Edit Distances and Factorisations of Even Permutations

Anthony Labarre[1]
alabarre@ulb.ac.be

Université libre de Bruxelles (U.L.B.)

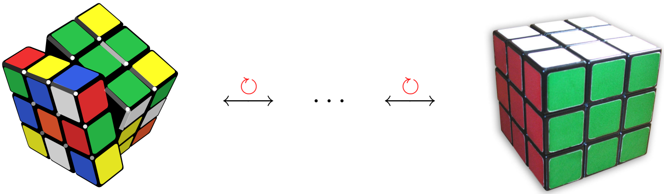
September 15th, 2008

Sixteenth European Symposium on Algorithms (ESA)

[1]Funded by the "Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture" (F.R.I.A.)

Edit distances

- *Edit operations*: given fixed set of allowed operations;
- *Edit distance*: minimum number of edit operations needed to transform X into Y;



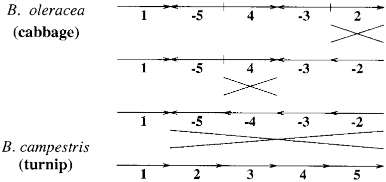
- Many applications:
  - spelling correction (example: type “dsitnace” in Google);
  - genome rearrangements;
  - interconnection networks;

    ⋮

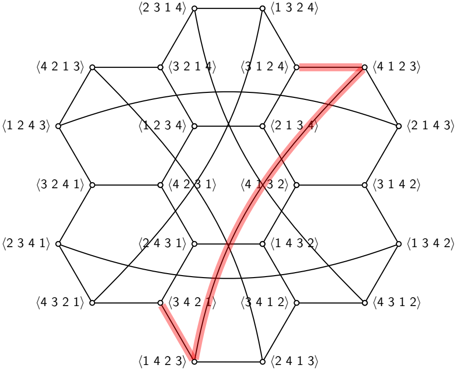
Permutations

- Permutations can model:

genomes and mutations
[Hannenhalli and Pevzner, 1999]



devices in interconnection networks

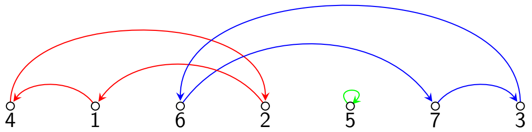


Permutations: basic definitions

- Permutation: linear ordering of $\{1, 2, \ldots, n\}$;
- Disjoint cycle decomposition:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 1 & 6 & 2 & 5 & 7 & 3 \end{pmatrix} = (1, 4, 2)(3, 6, 7)(5).$$

- The graph of permutation π, denoted by $\Gamma(\pi)$:



- π is *even* if $\Gamma(\pi)$ has an even number of even cycles;
- *Conjugacy class*: permutations with the same decomposition;
- 1-cycles (or *fixed points*) are often omitted;

The problem(s)

- Let:
  - π be a permutation of $\{1, 2, \ldots, n\}$;
  - $S = \{s_1, s_2, \ldots\}$ be a set of permutations of $\{1, 2, \ldots, n\}$ (the *edit operations*);
  - ι be the *identity permutation* $\langle 1\ 2\ \cdots\ n \rangle$;
- We want to:
  1. "**sort** π **by** S": find a sequence of elements of S that sorts π and is as short as possible:

     $$\pi \circ x_1 \circ x_2 \circ \cdots \circ x_t = \iota \text{ where } x_1, \ldots, x_t \in S \text{ and } t \text{ is minimal}$$

  2. "**compute the** S**-distance** $d_S(\pi, \iota)$": find the length of such a sequence;

Some edit operations

- From genome rearrangements:
  - reversals: $\langle 3\ 2\ \underline{5\ 4\ 1}\rangle \longrightarrow \langle\underline{3\ 2\ 1}\ 4\ 5\rangle$
  - transpositions: $\langle 3\ \boxed{2\ 5}\ \boxed{4\ 1}\rangle \longrightarrow \langle\boxed{3\ 4}\ \boxed{1\ 2}\ 5\rangle$
  - block-interchanges: $\langle\boxed{5}\ 4\ \boxed{3}\ 2\ 1\rangle \longrightarrow \langle\boxed{3\ 4\ 5}\ 2\ \boxed{1}\rangle$
- From interconnection networks:
  - prefix reversals: $\langle\underline{2\ 3\ 5\ 4}\ 1\rangle \longrightarrow \langle 4\ 5\ 3\ 2\ 1\rangle$
  - prefix transpositions: $\langle\boxed{3\ 2\ 5}\ \boxed{4\ 1}\rangle \longrightarrow \langle 4\ 1\ 3\ 2\ 5\rangle$

Background

	Operation	Sorting	Distance	Diameter
classical	reversals	NP-hard	NP-hard	$n-1$
	signed reversals	$O(n^{3/2})$	$O(n)$	$n+1$
	block-interchanges	$O(n \log n)$	$O(n)$	$n/2$
	transpositions[2]	?	?	$\frac{n}{2} \leq ? \leq \frac{2n}{3}$
prefix	reversals	?	?	$\frac{15n}{14} \leq ? \leq \frac{18n}{11}$
	signed reversals	?	?	$\frac{3n}{2} \leq ? \leq 2(n-1)$
	transpositions	?	?	$\frac{2n}{3} \leq ? \leq n - \log_8 n$

- All three prefix variants are 2-approximable;

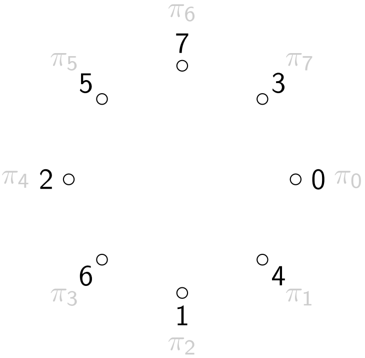
[2] 11/8-approximable [Elias and Hartman, 2006]

Results

- Expression of the "cycle graph" [Bafna and Pevzner, 1998] of π as an even permutation $\overline{\pi}$;
- Reformulation of **every** edit distance problem on π in terms of particular factorisations of $\overline{\pi}$;
- Simple recovery of previous results;
- New lower bound on the prefix transposition distance, which outperforms previous results;
- Improved lower bound on the maximal value of that distance ($\frac{2n}{3} \longrightarrow \left\lfloor \frac{3n+1}{4} \right\rfloor$);

The "cycle graph" [Bafna and Pevzner, 1998]

- The "cycle graph" of π, denoted by $G(\pi)$:

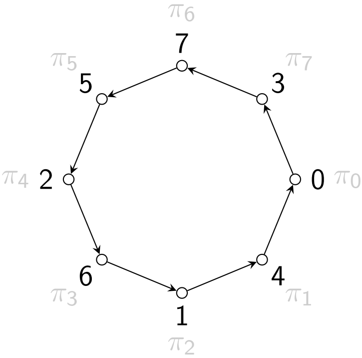


(here $\pi = \langle 4\ 1\ 6\ 2\ 5\ 7\ 3 \rangle$)

1. $V(G) = (\pi_0 = 0, \pi_1, \pi_2, \ldots, \pi_n)$;
2. $E(G) =$

The "cycle graph" [Bafna and Pevzner, 1998]

- The "cycle graph" of π, denoted by $G(\pi)$:

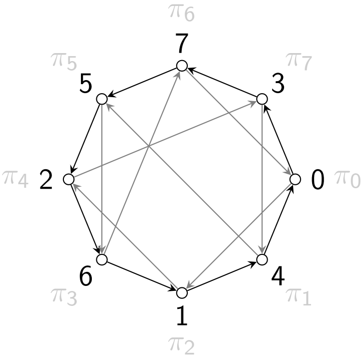


(here $\pi = \langle 4\ 1\ 6\ 2\ 5\ 7\ 3 \rangle$)

1. $V(G) = (\pi_0 = 0, \pi_1, \pi_2, \ldots, \pi_n)$;
2. $E(G) = \{$*black* arcs$\}$

The “cycle graph” [Bafna and Pevzner, 1998]

- The “cycle graph” of π, denoted by $G(\pi)$:

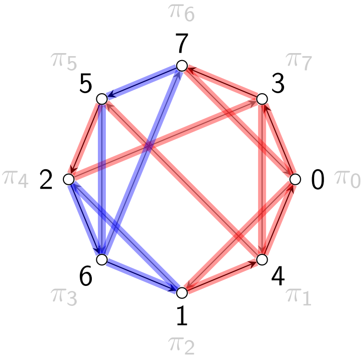


(here $\pi = \langle 4\ 1\ 6\ 2\ 5\ 7\ 3 \rangle$)

1. $V(G) = (\pi_0 = 0, \pi_1, \pi_2, \ldots, \pi_n)$;
2. $E(G) = \{$*black* arcs$\} \cup \{$*grey* arcs$\}$;

The "cycle graph" [Bafna and Pevzner, 1998]

- The "cycle graph" of π, denoted by $G(\pi)$:



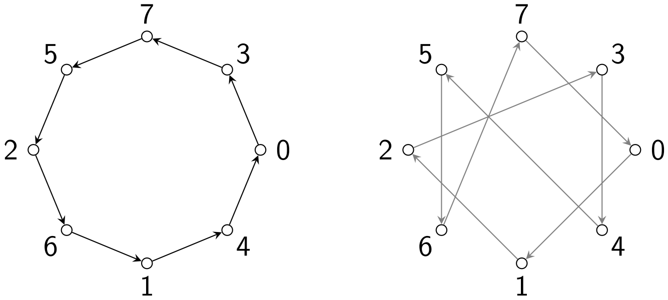
(here $\pi = \langle 4\ 1\ 6\ 2\ 5\ 7\ 3 \rangle$)

1. $V(G) = (\pi_0 = 0, \pi_1, \pi_2, \ldots, \pi_n)$;
2. $E(G) =$ {*black* arcs} $\cup$ {*grey* arcs};

- Unique decomposition into "alternating cycles";

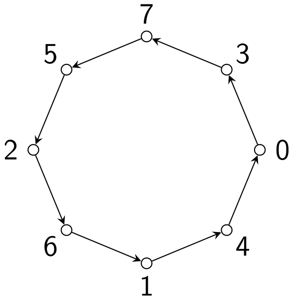
$G(\pi)$ as an even permutation

- "Monochrome" decomposition:

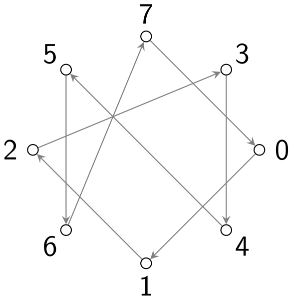


$G(\pi)$ as an even permutation

- "Monochrome" decomposition:



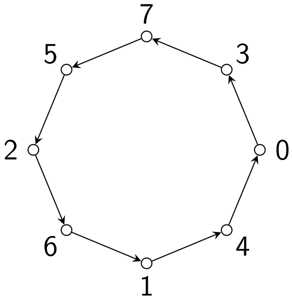
$\hat{\pi} = (0, 3, 7, 5, 2, 6, 1, 4)$

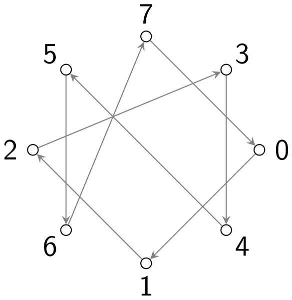


$\dot{\pi} = (0, 1, 2, 3, 4, 5, 6, 7)$

$G(\pi)$ as an even permutation

- "Monochrome" decomposition:

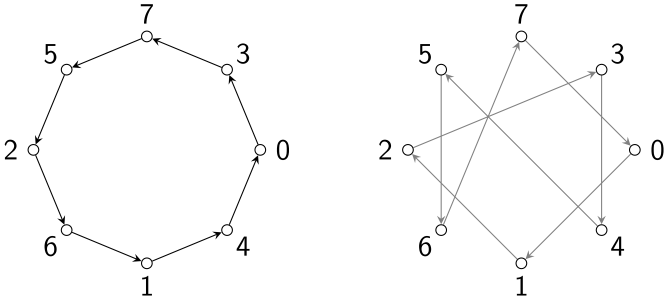




$$\overline{\pi} = \underbrace{\hat{\pi}}_{=\ (0,3,7,5,2,6,1,4)} \circ \underbrace{\dot{\pi}}_{=\ (0,1,2,3,4,5,6,7)}$$

$G(\pi)$ as an even permutation

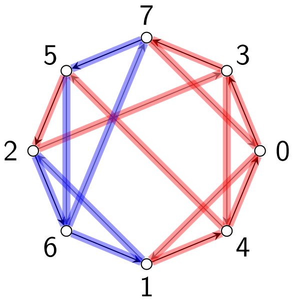
- "Monochrome" decomposition:



$$
\begin{array}{rccc}
 & \hat{\pi} & & \dot{\pi} \\
 & = & & = \\
\overline{\pi} = & (0,3,7,5,2,6,1,4) & \circ & (0,1,2,3,4,5,6,7) \\
= & & (0,4,2,7,3)(1,6,5) &
\end{array}
$$

$G(\pi)$ as an even permutation

- We have $\overline{\pi} = \hat{\pi} \circ \dot{\pi}$, with $\Gamma(\overline{\pi}) \simeq G(\pi)$; indeed:



$$
\begin{array}{rccc}
 & \hat{\pi} & & \dot{\pi} \\
 & = & & = \\
\overline{\pi} = & (0,3,7,5,2,6,1,4) & \circ & (0,1,2,3,4,5,6,7) \\
= & & (0,4,2,7,3)(1,6,5) &
\end{array}
$$

A general lower bounding technique

- Note that "sorting by S" is equivalent to "factorising by S":

$$\pi \circ \underbrace{x_1 \circ x_2 \circ \cdots \circ x_t}_{x_1, x_2, \ldots, x_t \in S} = \iota \Leftrightarrow \pi = \underbrace{x_t^{-1} \circ x_{t-1}^{-1} \circ \cdots \circ x_1^{-1}}_{x_1^{-1}, x_2^{-1}, \ldots, x_t^{-1} \in S}$$

Theorem 1

Let:

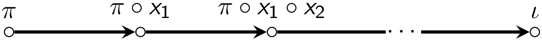
1. $S \subset S_n$, *with* $S = \{s_1, s_2, \ldots\}$,
2. $S' = \{\overline{s_1}, \overline{s_2}, \ldots\}$,
3. $\mathcal{C}$ *the set of conjugacy classes that intersect* S'.

Then for all π *in* S_n, *every factorisation of* π *into* t *elements of* S *yields a factorisation of* $\overline{\pi}$ *into* t *elements of* $\mathcal{C}$.

Theorem 1 in action

Our problem:

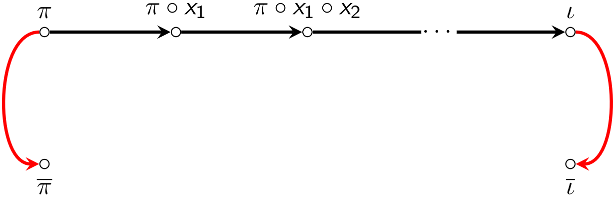
computing $d_S(\pi)$:



Theorem 1 in action

Our problem:
computing $d_S(\pi)$:

New problem:
computing $d_C(\overline{\pi})$:



Theorem 1 in action

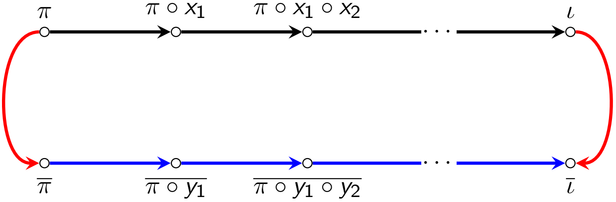
Our problem:

computing $d_S(\pi)$:

$d_{\mathcal{C}}(\overline{\pi}) \leq d_S(\pi)$

New problem:

computing $d_{\mathcal{C}}(\overline{\pi})$:

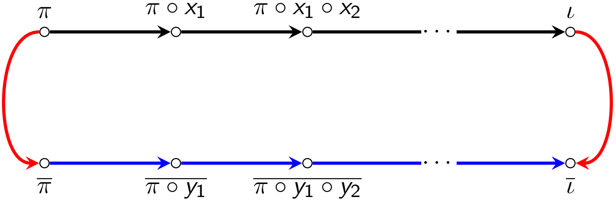


Theorem 1 in action

Our problem:
computing $d_S(\pi)$:

$\boxed{d_C(\overline{\pi}) \leq d_S(\pi)}$

New problem:
computing $d_C(\overline{\pi})$:



Lemma 2

For all π, σ in S_n:

$$\begin{aligned}\overline{\pi \circ \sigma} &= \pi \circ \overline{\sigma} \circ \pi^{-1} \circ \overline{\pi} \\ &= \overline{\sigma}^{\pi} \circ \overline{\pi}.\end{aligned}$$

A lower bound on the block-interchange distance

Example 3 (lower bound on $bid(\pi)$)

- $S = \{\text{block-interchanges}\}$, denoted by $\beta(i, j, k, l)$;

$$\begin{pmatrix} 1 & \cdots & i-1 & \boxed{i \ \cdots \ j-1} & j & j+1 & \cdots & k-1 & \boxed{k \ \cdots \ l-1} & l & l+1 & \cdots & n \\ 1 & \cdots & i-1 & \boxed{k \ \cdots \ l-1} & j & j+1 & \cdots & k-1 & \boxed{i \ \cdots \ j-1} & l & l+1 & \cdots & n \end{pmatrix}.$$

A lower bound on the block-interchange distance

Example 3 (lower bound on $bid(\pi)$)

- S ={block-interchanges}, denoted by $\beta(i, j, k, l)$; we have:

$$\overline{\beta(i, j, k, l)} = (i-1, k-1)(j-1, l-1) \quad (1 \leq i < j \leq k < l \leq n+1).$$

- We have $S' \subseteq \mathcal{C}$, where $\mathcal{C}$ contains all pairs of 2-cycles;
- We have $d_{\mathcal{C}}(\overline{\pi}) = \frac{|\overline{\pi}|-c(\Gamma(\overline{\pi}))}{2}$;
- Therefore, we recover the result of [Christie, 1996]:

$$\forall\ \pi \in S_n : bid(\pi) \geq \frac{n+1-c(\Gamma(\overline{\pi}))}{2}.$$

A lower bound on the transposition distance

Example 3 (lower bound on $td(\pi)$)

- S ={transpositions}, denoted by $\tau(i, j, k)$;

$$\left(\begin{array}{ccccccccccccccccc} 1 & \cdots & i-1 & i & i+1 & \cdots & j-2 & j-1 & j & j+1 & \cdots & k-1 & k & \cdots & n \\ 1 & \cdots & i-1 & j & j+1 & \cdots & k-1 & i & i+1 & \cdots & j-2 & j-1 & k & \cdots & n \end{array}\right).$$

A lower bound on the transposition distance

Example 3 (lower bound on $td(\pi)$)

- S ={transpositions}, denoted by $\tau(i,j,k)$; we have:

$$\overline{\tau(i,j,k)} = (i-1, k-1, j-1) \quad (1 \leq i < j < k \leq n+1).$$

- We have $S' \subseteq \mathcal{C}$, the set of all 3-cycles;
- We have $d_{\mathcal{C}}(\overline{\pi}) = \frac{|\overline{\pi}| - c_{odd}(\Gamma(\overline{\pi}))}{2}$;
- Therefore, we recover the result of [Bafna and Pevzner, 1998]:

$$\forall\ \pi \in S_n : td(\pi) \geq \frac{n+1-c_{odd}(\Gamma(\overline{\pi}))}{2}.$$

A new lower bound on the prefix transposition distance

Example 3 (lower bound on $ptd(\pi)$)

- S ={**prefix** transpositions}; we get:

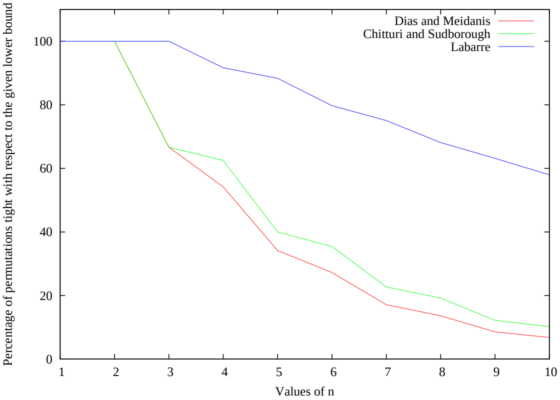
$$\overline{\tau(1, j, k)} = (0, k-1, j-1) \quad (1 < j < k \leq n+1).$$

- We have $S' \subseteq \mathcal{C}$, the set of all 3-cycles that contain 0;
- We can compute $d_{\mathcal{C}}(\overline{\pi})$, and this yields the following *new* lower bound :

$$\forall\, \pi \in S_n : ptd(\pi) \geq \frac{n+1+c(\Gamma(\overline{\pi}))}{2} - c_1(\Gamma(\overline{\pi})) - \begin{cases} 0 & \text{if } \pi_1 = 1, \\ 1 & \text{otherwise.} \end{cases}$$

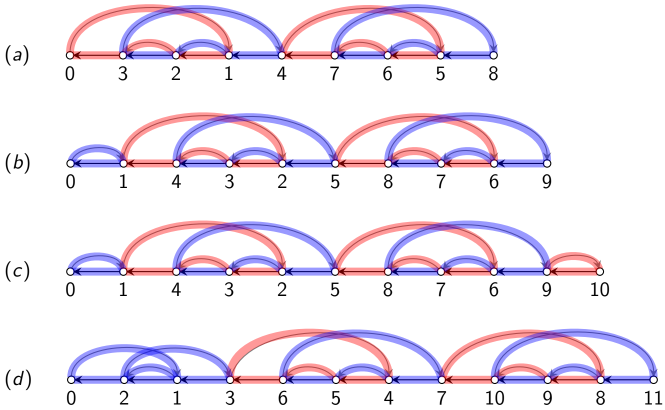
Quality of the results

- Block-interchanges: the lower bound is the exact distance;
- Prefix transpositions: the new result:
  - *always* outperforms [Dias and Meidanis, 2002];
  - "often" outperforms [Chitturi and Sudborough, 2008]:



The prefix transposition diameter of S_n

- These permutations satisfy $ptd(\pi) \geq \left\lfloor \frac{3n+1}{4} \right\rfloor$, thereby improving on the lower bound of $2n/3$ by [Chitturi and Sudborough, 2008]:



Future work

- Complexity/approximation issues (transpositions, prefix operations);
- Can the $\overline{\pi}$ model provide *upper* bounds?
- Extending the $\overline{\pi}$ model to *signed* permutations and/or other structures;

Thank you!

Bafna, V. and Pevzner, P. A. (1998).
Sorting by transpositions.
SIAM Journal on Discrete Mathematics, 11(2):224–240 (electronic).

Chitturi, B. and Sudborough, I. H. (2008).
Bounding prefix transposition distance for strings and permutations.
In *HICSS*, page 468, Los Alamitos, CA, USA.

Christie, D. A. (1996).
Sorting permutations by block-interchanges.
Information Processing Letters, 60(4):165–169.

Dias, Z. and Meidanis, J. (2002).
Sorting by prefix transpositions.
In *SPIRE*, volume 2476 of *LNCS*, pages 65–76. Springer-Verlag.

Elias, I. and Hartman, T. (2006).
A 1.375-approximation algorithm for sorting by transpositions.
IEEE/ACM Trans. Comput. Biol. Bioinform., 3(4):369–379.

Hannenhalli, S. and Pevzner, P. A. (1999).
Transforming cabbage into turnip: Polynomial algorithm for sorting signed permutations by reversals.
Journal of the ACM, 46(1):1–27.