Edit Distances and Factorisations of Even Permutations

Anthony Labarre⋆

Université libre de Bruxelles (U.L.B.),
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. A number of fields, including genome rearrangements and interconnection network design, are concerned with sorting permutations in "as few moves as possible", using a given set of allowed operations. These often act on just one or two segments of the permutation, e.g. by reversing one segment or exchanging two segments. The *cycle graph* of the permutation to sort is a fundamental tool in the theory of genome rearrangements. In this paper, we present an algebraic reinterpretation of the cycle graph as an even permutation, and show how to reformulate our sorting problems in terms of particular factorisations of the latter permutation. Using our framework, we recover known results in a simple and unified way, and obtain a new lower bound on the *prefix transposition distance* (where a *prefix transposition* displaces the initial segment of a permutation), which is shown to outperform previous results. Moreover, we use our approach to improve the best known lower bound on the *prefix transposition diameter* from $2n/3$ to $\lfloor \frac{3n+1}{4} \rfloor$.

1 Introduction

We study the problem of computing *edit distances* between permutations, i.e. the minimum number of operations needed to transform a permutation into another, using a given set of allowed operations. Those operations satisfy the property that the induced edit distance between any two permutations π and σ of the same set equals the distance between $\sigma^{-1} \circ \pi$ and the *identity permutation* $\iota = \langle 1\ 2\ \cdots\ n \rangle$, thereby allowing us to restrict our attention to *sorting permutations* using a minimum number of allowed operations. Two areas in which these problems have applications are the fields of *genome rearrangements* and *interconnection network design*, which we briefly review below.

In genome rearrangements (recently surveyed in [1]), the permutation to sort represents an ordering of genes in a given genome, and the allowed operations model *mutations* that are known to actually occur in evolution. Rearrangements

⋆ Funded by the "Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture" (F.R.I.A.). Research supported by "Communauté française de Belgique - Actions de Recherche Concertées".

D. Halperin and K. Mehlhorn (Eds.): ESA 2008, LNCS 5193, pp. 635–646, 2008.

studied in that context include *reversals* [2], which reverse a segment of the permutation, *block-interchanges* [3], which exchange two not necessarily contiguous segments, and *transpositions* [4], which displace a block of contiguous elements. While the complexity of the sorting and distance computation problems is known for the first two operations (NP-hard for reversals [5] and polynomial for block-interchanges [3]), it is open for transpositions, and the best polynomial time approximation algorithm to date has ratio 11/8 [6].

In interconnection network design (see [7] for a thorough survey), permutations stand e.g. for processors and form the vertex set of a graph whose edges correspond to physical connections between two devices. One wants to build a graph with small degree and small *diameter*, among other desirable properties, and this is often done by choosing a set of allowed operations on permutations, then connecting two permutations if there is an allowed operation that transforms one into the other [8]. In that setting, sorting algorithms for permutations correspond to *routing algorithms* for the corresponding networks. Two kinds of operations that received much attention in that context are *prefix reversals* [9], which reverse the initial segment of the permutation, and *prefix exchanges* [10], which swap the first element of the permutation with another element. Those operations gave birth to the *pancake network* and *star graph* topologies, respectively, which are extensively studied models in that field. We also mention *prefix transpositions* [11], which displace the initial segment of the permutation; they bear little relevance with biological problems, but they are hoped to shed light and give insight on the seemingly challenging problem of sorting by transpositions.

The *cycle graph* is a ubiquitous structure in the field of genome rearrangements. In this paper, we present a way of encoding the cycle graph as an even permutation, inspired by a previous work of ours [12], and show how to reformulate *any* sorting problem of the form described above in terms of particular factorisations of the latter permutation. We first illustrate the power of our framework by recovering known lower bounds on the block-interchange and transposition distances, and then use it to prove a new lower bound on the prefix transposition distance. We prove that our lower bound always outperforms the one proved in [11], and show experimentally that it is a significant improvement over both that result and the only other known lower bound [13]. Finally, we use this new result to improve the previously best known lower bound on the maximal value of the prefix transposition distance from $2n/3$ to $\left\lfloor \frac{3n+1}{4} \right\rfloor$.

2 Notation and Definitions

2.1 Basic Permutation Group Theory

Let us start with a quick reminder of basic notions on permutations (for details, see e.g. [14]). The *symmetric group* S_n is the set of all permutations of $\{1, 2, \ldots, n\}$, together with the usual function composition $\circ$, applied from right to left. Permutations are denoted by lower case Greek letters, typically $\pi = \langle \pi_1 \ \pi_2 \ \cdots \ \pi_n \rangle$, with $\pi_i = \pi(i)$. The usual *graph* $\Gamma(\pi)$ of the permutation π contains an arc (i, j) whenever $\pi_i = j$, and decomposes in a single way into disjoint cycles, leading to another notation for π based on its *disjoint cycle decomposition*. For instance, when $\pi = \langle 4\ 1\ 6\ 2\ 5\ 7\ 3 \rangle$, the disjoint cycle notation is $\pi = (1,4,2)(3,6,7)(5)$ (notice the parentheses and the commas). As in [15], we order the vertices of $\Gamma(\pi)$ by positions. The number of cycles in $\Gamma(\pi)$ is denoted by $c(\Gamma(\pi))$, and the *length* of a cycle is the number of elements it contains. It is common to drop the 1-cycles from that representation, and to call the permutation a *k-cycle* if the resulting decomposition consists of a single cycle of length $k > 1$. A permutation π is *even* if the number of even cycles in $\Gamma(\pi)$ is even or, equivalently, if it can be expressed as a product of an even number of 2-cycles. The *alternating group* A_n is the set of all even permutations in S_n. Finally, the *conjugate* of a permutation π by a permutation σ, both in S_n, is the permutation $\pi^\sigma = \sigma \circ \pi \circ \sigma^{-1}$. It has the same disjoint cycle decomposition as π, and can be obtained, if $\pi = (c_{1,1}, c_{1,2}, \ldots, c_{1,\ell_1}) \cdots (c_{m,1}, c_{m,2}, \ldots, c_{m,\ell_m})$, by replacing each element in each cycle of π with the element onto which it is mapped by σ, i.e. $\pi^\sigma = (\sigma_{c_{1,1}}, \sigma_{c_{1,2}}, \ldots, \sigma_{c_{1,\ell_1}}) \cdots (\sigma_{c_{m,1}}, \sigma_{c_{m,2}}, \ldots, \sigma_{c_{m,\ell_m}})$. All permutations that have the same disjoint cycle decomposition form a *conjugacy class* (of S_n).

2.2 Genome Rearrangements and Prefix Operations

We recall a number of operations on permutations, starting with the most general one, introduced in [3]. For any π in S_n, the *block-interchange* $\beta(i,j,k,l)$ with $1 \leq i < j \leq k < l \leq n+1$ applied to π exchanges the closed intervals determined respectively by i and $j-1$ and by k and $l-1$. It transforms π into $\pi \circ \beta(i,j,k,l)$, where $\beta(i,j,k,l)$ is the following permutation:

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

Two particular cases of block-interchanges are of interest: when $j = k$, the resulting operation exchanges two adjacent intervals, and is called a *transposition*, denoted by $\tau(i,j,l)$; when $j = i+1$ and $l = k+1$, the resulting operation swaps two not necessarily adjacent elements in respective positions i and k, and is called an *exchange*, denoted by $\varepsilon(i,k)$. Two generic problems are studied in connection to these operations: the problem of finding a sequence of transformations that sorts a permutation π and is of the shortest possible length, and the related problem of merely computing the length of such a sequence, called the *distance* of π (with respect to the given operation). It is easily seen that the sorting problem on π is equivalent to factorising π into the product of permutations that are allowed transformations, provided that the inverse of an edit operation is still an allowed edit operation (which is easily shown to be the case for all operations considered in this paper). Indeed, any sorting sequence for π, i.e. $\pi \circ x_1 \circ x_2 \circ \cdots \circ x_t = \iota$, where x_i belongs to the set S of allowed operations for $1 \leq i \leq t$, immediately yields the factorisation $\pi = x_t^{-1} \circ x_{t-1}^{-1} \circ \cdots \circ x_1^{-1}$, and vice versa. We denote $bid(\pi)$, $td(\pi)$ and $exc(\pi)$ the block-interchange distance, transposition distance and exchange distance of π, respectively. Moreover, the *diameter* of S_n with respect to a given set of edit operations is the maximal value that the corresponding edit distance can reach.

The following traditional tool introduced by Bafna and Pevzner [4] has proved most useful in the study of genome rearrangements. The *cycle graph* of π in S_n is the bicoloured directed graph $G(\pi)$, whose vertex set $(\pi_0 = 0, \pi_1, \ldots, \pi_n, \pi_{n+1} = n+1)$ is ordered by positions, and whose arc set consists of:

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

The set of black and grey arcs decomposes in a single way into *alternating cycles*, i.e. cycles which alternate black and grey arcs, and we note the number of such cycles $c(G(\pi))$. The *length* of an alternating cycle in $G(\pi)$ is the number of black arcs it contains, and a k-*cycle* in $G(\pi)$ is an alternating cycle of length k. Fig. 1 shows an example of a cycle graph, together with its decomposition into a 4-cycle and a 2-cycle.

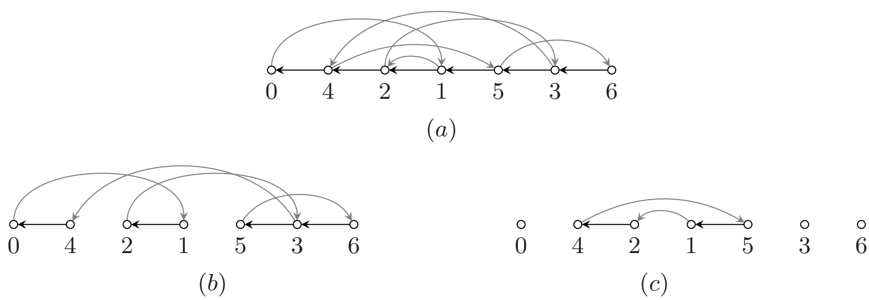

Fig. 1. (a) The cycle graph of $\langle 4\ 2\ 1\ 5\ 3 \rangle$, (b) and (c) the two cycles in its decomposition

Setting $i = 1$ in the rearrangement operations presented above turns them into "prefix rearrangements". The corresponding "prefix distances" are defined as before, and we denote $ptd(\pi)$ and $pexc(\pi)$ the *prefix transposition distance* and *prefix exchange distance* of π, respectively. While the computational complexity of sorting by transpositions or by prefix transpositions is unknown, a polynomial time algorithm for sorting by prefix exchanges is known [10], as well as a formula for computing the associated distance.

Theorem 1. *[10] For any π in S_n, we have*

$$pexc(\pi) = n + c(\Gamma(\pi)) - 2c_1(\Gamma(\pi)) - \begin{cases} 0 \text{ if } \pi_1 = 1, \\ 2 \text{ otherwise,} \end{cases}$$

where $c_1(\Gamma(\pi))$ denotes the number of 1-cycles in $\Gamma(\pi)$, or equivalently the number of fixed points of π.

Dias and Meidanis [11] initiated the study of sorting by prefix transpositions, and derived a lower bound on the corresponding distance using the following concepts. Given a permutation π in S_n, build the permutation $\widetilde{\pi} = \langle 0\ \pi_1\ \cdots\ \pi_n\ n+1\rangle$; a pair $(\widetilde{\pi}_i, \widetilde{\pi}_{i+1})$ with $0 \leq i \leq n$ is a *prefix transposition breakpoint* if $\widetilde{\pi}_{i+1} \neq \widetilde{\pi}_i + 1$ or if $i = 0$, and an *adjacency* otherwise. The number of prefix transposition breakpoints of π is denoted by $ptb(\pi)$. Noting that a prefix transposition can create at most two adjacencies and that ι is the only permutation with one prefix transposition breakpoint, they obtained the following lower bound.

Lemma 1. *[11] For any π in S_n:*

$$ptd(\pi) \geq \left\lceil \frac{ptb(\pi) - 1}{2} \right\rceil \ . \tag{1}$$

Finally, Chitturi and Sudborough [13] recently obtained new bounds on the prefix transposition distance. They used the following concepts, based on permutations of $\{0, 1, 2, \ldots, n-1\}$ rather than $\{1, 2, \ldots, n\}$. For a permutation π, an ordered pair (π_i, π_{i+1}) is an *anti-adjacency* if $\pi_{i+1} = \pi_i - 1 \pmod n$. A *strip* in a permutation π is a maximal interval of π that contains only adjacencies, and a *clan* is a maximal interval of π that contains only *anti*-adjacencies. They prove the following lower bound.

Lemma 2. *[13] For any π in S_n, let $\Upsilon(\pi)$ denote the set of all clans of π of length at least 3, and $s(\pi)$ denote the number of strips of π. Then*

$$ptd(\pi) \geq \frac{s(\pi) + \frac{\sum_{C \in \Upsilon(\pi)}(|C|-2)}{3}}{2} \ . \tag{2}$$

Using Lemma 2, Chitturi and Sudborough prove a lower bound of $2n/3$ on the prefix transposition distance of the reverse permutation $\langle n\ n-1\ \cdots\ 2\ 1\rangle$, and therefore on the prefix transposition diameter. They also prove an upper bound on the prefix transposition diameter.

Theorem 2. *[13] For all π in S_n, we have $ptd(\pi) \leq n - \log_8 n$.*

3 A General Lower Bounding Technique

In a previous paper [12], we introduced the following mapping:

$$f : S_n \rightarrow A_{n+1} : \pi \mapsto \overline{\pi} = (0, \pi_n, \pi_{n-1}, \ldots, \pi_1) \circ (0, 1, 2, \ldots, n) \ , \tag{3}$$

which in particular maps ι onto $\overline{\iota} = \langle 0\ 1\ 2\ \cdots\ n\rangle$. That mapping allowed us to encode a cycle graph $G(\pi)$ using an even permutation $\overline{\pi}$, as illustrated by the following example: let $\pi = \langle 4\ 2\ 1\ 5\ 3\rangle$, whose cycle graph is depicted in Fig. 1. Then

$$\overline{\pi} = (0, 3, 5, 1, 2, 4) \circ (0, 1, 2, 3, 4, 5) = (0, 2, 5, 3)(1, 4) \ ,$$

and the two disjoint cycles of $\overline{\pi}$ correspond to the two alternating cycles of $G(\pi)$, whose elements they list in the order they are encountered; indeed:

1. the first cycle of $G(\pi)$ (Fig. 1(b)) starts with 0, then visits 2 after following a grey-black path (i.e. a grey arc followed by a black arc), then visits 5 after following a grey-black path, and in the same way visits 3 after following a grey-black path before finally going back to 0, which corresponds to the first cycle of $\overline{\pi}$;
2. the second cycle of $G(\pi)$ (Fig. 1(c)) starts with 4, then visits 1 after following a grey-black path, which corresponds to the second cycle of $\overline{\pi}$.

Consequently, speaking about cycles of $\overline{\pi}$, of $\Gamma(\overline{\pi})$ or of $G(\pi)$ is equivalent. We will now demonstrate how f can be used to obtain bounds on sorting problems. The following result expresses how the action of *any* rearrangement operation σ on π is translated on $\overline{\pi}$. In the following, we identify permutations in S_n with their extended versions in S_{n+1} (i.e. we identify π with $\langle 0\ \pi_1\ \pi_2\ \cdots\ \pi_n\rangle$).

Lemma 3. *For all π, σ in S_n, we have $\overline{\pi \circ \sigma} = \overline{\sigma}^{\pi} \circ \overline{\pi}$.*

Proof. The following relation will be useful:

$$\pi = (0, \pi_n, \pi_{n-1}, \ldots, \pi_1) \circ \pi \circ (0, 1, \ldots, n) \ . \tag{4}$$

By definition, we have:

$$\begin{aligned}
\overline{\pi \circ \sigma} &= (0, (\pi \circ \sigma)_n, (\pi \circ \sigma)_{n-1}, \ldots, (\pi \circ \sigma)_1) \circ (0, 1, \ldots, n) \\
&= (0, \pi_{\sigma_n}, \pi_{\sigma_{n-1}}, \ldots, \pi_{\sigma_1}) \circ (0, 1, \ldots, n) \\
&= \pi \circ (0, \sigma_n, \sigma_{n-1}, \ldots, \sigma_1) \circ \pi^{-1} \circ (0, 1, \ldots, n) \\
&= \pi \circ (0, \sigma_n, \sigma_{n-1}, \ldots, \sigma_1) \circ (0, 1, \ldots, n) \circ (0, 1, \ldots, n)^{-1} \circ \pi^{-1} \\
&\quad \circ (0, 1, \ldots, n) \\
&= \pi \circ \overline{\sigma} \circ (\pi \circ (0, 1, \ldots, n))^{-1} \circ (0, 1, \ldots, n) \\
&= \pi \circ \overline{\sigma} \circ \left((0, \pi_n, \ldots, \pi_1)^{-1} \circ \pi\right)^{-1} \circ (0, 1, \ldots, n) \qquad \text{(using (4))} \\
&= \pi \circ \overline{\sigma} \circ \pi^{-1} \circ (0, \pi_n, \ldots, \pi_1) \circ (0, 1, \ldots, n) \\
&= \pi \circ \overline{\sigma} \circ \pi^{-1} \circ \overline{\pi} \ .
\end{aligned}$$

□

We are now ready to prove our main result.

Theorem 3. *Let X be a subset of S_n whose elements are mapped by f onto $X' \subseteq A_{n+1}$. Moreover, let $\mathscr{C}$ be the union of the conjugacy classes (of S_{n+1}) that intersect with X'; then for any π in S_n, any factorisation of π into t elements of X yields a factorisation of $\overline{\pi}$ into t elements of $\mathscr{C}$.*

Proof. Induction on t. The base case is $\pi \in X$, and clearly $\overline{\pi} \in X' \subseteq \mathscr{C}$. For the induction, let $\pi = g_t \circ g_{t-1} \circ \cdots \circ g_1$, where $g_i \in X$ for $1 \leq i \leq t$, and let $\sigma = g_{t-1} \circ \cdots \circ g_2 \circ g_1$; by Lemma 3, we have:

$$\overline{\pi} = \overline{g_t \circ g_{t-1} \circ \cdots \circ g_2 \circ g_1} = \overline{g_t \circ \sigma} = g_t \circ \overline{\sigma} \circ g_t^{-1} \circ \overline{g_t} \ .$$

By induction, $\overline{\sigma} = g'_{t-1} \circ g'_{t-2} \circ \cdots \circ g'_1$, where $g'_i \in \mathscr{C}$ for $1 \leq i \leq t$; therefore:

$$\begin{aligned} g_t \circ \overline{\sigma} \circ g_t^{-1} &= g_t \circ g'_{t-1} \circ g'_{t-2} \circ \cdots \circ g'_1 \circ g_t^{-1} \\ &= \underbrace{g_t \circ g'_{t-1} \circ g_t^{-1}}_{h_t} \circ \underbrace{g_t \circ g'_{t-2} \circ g_t^{-1}}_{h_{t-1}} \circ g_t \circ \cdots \circ g_t^{-1} \circ \underbrace{g_t \circ g'_1 \circ g_t^{-1}}_{h_1} , \end{aligned}$$

and $h_1, \ldots, h_{t-1} \in \mathscr{C}$, which completes the proof. □

As we briefly explain before applying our method in the next section, Theorem 3 allows us to prove lower bounds on our sorting problems: indeed, as we explained in Section 2.2, any sorting sequence of length t for π made of elements of X yields a factorisation of π into the product of t elements (of X, provided X contains both the transformations and their inverses, which is easily shown to be the case for all operations considered in this paper), which can in turn be converted, as in the proof of Theorem 3, into a factorisation of $\overline{\pi}$ into the product of t elements of $\mathscr{C}$. Therefore, the length of a shortest such factorisation of $\overline{\pi}$ into the product of elements of $\mathscr{C}$ is a lower bound on the length of a factorisation of π into the product of elements of X.

4 Recovering Previous Results

We illustrate how to use Theorem 3 to recover two previously known results on *bid* and *td*. First, we need to characterise the image of a block-interchange by our mapping.

Lemma 4. *For any block-interchange $\beta(i, j, k, l)$, we have*

$$\overline{\beta(i, j, k, l)} = (j - 1, l - 1) \circ (i - 1, k - 1) \ .$$

Proof. Using (3) and the definition of a block-interchange, we have

$$\begin{aligned} &(0, n, n-1, \ldots, l, j-1, j-2, \ldots, i, k-1, k-2, \ldots, j, l-1, l-2, \ldots, \\ &k, i-1, i-2, \ldots, 1) \circ (0, 1, 2, \ldots, n) \\ = &(0)(1) \cdots (i-2)(i-1, k-1)(i)(i+1) \cdots (j-2)(j-1, l-1)(j) \\ &(j+1) \cdots (k-2)(k)(k+1) \cdots (l-2)(l)(l+1) \cdots (n) \\ = &(j-1, l-1) \circ (i-1, k-1) \ . \end{aligned}$$ □

Note that $(j-1, l-1)$ and $(i-1, k-1)$ might not be disjoint, since by definition of $\beta(i, j, k, l)$ we may have $j = k$ (hence the use of $\circ$ in the expression of $\overline{\beta(i, j, k, l)}$). We can now recover a known lower bound on the block-interchange distance, which is actually the exact distance [3].

Theorem 4. *[3] For all π in S_n, we have $bid(\pi) \geq \frac{n+1-c(\Gamma(\overline{\pi}))}{2}$.*

Proof. By Theorem 3 and Lemma 4, a lower bound on $bid(\pi)$ is given by the length of a minimum factorisation of $\overline{\pi}$ into pairs of exchanges. Since this length equals $(n + 1 - c(\Gamma(\overline{\pi})))/2$ (see e.g. [16]), the proof follows. □

Let us now characterise the image of a transposition by our mapping.

Lemma 5. *For any transposition* $\tau(i,j,l)$, *we have*

$$\overline{\tau(i,j,l)} = (i-1, l-1, j-1) \ .$$

Proof. As noted in Section 2.2, we have $\tau(i,j,l) = \beta(i,j,j,l)$; Lemma 4 yields:

$$\overline{\tau(i,j,l)} = \overline{\beta(i,j,j,l)} = (j-1, l-1) \circ (i-1, j-1) = (i-1, l-1, j-1) \ . \quad \square$$

We recover the following known lower bound on the transposition distance, where $c_{odd}(\Gamma(\overline{\pi}))$ denotes the number of odd cycles in $\Gamma(\overline{\pi})$.

Theorem 5. *[4] For all* π *in* S_n, *we have* $td(\pi) \geq \frac{n+1-c_{odd}(\Gamma(\overline{\pi}))}{2}$.

Proof. By Theorem 3 and Lemma 5, a lower bound on $td(\pi)$ is given by the length of a minimum factorisation of $\overline{\pi}$ into 3-cycles. Since this length equals $(n+1-c_{odd}(\Gamma(\overline{\pi})))/2$ (see e.g. [16]), the proof follows. $\square$

5 An Improved Lower Bound on the Prefix Transposition Distance

Using our theory, we prove a *new* lower bound on $ptd(\pi)$ and show that it always outperforms (1). We will find it convenient to express $ptb(\pi)$ (defined after Theorem 1 page 638) as follows.

Lemma 6. *For any* π *in* S_n, *we have*

$$ptb(\pi) = n + 1 - c_1(\Gamma(\overline{\pi})) + \begin{cases} 1 \ \textit{if} \ \pi_1 = 1, \\ 0 \ \textit{otherwise.} \end{cases}$$

Proof. The formula results from the observation that, among the $n+1$ pairs of adjacent elements in $\widetilde{\pi}$, each adjacency in $\widetilde{\pi}$ gives rise to a 1-cycle in $\Gamma(\overline{\pi})$, and from the fact that if $\pi_1 = 1$, then we counted the 1-cycle that corresponds to $(0,1)$ as an adjacency, which we correct by adding 1. $\square$

Let $d_3^1(\pi)$ denote the length of a minimum factorisation of π in S_n into a product of 3-cycles, where each 3-cycle in the factorisation is further required to contain the first element.

Proposition 1. *For any* π *in* S_n, *we have* $ptd(\pi) \geq d_3^1(\overline{\pi})$.

Proof. Replace i with 1 in Lemma 5, and mimic the proof of Theorem 5. $\square$

Next, we show how to compute $d_3^1(\pi)$ for π in A_n. The following simple observation will be useful.

Observation 1. *For any* π *in* A_n, *we have* $n \equiv c(\Gamma(\pi)) \pmod 2$.

Lemma 7. *For any* π *in* A_n, *we have*

$$d_3^1(\pi) = \frac{n + c(\Gamma(\pi))}{2} - c_1(\Gamma(\pi)) - \begin{cases} 0 \ \textit{if} \ \pi_1 = 1, \\ 1 \ \textit{otherwise.} \end{cases}$$

Proof. Given a minimum factorisation of length ℓ of an even permutation π into prefix exchanges, we can construct a sequence of $\ell/2$ 3-cycles by noting that $(1,j) \circ (1,i) = (1,i,j)$. Therefore $d_3^1(\pi) \leq \ell/2$. On the other hand, assume there exists a shorter sequence of 3-cycles acting on the first element whose product is π; then one can split each of these 3-cycles into two prefix exchanges using the relation above and find a shorter expression for π as a product of prefix exchanges, a contradiction. The result follows from Theorem 1. □

As a corollary, we obtain the following lower bound on the prefix transposition distance:

Theorem 6. *For any π in S_n, we have*

$$ptd(\pi) \geq \frac{n+1+c(\Gamma(\overline{\pi}))}{2} - c_1(\Gamma(\overline{\pi})) - \begin{cases} 0 & \text{if } \pi_1 = 1, \\ 1 & \text{otherwise.} \end{cases} \quad (5)$$

Proof. Follows from Proposition 1 and Lemma 7. □

We conclude this section by proving that our lower bound always outperforms Dias and Meidanis' (Lemma 1).

Theorem 7. *Lower bound* (5) *is always at least as large as lower bound* (1).

Proof. Assume $\pi \neq \iota$ (otherwise the result trivially holds); this implies that $\Gamma(\overline{\pi})$ has at least one cycle of length at least 2, which means that $c(\Gamma(\overline{\pi})) - c_1(\Gamma(\overline{\pi})) \geq 1$. There are two cases to prove: if $\pi_1 = 1$, then lower bound (1) becomes

$$\left\lceil \frac{(n+1-c_1(\Gamma(\overline{\pi}))+1)-1}{2} \right\rceil = \left\lceil \frac{n+1-c_1(\Gamma(\overline{\pi}))}{2} \right\rceil ,$$

and lower bound (5) satisfies

$$\frac{n+1+c(\Gamma(\overline{\pi}))-2c_1(\Gamma(\overline{\pi}))}{2} \geq \frac{n+2-c_1(\Gamma(\overline{\pi}))}{2} \geq \left\lceil \frac{n+1-c_1(\Gamma(\overline{\pi}))}{2} \right\rceil .$$

On the other hand, if $\pi_1 \neq 1$, then lower bound (1) becomes

$$\left\lceil \frac{(n+1-c_1(\Gamma(\overline{\pi})))-1}{2} \right\rceil = \left\lceil \frac{n-c_1(\Gamma(\overline{\pi}))}{2} \right\rceil ,$$

and by Observation 1, lower bound (5) becomes

$$\begin{aligned} \frac{n+1+c(\Gamma(\overline{\pi}))}{2} - c_1(\Gamma(\overline{\pi})) - 1 &= \left\lceil \frac{n+1+c(\Gamma(\overline{\pi}))-2c_1(\Gamma(\overline{\pi}))-2}{2} \right\rceil \\ &\geq \left\lceil \frac{n-c_1(\Gamma(\overline{\pi}))}{2} \right\rceil . \end{aligned}$$

□

6 A Tighter Lower Bound on the Prefix Transposition Diameter

Dias and Meidanis [11] observed that the prefix transposition diameter lies between $n/2$ and $n-1$, and conjectured that it is equal to $n - \lfloor \frac{n}{4} \rfloor$. Recently, Chitturi and Sudborough [13] improved those bounds to $2n/3$ and $n - \log_8 n$, respectively. Using our new lower bound on the prefix transposition distance, we further improve the lower bound on the prefix transposition diameter.

Theorem 8. *For $n \geq 2$, the prefix transposition diameter of S_n is at least $\lfloor \frac{3n+1}{4} \rfloor$.*

Proof. We construct a family of permutations whose prefix transposition distance is at least $\lfloor \frac{3n+1}{4} \rfloor$. Let $\pi = \langle 3\ 2\ 1\ 4\ 7\ 6\ 5\ \cdots\ n-4\ n\ n-2\ n-3 \rangle$, or any other 2-*permutation*, i.e. a permutation such that $\Gamma(\overline{\pi})$ contains only cycles of length 2 (this requires that $n \equiv 3 \pmod 4$). There are four cases to examine, each of which relies on Theorem 6:

1. if $n \equiv 3 \pmod 4$, we have $ptd(\pi) \geq (n+1+(n+1)/2)/2 - 0 - 1 = \frac{3n-1}{4}$.
2. if $n \equiv 0 \pmod 4$, let σ be a permutation such that $\Gamma(\overline{\sigma})$ is obtained by inserting a fixed point at the beginning of $\Gamma(\overline{\pi})$; since $\overline{\sigma}$ fixes 0 and has $n/2$ 2-cycles, we have $ptd(\sigma) \geq (n+1+n/2+1)/2 - 1 - 0 = \frac{3n}{4}$.
3. if $n \equiv 1 \pmod 4$, let σ' be a permutation such that $\Gamma(\overline{\sigma'})$ is obtained by inserting a fixed point anywhere in $\Gamma(\overline{\sigma})$; we have $ptd(\sigma') \geq (n+1+\frac{n-2+1}{2}+2)/2 - 2 = \frac{3n+1}{4}$.
4. if $n \equiv 2 \pmod 4$, let σ'' be a permutation such that $\Gamma(\overline{\sigma''})$ is obtained by inserting a 3-cycle (a, c, b) with $a < b < c$ anywhere in $\Gamma(\overline{\pi})$. Since $\overline{\sigma''}$ has $(n+1-3)/2+1$ cycles of length at least 2, we have $ptd(\sigma'') \geq (n+1+\frac{n+1-3}{2}+1)/2 - 0 - 1 = \frac{3n-2}{4}$. □

7 Experimental Results

We generated all permutations in S_n, for $1 \leq n \leq 10$, along with their prefix transposition distance, and compared lower bounds (1), (2) and (5) to the actual distance. Table 1 shows the results. It can be observed that many more permutations are tight with respect to our lower bound (column 5) than with respect to Dias and Meidanis' (column 3) or Chitturi and Sudborough's (column 4).

We also examined how large the gap between our lower bound and the actual prefix transposition distance can get. The remaining columns of Table 1 list the number of permutations whose prefix transposition distance equals our lower bound plus Δ. We note that, for $n \leq 9$, all permutations have a prefix transposition distance that is at most our lower bound plus 2 (plus 3 for $n = 10$).

Table 1. Experimental results; column 3 lists the number of cases where (1) is tight [17], column 4 lists the number of cases where (2) is tight, and columns 5 to 8 list the number of cases where (5) underestimates $ptd(\pi)$ by Δ

n	$n!$	tight w.r.t. (1)	tight w.r.t. (2)	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	$\Delta = 3$
1	1	1	1	1	0	0	0
2	2	2	2	2	0	0	0
3	6	4	4	6	0	0	0
4	24	13	15	22	2	0	0
5	120	41	48	106	14	0	0
6	720	196	255	574	143	3	0
7	5 040	862	1 144	3 782	1 234	24	0
8	40 320	5 489	7 737	27 471	12 310	539	0
9	362 880	31 033	44 187	229 167	128 576	5 137	0
10	3 628 800	247 006	369 979	2 103 510	1 427 966	97 321	3

8 Conclusions

We presented a new framework for reformulating any edit distance problem on permutations as a minimum-length factorisation problem on a related even permutation, under the implicit assumption that the edit operations are revertible. This approach is based on a new representation of a structure known as the *cycle graph*, which pervades the field of genome rearrangements in several different forms; it previously allowed us to enumerate permutations whose cycle graph decomposes into a given number of alternating cycles [12], and allowed us in this work to recover two previously known results in a simple and unified way. Moreover, we used our approach to derive a new lower bound on the prefix transposition distance that, as we showed both theoretically and experimentally, is a significant improvement over previous results. From that result, we deduced an improved lower bound on the prefix transposition diameter of the symmetric group, whose exact value is still unknown.

Future research will need to focus on computational complexity issues, since the complexity of sorting permutations by transpositions or by any prefix operation (except prefix exchanges) is still open, as well as on finding improved approximations and upper bounds on the corresponding distances. We hope that our framework will provide further insight on various issues related to those edit distance problems and their variants, and will allow us to characterise polynomial time solvable cases, if the general problems indeed prove to be difficult.

References

1. Li, Z., Wang, L., Zhang, K.: Algorithmic approaches for genome rearrangement: a review. IEEE Transactions on Systems, Man and Cybernetics, Part C 36(5), 636–648 (2006)
2. Kececioglu, J., Sankoff, D.: Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement. Algorithmica 13(1-2), 180–210 (1995)

3. Christie, D.A.: Sorting permutations by block-interchanges. Information Processing Letters 60(4), 165–169 (1996)
4. Bafna, V., Pevzner, P.A.: Sorting by transpositions. SIAM Journal on Discrete Mathematics 11(2), 224–240 (1998)
5. Caprara, A.: Sorting permutations by reversals and eulerian cycle decompositions. SIAM Journal on Discrete Mathematics 12(1), 91–110 (1999)
6. Elias, I., Hartman, T.: A 1.375-approximation algorithm for sorting by transpositions. IEEE/ACM Transactions on Computational Biology and Bioinformatics 3(4), 369–379 (2006)
7. Lakshmivarahan, S., Jwo, J.S., Dhall, S.K.: Symmetry in interconnection networks based on Cayley graphs of permutation groups: A survey. Parallel Computing 19(4), 361–407 (1993)
8. Akers, S.B., Krishnamurthy, B.: A group-theoretic model for symmetric interconnection networks. IEEE Transactions on Computers 38(4), 555–566 (1989)
9. Gates, W.H., Papadimitriou, C.H.: Bounds for sorting by prefix reversal. Discrete Mathematics 27(1), 47–57 (1979)
10. Akers, S.B., Krishnamurthy, B., Harel, D.: The star graph: An attractive alternative to the n-cube. In: Proceedings of the Fourth International Conference on Parallel Processing, August 1987, pp. 393–400. Pennsylvania State University Press (1987)
11. Dias, Z., Meidanis, J.: Sorting by prefix transpositions. In: Laender, A.H.F., Oliveira, A.L. (eds.) SPIRE 2002. LNCS, vol. 2476, pp. 65–76. Springer, Heidelberg (2002)
12. Doignon, J.P., Labarre, A.: On Hultman numbers. Journal of Integer Sequences 10(6), Article 07.6.2, 13 pages (2007)
13. Chitturi, B., Sudborough, I.H.: Bounding prefix transposition distance for strings and permutations. In: Proceedings of the Forty-First Annual Hawaii International Conference on System Sciences, January 2008, p. 468. IEEE Computer Society Press, Los Alamitos (2008)
14. Wielandt, H.: Finite permutation groups. Translated from German by R. Bercov. Academic Press, New York (1964)
15. Labarre, A.: New bounds and tractable instances for the transposition distance. IEEE/ACM Transactions on Computational Biology and Bioinformatics 3(4), 380–394 (2006)
16. Jerrum, M.R.: The complexity of finding minimum-length generator sequences. Theoretical Computer Science 36(2-3), 265–289 (1985)
17. Fortuna, V.J.: Distâncias de transposição entre genomas. Master's thesis, Universidade Estadual de Campinas, São Paulo, Brazil (March 2005)