

A new tight upper bound on the transposition distance

Anthony Labarre

► **To cite this version:**

Anthony Labarre. A new tight upper bound on the transposition distance. Workshop on Algorithms in Bioinformatics (WABI), Sep 2005, Spain. pp.216-227, 10.1007/11557067_18 . hal-00728796

HAL Id: hal-00728796

<https://hal-upec-upem.archives-ouvertes.fr/hal-00728796>

Submitted on 6 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A New Tight Upper Bound on the Transposition Distance

Anthony Labarre*

Université Libre de Bruxelles,
Département de Mathématique, CP 216,
Service de Géométrie, Combinatoire et Théorie des Groupes,
Boulevard du Triomphe, B-1050 Bruxelles, Belgium
alabarre@ulb.ac.be

Abstract. We study the problem of computing the minimal number of adjacent, non-intersecting block interchanges required to transform a permutation into the identity permutation. In particular, we use the graph of a permutation to compute that number for a particular class of permutations in linear time and space, and derive a new tight upper bound on the so-called transposition distance.

1 Introduction

The problem we study is a particular case of a problem called *genome rearrangement* [1,2], which is motivated by applications in biology. The genome rearrangement problem can be formulated as follows: given two genomes, find the minimum number of evolutionary events transforming one into the other. This number is defined as the *distance* between the two genomes.

The model we are interested in applies to the case where the order of genes is known and where all genomes share the same set and number of genes (without duplications), which allows us to represent them by *permutations*. It is easy to show that what we have defined as a distance is indeed a distance on the set of all permutations (i.e. it satisfies the three usual axioms).

We will consider only one operation on permutations: biological *transpositions*, which consist in moving a block of contiguous elements from one place to another one. This problem was first introduced in 1995 by Bafna and Pevzner [3], and is generally considered harder than similar problems. In particular, neither its complexity, nor even the *diameter* of the transposition distance (i.e. the maximal value it can reach), is known, which has led several authors to design polynomial-time approximation algorithms (whose best known ratio¹ is $\frac{3}{2}$ [3,5,6]) as well as using and comparing heuristics [5,7,8,9]. An interesting property of this distance is that the transposition distance between any two permutations π , σ is the same as the distance between $\sigma^{-1} \circ \pi$ and the *identity permutation*

* Funded by the “Fonds pour la Formation à la Recherche dans l’Industrie et dans l’Agriculture” (F.R.I.A.).

¹ A 1.375-approximation has recently been proposed in [4].

$\iota = (1\ 2\ \dots\ n)$. Therefore, the problem of transforming a permutation into another one using as few transpositions as possible is the same as that of *sorting* a permutation using the minimum number of transpositions. In what follows, we refer to the latter number as the *distance of π* , noted $d(\pi)$. While other authors have tried to find the shortest possible sequences of transpositions that sort a permutation, we have chosen to focus on computing their length.

In this paper, we make use of the common graph of a permutation rather than of the “cycle graph” introduced in [3]. A step in this direction was mentioned in [10] and successfully used to compute another rearrangement distance in [11]. As we suspected, it proved fruitful for our problem too: we were able to show that the distance of some nicely characterized permutations, namely those who fix even or odd elements and another class derived from those two, can be computed in linear time, using a formula that completely bypasses any graph structure used so far. Furthermore, we use those permutations to derive a tight upper bound on the transposition distance of every permutation.

This paper is organized as follows. In Section 2, we review previous results and typical notations. In Section 3, we introduce a graph that we use in Section 4 to provide a formula for computing the distance of some special permutations in linear time. In Section 5, we use those to derive an upper bound on the transposition distance of every permutation. Finally, we discuss our results in Section 6 and suggest some open questions of interest.

2 Notations and Preliminaries

Permutations are denoted by lower case Greek letters, typically π , and S_n is the set of all permutations of $\{1, 2, \dots, n\}$. For any permutation π in S_n , the *transposition* $\tau(i, j, k)$ with $1 \leq i < j < k \leq n + 1$ applied to π exchanges the closed intervals determined respectively by i and $j - 1$ and by j and $k - 1$, transforming π into $\pi \circ \tau(i, j, k)$. So $\tau(i, j, k)$ is the following permutation:

$$\left(\begin{array}{c} 1 \ \dots \ i - 1 \ \boxed{i \ i + 1 \ \dots \ j - 2 \ j - 1} \ \boxed{j \ j + 1 \ \dots \ k - 1} \ k \ \dots \ n \\ 1 \ \dots \ i - 1 \ \boxed{j \ j + 1 \ \dots \ k - 1} \ \boxed{i \ i + 1 \ \dots \ j - 2 \ j - 1} \ k \ \dots \ n \end{array} \right).$$

The usual notation is shorter than the one we have just used to describe transpositions, i.e. we write a permutation π in S_n as $(\pi_1 \ \pi_2 \ \dots \ \pi_n)$. Bafna and Pevzner [3] define the *cycle graph* of π as the bicoloured directed graph $G(\pi)$, whose vertex set consists of the elements of π plus two new elements $\pi_0 = 0$ and $\pi_{n+1} = n + 1$, and whose edge set consists of:

- *black* edges (π_i, π_{i-1}) for $1 \leq i \leq n + 1$;
- *gray* edges $(i, i + 1)$ for $0 \leq i \leq n$.

The set of black and gray edges decomposes in a single way into *alternate cycles*, i.e. cycles which alternate black and gray edges, and we note the number of such cycles $c(G(\pi))$. An alternate cycle of $G(\pi)$ is *odd* (resp. *even*) if it contains an *odd* (resp. *even*) number of black edges, and we note $c_{\text{odd}}(G(\pi))$ (resp.

$c_{even}(G(\pi))$ the number of odd (resp. even) alternate cycles of $G(\pi)$. Bafna and Pevzner proved the following lower bound on the transposition distance.

Theorem 1. [3] $\forall \pi \in S_n : d(\pi) \geq \frac{n+1-c_{odd}(G(\pi))}{2}$.

For a permutation π , define an ordered pair (π_i, π_{i+1}) as a *breakpoint* if $\pi_{i+1} \neq \pi_i + 1$. The number of breakpoints of π is denoted by $b(\pi)$. Christie [5] decomposes permutations into *strips*, which he defines as maximal intervals containing no breakpoint. He denotes $gl(\pi)$ the *reduced* version of π , obtained as follows: assuming π has r strips, remove strip 1 if it begins with 1, strip r if it ends with n , replace every other strip with its minimal element and finally, renumber the resulting sequence so as to obtain a permutation of S_r ($r \leq n$).

Theorem 2. [5] $\forall \pi \in S_n : d(\pi) = d(gl(\pi))$.

We say that π and σ are *equivalent by reduction* if $gl(\pi) = gl(\sigma)$, which we also write as $\pi \equiv_r \sigma$. Since we are presenting a new upper bound on the transposition distance of every permutation, it is only fair that we conclude this section with the ones that were previously shown.

Theorem 3. [3] $\forall \pi \in S_n :$

$$d(\pi) \leq \frac{3(n+1-c_{odd}(G(\pi)))}{4} . \tag{1}$$

Theorem 4. [12] $\forall \pi \in S_n :$

$$d(\pi) \leq \frac{3}{4} b(\pi) . \tag{2}$$

Theorem 5. [13] $\forall \pi \in S_n :$

$$d(\pi) \leq \begin{cases} \lceil \frac{2n}{3} \rceil & \text{if } n < 9 ; \\ \lfloor \frac{2n-2}{3} \rfloor & \text{if } n \geq 9 . \end{cases} \tag{3}$$

3 Another Useful Graph

We will make use of a variant of the well-known *graph of a permutation*. The Γ -*graph* of a permutation π in S_n is the directed graph $\Gamma(\pi)$ with vertex set $\{(1, \pi_1), (2, \pi_2), \dots, (n, \pi_n)\}$ and edge set $\{(i, \pi_i), (j, \pi_j) \mid j = \pi_i\}$.

If $C = (i_1, i_2, \dots, i_k)$ is a cycle of π (i.e. π maps i_l onto i_{l+1} for $1 \leq l \leq k-1$ and i_k onto i_1), we obtain a cycle $(i_1, \pi_{i_1}), (i_2, \pi_{i_2}), \dots, (i_k, \pi_{i_k})$, which we also denote C , in $\Gamma(\pi)$, and call it a k -*cycle*. We say that such a cycle is *positively oriented* if $k \geq 3$ and its elements can be written as a strictly increasing sequence, *negatively oriented* if $k \geq 3$ and its elements can be written as a strictly decreasing sequence, and *unoriented* otherwise.

For instance, in Fig. 1, cycle (4, 2, 1) is negatively oriented, cycle (5) is unoriented, and cycle (3, 6, 7) is positively oriented. Note that every 1-cycle

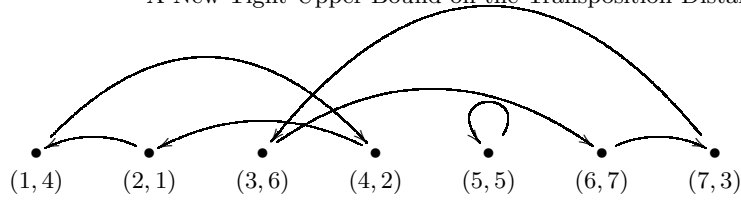


Fig. 1. The Γ -graph of the permutation (4 1 6 2 5 7 3)

and every 2-cycle is unoriented. In a quite similar fashion to the parity of cycles defined in the context of $G(\pi)$, we say that a k -cycle of $\Gamma(\pi)$ is *odd* (resp. *even*) if k is odd (resp. even). Likewise, we note $c(\Gamma(\pi))$ the number of cycles of $\Gamma(\pi)$, and $c_{odd}(\Gamma(\pi))$ (resp. $c_{even}(\Gamma(\pi))$) the number of odd (resp. even) cycles of $\Gamma(\pi)$.

4 An Explicit Formula for Some Permutations

We define a γ -permutation as a reduced permutation that fixes even elements (thus n must be odd), and show (Theorem 6) that the distance of such a permutation, and several others, can be computed quickly, without the need of the cycle graph.

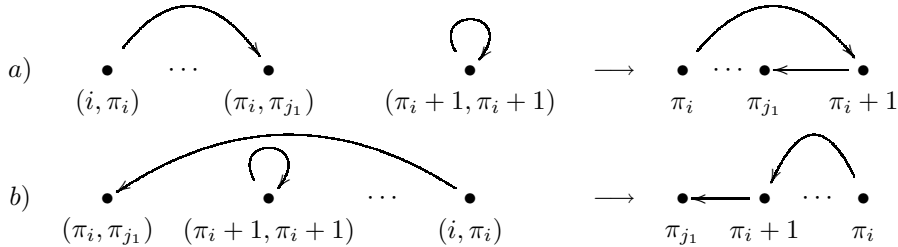
Proposition 1. For every γ -permutation π in S_n :

$$\begin{cases} c_{even}(G(\pi)) = 2 c_{even}(\Gamma(\pi)) ; \\ c_{odd}(G(\pi)) = 2 (c_{odd}(\Gamma(\pi)) - \frac{n-1}{2}) . \end{cases}$$

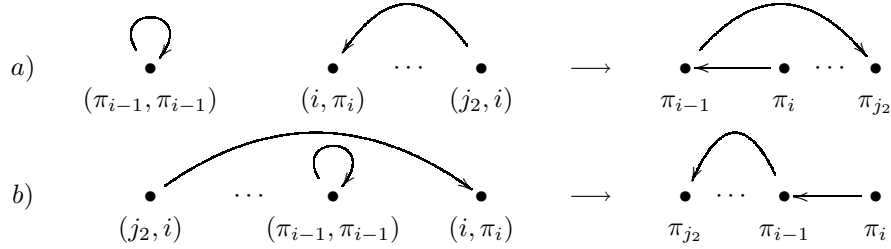
Proof. Each vertex (i, π_i) of $\Gamma(\pi)$ corresponding to an odd element π_i is both the starting point of an edge $((i, \pi_i), (\pi_i, \pi_{j_1}))$ and the ending point of an edge $((j_2, i), (i, \pi_i))$. Since π_i is odd and π is a γ -permutation, $\pi_i + 1$ is mapped onto itself, and π_{j_1} precedes $\pi_i + 1$ in π . In $G(\pi)$, those edges are each transformed into one sequence of two edges (gray-black for the first one, black-gray for the second one):

$$\begin{cases} ((i, \pi_i), (\pi_i, \pi_{j_1})) \text{ becomes } (\pi_i, \pi_i + 1), (\pi_i + 1, \pi_{j_1}) ; \\ ((j_2, i), (i, \pi_i)) \text{ becomes } (\pi_i, \pi_{i-1}), (\pi_{i-1}, \pi_{j_2}) . \end{cases}$$

I.e. the outgoing edge of (i, π_i) in $\Gamma(\pi)$ is transformed in one of the following ways (according to the relative positions of π_i and π_{j_1}):



Similarly, the incoming edge of (i, π_i) in $\Gamma(\pi)$ is transformed in one of the following ways (according to the relative positions of π_i and π_{j_2}):



Each k -cycle ($k \geq 2$) of $\Gamma(\pi)$ provides an alternate cycle with k black edges in $G(\pi)$; moreover, for every such cycle of $\Gamma(\pi)$, a new cycle is created in $G(\pi)$, which actually corresponds to the cycle of $\Gamma(\pi)$ followed in the opposite direction. Parity of cycles of $\Gamma(\pi)$ is obviously preserved in $G(\pi)$, since to each vertex of a k -cycle ($k \geq 2$) of $\Gamma(\pi)$ corresponds a black edge in $G(\pi)$. Finally, 1-cycles of $\Gamma(\pi)$ are not preserved in $G(\pi)$, and there are $\frac{n-1}{2}$ of them. \square

We derive the following lower bound from Proposition 1 and Theorem 1.

Lemma 1. *For every γ -permutation π in S_n , we have $d(\pi) \geq n - c_{\text{odd}}(\Gamma(\pi))$.*

Proof. Straightforward. \square

We will first study permutations such that odd elements form only one cycle in $\Gamma(\pi)$, distinguishing the case of oriented cycles and that of unoriented ones.

4.1 Oriented Cycles

We define an α -permutation as a reduced permutation that fixes even elements and whose odd elements form one oriented cycle in the graph Γ , and refer to the long cycle formed by the $\frac{n+1}{2}$ odd elements as its *main cycle*.

Proposition 2. *For every α -permutation π in S_n , we have $d(\pi) = \frac{n+1}{2} - (\frac{n+1}{2} \bmod 2)$.*

Proof. Since every α -permutation is a γ -permutation, Lemma 1 yields $d(\pi) \geq \frac{n+1}{2} - (\frac{n+1}{2} \bmod 2)$. We assume that the main cycle of $\Gamma(\pi)$ is positively oriented (a similar proof is easily obtained in the negative case).

1. if $\frac{n+1}{2}$ is odd, consider transpositions $\tau_1(2, 4, n + 1)$, $\tau_2(1, 3, n)$; then an optimal sequence of $\frac{n+1}{2} - 1$ transpositions that sorts π is

$$(\tau_1 \circ \tau_2)^{\frac{n+1}{2} - 1} .$$

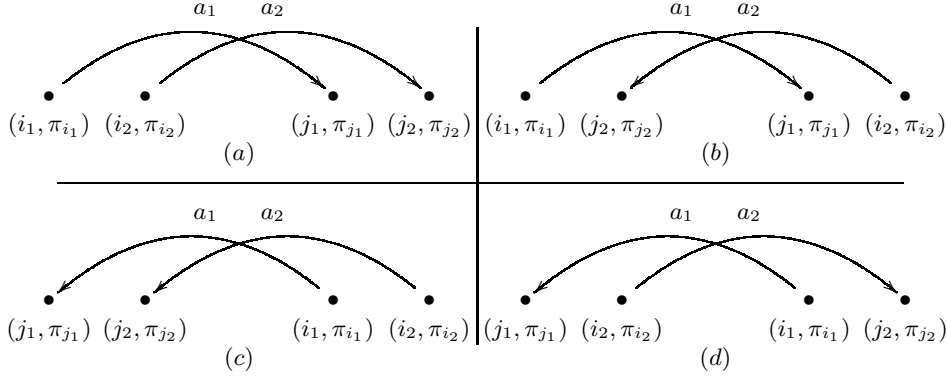


Fig. 2. The four possible configurations for crossing edges in $\Gamma(\pi)$

- if $\frac{n+1}{2}$ is even, consider again transpositions τ_1, τ_2 defined above, and also transpositions $\tau_3(2, 3, n + 1), \tau_4(1, 2, n + 1)$; then an optimal sequence of $\frac{n+1}{2}$ transpositions that sorts π is

$$(\tau_1 \circ \tau_2)^{\frac{n+1}{2}-2} \circ \tau_3 \circ \tau_4 .$$

Short of space, we omit the proof that those sequences indeed sort our permutations, but this can be easily shown by induction. \square

4.2 Unoriented Cycles

We now show that the orientation of a cycle does not matter, i.e. Proposition 2 still holds if the main cycle of $\Gamma(\pi)$ is unoriented. We will make use of so-called *exchanges* to simplify the proofs, namely bypassing the construction of optimal sequences. An *exchange* $exc(i, j)$ is the permutation that exchanges elements in positions i and j , thus transforming every permutation π into the permutation $\pi \circ exc(i, j)$.

$$exc(i, j) = \begin{pmatrix} 1 \cdots i - 1 \boxed{i} i + 1 \cdots j - 1 \boxed{j} j + 1 \cdots n \\ 1 \cdots i - 1 \boxed{j} i + 1 \cdots j - 1 \boxed{i} j + 1 \cdots n \end{pmatrix} .$$

We will only use exchanges of the form $exc(i, i + 2k)$ with $k \geq 1$; such an exchange can be simulated by two transpositions, but the correspondence between those two types of operations is not that straightforward when exchanges are composed.

We say that two edges $a_1 = ((i_1, \pi_{i_1}), (j_1, \pi_{j_1}))$ and $a_2 = ((i_2, \pi_{i_2}), (j_2, \pi_{j_2}))$ of $\Gamma(\pi)$ *cross* if intervals $[i_1, j_1]$ and $[i_2, j_2]$ do not contain each other and have a non-empty intersection. The four possible configurations for crossing edges are shown in Fig. 2.

We define a β -permutation as a reduced permutation that fixes even elements and whose odd elements form one unoriented cycle in the graph Γ .

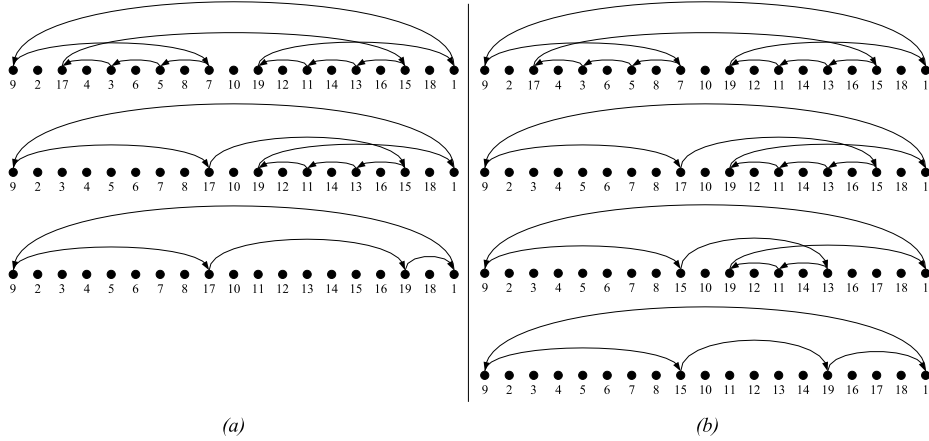


Fig. 3. Two possible ways of contracting paths in a β -permutation (1-cycles and indices omitted for clarity)

Clearly, the main cycle of the Γ -graph of every β -permutation π (except $(3\ 2\ 1)$) contains crossing edges. We are going to transform π into a permutation σ that reduces to an α -permutation, and this will be achieved through the removal of crossing edges using a certain sequence \mathcal{E} of exchanges. We thus get the following upper bound on the distance of a β -permutation π :

$$d(\pi) \leq f(\mathcal{E}) + d(\sigma)$$

where $f(\mathcal{E})$ gives the minimum number of transpositions having the same effect on π as \mathcal{E} does. Finding σ is not difficult, but we have to find a σ such that our upper bound is minimized.

To eliminate a crossing, we just have to make the ending point of one edge become the starting point of the one it crosses, and this will be achieved using a sequence of exchanges of the form described in the following proposition. By a *path*, we mean the sequence of edges joining the extremities of the crossing edges as mentioned above, and we will refer to the elimination of this path as the *contraction* of it. The following proposition will be useful.

Proposition 3. For both sequences $\mathcal{E} = exc(i, i+2) \circ exc(i, i+4) \circ \dots \circ exc(i, i+2t)$ and $\mathcal{F} = exc(i, i+2t) \circ \dots \circ exc(i, i+4) \circ exc(i, i+2)$ of t exchanges:

$$f(\mathcal{E}) = f(\mathcal{F}) = t + (t \bmod 2) .$$

Proof. For any valid i, t , \mathcal{E} and \mathcal{F} reduce to an α -permutation π whose main cycle is a $(t+1)$ -cycle, and by Proposition 2:

$$d(\pi) = t + 1 - ((t + 1) \bmod 2) = t + (t \bmod 2) . \quad \square$$

Proposition 4. *For every β -permutation π in S_n , we have $d(\pi) = \frac{n+1}{2} - \left(\frac{n+1}{2} \bmod 2\right)$.*

Proof. Since every β -permutation is a γ -permutation, Lemma 1 yields $d(\pi) \geq \frac{n+1}{2} - \left(\frac{n+1}{2} \bmod 2\right)$. If $\pi = (3\ 2\ 1)$, the thesis is easily verified; else the main cycle of $\Gamma(\pi)$ contains at least one crossing.

If the main cycle of $\Gamma(\pi)$ contains only one crossing, then there is a path of t edges joining the two crossing edges; this path can be contracted by a sequence of t exchanges, sorting the elements belonging to that part of the cycle. For instance, in case (a) of Fig. 2, it suffices to apply the sequence $exc(i_2, j_1) \circ \dots \circ exc(i_2, i_2 + 4) \circ exc(i_2, i_2 + 2)$, and those t exchanges correspond to exactly $t + (t \bmod 2)$ transpositions (Proposition 3).

Once this path has been contracted, t vertices have been removed from the main cycle of $\Gamma(\pi)$ and this results in a permutation σ reducible to an α -permutation. We have:

$$\begin{aligned} d(\pi) &\leq d(\pi, \sigma) + d(\sigma) = t + (t \bmod 2) + \frac{n+1}{2} - t - \left(\left(\frac{n+1}{2} - t\right) \bmod 2\right) \\ &= \frac{n+1}{2} - \left(\frac{n+1}{2} \bmod 2\right) \end{aligned}$$

which verifies our thesis.

In the case where several crossings exist, one must be careful not to contract paths “individually”. Indeed, if we were to contract p such paths of t_g edges ($1 \leq g \leq p$) in that way, we would have to use $\sum_{g=1}^p t_g$ exchanges to contract them all, which would correspond to $\sum_{g=1}^p (t_g + (t_g \bmod 2))$ transpositions. This can actually be improved by exchanging the last exchanged element in the first contracted path with the first element of the next path to contract, then continue the contraction of the latter with dependent exchanges as before, and repeating the same process whenever need be. For instance, Fig. 3 shows the transformation of a β -permutation into a permutation reducible to an α -permutation in two different ways. Scenario (a) uses 3+3 exchanges = 8 transpositions (Proposition 3), whereas scenario (b) uses the same number of exchanges, but requiring this time only 6 transpositions.

Every β -permutation π containing p paths of t_g edges to contract ($1 \leq g \leq p$) can thus be transformed into a permutation σ reducible to an α -permutation such that $d(\pi, \sigma) = T + T \bmod 2$, where $T = \sum_{g=1}^p t_g$. The transpositions representing those exchanges will eliminate T vertices from the main cycle of $\Gamma(\pi)$, which yields the following upper bound:

$$\begin{aligned} d(\pi) &\leq d(\pi, \sigma) + d(\sigma) \\ &= T + (T \bmod 2) + \left(\frac{n+1}{2} - T - \left(\left(\frac{n+1}{2} - T\right) \bmod 2\right)\right) \\ &= \frac{n+1}{2} - \left(\frac{n+1}{2} \bmod 2\right) \end{aligned}$$

which equals the lower bound given above. □

4.3 Distance of γ -Permutations

Every permutation π can be sorted by eliminating each cycle of $\Gamma(\pi)$ individually using exchanges (and therefore also using only transpositions), so that each elimination does not modify other cycles. This strategy yields the following upper bound on $d(\pi)$.

Lemma 2. *For every permutation π , consider its disjoint cycle decomposition $\Gamma(\pi) = C_1 \cup C_2 \cup \dots \cup C_{c(\Gamma(\pi))}$. Denote $d(C)$ the minimum number of transpositions required to transform $C = (i_1, i_2, \dots, i_k)$ into $(i_1), (i_2), \dots, (i_k)$; then:*

$$d(\pi) \leq \sum_{i=1}^{c(\Gamma(\pi))} d(C_i) .$$

We now show that the lower bound of Lemma 1 is reached.

Proposition 5. *For every γ -permutation π in S_n :*

$$d(\pi) = n - c_{\text{odd}}(\Gamma(\pi)) . \tag{4}$$

Proof. Each cycle of $\Gamma(\pi)$ is either oriented or unoriented, and the distance of both kinds of cycles is known (Propositions 2 and 4). Denote $\text{odd}(\Gamma(\pi))$ (resp. $\text{even}(\Gamma(\pi))$) the set of odd (resp. even) cycles of $\Gamma(\pi)$; by Lemma 2, we have:

$$\begin{aligned} d(\pi) &\leq \sum_{i=1}^{c(\Gamma(\pi))} |C_i| - (|C_i| \bmod 2) \\ &= \sum_{C_{i_1} \in \text{odd}(\Gamma(\pi))} (|C_{i_1}| - 1) + \sum_{C_{i_2} \in \text{even}(\Gamma(\pi))} |C_{i_2}| \\ &= \sum_{i=1}^{c(\Gamma(\pi))} |C_i| - c_{\text{odd}}(\Gamma(\pi)) . \end{aligned}$$

And since every element belongs to exactly one cycle, the last sum equals n and Lemma 1 verifies the thesis. \square

Note that Proposition 5 can be expressed in a more general way: by Theorem 2, we know that π needs not be reduced, and adding k 1-cycles to $\Gamma(\pi)$ at any position increases both n and $c_{\text{odd}}(\Gamma(\pi))$ by k , so Equation 4 still holds. The same Theorem allows us to ask for *odd* elements to be fixed, instead of *even* ones; we then have $\pi_1 = 1$, and we can reduce π to a γ -permutation (e.g. $(1\ 4\ 3\ 6\ 5\ 8\ 7\ 2) \equiv_r (3\ 2\ 5\ 4\ 7\ 6\ 1)$). This result can also be extended using *toric permutations*, introduced in [13] and further studied in [14]. We borrow the latter author's notations; let us note, for $x \in \{0, 1, 2, \dots, n\}$:

1. $\overline{x}^m = (x + m) \pmod{n + 1}$;
2. $\underline{x}_m = (x - m) \pmod{n + 1}$.

Define a *circular permutation* obtained from a permutation π in S_n as $\pi^\circ = 0 \pi_1 \pi_2 \cdots \pi_n$, where $0 = \pi_0^\circ = \pi_{n+1}^\circ$. This circular permutation can be read starting from any position, and the original linear permutation is reconstructed by taking the element following 0 as π_1 and removing 0. Define the following operation on circular permutations:

$$m + \pi^\circ = \overline{0}^m \overline{\pi_1}^m \overline{\pi_2}^m \cdots \overline{\pi_n}^m .$$

Then for π in S_n , the corresponding *toric permutation* is π° , which is the set of permutations obtained from $m + \pi^\circ$ with $0 \leq m \leq n$.

Lemma 3. [13] $\forall \pi, \sigma \in S_n : \sigma \in \pi^\circ \Rightarrow d(\sigma) = d(\pi) .$

Therefore, if n is odd and all odd elements of π in S_n occupy odd positions and form an increasing subsequence modulo $n + 1$, then $\pi \in \sigma^\circ$ where σ satisfies the conditions described right after Proposition 5. Indeed, if n is odd, adding 1 (mod $n + 1$) to each element of π° transforms n into 0, i.e. the new starting point of the resulting permutation, and all odd elements into even ones; therefore π is transformed into a permutation whose even elements are all fixed, and we have the following.

Theorem 6. *For every π in S_n that fixes even or odd elements:*

$$d(\pi) = n - c_{\text{odd}}(\Gamma(\pi)) .$$

Moreover, every permutation σ with n odd and whose odd elements occupy odd positions and form an increasing subsequence modulo $n + 1$ can be transformed in linear time into a permutation π such that $d(\sigma) = d(\pi) = n - c_{\text{odd}}(\Gamma(\pi)) .$

5 A New Upper Bound

We now show that the right-hand side of Equation 5 is an upper bound on the transposition distance. First we show why γ -permutations are so important.

Theorem 7. *Every permutation π in S_n , except ι , can be obtained from a γ -permutation σ in S_{n+k} by removing k even elements in σ .*

Proof. If $\pi \neq \iota$ is no γ -permutation, just add a 1-cycle to $\Gamma(\pi)$ between every ordered pair (π_i, π_{i+1}) ($1 \leq i \leq n - 1$) of non-fixed elements and reduce the resulting permutation π' in order to obtain a γ -permutation $\sigma \in S_{n+k}$. This operation can clearly be reverted, and this completes the proof. \square

We can now prove our main result.

Theorem 8. $\forall \pi \in S_n :$

$$d(\pi) \leq n - c_{\text{odd}}(\Gamma(\pi)) . \tag{5}$$

Proof. Let σ be the γ -permutation from which $\pi \neq \iota$ is obtained by deletion of k even elements. Cycles in $\Gamma(\sigma)$ can all be sorted individually by sequences of exchanges for which we know the corresponding minimal number of transpositions (Proposition 5), and all these exchanges still work on $\Gamma(\pi)$ (after accordingly adapting some of them). Therefore, we can claim that $d(\pi) \leq d(\sigma)$, and:

$$d(\pi) \leq d(\sigma) = n + k - c_{\text{odd}}(\Gamma(\sigma)) = n + k - c_{\text{odd}}(\Gamma(\pi)) - k = n - c_{\text{odd}}(\Gamma(\pi)).$$

And even though ι cannot be obtained from a γ -permutation, it is clear that our thesis holds for it too, since $d(\iota) = 0 \leq n - n$. \square

6 Conclusions and Future Plans

Using a well-known graph, we were able to show that the transposition distance of some nicely characterized permutations can be computed in linear time, bypassing the classical structure introduced in [3]. In fact, no graph at all is needed, since decomposing a permutation into “classical cycles” is quite a trivial algorithm, running in linear time. Such an approach has proved most successful for computing another rearrangement distance in [11], and we are confident it is of great interest, certainly not just for the transposition distance. Furthermore, we also proved that the formula used to compute this distance is actually an upper bound on the transposition distance of every permutation.

Several questions arise.

Firstly, this new upper bound in Equation (5) is sometimes better, sometimes worse than the bounds in Equations (1), (2) and (3), and we want to tighten it. Table 1 compares our result with previous ones, giving the number of cases where it is at least as good as that of Theorems 3, 4, and 5. Apart from a fast approximation of the transposition distance, this could also help determine the maximal value of $d(\pi)$, which is still an open problem, as is the complexity of sorting by transpositions.

Secondly, the permutations characterized in Theorem 6 are not the only ones to reach our upper bound. Can the set of all such permutations be characterized?

Finally, there might be other useful permutation-related notions in combinatorics that are as well-known and eluded in the theory of genome rearrangement as is the graph we used. Although we do not think that those classical notions

Table 1. Comparison of our new upper bound with previous results

n	Number of permutations	$(5) \leq (1)$	$(5) \leq (2)$	$(5) \leq (3)$
3	6	2	1	6
4	24	8	8	15
5	120	45	24	31
6	720	304	49	495
7	5040	2055	722	1611
8	40320	17879	3094	4355
9	362880	104392	60871	10243

can model each and every notion of this problem (in particular, it has been shown [14] that the structure of $G(\pi)$ is much more stable than that of $\Gamma(\pi)$ under the toric equivalence class), we feel that some of them could be of interest.

References

1. Meidanis, J., Setubal, J.: Introduction to Computational Molecular Biology. Brooks-Cole (1997)
2. Pevzner, P.A.: Computational molecular biology. MIT Press, Cambridge, MA (2000)
3. Bafna, V., Pevzner, P.A.: Sorting by transpositions. *SIAM J. Discrete Math.* **11** (1998) 224–240 (electronic)
4. Elias, I., Hartman, T.: A 1.375–Approximation Algorithm for Sorting by Transpositions (2005) Submitted.
5. Christie, D.A.: Genome Rearrangement Problems. PhD thesis, University of Glasgow, Scotland (1998)
6. Hartman, T.: A simpler 1.5-approximation algorithm for sorting by transpositions. In: Combinatorial pattern matching. Volume 2676 of Lecture Notes in Computer Science. Springer, Berlin (2003) 156–169
7. Guyer, S.A., Heath, L.S., Vergara, J.P.: Subsequence and run heuristics for sorting by transpositions. In: Fourth DIMACS Algorithm Implementation Challenge, Rutgers University (1995)
8. Vergara, J.P.C.: Sorting by Bounded Permutations. PhD thesis, Virginia Polytechnic Institute, Blacksburg, Virginia, USA (1997)
9. Walter, M.E.M.T., Curado, L.R.A.F., Oliveira, A.G.: Working on the problem of sorting by transpositions on genome rearrangements. In: Combinatorial pattern matching. Volume 2676 of Lecture Notes in Computer Science. Springer, Berlin (2003) 372–383
10. Dias, Z., Meidanis, J.: An Alternative Algebraic Formalism for Genome Rearrangements. *Comparative Genomics* (2000) 213–223
11. Dias, Z., Meidanis, J.: Genome Rearrangements Distance by Fusion, Fission, and Transposition is Easy. In: Proceedings of SPIRE’2001 - String Processing and Information Retrieval, Laguna de San Rafael, Chile (2001) 250–253
12. Dias, Z., Meidanis, J., Walter, M.E.M.T.: A New Approach for Approximating The Transposition Distance. In: Proceedings of SPIRE’2000 - String Processing and Information Retrieval, La Coruna, Espagne (2000)
13. Eriksson, H., Eriksson, K., Karlander, J., Svensson, L., Wästlund, J.: Sorting a bridge hand. *Discrete Mathematics* **241** (2001) 289–300 Selected papers in honor of Helge Tverberg.
14. Hultman, A.: Toric Permutations. Master’s thesis, Dept. of Mathematics, KTH, Stockholm, Sweden (1999)