Whole Mirror Duplication Random Loss Model and Pattern Avoiding Permutations

Jean-Luc Baril and Rémi Vernay
barjl@u-bourgogne.fr
http://jl.baril.u-bourgogne.fr

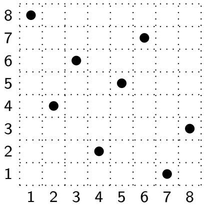
Laboratory LE2I – University of Burgundy – Dijon

Some definitions and notations

- genome = set of chromosomes
- chromosome = sequence of genes
- gene = sequence of Adénine, Guanine, Cytosine, Thynmine (AGCT)
- genome $\rightarrow$ n-length permutation $\sigma = \sigma_1\sigma_2\sigma_3\ldots\sigma_n$
- S_n = the set of n-length permutations

Graphical representation of the permutation

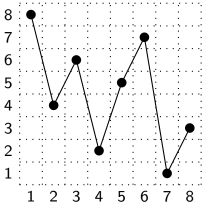
$\sigma = 8\ 4\ 6\ 2\ 5\ 7\ 1\ 3$



8 4 6 2 5 7 1 3

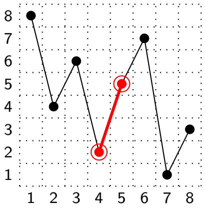
Let $\sigma = \sigma_1\sigma_2\ldots\sigma_n$ be a permutation:

- ascent $\rightarrow \sigma_i < \sigma_{i+1}$
- run up $\rightarrow \sigma_i < \sigma_{i+1} < \cdots < \sigma_j$
- descent, run-down
- valley $\rightarrow \sigma_{i-1} > \sigma_i < \sigma_{i+1}$
- Alternating permutation $\rightarrow \sigma_1 > \sigma_2 < \sigma_3 > \sigma_4 < \sigma_5 > \cdots$



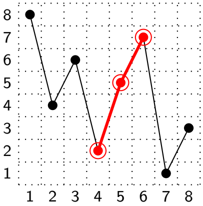
Let $\sigma = \sigma_1\sigma_2\ldots\sigma_n$ be a permutation:

- ascent → $\sigma_i < \sigma_{i+1}$
- run up → $\sigma_i < \sigma_{i+1} < \cdots < \sigma_j$
- descent, run-down
- valley → $\sigma_{i-1} > \sigma_i < \sigma_{i+1}$
- Alternating permutation → $\sigma_1 > \sigma_2 < \sigma_3 > \sigma_4 < \sigma_5 > \cdots$



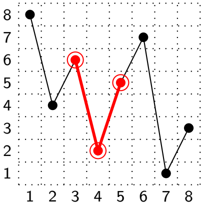
Let $\sigma = \sigma_1\sigma_2\ldots\sigma_n$ be a permutation:

- ascent $\rightarrow \sigma_i < \sigma_{i+1}$
- run up $\rightarrow \sigma_i < \sigma_{i+1} < \cdots < \sigma_j$
- descent, run-down
- valley $\rightarrow \sigma_{i-1} > \sigma_i < \sigma_{i+1}$
- Alternating permutation $\rightarrow \sigma_1 > \sigma_2 < \sigma_3 > \sigma_4 < \sigma_5 > \cdots$



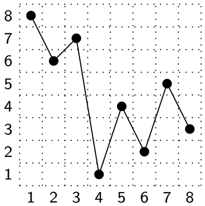
Let $\sigma = \sigma_1\sigma_2\dots\sigma_n$ be a permutation:

- ascent → $\sigma_i < \sigma_{i+1}$
- run up → $\sigma_i < \sigma_{i+1} < \cdots < \sigma_j$
- descent, run-down
- valley → $\sigma_{i-1} > \sigma_i < \sigma_{i+1}$
- Alternating permutation → $\sigma_1 > \sigma_2 < \sigma_3 > \sigma_4 < \sigma_5 > \cdots$



Let $\sigma = \sigma_1\sigma_2\ldots\sigma_n$ be a permutation:

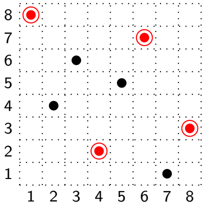
- ascent $\rightarrow \sigma_i < \sigma_{i+1}$
- run up $\rightarrow \sigma_i < \sigma_{i+1} < \cdots < \sigma_j$
- descent, run-down
- valley $\rightarrow \sigma_{i-1} > \sigma_i < \sigma_{i+1}$
- Alternating permutation $\rightarrow \sigma_1 > \sigma_2 < \sigma_3 > \sigma_4 < \sigma_5 > \cdots$



Definition:
$\sigma \in S_n$ contains the pattern $\pi \in S_k$ ($\pi \preceq \sigma$) if:
$\exists 1 \leq i_1 < i_2 < \cdots < i_k \leq n$ such that $\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_k}$ is order-isomorphic to π, *i.e.*,

$$\forall 1 \leq u, v \leq k, \qquad \sigma_{i_u} < \sigma_{i_v} \Leftrightarrow \pi_u < \pi_v .$$

Example: $\sigma = 8\ 4\ 6\ 2\ 5\ 7\ 1\ 3$ contains the pattern $\pi = 4132$



8 4 6 2 5 7 1 3

Class of permutations

$\mathcal{C}$ is a **class of permutations** if $\mathcal{C}$ is **stable** for the relation $\preceq$

$$\sigma \in \mathcal{C} \text{ and } \pi \preceq \sigma \Rightarrow \pi \in \mathcal{C}.$$

Basis for a class of permutations

A class $\mathcal{C}$ of permutations is characterized by its basis B:

$$\mathcal{B} = \{\sigma \notin \mathcal{C}, \forall \pi \prec \sigma \text{ with } \pi \neq \sigma, \pi \in \mathcal{C}\}$$

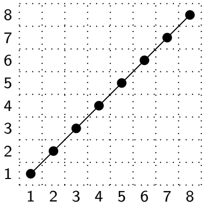
We have $\mathcal{C} = S(B)$ where $S(B)$ is the class of permutations avoiding all patterns in B.

The tandem duplication random-loss process

* This model is well-known in the evolutionary biology literature.

* Used for vertebrate mitochondrial genomes

1 2 3 4 5 6 7 8



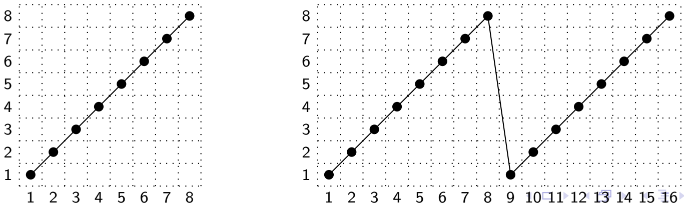
The tandem duplication random-loss process

* This model is well-known in the evolutionary biology literature.

* Used for vertebrate mitochondrial genomes

$$\overline{1\,2\,3\,4\,5\,6\,7\,8} \quad \rightsquigarrow \quad \overline{1\,2\,3\,4\,5\,6\,7\,8}\,\overline{1\,2\,3\,4\,5\,6\,7\,8}$$

(duplication)



The tandem duplication random-loss process

* This model is well-known in the evolutionary biology literature.

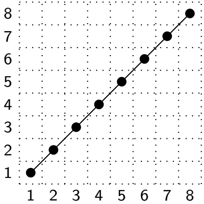
* Used for vertebrate mitochondrial genomes

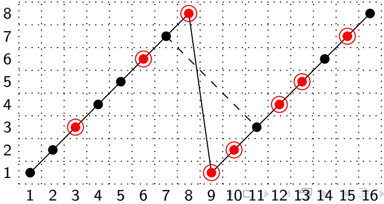
1 2 3 4 5 6 7 8 ⇝ 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8

(duplication)

⇝ 1 2 ~~3~~ 4 5 ~~6~~ 7 ~~8~~ ~~1~~ ~~2~~ 3 ~~4~~ ~~5~~ 6 ~~7~~ 8

(random loss)





The tandem duplication random-loss process

* This model is well-known in the evolutionary biology literature.

* Used for vertebrate mitochondrial genomes

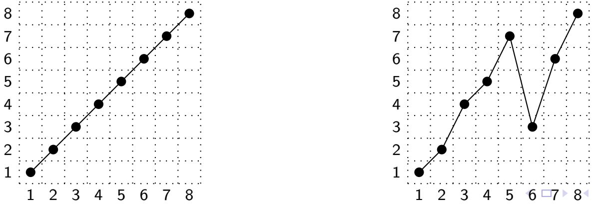
$$\overbrace{1\ 2\ 3\ 4\ 5\ 6\ 7\ 8} \quad \rightsquigarrow \quad \overbrace{1\ 2\ 3\ 4\ 5\ 6\ 7\ 8}\,\overbrace{1\ 2\ 3\ 4\ 5\ 6\ 7\ 8}$$

(duplication)

$$\rightsquigarrow \quad 1\ 2\ \cancel{3}\ 4\ 5\ \cancel{6}\ 7\ \cancel{8}\ \cancel{1}\ \cancel{2}\ 3\ \cancel{4}\ \cancel{5}\ 6\ \cancel{7}\ 8$$

(random loss)

$$\rightsquigarrow \quad 1\ 2\ 4\ 5\ 7\ 3\ 6\ 8$$



2006 K. Chaudhuri, K. Chen, R. Mihaescu and S. Rao
On the tandem duplication-random loss model of genome rearrangement,*SODA*
Tandem duplication random-loss process of an interval of size K;
Efficient algorithm for the distance between 2 genomes

2006 K. Chaudhuri, K. Chen, R. Mihaescu and S. Rao
On the tandem duplication-random loss model of genome rearrangement,*SODA*

2009 M. Bouvel and D. Rossin
A variant of the tandem duplication-random loss model of genome rearrangement, *TCS*
Permutations obtained after p duplications of an interval of size K define a class of permutations avoiding some patterns in B.
$B =$ set of minimal permutations with $d = 2^p$ descents.

2006 K. Chaudhuri, K. Chen, R. Mihaescu and S. Rao
On the tandem duplication-random loss model of genome rearrangement, *SODA*

2009 M. Bouvel and D. Rossin

A variant of the tandem duplication-random loss model of genome rearrangement, *TCS*

2010 M. Bouvel and E. Pergola
Posets and permutations in the duplication-loss model: minimal permutations with d descents,
Theoretical Computer Science
enumeration minimal permutations of size
$n = d + 1, d + 2, 2d$;

- 2006 K. Chaudhuri, K. Chen, R. Mihaescu and S. Rao
  On the tandem duplication-random loss model of genome rearrangement,*SODA*
- 2009 M. Bouvel and D. Rossin
  A variant of the tandem duplication-random loss model of genome rearrangement, *TCS*
- 2010 M. Bouvel and E. Pergola
  Posets and permutations in the duplication-loss model: minimal permutations with d descents, *Theoretical Computer Science*
- 2010 T. Mansour and S. H.F. Yan
  Minimal permutations with d descents, *European Journal of Combinatorics*
  enumeration minimal permutations of size $n = 2d - 1$

2006 K. Chaudhuri, K. Chen, R. Mihaescu and S. Rao
On the tandem duplication-random loss model of genome rearrangement,*SODA*

2009 M. Bouvel and D. Rossin
A variant of the tandem duplication-random loss model of genome rearrangement, *TCS*

2010 M. Bouvel and E. Pergola
Posets and permutations in the duplication-loss model: minimal permutations with d descents, *Theoretical Computer Science*

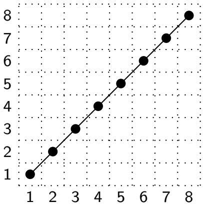
2010 T. Mansour and S. H.F. Yan

Minimal permutations with d descents, *European Journal of Combinatorics*

2010 M. Bouvel and L. Ferrari
On the enumeration of d-minimal permutations, *Arxiv*
prove that the number of minimal permutations with d descents can be obtained by computing some determinants, but they cannot provide a general closed formula

The whole mirror duplication random-loss process

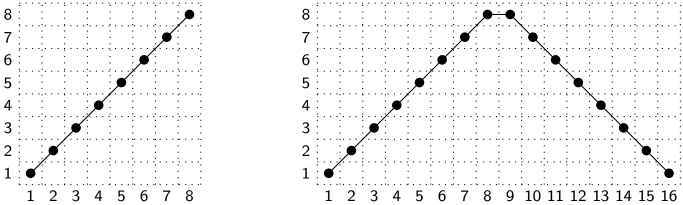
1 2 3 4 5 6 7 8



The whole mirror duplication random-loss process

$$\overline{1\;2\;3\;4\;5\;6\;7\;8} \quad\rightsquigarrow\quad \overline{1\;2\;3\;4\;5\;6\;7\;8}\;\overline{8\;7\;6\;5\;4\;3\;2\;1}$$

(mirror duplication)



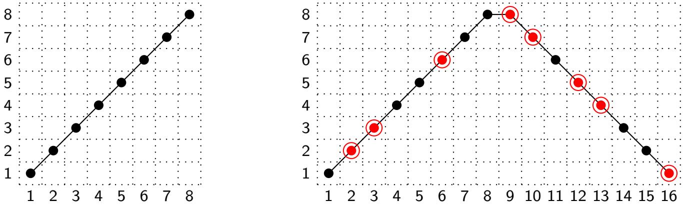
The whole mirror duplication random-loss process

$1\ 2\ 3\ 4\ 5\ 6\ 7\ 8 \quad \rightsquigarrow \quad 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 8\ 7\ 6\ 5\ 4\ 3\ 2\ 1$

(mirror duplication)

$\rightsquigarrow \quad 1\ \cancel{2}\ \cancel{3}\ 4\ 5\ \cancel{6}\ 7\ 8\ \cancel{8}\ \cancel{7}\ 6\ \cancel{5}\ \cancel{4}\ 3\ 2\ \cancel{1}$

(random loss)



The whole mirror duplication random-loss process

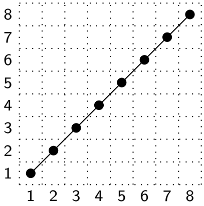
$$\overline{1\ 2\ 3\ 4\ 5\ 6\ 7\ 8} \quad \rightsquigarrow \quad \overline{1\ 2\ 3\ 4\ 5\ 6\ 7\ 8}\ \overline{8\ 7\ 6\ 5\ 4\ 3\ 2\ 1}$$

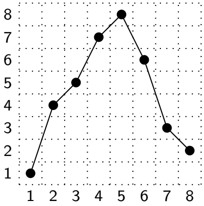
(mirror duplication)

$$\rightsquigarrow \quad 1\ \cancel{2}\ \cancel{3}\ 4\ 5\ \cancel{6}\ 7\ 8\ \cancel{8}\ \cancel{7}\ 6\ \cancel{5}\ \cancel{4}\ 3\ 2\ \cancel{1}$$

(random loss)

$$\rightsquigarrow \quad 1\ 4\ 5\ 7\ 8\ 6\ 3\ 2$$





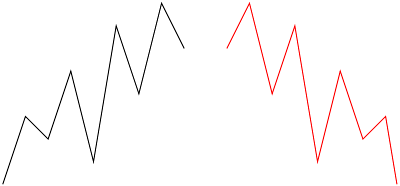
Theorem 1

The class $\mathcal{C}(p)$ of permutations obtained from the identity after a given number p of whole mirror duplications is the class of permutations with at most $2^{p-1}-1$ valleys.

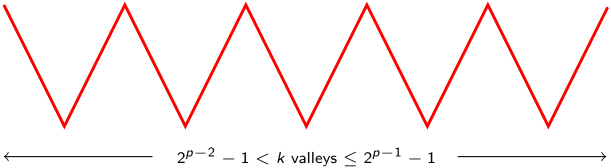
$2^{p-2}-1$ valleys

Theorem 1

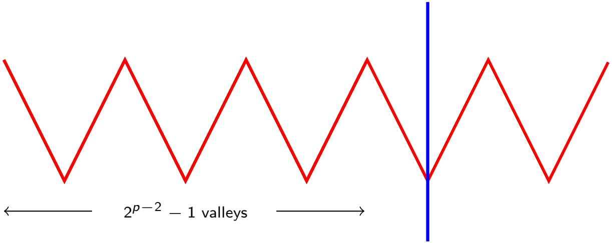
The class $\mathcal{C}(p)$ of permutations obtained from the identity after a given number p of whole mirror duplications is the class of permutations with at most $2^{p-1} - 1$ valleys.

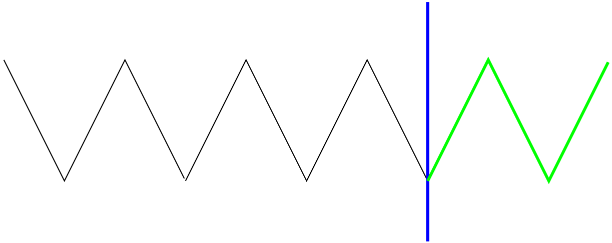


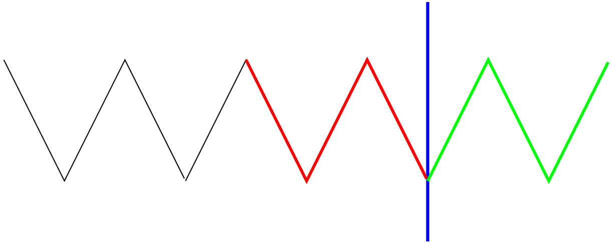
$2^{p-1} - 1 = 2^{p-2} - 1 + 2^{p-2} - 1 + 1$ valleys

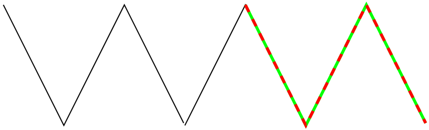


$\longleftarrow$ $2^{p-2} - 1 < k$ valleys $\leq 2^{p-1} - 1$ $\longrightarrow$


$2^{p-2} - 1$ valleys







Theorem 2

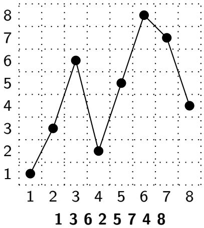
The class $\mathcal{C}(p)$ of permutations obtained after a given number p of whole mirror duplications is the class of permutations avoiding the alternating permutations of length $2^p + 1$.

For $p = 1$, $\mathcal{C}(1) = S(213, 312)$
For $p = 2$, $\mathcal{C}(2) =$
$S(21435, 31425, 41325, 32415, 42315, 21534, 31524, 51324, 32514,$
$52314, 41523, 51423, 42513, 52413, 43512, 53412)$

$|\mathcal{C}(p)|$ given by the generating function:

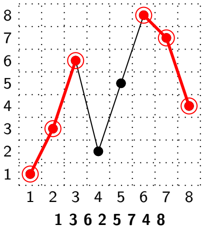
$$\frac{1}{1-y}\left(1 - \frac{1}{y} + \frac{1}{y}\sqrt{y-1}\cdot\tan\left(x\sqrt{y-1} + \arctan\left(\frac{1}{\sqrt{y-1}}\right)\right)\right)$$

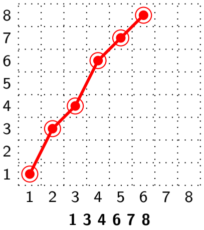
Algorithm 1 for a shortest path from $12\cdots n$ to $\sigma \in S_n$.



Complexity: $\mathcal{O}(n \cdot \log \text{val}(\sigma)) < \mathcal{O}(n \cdot \log n)$

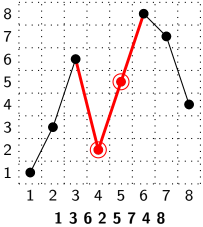
Algorithm 1 for a shortest path from $12 \cdots n$ to $\sigma \in S_n$.

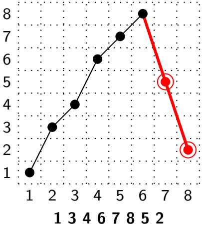




Complexity: $\mathcal{O}(n \cdot \log \text{val}(\sigma)) < \mathcal{O}(n \cdot \log n)$

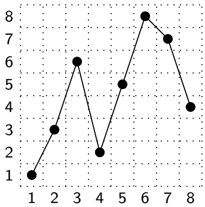
Algorithm 1 for a shortest path from $12\cdots n$ to $\sigma \in S_n$.





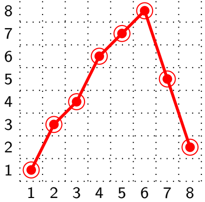
Complexity: $\mathcal{O}(n \cdot \log \mathrm{val}(\sigma)) < \mathcal{O}(n \cdot \log n)$

Algorithm 1 for a shortest path from $12\cdots n$ to $\sigma \in S_n$.

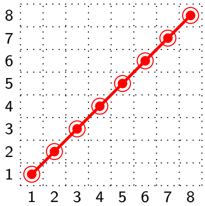


1 3 6 2 5 7 4 8

$\Rightarrow \mathcal{O}(n)$



1 3 6 8 4 2 5 7



1 2 3 4 5 6 7 8

Complexity: $\mathcal{O}(n \cdot \log \text{val}(\sigma)) < \mathcal{O}(n \cdot \log\ n)$

Algorithm 2 for a shortest path from $12\cdots n$ to $\sigma \in S_n$.

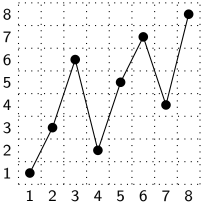
Step 1 – We label the runs up and runs down with the Binary Reflected Gray Code (F. Gray [1953], Bitner, Ehrlich, Reingold [1976])

$$B_n = \mathbf{0}B_{n-1} \circ \mathbf{1}\overline{B_{n-1}}$$

$$\begin{array}{c}
\mathbf{0}00\\
\mathbf{0}01\\
\mathbf{0}11\\
\mathbf{0}10\\
\hline
\mathbf{1}10\\
\mathbf{1}11\\
\mathbf{1}01\\
\mathbf{1}00
\end{array}$$

Algorithm 2 for a shortest path from $12\cdots n$ to $\sigma \in S_n$.

Step 1 – We label the runs up and runs down with the Binary Reflected Gray Code (F. Gray [1953], Bitner, Ehrlich, Reingold [1976])



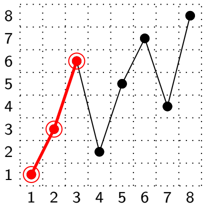
$$B_n = \mathbf{0}B_{n-1} \circ \mathbf{1}\overline{B_{n-1}}$$

000
001
011
010

110
111
101
100

Algorithm 2 for a shortest path from $12\cdots n$ to $\sigma \in S_n$.

Step 1 – We label the runs up and runs down with the Binary Reflected Gray Code (F. Gray [1953], Bitner, Ehrlich, Reingold [1976])



136	000

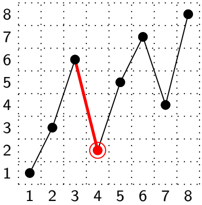
$$B_n = \mathbf{0}B_{n-1} \circ \mathbf{1}\overline{B_{n-1}}$$

000
001
011
010

110
111
101
100

Algorithm 2 for a shortest path from $12\cdots n$ to $\sigma \in S_n$.

Step 1 – We label the runs up and runs down with the Binary Reflected Gray Code (F. Gray [1953], Bitner, Ehrlich, Reingold [1976])



136	000
2	001

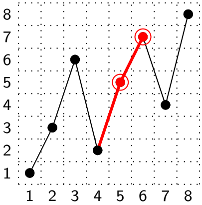
$$B_n = \mathbf{0}B_{n-1} \circ \mathbf{1}\overline{B_{n-1}}$$

000
001
011
010

110
111
101
100

Algorithm 2 for a shortest path from $12 \cdots n$ to $\sigma \in S_n$.

Step 1 – We label the runs up and runs down with the Binary Reflected Gray Code (F. Gray [1953], Bitner, Ehrlich, Reingold [1976])



136	000
2	001
57	011

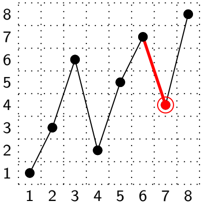
$$B_n = \mathbf{0}B_{n-1} \circ \mathbf{1}\overline{B_{n-1}}$$

000
001
011
010

110
111
101
100

Algorithm 2 for a shortest path from $12\cdots n$ to $\sigma \in S_n$.

Step 1 – We label the runs up and runs down with the Binary Reflected Gray Code (F. Gray [1953], Bitner, Ehrlich, Reingold [1976])



136	000
2	001
57	011
4	010

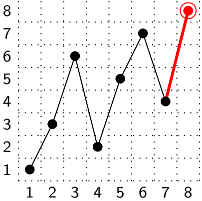
$$B_n = \mathbf{0}B_{n-1} \circ \mathbf{1}\overline{B_{n-1}}$$

000
001
011
010

110
111
101
100

Algorithm 2 for a shortest path from $12\cdots n$ to $\sigma \in S_n$.

Step 1 – We label the runs up and runs down with the Binary Reflected Gray Code (F. Gray [1953], Bitner, Ehrlich, Reingold [1976])



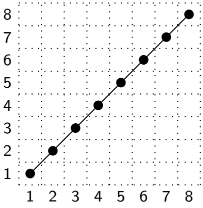
136	000
2	001
57	011
4	010
8	110

$$B_n = \mathbf{0}B_{n-1} \circ \mathbf{1}\overline{B_{n-1}}$$

000
001
011
010
110
111
101
100

Step 2 – We construct the path

136	000
2	001
57	011
4	010
8	110

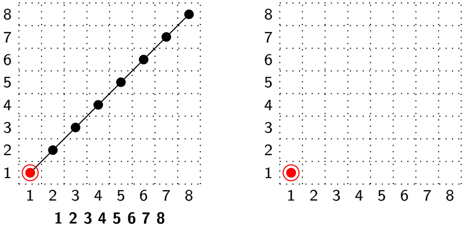


1 2 3 4 5 6 7 8

Step 2 – We construct the path

⇓

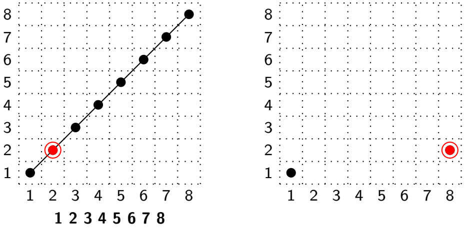
136	000
2	001
57	011
4	010
8	110



Step 2 – We construct the path

⇓

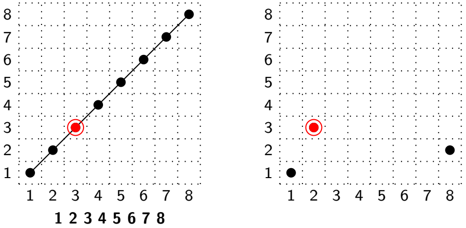
136	000
2	001
57	011
4	010
8	110



Step 2 – We construct the path

⇓

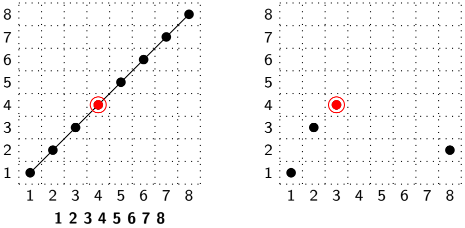
136	000
2	001
57	011
4	010
8	110



Step 2 – We construct the path

⇓

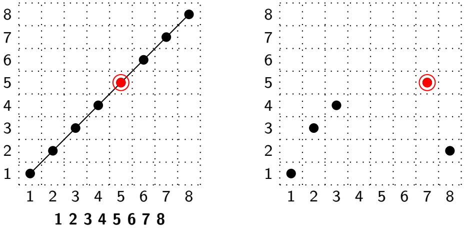
136	000
2	001
57	011
4	010
8	110



Step 2 – We construct the path

$\Downarrow$

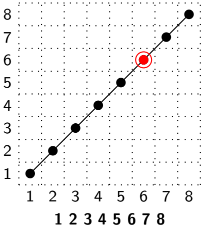
136	000
2	001
57	011
4	010
8	110

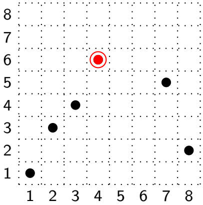


Step 2 – We construct the path

⇓

136	000
2	001
57	011
4	010
8	110

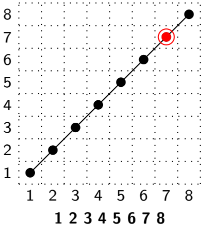


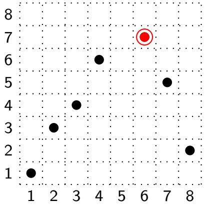


Step 2 – We construct the path

⇓

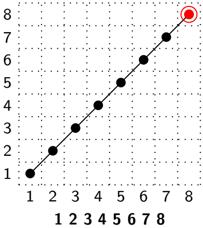
136	000
2	001
57	011
4	010
8	110

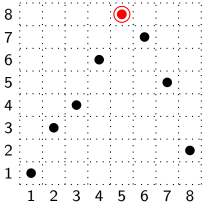




Step 2 – We construct the path

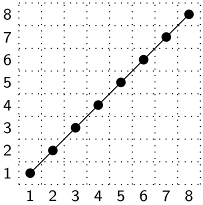
	⇓
136	000
2	001
57	011
4	010
8	110



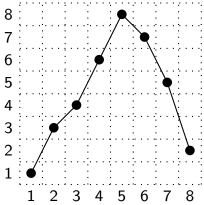


Step 2 – We construct the path

136	000
2	001
57	011
4	010
8	110



1 2 3 4 5 6 7 8

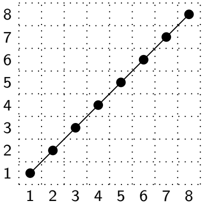


1 3 4 6 8 7 5 2

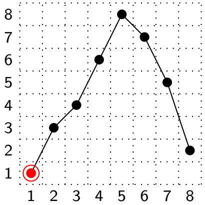
Step 2 – We construct the path

$\Downarrow$

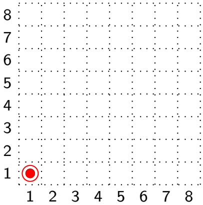
136	000
2	001
57	011
4	010
8	110



1 2 3 4 5 6 7 8

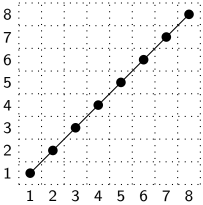


1 3 4 6 8 7 5 2

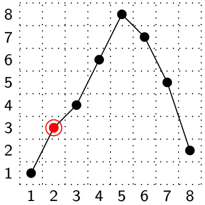


Step 2 – We construct the path

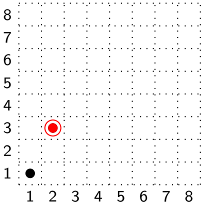
	⇓
136	000
2	001
57	011
4	010
8	110



1 2 3 4 5 6 7 8



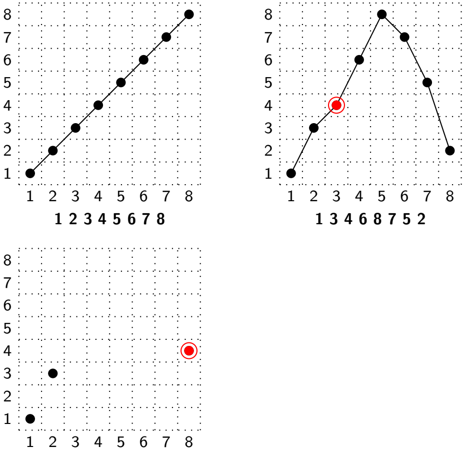
1 3 4 6 8 7 5 2



Step 2 – We construct the path

⇓

136	000
2	001
57	011
4	010
8	110



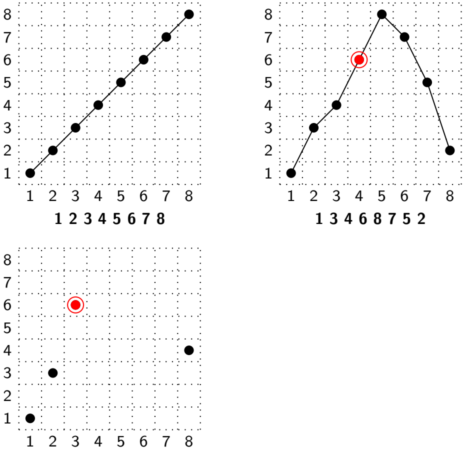
1 2 3 4 5 6 7 8

1 3 4 6 8 7 5 2

Step 2 – We construct the path

⇓

136	000
2	001
57	011
4	010
8	110



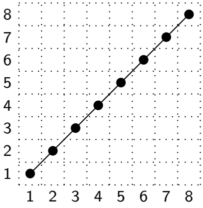
1 2 3 4 5 6 7 8

1 3 4 6 8 7 5 2

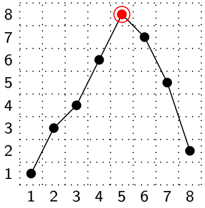
Step 2 – We construct the path

$\Downarrow$

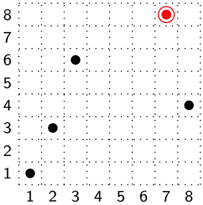
136	000
2	001
57	011
4	010
8	110



1 2 3 4 5 6 7 8



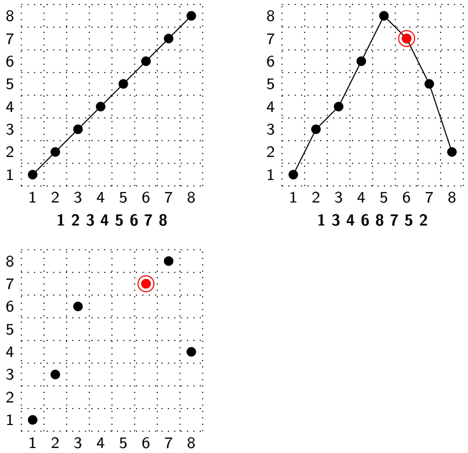
1 3 4 6 8 7 5 2



Step 2 – We construct the path

⇓

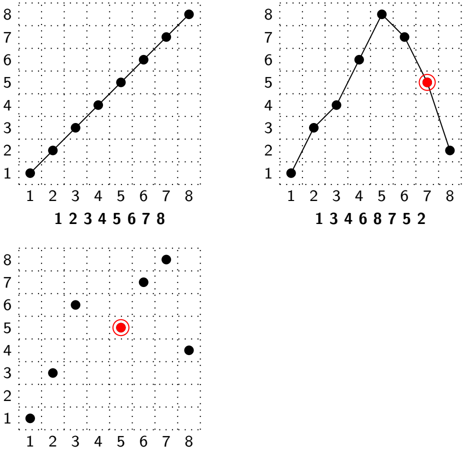
136	000
2	001
57	011
4	010
8	110



Step 2 – We construct the path

⇓

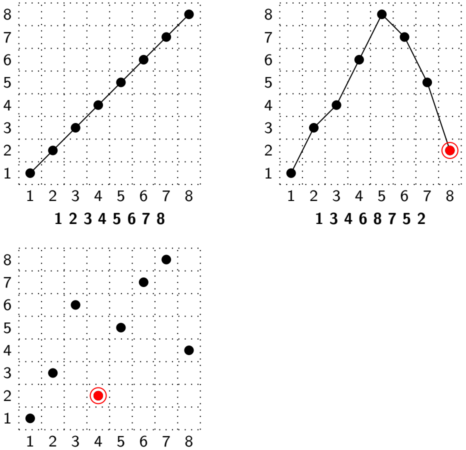
136	000
2	001
57	011
4	010
8	110



Step 2 – We construct the path

⇓

136	000
2	001
57	011
4	010
8	110

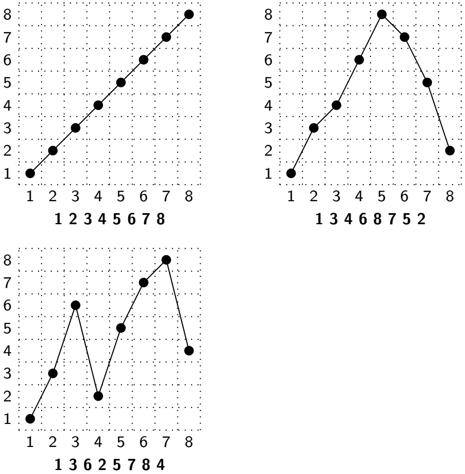


1 2 3 4 5 6 7 8

1 3 4 6 8 7 5 2

Step 2 – We construct the path

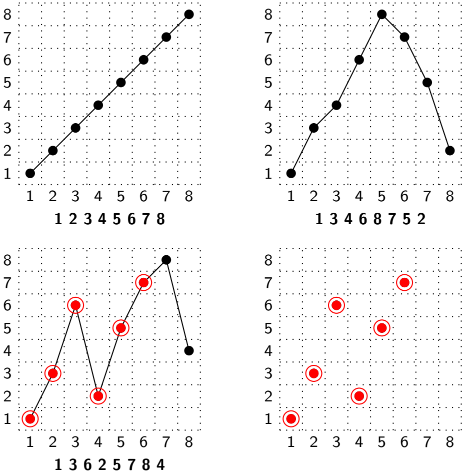
136	000
2	001
57	011
4	010
8	110



Step 2 – We construct the path

$\Downarrow$

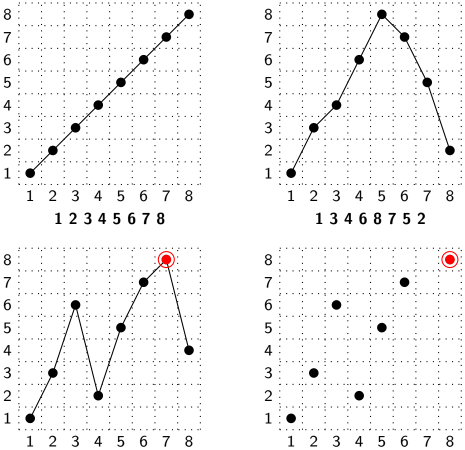
136	000
2	001
57	011
4	010
8	110



Step 2 – We construct the path

⇓

136	000
2	001
57	011
4	010
8	110



1 2 3 4 5 6 7 8

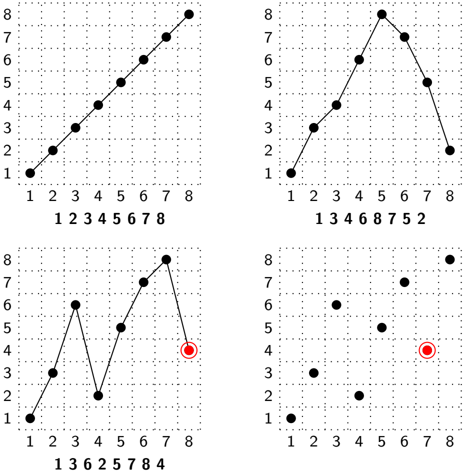
1 3 4 6 8 7 5 2

1 3 6 2 5 7 8 4

Step 2 – We construct the path

⇓

136	000
2	001
57	011
4	010
8	110



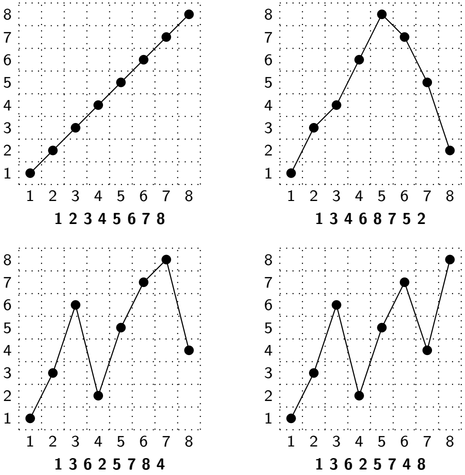
1 2 3 4 5 6 7 8

1 3 4 6 8 7 5 2

1 3 6 2 5 7 8 4

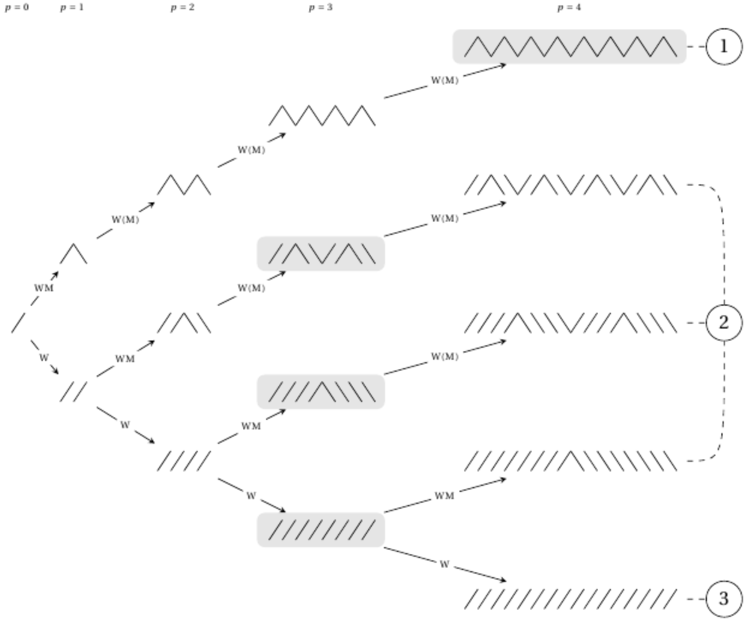
Step 2 – We construct the path

136	000
2	001
57	011
4	010
8	110



Complexity

One step requires : $\mathcal{O}(n)$
Whole process : $\mathcal{O}(n \cdot \log \text{val}(\sigma)) < \mathcal{O}(n \cdot \log n)$


p = 0
p = 1
p = 2
p = 3
p = 4
WM
W
W(M)
WM
W
W(M)
W(M)
WM
W
W(M)
W(M)
W(M)
WM
W
1
2
3