac{m}{2}-1}(x_2).
\end{aligned}$$

□

Corollary 5.

$$\begin{aligned}
T(K_n - C_9^{2,4}) &= n^{n-10}(n-6)^2(n^3 - 12n^2 + 45n - 51)^2, \ n \geq 9; \\
T(K_n - C_{10}^{2,4}) &= n^{n-10}(n-5)^8, \ n \geq 10; \\
T(K_n - C_{11}^{2,4}) &= n^{n-12}(n^5 - 22n^4 + 187n^3 - 759n^2 + 1441n - 979)^2, \ n \geq 11.
\end{aligned}$$

We now discuss the general technique for calculating $T(K_n - C_m^{s_1,s_2,\cdots,s_k})$ when $\gcd(s_1, s_2, \cdots, s_k, m) = 1$ (the case $\gcd(s_1, s_2, \cdots, s_k, m) \neq 1$ can then be dealt with similarly to the case "m is even" in the proof of Theorem 7). In the following paragraphs let $\varepsilon = e^{\frac{2\pi i}{m}}$, then from Corollary 1,

$$\begin{aligned}
&T(K_n - C_m^{s_1,s_2,\cdots,s_k}) \\
&= n^{n-m-1} \prod_{j=1}^{m-1} (n - 2k + \varepsilon^{-s_1 j} + \varepsilon^{-s_2 j} + \cdots + \varepsilon^{-s_k j} + \varepsilon^{s_1 j} + \varepsilon^{s_2 j} + \cdots + \varepsilon^{s_k j}) \\
&= n^{n-m-1} \prod_{j=1}^{m-1} (n - 2k + 2\cos(2s_1\pi/m) + 2\cos(2s_2\pi/m) + \cdots + 2\cos(2s_k\pi/m)).
\end{aligned}$$

It is easy to prove by induction that $\cos(kx)$ can be expressed as a polynomial in $\cos x$ of order k. Using this fact, for any integer s, $\cos(2sj\pi/m)$ can be written as a polynomial in $\cos^2(j\pi/m)$ of order s. So,

$$T(K_n - C_m^{s_1,s_2,\cdots,s_k}) = n^{n-m-1} \prod_{j=1}^{m-1} (n - 2k + g(\cos^2(j\pi/m))),$$

where $g(x)$ is a polynomial of order s_k. Thus,

$$T(K_n - C_m^{s_1,s_2,\cdots,s_k}) = n^{n-m-1} \prod_{j=1}^{m-1} f(\cos^2(\pi j/m)),$$

where $f(x)$ is a polynomial with degree s_k and the constant term is a linear function of n. Even more, by explicit calculation we can see that the coefficient of x^{s_k} in $f(x)$ is 4^{s_k}. We can therefore write

$$f(x) = (-4)^{s_k} \prod_{i=1}^{s_k} (x_i - \cos^2(\pi j/m)),$$

where $x_1, x_2, \cdots, x_k$ are zeros of $f(x)$. Then, combining formula (8) with last two equations we have

$$T(K_n - C_m^{s_1,s_2,\cdots,s_k}) = (-1)^{s_k} n^{n-m-1} \prod_{i=1}^{s_k} U_{m-1}^2(\sqrt{x_i}).$$

From this equation the exact formula for $T(K_n - C_m^{s_1,s_2,\cdots,s_k})$ can be derived from the properties of *Chebyshev* polynomials.

In the special cases $s_k \leq 4$, the polynomial $f(x)$ can be explicitly factored and an exact formula for the number of spanning trees in $K_n - C_m^{s_1,s_2,\cdots,s_k}$ as a function of n, m can therefore be derived. We have done this and the details are included in the full paper; in this extended abstract we omit these formulae.

From Corollary 2 and the properties of *Chebyshev* polynomials, we also can derive the following closed formulae for the numbers of spanning trees in complete graphs with circulant graphs added. As before we start by adding C_m^s.

Theorem 8. *For $n \geq m$, if $(m, s) = d$, then*

$$T(K_n + C_m^s) = n^{n-m-2} \left[\left(\sqrt{\frac{n+4}{4}} + \sqrt{\frac{n}{4}} \right)^{m/d} - \left(\sqrt{\frac{n+4}{4}} - \sqrt{\frac{n}{4}} \right)^{m/d} \right]^{2d}.$$

Proof. It's similar to the proof of Theorem 5. By Corollary 2, we have

$$\begin{aligned} T(K_n + C_m^s) &= T(K_n + \bigcup_{u=1}^{d} C_{m/d}^1) \\ &= n^{n-m+d-2} \prod_{u=1}^{d} \prod_{j=1}^{\frac{m}{d}-1} (n + 2 - \varepsilon_1^{-j} - \varepsilon_1^{j}) \\ &= n^{n-m+d-2} \prod_{u=1}^{d} \prod_{j=1}^{\frac{m}{d}-1} (n + 2 - 2\cos(2dj\pi/m)) \\ &= n^{n-m+d-2} \prod_{u=1}^{d} \left[4^{\frac{m}{d}-1} \prod_{j=1}^{\frac{m}{d}-1} \left(\frac{n+4}{4} - \cos^2(dj\pi/m) \right) \right]. \end{aligned}$$

By using the formulae (8) and then (5),

$$\begin{aligned} T(K_n + C_m^s) &= n^{n-m+d-2} \prod_{u=1}^{d} \left[U_{\frac{m}{d}-1}^2 \left(\sqrt{\frac{n+4}{4}} \right) \right] \\ &= n^{n-m-2} \left[\left(\sqrt{\frac{n+4}{4}} + \sqrt{\frac{n}{4}} \right)^{m/d} - \left(\sqrt{\frac{n+4}{4}} - \sqrt{\frac{n}{4}} \right)^{m/d} \right]^{2d}. \end{aligned}$$

□

This then immediately gives us, for example,

Corollary 6.

$$\begin{aligned}
T(K_n + C_2^1) &= n^{n-3}(n+4),\ n \geq 2;\\
T(K_n + C_3^1) &= n^{n-4}(n+3)^2,\ n \geq 3;\\
T(K_n + C_4^1) &= n^{n-5}(n+4)(n+2)^2,\ n \geq 4;\\
T(K_n + C_4^2) &= n^{n-4}(n+4)^2,\ n \geq 4;\\
T(K_n + C_5^1) &= n^{n-6}(n^2+5n+5)^2,\ n \geq 5.
\end{aligned}$$

We can also derive results for general $K_n + C_m^{s_1,s_2,\cdots,s_k}$ that are analogous to the ones previously derived for $K_n - C_m^{s_1,s_2,\cdots,s_k}$. Since the proofs are so similar, we omit them.

Theorem 9. *For $n \geq m$,*

$$\begin{aligned}
T(K_n + C_m^{1,2}) = {} & (-1)^m n^{n-m-2}\left[\left(x_1+\sqrt{x_1^2-1}\right)^m - \left(x_1-\sqrt{x_1^2-1}\right)^m\right]^2 \cdot\\
& \left[\left(x_2+\sqrt{x_2^2-1}\right)^m - \left(x_2-\sqrt{x_2^2-1}\right)^m\right]^2,
\end{aligned}$$

where $x_1 = \sqrt{\frac{3}{8}+\frac{1}{8}\sqrt{25+4n}}$, $x_2 = \sqrt{\frac{3}{8}-\frac{1}{8}\sqrt{25+4n}}$.

Corollary 7.

$$\begin{aligned}
T(K_n + C_3^{1,2}) &= n^{n-4}(n+6)^2,\ n \geq 3;\\
T(K_n + C_4^{1,2}) &= n^{n-5}(n+4)(n+6)^2,\ n \geq 4;\\
T(K_n + C_5^{1,2}) &= n^{n-6}(n+5)^2,\ n \geq 5;\\
T(K_n + C_6^{1,2}) &= n^{n-7}(n+6)^2(n+4)^3,\ n \geq 6;\\
T(K_n + C_7^{1,2}) &= n^{n-8}(n^3+14n^2+63n+91)^2,\ n \geq 7.
\end{aligned}$$

Theorem 10. *For $n \geq m$, if m is odd, then*

$$T(K_n + C_m^{2,4}) = T(K_n + C_m^{1,2});$$

Otherwise m is even, then

$$\begin{aligned}
T(K_n + C_m^{2,4}) = {} & n^{n-m-2}\left[\left(x_1+\sqrt{x_1^2-1}\right)^{m/2} - \left(x_1-\sqrt{x_1^2-1}\right)^{m/2}\right]^4 \cdot\\
& \left[\left(x_2+\sqrt{x_2^2-1}\right)^{m/2} - \left(x_2-\sqrt{x_2^2-1}\right)^{m/2}\right]^4,
\end{aligned}$$

where x_1 and x_2 are defined as in Theorem 9.

Corollary 8.

$$\begin{aligned}
T(K_n + C_6^{2,4}) &= n^{n-6}(n+6)^4,\ n \geq 6;\\
T(K_n + C_7^{2,4}) &= n^{n-8}(n^3+14n^2+63n+91)^2,\ n \geq 7;\\
T(K_n + C_8^{2,4}) &= n^{n-8}(n+4)^2(n+6)^4,\ n \geq 8;\\
T(K_n + C_9^{2,4}) &= n^{n-10}(n+6)^2(n^3+12n^2+45n+51)^2,\ n \geq 9.
\end{aligned}$$

4. Conclusion

In this paper we discussed how to use properties of *Chebyshev* polynomials to derive closed formulae for the number of spanning trees in $K_n \pm S$ where $S = C_m^{s_1,s_2,\cdots,s_k}$ is a circulant graph. Our key step was to factorize a polynomial of order s_k and then express the number of spanning trees in terms of *Chebyshev* polynomials evaluated at functions of the roots of the polynomial. In particular, when $s_k \leq 4$, we could explicitly factorize the polynomial and derive a"closed" form for the number of spanning trees.

One last thing that we should point it is that, in all the formulae we derived, we assumed that $s_1 < s_2 < \cdots < s_k$. This was just for the sake of convenience, though, and was not necessary for our proofs. The technique still works for reapeated s_i values, e.g., we could use it to evaluate $T(K_n + C_m^{1,1})$ $(m \leq n)$ where $C_m^{1,1}$ is the doubly-linked cycle.

References

[1] S. D. Bedrosian, Formulas for the number of trees in a networks, CT-8, (1961) 363-364.

[2] S. D. Bedrosian, Generating formulas for the number of trees in a graph, J. Franklin Inst., 277 (1964) 313-326.

[3] S. D. Bedrosian, Formulas for the number of trees in certain incomplete graphs, J. Franklin Inst., 289 (1970) 67-69.

[4] S. D. Bedrosian, Tree counting polynomials for labelled graphs, J. Franklin Inst., 312 (1981) 417-430.

[5] N. Biggs, Algebraic Graph Theory, London: Cambridge University Press, Second Edition, 1993.

[6] F. T. Boesch, H. Prodinger, Spanning Tree Formulas and Chebyshev Polynomials, Graph and Combinatorics 2 (1986) 191-200.

[7] G. Chartrand, L. Lesniak, Graphs & digraphs, 3rd edition, Chapman & Hall, 1996.

[8] D. Cvetkovič, M. Doob, H. Sachs, Spectra of Graphs: Theory and Applications, Third Edition, Johann Ambrosius Barth, Heidelberg, 1995.

[9] B. Gilbert, W. Myrvold, Maximizing spanning trees in almost complete graphs, Networks 30 (1997) 97-104.

[10] Z. J. Huang, X. M. Li, A general method for finding the number of spanning trees of some types of composite graphs, (Chinese) Acta Math. Sci. 15 (3)(1995) 259-268.

[11] A. K. Kel'mans, V. M. Chelnokov, A certain polynomial of a graph and graphs with an extremal number of trees, J. Comb. Theory (B) 16(1974) 197-214.

[12] G. Kirchhoff, Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird, Ann. Phys. Chem. 72 (1847) 497-508.

[13] S. D. Nikolopoulos, P. Rondogiannis, On the number of spanning trees of multi-star related graph, Inform. Process. Lett. 65 (1998) 183-188.

[14] P. V. O'Neil, The number of spanning trees in a certain network, Notices Amer. Math. Soc. 10 (1963) 569.

[15] P. V. O'Neil, P. Slepian, The number of trees in a network, IEEE Trans. Circuit Theory, CT-13 (1966) 271-281.

[16] P. V. O'Neil, Enumeration of spanning trees in certain graphs, IEEE Trans. Circuit Theory, CT-17 (1970) 250.

[17] M. Snyder, *Chebyshev* Methods in Numerical Analysis, Englewood Cliffs, Prentice-Hall, 1966.

[18] L. Weinberg, Number of trees in graph, Proc. IRE, 46 (1958) 1954-1955.

[19] W. M. Yan, W. Myrvold, K. L. Chung, A formula for the number of spanning trees of a multi-star related graph, Inform. Process. Lett. 68 (1998) 295-298.

[20] Y. P. Zhang, X. R. Yong, M. J. Golin, The number of spanning trees in circulant graphs, Discrete Math. 223 (2000) 337-350.

Department of Computer Science, The Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong

E-mail address: {ypzhang, golin}@cs.ust.hk

Author index