Rewrite systems on free clones and realizations of algebraic structures

Samuele Giraudo

LIGM, Université Gustave Eiffel

Séminaire PPS
IRIF, Paris

February 18, 2021

Outline

1. Algebraic combinatorics
2. Universal algebra and terms
3. Equivalence of terms and rewrite systems
4. Composition and clones
5. Clones of colored words
6. Appendix

Outline

1. Algebraic combinatorics

Combinatorial collections

A combinatorial collection is a set C endowed with a map

$$|-|: C \to \mathbb{N}$$

such that for any $n \in \mathbb{N}$, $C(n) := \{x \in C : |x| = n\}$ is finite.

For any $x \in C$, we call $|x|$ the size of x.

– Classical questions –

1. Enumerate the objects of C of size n.
2. Generate all the objects of C of size n.
3. Randomly generate an object of C of size n.
4. Establish transformations between C and other combinatorial collections D.

Operations and algebraic structures

– Main idea –

Endow C with operations to form an algebraic structure.

The algebraic study of C helps to discover combinatorial properties.

In particular,

1. minimal generating families of C
   $\rightsquigarrow$ highlighting of elementary pieces of assembly;

2. morphisms involving C
   $\rightsquigarrow$ transformation algorithms and revelation of symmetries.

Some algebraic structures arising in this context are

- monoids;
- groups;
- lattices;
- associative alg.;
- Hopf bialg.;
- Lie alg.;
- pre-Lie alg.;
- dendriform alg.;
- duplicial alg.

Some connected fields

(A) Enumerative combinatorics:

- Pattern avoidance in trees;
- Colored operads and generation;
- Generalized formal power series.

(B) Constructions of algebraic structures:

- Operads, clones*, and pros ;
- Posets and lattices;
- Universal algebra*.

(C) Algorithms and complexity:

- Detection of square permutations;
- Statistics on permutations;
- Random generation.

(D) Term rewrite systems:

- Free clones*;
- Combinatory logic;
- Models of computation.

*: intervenes in the sequel.

2. Universal algebra and terms

Universal algebra

Universal algebra is a formalism to work with algebraic structures.

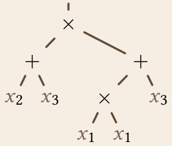
A signature is a graded set $\mathfrak{G} := \bigsqcup_{k \geqslant 0} \mathfrak{G}(k)$ wherein each $\mathbf{a} \in \mathfrak{G}(k)$ is an operation of arity k.

A $\mathfrak{G}$-term is

- either a variable x from the set $\mathbb{X} := \{x_1, x_2, \dots\}$;
- either a pair $(\mathbf{a}, (\mathfrak{t}_1, \dots, \mathfrak{t}_k))$ where $\mathbf{a} \in \mathfrak{G}(k)$ and each $\mathfrak{t}_i$ is a $\mathfrak{G}$-term.

The set of all $\mathfrak{G}$-terms is denoted by $\mathfrak{T}(\mathfrak{G})$.

– Example –



This is the tree representation of the $\mathfrak{G}$-term

$$(\times, ((+, (x_1, x_2)), (+, ((\times, (x_1, x_1)), x_3))))$$

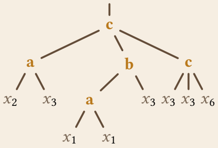
where $\mathfrak{G} := \mathfrak{G}(2) := \{+, \times\}$.

More on terms

Let $\mathfrak{t}$ be a $\mathfrak{G}$-term.

The frontier of $\mathfrak{t}$ is the sequence of all variables appearing in $\mathfrak{t}$.

– Example –



The frontier of this term is $(x_2, x_3, x_1, x_1, x_3, x_3, x_3, x_6)$ and its ground arity is 6.

The ground arity of $\mathfrak{t}$ is the greatest integer n such that x_n is a variable appearing in $\mathfrak{t}$.

The term $\mathfrak{t}$ is

- planar if its frontier is $(x_1, \dots, x_n)$;
- standard if its frontier is a permutation of $(x_1, \dots, x_n)$;
- linear if there are no multiple occurrences of the same variable in the frontier of $\mathfrak{t}$.

Varieties

A $\mathfrak{G}$-equation is a pair $(\mathfrak{t}, \mathfrak{t}')$ where $\mathfrak{t}$ and $\mathfrak{t}'$ are both $\mathfrak{G}$-terms.

A variety is a pair $(\mathfrak{G}, \mathfrak{R})$ where $\mathfrak{G}$ is a signature and $\mathfrak{R}$ is a set of $\mathfrak{G}$-equations. We denote by $\mathfrak{t} \; \mathfrak{R} \; \mathfrak{t}'$ the fact that $(\mathfrak{t}, \mathfrak{t}') \in \mathfrak{R}$.

– Example –

The variety of groups is the pair $(\mathfrak{G}, \mathfrak{R})$ where $\mathfrak{G} := \mathfrak{G}(0) \sqcup \mathfrak{G}(1) \sqcup \mathfrak{G}(2)$ with $\mathfrak{G}(0) := \{\mathbb{1}\}$, $\mathfrak{G}(1) := \{\mathrm{i}\}$, and $\mathfrak{G}(2) := \{\star\}$, and $\mathfrak{R}$ is the set of $\mathfrak{G}$-equations satisfying

$$\star(\star(x_1, x_2), x_3) \;\mathfrak{R}\; \star(x_1, \star(x_2, x_3)), \qquad \star(x_1, \mathbb{1}) \;\mathfrak{R}\; x_1 \;\mathfrak{R}\; \star(\mathbb{1}, x_1), \qquad \star(\mathrm{i}(x_1), x_1) \;\mathfrak{R}\; \mathbb{1} \;\mathfrak{R}\; \star(x_1, \mathrm{i}(x_1)).$$

– Example –

The variety of semilattices is the pair $(\mathfrak{G}, \mathfrak{R})$ where $\mathfrak{G} := \mathfrak{G}(2) := \{\wedge\}$, and $\mathfrak{R}$ is the set of $\mathfrak{G}$-equations satisfying

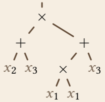
$$\wedge(\wedge(x_1, x_2), x_3) \;\mathfrak{R}\; \wedge(x_1, \wedge(x_2, x_3)), \qquad \wedge(x_1, x_2) \;\mathfrak{R}\; \wedge(x_2, x_1), \qquad \wedge(x_1, x_1) \;\mathfrak{R}\; x_1.$$

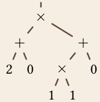
Algebras of a variety

Let $\mathcal{A}$ be a nonempty set. An $\mathcal{A}$-substitution is a map $\sigma : \mathbb{X} \to \mathcal{A}$.
An $\mathcal{A}$-interpretation of a signature $\mathfrak{G}$ is a set $\mathfrak{G}_{\mathcal{A}} := \left\{ \mathbf{a}_{\mathcal{A}} : \mathcal{A}^k \to \mathcal{A} : \mathbf{a} \in \mathfrak{G}(k) \text{ for a } k \geqslant 0 \right\}$.
The evaluation of a $\mathfrak{G}$-term $\mathfrak{t}$ under an $\mathcal{A}$-substitution σ and an $\mathcal{A}$-interpretation $\mathfrak{G}_{\mathcal{A}}$ is defined by induction as

$$\mathrm{ev}^{\sigma}_{\mathcal{A}}(\mathfrak{t}) := \begin{cases} \sigma(x) & \text{if } \mathfrak{t} = x \text{ is a variable,} \\ \mathbf{a}_{\mathcal{A}}(\mathrm{ev}^{\sigma}_{\mathcal{A}}(\mathfrak{t}_1), \dots, \mathrm{ev}^{\sigma}_{\mathcal{A}}(\mathfrak{t}_k)) & \text{otherwise, where } \mathfrak{t} = (\mathbf{a}, (\mathfrak{t}_1, \dots, \mathfrak{t}_k)). \end{cases}$$

– Example –

$\mathfrak{t} :=$ 

$\rightsquigarrow$ 

With $\mathcal{A} := \mathbb{N}$, $\mathfrak{G}_{\mathcal{A}}$ defined naturally, and σ satisfying $\sigma(x_1) := 1$, $\sigma(x_2) := 2$, and $\sigma(x_3) := 0$, one obtains $\mathrm{ev}^{\sigma}_{\mathcal{A}}(\mathfrak{t}) = 2$.

An algebra of a variety $(\mathfrak{G}, \mathfrak{R})$ is a pair $(\mathcal{A}, \mathfrak{G}_{\mathcal{A}})$ where for any $(\mathfrak{t}, \mathfrak{t}') \in \mathfrak{R}$ and $\mathcal{A}$-substitution σ, $\mathrm{ev}^{\sigma}_{\mathcal{A}}(\mathfrak{t}) = \mathrm{ev}^{\sigma}_{\mathcal{A}}(\mathfrak{t}')$.

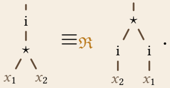
3. Equivalence of terms and rewrite systems

Equivalent terms

Two $\mathfrak{G}$-terms $\mathfrak{t}$ and $\mathfrak{t}'$ are $\mathfrak{R}$-equivalent if for all algebras $(\mathcal{A}, \mathfrak{G}_{\mathcal{A}})$ of $(\mathfrak{G}, \mathfrak{R})$ and for all $\mathcal{A}$-substitutions σ, one has $\mathrm{ev}^{\sigma}_{\mathcal{A}}(\mathfrak{t}) = \mathrm{ev}^{\sigma}_{\mathcal{A}}(\mathfrak{t}')$. This property is denoted by $\mathfrak{t} \equiv_{\mathfrak{R}} \mathfrak{t}'$.

– Example –

In the variety of groups,



– Some usual questions –

1. Design an algorithm to decide if two $\mathfrak{G}$-terms are $\equiv_{\mathfrak{R}}$-equivalent. This is known as the word problem [**Baader, Nipkow**, 1998].
2. Construct a system of representatives $\mathcal{C}$ of the $\equiv_{\mathfrak{R}}$-equivalence classes. The set $\mathcal{C}$ is a partial combinatorial realization of the variety.
3. Enumerate the $\equiv_{\mathfrak{R}}$-equivalence classes of (planar/standard/linear) $\mathfrak{G}$-terms w.r.t. their ground arity.

Rewrite systems on terms

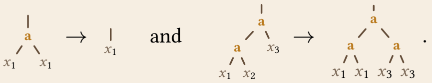
Rewrite systems on terms are tools to tackle these questions.

A rewrite relation on $\mathfrak{T}(\mathfrak{G})$ is a binary relation $\to$ on $\mathfrak{T}(\mathfrak{G})$ such that if $\mathfrak{s} \to \mathfrak{s}'$, then all variables of $\mathfrak{s}'$ appear in $\mathfrak{s}$.

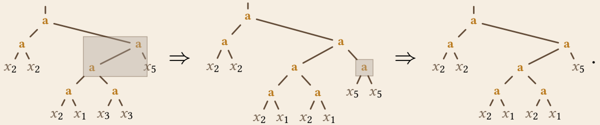
The context closure of $\to$ is the binary relation $\Rightarrow$ satisfying $\mathfrak{t} \Rightarrow \mathfrak{t}'$ whenever $\mathfrak{t}'$ is obtained by replacing in $\mathfrak{t}$ a factor $\mathfrak{s}$ by $\mathfrak{s}'$ provided that $\mathfrak{s} \to \mathfrak{s}'$.

– Example –

For $\mathfrak{G} := \mathfrak{G}(2) := \{\mathsf{a}\}$, let the rewrite relation $\to$ defined by



We have



Rewrite systems on terms and varieties

Let $(\mathfrak{G}, \mathfrak{R})$ be a variety. A rewrite relation $\to$ of $\mathfrak{T}(\mathfrak{G})$ is an orientation of $\mathfrak{R}$ if the reflexive, symmetric, and transitive closure of $\to$ is $\mathfrak{R}$.

– Proposition [Straightforward, –] –

Let $\to$ be an orientation of $\mathfrak{R}$.

For any two $\mathfrak{G}$-terms $\mathfrak{t}$ and $\mathfrak{t}'$, $\mathfrak{t} \equiv_{\mathfrak{R}} \mathfrak{t}'$ iff $\mathfrak{t} \overset{*}{\Leftrightarrow} \mathfrak{t}'$.

When there is no infinite chain $\mathfrak{t}_0 \Rightarrow \mathfrak{t}_1 \Rightarrow \mathfrak{t}_2 \Rightarrow \cdots$, the rewrite relation $\to$ is terminating.

If $\mathfrak{t} \overset{*}{\Rightarrow} \mathfrak{s}_1$ and $\mathfrak{t} \overset{*}{\Rightarrow} \mathfrak{s}_2$ implies the existence of $\mathfrak{t}'$ such that $\mathfrak{s}_1 \overset{*}{\Rightarrow} \mathfrak{t}'$ and $\mathfrak{s}_2 \overset{*}{\Rightarrow} \mathfrak{t}'$, then $\to$ is confluent.

A normal form for $\to$ is a $\mathfrak{G}$-term $\mathfrak{t}$ such that $\mathfrak{t} \overset{*}{\Rightarrow} \mathfrak{t}'$ implies $\mathfrak{t} = \mathfrak{t}'$.

– Proposition [Straightforward, –] –

Let $\to$ be an orientation of $\mathfrak{R}$.

If $\to$ is terminating and confluent, then $\mathfrak{t} \equiv_{\mathfrak{R}} \mathfrak{t}'$ iff there is a normal form $\mathfrak{s}$ such that $\mathfrak{t} \overset{*}{\Rightarrow} \mathfrak{s}$ and $\mathfrak{t}' \overset{*}{\Rightarrow} \mathfrak{s}$.

In this context, completion algorithms are important [**Knuth, Bendix**, 1970].

Duplicial algebras

A duplicial algebra [**Brouder, Frabetti**, 2003] is a set $\mathcal{A}$ endowed with two binary operations

$$\ll, \gg : \mathcal{A}^2 \to \mathcal{A}$$

satisfying the three relations

$$(x_1 \ll x_2) \ll x_3 = x_1 \ll (x_2 \ll x_3),$$
$$(x_1 \gg x_2) \ll x_3 = x_1 \gg (x_2 \ll x_3),$$
$$(x_1 \gg x_2) \gg x_3 = x_1 \gg (x_2 \gg x_3).$$

– Example –

On $\mathbb{N}^+$, let $\ll$ and $\gg$ be the operations defined by

$$u \ll v := u(v \uparrow_{\max(u)}), \qquad u \gg v := u(v \uparrow_{|u|}).$$

Then, for instance,

$$0211 \ll 14 = 021136, \qquad 0211 \gg 14 = 021158.$$

This structure is a duplicial algebra [**Novelli, Thibon**, 2013].

Equivalence of duplicial operations

Let us describe an algorithm to test if two planar duplicial operations are equivalent.

By the duplicial relations, we have

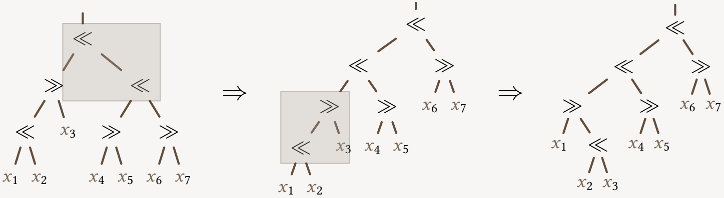
$$(x_1 \ll x_2) \ll x_3 \;\; \mathfrak{R} \;\; x_1 \ll (x_2 \ll x_3), \qquad (x_1 \ll x_2) \gg x_3 \;\; \mathfrak{R} \;\; x_1 \ll (x_2 \gg x_3), \qquad (x_1 \gg x_2) \gg x_3 \;\; \mathfrak{R} \;\; x_1 \gg (x_2 \gg x_3).$$

Let the orientation $\to$ of $\mathfrak{R}$ defined by

$$(x_1 \ll x_2) \ll x_3 \longleftarrow x_1 \ll (x_2 \ll x_3), \qquad (x_1 \ll x_2) \gg x_3 \longrightarrow x_1 \ll (x_2 \gg x_3), \qquad (x_1 \gg x_2) \gg x_3 \longleftarrow x_1 \gg (x_2 \gg x_3).$$

Testing equivalence of duplicial operations

We have for instance the sequence



of rewritings.

– Proposition [Loday, 2008] –

The rewrite relation $\to$ is terminating and confluent.

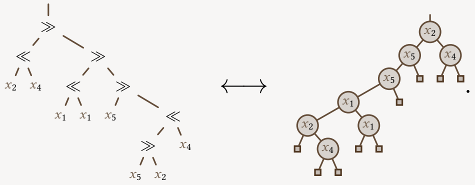
Therefore, two duplicial operations $\mathfrak{t}$ and $\mathfrak{t}'$ are $\mathfrak{R}$-equivalent iff the procedure consisting in rewriting $\mathfrak{t}$ and $\mathfrak{t}'$ as far as possible produces in both cases the same term.

Encoding duplicial operations

– Proposition [Loday, 2008] –

The set of normal forms for $\to$ of duplicial operations with $n \geqslant 0$ inputs is in one-to-one correspondence with the set of all binary trees with n internal nodes where internal nodes are labeled on $\mathbb{X}$.

A possible bijection puts the following two trees in correspondence:



Therefore, there are

$$\frac{1}{n+1}\binom{2n}{n}k^n$$

pairwise nonequivalent duplicial operations with n inputs on variables of $\{x_1, \ldots, x_k\}$.

Distributive lattices

A distributive lattice is a set $\mathcal{A}$ endowed with two binary operations

$$\wedge, \vee : \mathcal{A}^2 \to \mathcal{A}$$

satisfying the relations

$$\begin{aligned} (x_1 \wedge x_2) \wedge x_3 &= x_1 \wedge (x_2 \wedge x_3), & (x_1 \vee x_2) \vee x_3 &= x_1 \vee (x_2 \vee x_3), \\ x_1 \wedge x_2 &= x_2 \wedge x_1, & x_1 \vee x_2 &= x_2 \vee x_1, \\ x_1 \wedge (x_1 \vee x_2) &= x_1, & x_1 \vee (x_1 \wedge x_2) &= x_1, \\ x_1 \vee (x_2 \wedge x_3) &= (x_1 \vee x_2) \wedge (x_1 \vee x_3), & x_1 \wedge (x_2 \vee x_3) &= (x_1 \wedge x_2) \vee (x_1 \wedge x_3). \end{aligned}$$

– Examples –

- On the subsets of $[n]$, $\vee$ defined as the union and $\wedge$ as the intersection of sets is a finite distributive lattice.
- The set of all Young diagrams is an infinite distributive lattice for the intersection and the union of Young diagrams [**Kreweras**, 1965] [**Stanley**, 1988].

Encoding distributive lattice operations

A normal term is a term $\mathfrak{t}$ expressing as

$$\mathfrak{t} = \mathfrak{s}_1 \vee \ldots \vee \mathfrak{s}_m, \quad m \geqslant 0, \quad \text{where} \quad \mathfrak{s}_i = x_{f_{i,1}} \wedge \ldots \wedge x_{f_{i,k_i}}, \quad k_i \geqslant 1,$$

for any $i, i' \in [k]$, $x_{f_{i,r}} = x_{f_{i,r'}}$ implies $r = r'$, and $\{f_{i,1}, \ldots, f_{i,k_i}\} \subseteq \{f_{i',1}, \ldots, f_{i',k_{i'}}\}$ implies $i = i'$.

– Examples –

- $(x_2 \wedge x_3 \wedge x_5) \vee (x_3 \wedge x_7) \vee (x_3 \wedge x_4) \vee x_6$ is a normal term —as polynomial: $x_2x_3x_5 + x_3x_7 + x_3x_4 + x_6$;
- $(x_2 \wedge x_3 \wedge x_5) \vee (x_2 \wedge x_5)$ is not —as polynomial: $x_2x_3x_5 + x_2x_5$.

– Proposition [Peterson, Stickel, 1977] –

The set of all normal terms is a partial combinatorial realization of the variety of distributive lattices.

Pairwise nonequivalent distributive lattice operations with n inputs are enumerated by the Dedekind numbers whose sequence begins with (only these few terms are known today)

$$1, 2, 5, 19, 167, 7580, 7828353, 2414682040997, 56130437228687557907787.$$

4. Composition and clones

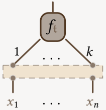
Abstract operations

To have a complete combinatorial realization of a variety, we need to describe an algorithm to compute the composition of the representations of two operations.

An $\mathfrak{G}$-term $\mathfrak{t}$ on the variables $\{x_1, \dots, x_n\}$ can be seen as an abstract operation

$$(x_1, \dots, x_n) \mapsto f_{\mathfrak{t}}(x_1, \dots, x_n)$$

depicted as

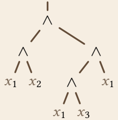


where k is the length of the frontier of $\mathfrak{t}$.

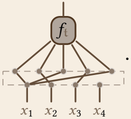
– Example –

For the signature $\mathfrak{G}$ of the variety of semi-lattices, here is a $\mathfrak{G}$-term seen on the set $\{x_1, \dots, x_4\}$ of variables and the abstract operation it denotes:

$\mathfrak{t} :=$



,



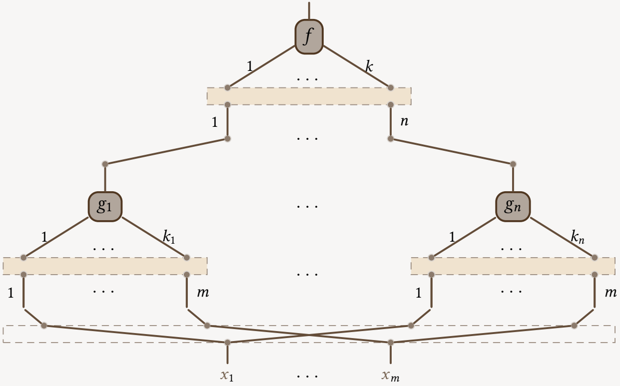
.

Composition of abstract operations

If f is an abstract operation of arity n and $g_1, \ldots, g_n$ are abstract operations all of arity m, then $f \circledcirc [g_1, \ldots, g_n]$ is the operation satisfying

$$(x_1, \ldots, x_m) \mapsto f(g_1(x_1, \ldots, x_m), \ldots, g_n(x_1, \ldots, x_m)).$$

This is the abstract operation depicted as



.

Clones

Abstract clones [**Cohn**, 1965] provide a formalization of general operations with their composition.

An abstract clone is a triple $(\mathcal{C}, \circledcirc, \mathbb{1}_{i,n})$ where

- $\mathcal{C}$ is a graded set
$$\mathcal{C} = \bigsqcup_{n \geqslant 0} \mathcal{C}(n);$$
- $\circledcirc$ is a map
$$\circledcirc : \mathcal{C}(n) \times \mathcal{C}(m)^n \to \mathcal{C}(m)$$
called superposition map;
- for each $n \geqslant 0$ and $i \in [n]$, $\mathbb{1}_{i,n}$ is an element of $\mathcal{C}(n)$ called projection.

This data has to satisfy some axioms.

Clone axioms

The following relations have to be satisfied:

- For all $x_i \in \mathcal{C}(m)$,

$$\mathbb{1}_{i,n} \circledcirc [x_1, \ldots, x_n] = x_i.$$

  This says that $\mathbb{1}_{i,n}$ is the operation returning its i-th input.

- For all $x \in \mathcal{C}(n)$,

$$x \circledcirc [\mathbb{1}_{1,n}, \ldots, \mathbb{1}_{n,n}] = x.$$

  This says that each $\mathbb{1}_{j,n}$, put as j-th input, is an identity operation.

- For all $x \in \mathcal{C}(n)$, $y_i \in \mathcal{C}(m)$, and $z_j \in \mathcal{C}(k)$,

$$(x \circledcirc [y_1, \ldots, y_n]) \circledcirc [z_1, \ldots, z_m] = x \circledcirc [y_1 \circledcirc [z_1, \ldots, z_m], \ldots, y_n \circledcirc [z_1, \ldots, z_m]].$$

  This says that the two ways to compose elements to form an operation having three layers (by starting from top or by starting from bottom) give the same operation.

Free clones

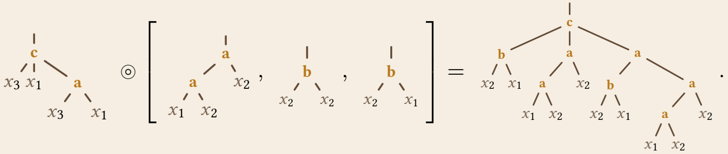
Let $\mathfrak{G}$ be a signature.

The free clone on $\mathfrak{G}$ is the clone $(\mathfrak{T}(\mathfrak{G}), \circledcirc, \mathbb{1}_{i,n})$ where

- $\mathfrak{T}(\mathfrak{G})$ is the set of all $\mathfrak{G}$-terms. Each $\mathfrak{G}$-term $\mathfrak{t}$ is endowed with an integer equal as or greater than its ground arity and called arity;
- $\circledcirc$ is defined as follows. The $\mathfrak{G}$-term $\mathfrak{t} \circledcirc [\mathfrak{s}_1, \ldots, \mathfrak{s}_n]$ is obtained by replacing each occurrence of a variable x_i of $\mathfrak{t}$ by the root of $\mathfrak{s}_i$;
- $\mathbb{1}_{i,n}$ is the term $\underset{x_i}{|}$ of arity n.

– Example –

By setting $\mathfrak{G} := \mathfrak{G}(2) \sqcup \mathfrak{G}(3)$ where $\mathfrak{G}(2) := \{\mathtt{a}, \mathtt{b}\}$ and $\mathfrak{G}(3) := \{\mathtt{c}\}$, in the free clone $\mathfrak{T}(\mathfrak{G})$, one has



5. Clones of colored words

A variety from a monoid

Let $(\mathcal{M}, \cdot, \epsilon)$ be a monoid.

– Definition [G., 2015] –

Let $(\mathfrak{G}_{\mathcal{M}}, \mathfrak{R}_{\mathcal{M}})$ be the variety such that

- $\mathfrak{G}_{\mathcal{M}} := \mathfrak{G}_{\mathcal{M}}(1) \sqcup \mathfrak{G}_{\mathcal{M}}(2)$ where $\mathfrak{G}_{\mathcal{M}}(1) := \mathcal{M}$ and $\mathfrak{G}_{\mathcal{M}}(2) := \{\mathtt{a}\}$;
- $\mathfrak{R}_{\mathcal{M}}$ s the set of $\mathfrak{G}_{\mathcal{M}}$-equations satisfying

$$\mathtt{a}(\mathtt{a}(x_1, x_2), x_3) \;\mathfrak{R}_{\mathcal{M}}\; \mathtt{a}(x_1, \mathtt{a}(x_2, x_3)), \qquad \mathtt{a}(\alpha(x_1), \alpha(x_2)) \;\mathfrak{R}_{\mathcal{M}}\; \alpha(\mathtt{a}(x_1, x_2)), \qquad \alpha_1(\alpha_2(x_1)) \;\mathfrak{R}_{\mathcal{M}}\; (\alpha_1 \cdot \alpha_2)(x_1), \qquad \epsilon(x_1) \;\mathfrak{R}_{\mathcal{M}}\; x_1,$$

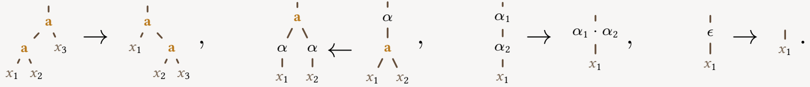
for any $\alpha, \alpha_1, \alpha_2 \in \mathcal{M}$.

Any algebra of this variety is a semigroup $(\mathcal{A}, \mathtt{a})$ endowed with semigroup endomorphisms $\phi_\alpha : \mathcal{A} \to \mathcal{A}$ with $\alpha \in \mathcal{M}$ such that ϕ_ϵ is the identity map and

$$\phi_{\alpha_1} \circ \phi_{\alpha_2} = \phi_{\alpha_1 \cdot \alpha_2}, \qquad \alpha_1, \alpha_2 \in \mathcal{M}.$$

Orientation of the equations

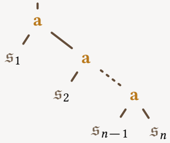
Let the orientation $\to$ of $\mathfrak{R}_{\mathcal{M}}$ satisfying



– Proposition [G., 2020–] –

For any monoid $\mathcal{M}$, the orientation $\to$ of $\mathfrak{R}_{\mathcal{M}}$ is terminating and confluent.

The set of normal forms for $\to$ of planar $\mathfrak{G}_{\mathcal{M}}$-terms is the set of the terms avoiding the left members of $\to$. These are the terms of the form



where $\quad \mathfrak{s}_i \in \left\{ x_{k_i}, \; \begin{matrix} \alpha_{k_i} \\ | \\ x_{k_i} \end{matrix} \right\}, \quad \alpha_{k_i} \in \mathcal{M} \setminus \{\epsilon\}.$

Colored words

Let $(\mathcal{M}, \cdot, \epsilon)$ be a monoid.

Let $\mathrm{W}\mathcal{M}$ be the graded set of all $\mathcal{M}$-colored words defined, for any $n \geqslant 0$, by

$$\mathrm{W}\mathcal{M}(n) := \bigsqcup_{n \geqslant 0} \left\{ \begin{pmatrix} u \\ c \end{pmatrix} : u \in [n]^{\ell}, c \in \mathcal{M}^{\ell}, \ell \geqslant 0 \right\}.$$

– Example –

$\begin{pmatrix} 1 & 2 & 1 & 6 \\ 0 & 0 & 1 & 0 \end{pmatrix}$ is a $\mathbb{Z}/2\mathbb{Z}$-colored word of arity 6 (or greater).

Let $\circledcirc$ be the superposition map defined by

$$\begin{pmatrix} u \\ c \end{pmatrix} \circledcirc \left[\begin{pmatrix} v_1 \\ d_1 \end{pmatrix}, \dots, \begin{pmatrix} v_n \\ d_n \end{pmatrix} \right] := \begin{pmatrix} v_{u(1)} & \dots & v_{u(\ell)} \\ \left(c(1) \,\bar{\cdot}\, d_{u(1)}\right) & \dots & \left(c(\ell) \,\bar{\cdot}\, d_{u(\ell)}\right) \end{pmatrix}$$

where for any $\alpha \in \mathcal{M}$ and $w \in \mathcal{M}^*$, $\alpha \,\bar{\cdot}\, w := (\alpha \cdot w(1)) \ \dots \ (\alpha \cdot w(|w|))$.

Let also $\mathbb{1}_{i,n} := \begin{pmatrix} i \\ \epsilon \end{pmatrix}$.

Clone of colored words

– Example –

In $\mathrm{W}(\mathbb{N}, +, 0)$,

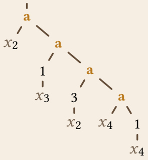
$$\begin{pmatrix} 2 & 2 & 3 \\ 0 & 1 & 0 \end{pmatrix} \circledcirc \left[\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 2 \\ 3 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 2 \\ 1 & 0 \end{pmatrix} \right] = \begin{pmatrix} 1 & 1 & 2 & & 1 & 1 & 2 & & 2 & 2 \\ 3 & 0 & 0 & & 4 & 1 & 1 & & 1 & 0 \end{pmatrix}.$$

– Theorem [G., 2020–] –

For any monoid $\mathcal{M}$, $(\mathrm{W}\mathcal{M}, \circledcirc, \mathbb{1}_{i,n})$ is a clone and is a combinatorial realization of the variety $(\mathfrak{G}_{\mathcal{M}}, \mathfrak{R}_{\mathcal{M}})$.

– Example –

Here is a normal for $\to$ of the variety $(\mathfrak{G}_{\mathcal{M}}, \mathfrak{R}_{\mathcal{M}})$ where $\mathcal{M}$ is the monoid $(\mathbb{N}, +, 0)$ and the $\mathcal{M}$-colored word in correspondence:



$$\longleftrightarrow \begin{pmatrix} 2 & 3 & 2 & 4 & 4 \\ 0 & 1 & 3 & 0 & 1 \end{pmatrix}.$$

Clone of words and congruences

Let us focus on the case where $\mathcal{M}$ is the trivial monoid $\{\epsilon\}$.

Let **Word** $:= \mathrm{W}\{\epsilon\}$. We can forget about the colors of the elements of **Word** without any loss of information.

- Let $\equiv_{\mathrm{s}}$ be the equivalence relation on **Word** wherein $u \equiv_{\mathrm{s}} v$ if u and v have both the same sorted version.
- Let $\equiv_{\mathrm{l}}$ (resp. $\equiv_{\mathrm{r}}$) be the equivalence relation on **Word** wherein $u \equiv_{\mathrm{l}} v$ (resp. $u \equiv_{\mathrm{r}} v$) if the versions of u and v obtained by keeping only the leftmost (resp. rightmost) among the multiple occurrences of a same letter are equal.

– Examples –

We have $311322 \equiv_{\mathrm{s}} 131232$, $\quad 223111352 \equiv_{\mathrm{l}} 2333315$, $\quad 5142144 \equiv_{\mathrm{r}} 552214$.

– Proposition [G., 2020–] –

The equivalence relations $\equiv_{\mathrm{s}}$, $\equiv_{\mathrm{l}}$, and $\equiv_{\mathrm{r}}$ are clone congruences of **Word**.

Quotients of Word

Clone	Combinatorial objects	Realized variety
$\mathbf{MSet} := \mathbf{Word}/_{\equiv_{\mathrm{s}}}$	Multisets	Commutative semigroups
$\mathbf{Arr}_{\mathrm{l}} := \mathbf{Word}/_{\equiv_{\mathrm{l}}}$	Arrangements	Left-regular bands
$\mathbf{Set} := \mathbf{Word}/_{\equiv_{\mathrm{s}} \circ \equiv_{\mathrm{l}}}$	Sets	Semilattices
$\mathbf{ArrB}_{\mathrm{l}} := \mathbf{Word}/_{\equiv_{\mathrm{s}} \cap \equiv_{\mathrm{l}}}$	Arrangements of blocks	Ass. and $x_1 x_2 x_1 = x_1 x_1 x_2$
$\mathbf{PArr} = \mathbf{Word}/_{\equiv_{\mathrm{l}} \cap \equiv_{\mathrm{r}}}$	Pairs of compatible arr.	Regular bands
$\mathbf{PArrB} := \mathbf{Word}/_{\equiv_{\mathrm{s}} \cap \equiv_{\mathrm{l}} \cap \equiv_{\mathrm{r}}}$	Pairs of comp. arr. of blocks	Ass. and $x_1 x_1 x_2 x_3 x_1 = x_1 x_2 x_1 x_3 x_1 = x_1 x_2 x_3 x_1 x_1$

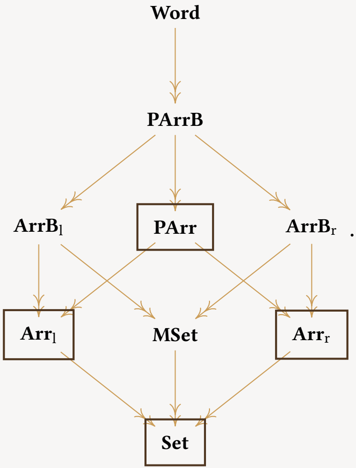
The congruence $\equiv_{\mathrm{s}} \cap \equiv_{\mathrm{l}}$ is the stalactic congruence [**Hivert, Novelli, Thibon**, 2007].

The superposition maps of these clones can be described by simple algorithms.

All this provides combinatorial realizations of some varieties of semigroups.

Conclusion and future work

The previous clones on noncolored words fit in the diagram (squared clones are combinatorial):



Some open questions:

1. Enrich this diagram with quotients of **Word** by other congruences.
2. Describe the analogous hierarchy when $\mathcal{M}$ is one of the two finite monoids on two elements.
3. Construct other combinatorial realizations of varieties as clones of $\mathcal{M}$-colored words for suitable monoids $\mathcal{M}$.
4. Describe the subclones of a clone of $\mathcal{M}$-colored words generated by finite sets of generators.

6. Appendix

Multisets

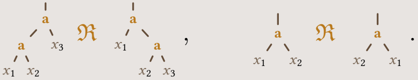
Let **MSet** $:=$ **Word**$/_{\equiv_s}$.

The elements of **MSet** can be seen as multisets of positive integers. By encoding any such multiset $u = \wr 1^{a(1)}, \dots, n^{a(n)} \int$ by the tuple $a = (a(1), \dots, a(n))$, the superposition map of **MSet** expresses as a matrix multiplication

$$a \circledcirc [b_1, \dots, b_n] = \begin{pmatrix} a(1) & \dots & a(n) \end{pmatrix} \begin{pmatrix} b_1(1) & \dots & b_1(m) \\ \vdots & \dots & \vdots \\ b_n(1) & \dots & b_n(m) \end{pmatrix}.$$

– Proposition [G., 2020–] –

The clone **MSet** admits the presentation $(\mathfrak{G}, \mathfrak{R})$ where $\mathfrak{G} := \mathfrak{G}(2) := \{\mathtt{a}\}$ and $\mathfrak{R}$ satisfies



Therefore, **MSet** is a combinatorial realization of the variety of commutative semigroups.

Arrangements

Let $\mathbf{Arr}_{\mathrm{l}} := \mathbf{Word}/_{\equiv_{\mathrm{l}}}$.

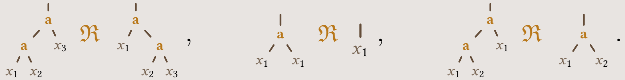
The elements of $\mathbf{Arr}_{\mathrm{l}}(n)$ can be seen as arrangements (words without repetitions) on $[n]$. For any $n \geqslant 0$,

$$\#\mathbf{Arr}_{\mathrm{l}}(n) = \sum_{0 \leqslant k \leqslant n} \frac{n!}{k!}$$

and this sequence starts by 1, 2, 5, 16, 65, 326, 1957, 13700, 109601.

– Proposition [G., 2020–] –

The clone $\mathbf{Arr}_{\mathrm{l}}$ admits the presentation $(\mathfrak{G}, \mathfrak{R})$ where $\mathfrak{G} := \mathfrak{G}(2) := \{\mathsf{a}\}$ and $\mathfrak{R}$ satisfies



The algebra of this variety are left-regular bands, that are idempotent semigroups wherein the operation a satisfies the relation $x_1 \;\mathsf{a}\; x_2 \;\mathsf{a}\; x_1 = x_1 \;\mathsf{a}\; x_2$.

Analog properties hold for the quotient $\mathbf{Arr}_{\mathrm{r}} := \mathbf{Word}/_{\equiv_{\mathrm{r}}}$, leading to right-regular bands.

– Lemma [G., 2020–] –

$\equiv_{\mathrm{s}} \circ \equiv_{\mathrm{l}} = \equiv_{\mathrm{l}} \circ \equiv_{\mathrm{s}}$

Therefore, this composition is a clone congruence of **Word**.

Let us set it as $\equiv_{\mathrm{i}}$ and let **Set** $:=$ **Word**$/_{\equiv_{\mathrm{i}}}$.

The elements of **Set** can be seen as sets of positive integers. On such objects, the superposition map of **Set** expresses as

$$U \circledcirc [V_1, \dots, V_n] = \bigcup_{j \in U} V_j.$$

Moreover, for any $n \geqslant 0$, $\#\mathbf{Set}(n) = 2^n$.

– Proposition [G., 2020–] –

The clone **Set** admits the presentation $(\mathfrak{G}, \mathfrak{R})$ where $\mathfrak{G} := \mathfrak{G}(2) := \{\mathtt{a}\}$ and $\mathfrak{R}$ satisfies

$$\mathtt{a}(\mathtt{a}(x_1, x_2), x_3) \;\mathfrak{R}\; \mathtt{a}(x_1, \mathtt{a}(x_2, x_3)), \qquad \mathtt{a}(x_1, x_2) \;\mathfrak{R}\; \mathtt{a}(x_2, x_1), \qquad \mathtt{a}(x_1, x_1) \;\mathfrak{R}\; x_1.$$

Therefore, **Set** is a combinatorial realization of the variety of semilattices.

Arrangements of blocks

Let us consider some intersections involving the congruences $\equiv_{\mathrm{s}}$, $\equiv_{\mathrm{l}}$, and $\equiv_{\mathrm{r}}$.

Let $\equiv_{\mathrm{sl}} := \equiv_{\mathrm{s}} \cap \equiv_{\mathrm{l}}$ and $\mathbf{ArrB}_{\mathrm{l}} := \mathbf{Word}/_{\equiv_{\mathrm{sl}}}$.

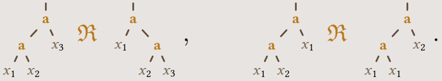
The elements of $\mathbf{ArrB}_{\mathrm{l}}(n)$ can be seen as arrangements of possibly empty blocks of repeated letters of $[n]$.

– Examples –

The word 3311115526 is such an element of $\mathbf{ArrB}_{\mathrm{l}}(9)$. The word 22222333112 is not an element of $\mathbf{ArrB}_{\mathrm{l}}$.

– Proposition [G., 2020–] –

The clone $\mathbf{ArrB}_{\mathrm{l}}$ admits the presentation $(\mathfrak{G}, \mathfrak{R})$ where $\mathfrak{G} := \mathfrak{G}(2) := \{\mathsf{a}\}$ and $\mathfrak{R}$ satisfies



Analog properties hold for the quotient $\mathbf{ArrB}_{\mathrm{r}} := \mathbf{Word}/_{\equiv_{\mathrm{sr}}}$, where $\equiv_{\mathrm{sr}} := \equiv_{\mathrm{s}} \cap \equiv_{\mathrm{r}}$.

Pairs of compatible arrangements

Let $\equiv_{\mathrm{lr}} := \equiv_{\mathrm{l}} \cap \equiv_{\mathrm{r}}$ and **PArr** $=$ **Word**$/_{\equiv_{\mathrm{lr}}}$.

The elements of $\mathbf{PArr}(n)$ can be seen as pairs (u, v) such that u and v are arrangements on $[n]$, such that j appears in u iff j appears in v.

– Example –

$(3261, 1263)$ is such an element of $\mathbf{PArr}(6)$.

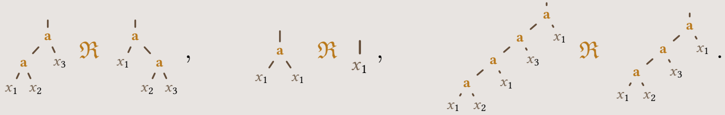
For any $n \geqslant 0$,

$$\#\mathbf{PArr}(n) = \sum_{0 \leqslant k \leqslant n} \frac{n!k!}{(n-k)!}$$

and this sequence starts by 1, 2, 7, 52, 749, 17686, 614227, 29354312, 1844279257.

– Proposition [G., 2020–] **–**

The clone **PArr** admits the presentation $(\mathfrak{G}, \mathfrak{R})$ where $\mathfrak{G} := \mathfrak{G}(2) := \{\mathtt{a}\}$ and $\mathfrak{R}$ satisfies



Therefore, **PArr** is a combinatorial realization of the variety of regular bands.

Pairs of compatible arrangements of blocks

Let $\equiv_{\mathrm{slr}} := \equiv_{\mathrm{s}} \cap \equiv_{\mathrm{l}} \cap \equiv_{\mathrm{r}}$ and $\mathbf{PArrB} := \mathbf{Word}/_{\equiv_{\mathrm{slr}}}$.

The elements of $\mathbf{PArrB}(n)$ can be seen as pairs (u, v) such that u and v are arrangements of possibly empty blocks of repeated letters on $[n]$, with u and v having the same number of occurrences of any letter.

– Example –

$(3222611, 22211263)$ is such an element of $\mathbf{PArrB}(6)$.

– Proposition [G., 2020–] **–**

The clone **PArrB** admits the presentation $(\mathfrak{G}, \mathfrak{R})$ where $\mathfrak{G} := \mathfrak{G}(2) := \{\mathtt{a}\}$ and $\mathfrak{R}$ satisfies