Posets and Permutations in the Duplication-Loss Model: Minimal Permutations with d Descents.

Mathilde Bouvel   Elisa Pergola

GASCom 2008

université PARIS DIDEROT PARIS 7
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
LIAFA
STUDIORUM UNIVERSITAS FLORENTINA

Outline of the talk

1. Pattern involvement and **minimal permutations with d descents**
2. Motivation: the duplication-loss model
3. **Local characterization** of minimal permutations with d descents
4. **Poset representation** of minimal permutations with d descents
5. **Enumeration**: partial results for subclasses of fixed size
6. Open problems and perspectives

Patterns in permutations

Definition (Pattern relation $\preccurlyeq$)

$\pi \in S_k$ is a pattern of $\sigma \in S_n$ when $\exists\ 1 \leq i_1 < \ldots < i_k \leq n$ such that $\sigma_{i_1} \ldots \sigma_{i_k}$ is order-isomorphic to π. We write $\pi \preccurlyeq \sigma$.

Equivalently: Normalizing $\sigma_{i_1} \ldots \sigma_{i_k}$ on $[1..k]$ yields π.

Example

$1\,2\,3\,4 \preccurlyeq 3\,\mathbf{1}\,\mathbf{2}\,8\,\mathbf{5}\,4\,\mathbf{7}\,9\,6$ since $1\,2\,5\,7 \equiv 1\,2\,3\,4$.

$Av(B)$: the class of permutations avoiding all the patterns in the basis B.

$Av(231)$ = Stack sortable ; $Av(2413, 3142)$ = Separable ; ...

Classes of permutations

Basis of excluded patterns

Definition (Permutation class)

$\mathcal{C}$ is a permutation class when it is stable for $\preccurlyeq$, *i.e.* when $\forall\sigma \in \mathcal{C}, \forall\pi \preccurlyeq \sigma, \pi \in \mathcal{C}$.

Theorem (Basis of excluded patterns)

Every permutation class $\mathcal{C}$ is characterized by a (finite or infinite) basis B of excluded patterns: $\mathcal{C} = Av(B)$.

Basis: $B = \{\sigma : \sigma \notin \mathcal{C} \text{ but } \forall\pi \prec \sigma, \pi \in \mathcal{C}\}$.
B is the set of minimal patterns not in $\mathcal{C}$.

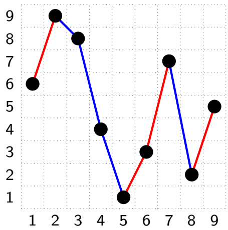
Minimal is intented in the sense of $\preccurlyeq$.

Descents in permutations

Grid representation

Definition (Descents and ascents in a permutation)

There is a descent (resp. ascent) in $\sigma \in S_n$ at position $i \in [1..n-1]$ when $\sigma_i > \sigma_{i+1}$ (resp. $\sigma_i < \sigma_{i+1}$).
desc(σ): the number of descents of σ.



The grid representation of the permutation $\sigma = 6\,9\,8\,4\,1\,3\,7\,2\,5$

ascents
descents

Minimal permutations with d descents

$\mathcal{D}_d$ = the set of permutations with at most $d-1$ descents.

Theorem

$\mathcal{D}_d$ is stable for $\preccurlyeq$, hence is a permutation class.
Basis of $\mathcal{D}_d$: the minimal (for $\preccurlyeq$) permutations not in $\mathcal{D}_d$

B_d = the set of minimal (for $\preccurlyeq$) permutations with d descents.
Rem.: In this context, **exactly** d descents $\Leftrightarrow$ **at least** d descents.

Theorem

The basis of excluded patterns characterizing $\mathcal{D}_d$ is B_d.
$\mathcal{D}_d = Av(B_d)$.

The (whole genome) duplication - (random) loss model

Definition (Duplication-loss step)

One duplication-loss step starting from a permutation σ:

- duplication of σ after itself
- loss of one of the two copies of every element

1 2 3 4 5 6 7 $\rightsquigarrow$ 1 2 3 4 5 6 7 1 2 3 4 5 6 7 $\rightsquigarrow$ ~~1~~ 2 3 ~~4~~ 5 6 ~~7~~ 1 ~~2~~ ~~3~~ 4 ~~5~~ ~~6~~ 7 $\rightsquigarrow$ 2 3 5 6 1 4 7

Cost of any step $= 1$.

Specialization of the tandem duplication-random loss model[1]:

- duplication: only of a fragment of the permutation
- cost of a step: depends on the number of elements duplicated

[1] Chaudhuri, Chen, Mihaescu and Rao, *On the tandem duplication-random loss model of genome rearrangement*, SODA06

Permutations obtained after p steps

Basis of this permutation class

What are the permutations obtainable from $1\,2\dots n$ (for any n) with a cost at most p ?

Specialized model $\rightsquigarrow$ Permutations obtained after p steps ?
Prop. σ is obtained in at most p steps $\Leftrightarrow$ $\mathrm{desc}(\sigma) \leq 2^p - 1$.

For $d = 2^p$, {Permutations obtained in at most p steps} $= \mathcal{D}_d$.

Theorem (Permutations obtained after p steps[2])

{Permutations obtained after p steps} is a class.
Basis = {minimal permutations with 2^p descents } = B_d.

[2] Bouvel and Rossin, *A variant of the tandem duplication - random loss model of genome rearrangement*

Study of B_d

What we know:

- Class $\mathcal{D}_d$ arise from biological motivations (for $d = 2^p$)
- $\mathcal{D}_d = Av(B_d)$
  $\hookrightarrow$ $B_d = \{$minimal permutations with d descents $\}$

What we want:

- Properties of the basis $B_d \Rightarrow$ Properties of the class $\mathcal{D}_d$

What we do:

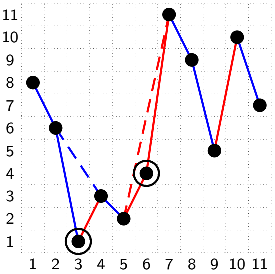
- Characterization of the permutations in B_d
- Size of the permutations in B_d
- Enumeration of the permutations of min. and max. size in B_d

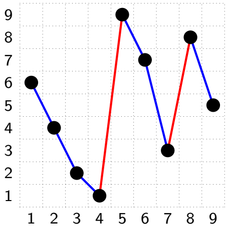
A necessary condition
for being minimal with d descents

Prop.: σ minimal with d descents $\Rightarrow$ no consecutive ascents in σ

Rem. This condition is **not** sufficient !

Proof:





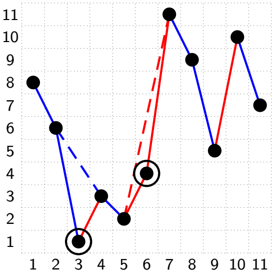
Consequence: σ minimal with d descents $\Rightarrow d+1 \leq |\sigma| \leq 2d$

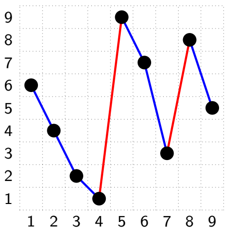
A necessary condition
for being minimal with d descents

Prop.: σ minimal with d descents $\Rightarrow$ no consecutive ascents in σ

Rem. This condition is **not** sufficient !

Proof:



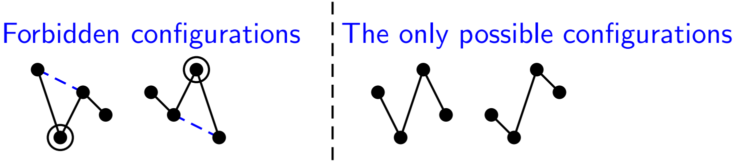


Consequence: σ minimal with d descents $\Rightarrow d+1 \leq |\sigma| \leq 2d$

A necessary **and sufficient** condition
for being minimal with d descents

Theorem (NSC for being minimal with d descents)

σ is minimal with d descents $\Leftrightarrow$ $desc(\sigma) = d$ and the 4 elements around each ascent of σ are ordered as 2143 or 3142.



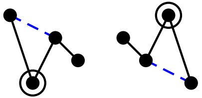
$\Rightarrow$ **Local characterization**

A necessary **and sufficient** condition
for being minimal with d descents

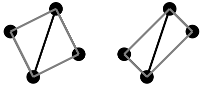
Theorem (NSC for being minimal with d descents)

σ is minimal with d descents $\Leftrightarrow$ $desc(\sigma) = d$ and the 4 elements around each ascent of σ are ordered as 2143 or 3142.

Forbidden configurations



The only possible configurations

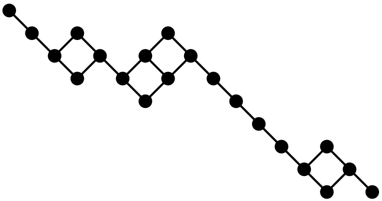


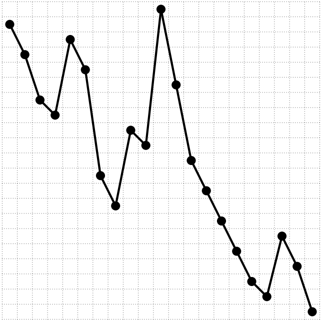
Diamonds

$\Rightarrow$ **Local characterization**

A poset for a set of minimal permutations with d descents

Same d, same size, and same positions of ascents and descents





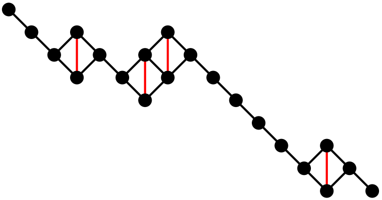
$d = 16$, size $= 21$

dddadddadadddddddadd

Bijection: Permutation $\Leftrightarrow$ Authorized labelling of the poset

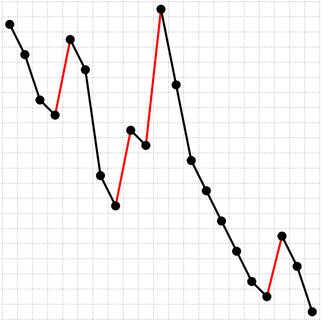
A poset for a set of minimal permutations with d descents

Same d, same size, and same positions of ascents and descents



$d = 16$, size $= 21$

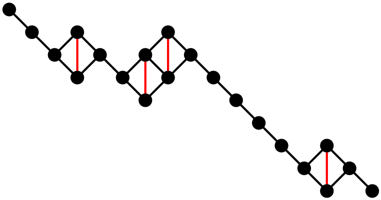
dddadddadaddddddddadd



Bijection: Permutation $\Leftrightarrow$ Authorized labelling of the poset

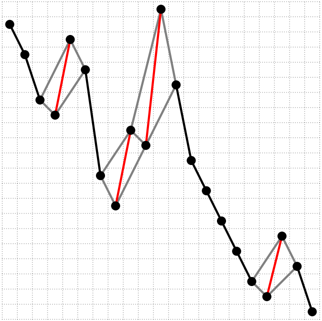
A poset for a set of minimal permutations with d descents

Same d, same size, and same positions of ascents and descents



$d = 16$, size $= 21$

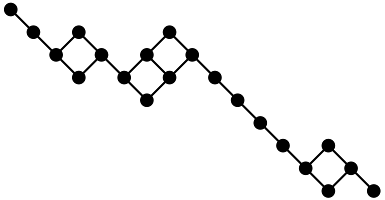
dddadddadadddddddadd



Bijection: Permutation $\Leftrightarrow$ Authorized labelling of the poset

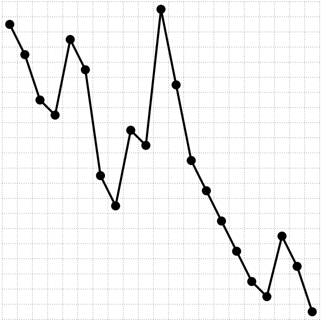
A poset for a set of minimal permutations with d descents

Same d, same size, and same positions of ascents and descents



$d = 16$, size $= 21$

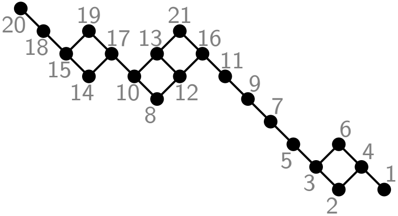
dddadddadadddddddadd

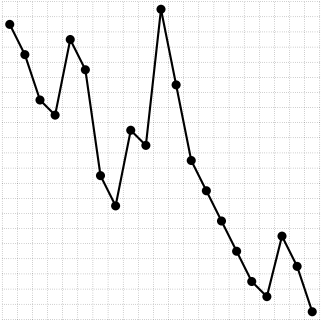


Bijection: Permutation $\Leftrightarrow$ Authorized labelling of the poset

A poset for a set of minimal permutations with d descents

Same d, same size, and same positions of ascents and descents





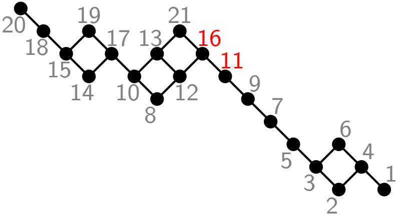
$d = 16$, size $= 21$

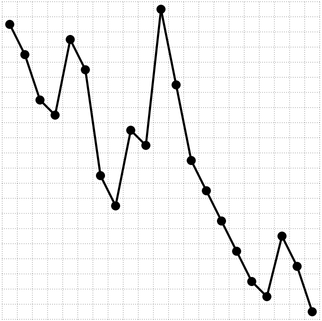
dddadddadaddddddddadd

Bijection: Permutation $\Leftrightarrow$ Authorized labelling of the poset

A poset for a set of minimal permutations with d descents

Same d, same size, and same positions of ascents and descents





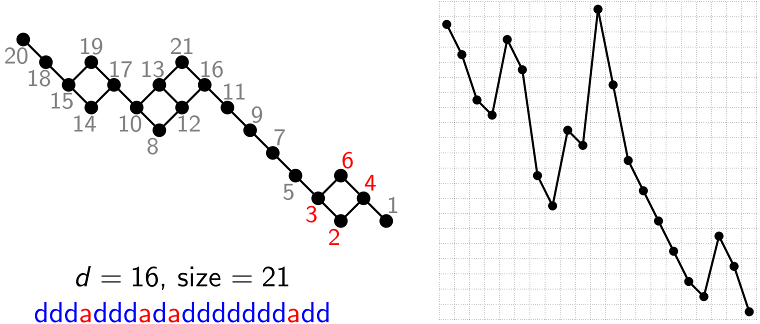
$d = 16$, size $= 21$

dddadddadadddddddadd

Bijection: Permutation $\Leftrightarrow$ Authorized labelling of the poset

A poset for a set of minimal permutations with d descents

Same d, same size, and same positions of ascents and descents



$d = 16$, size $= 21$

dddadddadadddddddadd

Bijection: Permutation $\Leftrightarrow$ Authorized labelling of the poset

Summary of the enumeration results obtained

Fact: $d+1 \leq |\sigma| \leq 2d$ for each σ minimal with d descents

Theorem (Partial enumeration of minimal permutation with d descents:)

Minimal size: 1 of size $d+1$

$\hookrightarrow$ the reverse identity of size $d+1$: $(d+1)d(d-1)\dots 321$

Minimal non-trivial size: $2^{d+2}-(d+1)(d+2)-2$ of size $d+2$

$\hookrightarrow$ Computational method

$\hookrightarrow$ Bijection with two copies of non-interval subsets of $\{1,2,\dots,d+1\}$

Maximal size: $C_d = \frac{1}{d+1}\binom{2d}{d}$ of size $2d$

$\hookrightarrow$ Using the ECO method

$\hookrightarrow$ Bijection with Dyck paths

Proof: Using poset representation

Minimal permutations with d descents of size $2d$

A unique poset represents all permutations

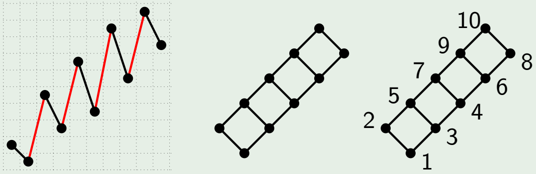
Facts:

- $2d$ elements, d descents $\Rightarrow d-1$ ascents
- Minimal $\Rightarrow$ ascents $=$ diamonds between two descents

Consequence: Poset $=$ ladder poset with d steps

Def.: Ladder poset $=$ sequence of $d-1$ diamonds linked by an edge

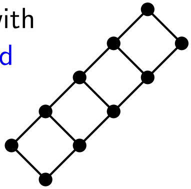
Example (for $d = 5$; Sequence dadadadad)



Size $2d$: proof by enumeration

ECO construction for authorized labelling of the ladder poset with d steps

Minimal permutation with d descents $\equiv$ authorized labelling of the ladder poset with d steps



Label k = number of children $= 2d - i + 1$

Labels of the children $= 2(d+1) - i' + 1$ for $i' \in [(i+1)..(2d+1)]$

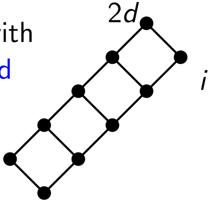
Succession rule: $\begin{cases} (2) \\ (k) \rightsquigarrow (2)(3)\cdots(k)(k+1) \end{cases}$

Enumerating sequence = Catalan numbers $C_d = \frac{1}{d+1}\binom{2d}{d}$

Size $2d$: proof by enumeration

ECO construction for authorized labelling of the ladder poset with d steps

Minimal permutation with d descents $\equiv$ authorized labelling of the ladder poset with d steps



Label k = number of children $= 2d - i + 1$

Labels of the children $= 2(d+1) - i' + 1$ for $i' \in [(i+1)..(2d+1)]$

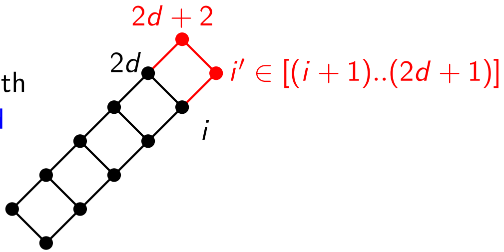
Succession rule: $\left\{ \begin{array}{l} (2) \\ (k) \rightsquigarrow (2)(3)\cdots(k)(k+1) \end{array} \right.$

Enumerating sequence = Catalan numbers $C_d = \frac{1}{d+1}\binom{2d}{d}$

Size $2d$: proof by enumeration

ECO construction for authorized labelling of the ladder poset with d steps

Minimal permutation with d descents $\equiv$ authorized labelling of the ladder poset with d steps



Label k = number of children $= 2d - i + 1$

Labels of the children $= 2(d+1) - i' + 1$ for $i' \in [(i+1)..(2d+1)]$

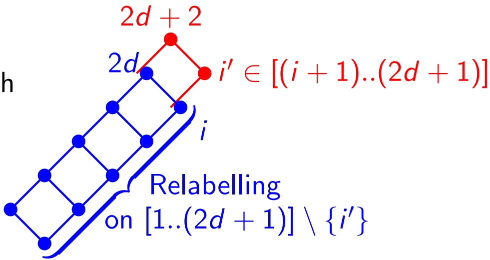
Succession rule: $\begin{cases} (2) \\ (k) \rightsquigarrow (2)(3)\cdots(k)(k+1) \end{cases}$

Enumerating sequence = Catalan numbers $C_d = \frac{1}{d+1}\binom{2d}{d}$

Size $2d$: proof by enumeration

ECO construction for authorized labelling of the ladder poset with d steps

Minimal permutation with d descents $\equiv$ authorized labelling of the ladder poset with d steps



Label k = number of children $= 2d - i + 1$

Labels of the children $= 2(d+1) - i' + 1$ for $i' \in [(i+1)..(2d+1)]$

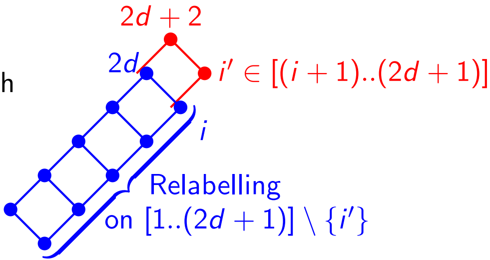
Succession rule: $\begin{cases} (2) \\ (k) \rightsquigarrow (2)(3)\cdots(k)(k+1) \end{cases}$

Enumerating sequence = Catalan numbers $C_d = \frac{1}{d+1}\binom{2d}{d}$

Size $2d$: proof by enumeration

ECO construction for authorized labelling of the ladder poset with d steps

Minimal permutation with d descents $\equiv$ authorized labelling of the ladder poset with d steps



Label k = number of children $= 2d - i + 1$

Labels of the children $= 2(d+1) - i' + 1$ for $i' \in [(i+1)..(2d+1)]$

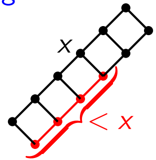
Succession rule: $\begin{cases} (2) \\ (k) \rightsquigarrow (2)(3)\cdots(k)(k+1) \end{cases}$

Enumerating sequence = Catalan numbers $C_d = \frac{1}{d+1}\binom{2d}{d}$

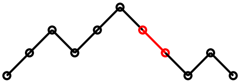
Size $2d$: proof by bijection

Bijection between Dyck paths and authorized labellings of the ladder poset

Labellings of the ladder poset

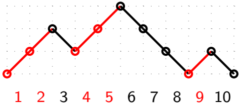


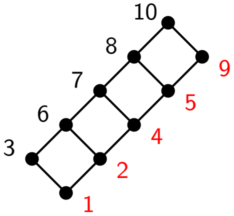
Dyck paths at least i up steps before the i-th down step



Bijection:

- lower line ≡ up step
- upper line ≡ down step





Size $d+2$: computational and bijective approaches

Theorem

There are $s_d = 2^{d+2} - (d+1)(d+2) - 2$ minimal permutations with d descents and of size $d+2$.

Computational proof

Fact: Only one diamond

- Choose the pattern of the diamond: 2143 or 3142
- Choose the elements labelling the diamond
- Choose (or remark) where the other labels are placed

$\Rightarrow$ Summation that simplifies into s_d

Size $d+2$: computational and bijective approaches

Theorem

There are $s_d = 2^{d+2} - (d+1)(d+2) - 2$ minimal permutations with d descents and of size $d+2$.

Bijective proof

Fact: $r_d = \frac{s_d}{2} =$ number of non-interval subsets of $\{1, 2, \ldots, (d+1)\}$

- Partition the set of permutations into $S_1 \uplus S_2$
- Simple bijection between S_1 and non-interval subsets
- More tricky bijection between S_2 and non-interval subsets
  - $\hookrightarrow$ Classification of permutations in S_2 into 5 types of permutations

Permutations with at most $d-1$ descents:

- Motivations in bio-informatics
- Define a permutation class by a property

Minimal permutations with d descents:

- Basis of the above
- Local characterization

Enumeration:

- Done for $n \in \{d+1, d+2, 2d\}$
- Open for $n \in [(d+3)..(2d-1)]$: computational method, with automated examination of (numerous) cases ?

Classes $\mathcal{C}$ defined by a property:

- Literature (stack sortable, separable, ...): simple basis
- Properties of the basis $B \Rightarrow$ Properties of $\mathcal{C}$