Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in Theoretical Computer Science 346 (2019) 29–40

Electronic Notes in Theoretical Computer Science

www.elsevier.com/locate/entcs

Approximation Algorithms for Sorting Permutations by Length-Weighted Short Rearrangements

Alexsandro Oliveira Alexandrino[a,1,2]
Guilherme Henrique Santos Miranda[a,1,3]
Carla Negri Lintzmayer[b,1,4] Zanoni Dias[a,1,5]

[a] *Institute of Computing, University of Campinas (Unicamp), Brazil*

[b] *Center for Mathematics, Computation and Cognition, Federal University of ABC (UFABC), Brazil*

Abstract

Genome rearrangements are events that affect large portions of a genome. When using the rearrangement distance to compare two genomes, one wants to find a minimum cost sequence of rearrangements that transforms one into another. Since we represent genomes as permutations, we can reduce this problem to the problem of sorting a permutation with a minimum cost sequence of rearrangements. In the traditional approach, we consider that all rearrangements are equally likely to occur and we set a unitary cost for all rearrangements. However, there are two variations of the problem motivated by the observation that rearrangements involving large segments of a genome rarely occur. The first variation adds a restriction to the rearrangement's length. The second variation uses a cost function based on the rearrangement's length. In this work, we present approximation algorithms for five problems combining both variations, that is, problems with a length-limit restriction and a cost function based on the rearrangement's length.

Keywords: Genome Rearrangements, Approximation Algorithms, Sorting Permutations.

1 Introduction

Understanding the evolutionary process and finding the evolutionary distance between two distinct organisms are challenging tasks in Computational Biology. A

[1] This work was supported by the National Counsel of Technological and Scientific Development, CNPq (grants 400487/2016-0 and 425340/2016-3), São Paulo Research Foundation, FAPESP (grants 2013/08293-7, 2015/11937-9, 2017/12646-3, 2017/16246-0, and 2017/16871-1), the Brazilian Federal Agency for the Support and Evaluation of Graduate Education, CAPES, and the CAPES/COFECUB program (grant 831/15).

[2] Email: alexsandro@ic.unicamp.br

[3] Email: guilherme.miranda@students.ic.unicamp.br

[4] Email: carla.negri@ufabc.edu.br

[5] Email: zanoni@ic.unicamp.br

https://doi.org/10.1016/j.entcs.2019.08.004
1571-0661/© 2019 The Author(s). Published by Elsevier B.V.

well-accepted way to estimate such distance is to use the rearrangement distance between the genomes of the two organisms, which is given by the smallest number of rearrangements required to transform one genome into another.

Genome rearrangements are events that modify segments of genomes. A rearrangement model M is the set of valid rearrangements that can be used to estimate the evolutionary distance. Two well known types of genome rearrangements are reversals, which revert and flip the orientation of elements in a genome, and transpositions, which exchange the positions of two adjacent segments in a genome.

We represent a genome as a permutation of integers, where each element corresponds to a genomic fragment shared by the genomes we are comparing. This representation assumes that there are no duplicate genes. If the orientation of the genes is known, we use a signed permutation, where each element has a sign to represent its orientation. If the orientation is unknown, the signs are omitted and we use an unsigned permutation.

Due to algebraic properties, it is possible to represent one of the genomes as the identity permutation and, in this way, we can reduce the problem of estimating the rearrangement distance to the problem of finding the minimum number of rearrangements to sort a permutation.

The problem of Sorting Unsigned Permutations by Reversals is NP-Hard [3]. The best-known result is an algorithm with approximation factor of 1.375, given by Berman *et al.* [1]. On the other hand, as shown by Hannenhalli and Pevzner [6], the problem of Sorting Signed Permutations by Reversals is polynomial. The problem of Sorting (Unsigned) Permutations by Transpositions is NP-Hard [2] and the best-known result is an algorithm with approximation factor of 1.375, given by Elias and Hartman [4].

When the rearrangement model contains reversals and transpositions, we have the problems of Sorting Signed and Unsigned Permutations by Reversals and Transpositions. The complexity of both problems are unknown. For unsigned permutations, the best-known result is a $2k$-approximation algorithm [14], where k is the approximation factor for cycle decomposition of breakpoints graphs. For signed permutations, Walter *et al.* [16] presented a 2-approximation algorithm.

Many variants of sorting permutations problems were inspired by different biological relevance [5, 7–9, 12, 13, 15]. Some of these variants [5, 7–9, 15] add an extra restriction that limits the length of the rearrangements. This restriction is motivated by the observation that rearrangements involving large segments of a genome rarely occur [11]. For length-limits equal to 2 and 3, we call these rearrangements super short and short, respectively. The problems of Sorting by Super Short Operations are polynomial [5, 8]. The best-known algorithms for Sorting Signed and Unsigned Permutations by Short Reversals have approximation factors of 5 [5] and 2 [7], respectively. The best-known algorithm for the problem of Sorting (Unsigned) Permutations by Short Transpositions has approximation factor of 5/4 [9]. For the same problem, there exists a $(1+n/x)$-approximation [10], where n is the number of elements and x is the number of inversions in the permutation. This algorithm gives a lower approximation factor for permutations with many inversions. For the problems of Sorting Signed and Unsigned Permutations by Short Reversals and Short Transpositions, the best-known algorithms have approximation factors of 3 [5] and 2 [15], respectively.

The same motivation for the length-limit restriction inspired some authors [12, 13] to consider length-weighted rearrangements, that is, a rearrangement has a cost based on its length. In this approach, the goal of the problem is to find a minimum-cost rearrangement sequence that sorts a permutation. We present approximation algorithms for five such problems: Sorting by Length-Weighted Short Reversals (SbR), Sorting by Length-Weighted Short Signed Reversals (Sb$\bar{\text{R}}$), Sorting by Length-Weighted Short Transpositions (SbT), Sorting by Length-Weighted Short Reversals and Short Transpositions (SbRT), and Sorting by Length-Weighted Short Signed Reversals and Short Transpositions (Sb$\bar{\text{R}}$T). To the best of our knowledge, there are no results in the literature considering length-weighted short rearrangements.

This work is organized as follows. Section 2 presents all concepts and notations related to the problems. Sections 3, 4, and 5 describe the approximation algorithms we developed. Section 6 concludes our work.

2 Definitions

A *signed permutation* of size n is represented as $\pi = (\pi_1\ \pi_2\ \dots\ \pi_n)$, such that $\pi_i \in \{-n, \dots, -2, -1, 1, 2, \dots, n\}$ and $|\pi_i| \neq |\pi_j|$ if and only if $i \neq j$, for all i and j. An *unsigned permutation* of size n is represented similarly, but $\pi_i \in \{1, 2, \dots, n\}$ and $\pi_i \neq \pi_j$ if and only if $i \neq j$, for all i and j. The *identity permutation*, $\iota = (1\ 2\ \dots\ n)$, is the goal of the sorting problems we will study. For signed permutations, all elements of ι have positive sign.

A *signed reversal* $\bar{\rho}(i, j)$, with $1 \leq i \leq j \leq n$, reverts the segment $\pi_i, \pi_{i+1}, \dots, \pi_j$ and flips the sign of each affected element. An *unsigned reversal* $\rho(i, j)$, with $1 \leq i < j \leq n$, only reverts the segment $\pi_i, \pi_{i+1}, \dots, \pi_j$. So, when π is a signed permutation, we have that $(\pi_1\ \dots\ \underline{\pi_i\ \dots\ \pi_j}\ \dots\ \pi_n) \cdot \bar{\rho}(i, j) = (\pi_1\ \dots\ \underline{-\pi_j\ \dots\ -\pi_i}\ \dots\ \pi_n)$ and, when π is an unsigned permutation, we have that $(\pi_1\ \dots\ \underline{\pi_i\ \dots\ \pi_j}\ \dots\ \pi_n) \cdot \rho(i, j)$ $= (\pi_1\ \dots\ \underline{\pi_j\ \dots\ \pi_i}\ \dots\ \pi_n)$. A signed (or unsigned) reversal $\bar{\rho}(i, j)$ (or $\rho(i, j)$) is called a signed (or unsigned) k-reversal, where $k = |\bar{\rho}| = j - i + 1$ (or $k = |\rho| = j - i + 1$) is the *rearrangement's length.*

A *transposition* $\tau(i, j, k)$, with $1 \leq i < j < k \leq n + 1$, exchanges the two adjacent segments $\pi_i, \pi_{i+1}, \dots, \pi_{j-1}$ and $\pi_j, \pi_{j+1}, \dots, \pi_{k-1}$ without changing signs of their elements. Therefore, we have that $(\pi_1\ \dots\ \pi_{i-1}\ \underline{\pi_i\ \dots\ \pi_{j-1}}\ \underline{\pi_j\ \dots\ \pi_{k-1}}\ \pi_k\ \dots\ \pi_n) \cdot \tau(i, j, k) =$ $(\pi_1\ \dots\ \pi_{i-1}\ \underline{\pi_j\ \dots\ \pi_{k-1}}\ \underline{\pi_i\ \dots\ \pi_{j-1}}\ \pi_k\ \dots\ \pi_n)$. A transposition $\tau(i, j, k)$ has length $|\tau| = k - i$.

For transpositions and reversals, we say that a rearrangement is *super short* if its length is less than or equal to 2 and *short* if its length is less than or equal to 3. Furthermore, considering the length-weighted approach, a rearrangement β has cost $f(\beta) = |\beta|^\alpha$, where $\alpha \geq 0$. Given a sequence of rearrangements $S = \beta_1, \beta_2, \dots, \beta_m$, we have that $f(S) = \sum_{i=1}^{m} f(\beta_i)$. Note that the unweighted approach is equivalent to set $\alpha = 0$. In this work, we will consider that $\alpha = 1$, that is, $f(\beta) = |\beta|$.

Given a rearrangement model M with only short rearrangements, the cost function f, and a permutation π, the *short rearrangement distance* $c^f_M(\pi)$ is equal to $\sum_{i=1}^{m} f(\beta_i)$ such that every rearrangement $\beta_1, \beta_2, \ldots, \beta_m$ is in M, $\pi \cdot \beta_1 \cdot \beta_2 \ldots \cdot \beta_m = \iota$, and $\sum_{i=1}^{m} f(\beta_i)$ is minimum.

We use $sr, s\bar{r}, st, srt, s\bar{r}t$ to denote unsigned short reversals, signed short reversals, short transpositions, unsigned short reversals and short transpositions, and signed short reversals and short transpositions, respectively. For instance, the signed short reversals distance of permutation π is denoted by $c^f_{s\bar{r}}(\pi)$.

2.1 Inversions and Inversion Graph

A pair of elements (π_i, π_j) of π is called an *inversion* if $i < j$ and $|\pi_i| > |\pi_j|$. We use $\text{Inv}(\pi)$ to denote the number of inversions in a permutation π. For any permutation π, $\text{Inv}(\pi) = 0$ if $|\pi_1| < |\pi_2| < \ldots < |\pi_n|$. Note that considering unsigned permutations, only the identity permutation has $\text{Inv}(\pi) = 0$.

Lemma 2.1 *For any permutation π, if $\text{Inv}(\pi) > 0$, then there exists an inversion (π_i, π_{i+1}).*

Proof. Let $\pi_1, \ldots, \pi_i$ be a maximal subsequence such that $|\pi_j| < |\pi_{j+1}|$ for all $1 \leq j < i$. Since $\text{Inv}(\pi) > 0$, we have that $i < n$. So, $|\pi_i| > |\pi_{i+1}|$ and (π_i, π_{i+1}) is an inversion. □

Given a permutation π and a rearrangement β, let $\Delta\text{Inv}(\pi, \beta) = \text{Inv}(\pi) - \text{Inv}(\pi \cdot \beta)$, that is, $\Delta\text{Inv}(\pi, \beta)$ denotes the change caused by β in the number of inversions in π. From Lemma 2.1, we can conclude that for any permutation π such that $\text{Inv}(\pi) > 0$, there exists a super short rearrangement β with $\Delta\text{Inv}(\pi, \beta) = 1$. Lemma 2.2 provides bounds on the value $\Delta\text{Inv}(\pi, \beta)$, where $|\beta| \leq 3$.

Lemma 2.2 *(Galvão et al. [5]) Given a permutation π and a short rearrangement β, we have that*

i) $-1 \leq \Delta\text{Inv}(\pi, \beta) \leq 1$ *if β has length* 2,

ii) $-2 \leq \Delta\text{Inv}(\pi, \beta) \leq 2$ *if β is a transposition of length* 3, *and*

iii) $-3 \leq \Delta\text{Inv}(\pi, \beta) \leq 3$ *if β is a signed or unsigned reversal of length* 3.

The *permutation graph* or *inversion graph* of a permutation π is the undirected graph $G(\pi) = (V, E)$, where $V = \{\pi_1, \pi_2, \ldots, \pi_n\}$ and $E = \{(\pi_i, \pi_j)$: the pair (π_i, π_j) is an inversion in $\pi\}$. Figure 1 shows an example of a permutation graph.

Given a signed permutation π, a connected component C of $G(\pi)$ is *odd* if it contains an odd number of negative elements (vertices) and it is *even* otherwise. We use $c(\pi)$ and $c_{\text{odd}}(\pi)$ to denote the number of connected components and the number of odd connected components of $G(\pi)$, respectively. An edge e of $G(\pi)$ whose deletion increases the number of connected components of $G(\pi)$ is called a *cut-edge*. We use $\gamma(\pi)$ to denote the number of non-isolated vertices in $G(\pi)$, that is, the number of elements in π belonging to at least one inversion.

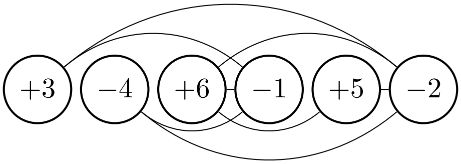

Fig. 1. Permutation graph $G(\pi)$ of the signed permutation $\pi = (+3\ {-4}\ {+6}\ {-1}\ {+5}\ {-2})$. In this example, $c(\pi) = c_{\text{odd}}(\pi) = 1$ and $\gamma(\pi) = 6$.

2.2 Entropy

Given a permutation π, the *entropy* of π_i is defined as $\text{ent}(\pi_i) = ||\pi_i| - i|$, for $1 \leq i \leq n$. In other words, the entropy of an element π_i is the distance between π_i and its position in the identity permutation. We define the sets $E_\pi^{\text{even}^-} = \{\pi_i : \pi_i < 0$ and $\text{ent}(\pi_i)$ is even$\}$ and $E_\pi^{\text{odd}^+} = \{\pi_i : \pi_i > 0$ and $\text{ent}(\pi_i)$ is odd$\}$. These sets will be used in lower bounds for Sb$\bar{\text{R}}$ and Sb$\bar{\text{R}}$T.

As an example, for $\pi = (+5\ {-4}\ {+3}\ {-1}\ {+2})$, we have $\text{ent}(+5) = 4$, $\text{ent}(-4) = 2$, $\text{ent}(+3) = 0$, $\text{ent}(-1) = 3$, and $\text{ent}(+2) = 3$, and the sets $E_\pi^{\text{even}^-} = \{-4\}$ and $E_\pi^{\text{odd}^+} = \{+2\}$.

3 Approximation Algorithms for Unsigned Permutations

In this section, we present a $\frac{4}{3}$-approximation algorithm for SbT and a 2-approximation algorithm for SbR and SbRT. They are greedy algorithms whose choice is to prioritize rearrangements that reduce the most number of inversions with the lowest cost. These algorithms are named $\frac{4}{3}$-T, 2-R, and 2-RT for the problems SbT, SbR, and SbRT, respectively.

To prove the approximation factors, Lemma 3.1 presents bounds on the ratio $\frac{\Delta\text{Inv}(\pi,\beta)}{|\beta|}$ considering that β is a short rearrangement. With these results, Lemmas 3.2 and 3.3 present lower bounds for the problems SbT, SbR, and SbRT relating the sorting distance to the number of inversions in a permutation.

Lemma 3.1 *Given a permutation π, any short reversal ρ has ratio $\frac{\Delta\text{Inv}(\pi,\rho)}{|\rho|} \leq 1$ and any short transposition τ has ratio $\frac{\Delta\text{Inv}(\pi,\tau)}{|\tau|} \leq \frac{2}{3}$.*

Lemma 3.2 *For any permutation π, $c_{st}^f(\pi) \geq \frac{3}{2}\text{Inv}(\pi)$.*

Proof. By Lemma 3.1, for any transposition τ, the ratio $\frac{\Delta\text{Inv}(\pi,\tau)}{|\tau|}$ is at most $\frac{2}{3}$, which means that the minimum cost to decrease $\text{Inv}(\pi)$ by 1 is $\frac{3}{2}$. Since the identity permutation has $\text{Inv}(\iota) = 0$, any sorting sequence needs to decrease $\text{Inv}(\pi)$ inversions in π to sort the permutation, resulting in the lower bound $\frac{3}{2}\text{Inv}(\pi)$. □

Lemma 3.3 *For any permutation π, $c_{sr}^f(\pi) \geq c_{srt}^f(\pi) \geq \text{Inv}(\pi)$.*

Proof. Since any sorting sequence for the problem SbR is a valid sorting sequence for SbRT, we have that $c_{sr}^f(\pi) \geq c_{srt}^f(\pi)$. The proof of $c_{srt}^f(\pi) \geq \text{Inv}(\pi)$ is analogous to the proof of Lemma 3.2. Note that for any rearrangement β, the ratio $\frac{\Delta\mathrm{Inv}(\pi,\beta)}{|\beta|}$ is at most 1 (Lemma 3.1). □

The following lemma describes upper bounds for the distance given by the greedy algorithms.

Lemma 3.4 *For any permutation π, $c^f_{st}(\pi) \leq 2\,\mathrm{Inv}(\pi)$ and $c^f_{srt}(\pi) \leq c^f_{sr}(\pi) \leq 2\,\mathrm{Inv}(\pi)$.*

Proof. We need to prove that it is always possible to decrease the number of inversions by 1 with a cost less than or equal to 2, considering the three rearrangement models. If $\mathrm{Inv}(\pi) = 0$, then π is already sorted. If $\mathrm{Inv}(\pi) > 0$, then there exists an inversion (π_i, π_{i+1}) in π (Lemma 2.1). So, the transposition $\tau(i, i{+}1, i{+}2)$ and the reversal $\rho(i, i{+}1)$ both have cost 2 and $\Delta\mathrm{Inv}(\pi, \tau(i, i{+}1, i{+}2)) = \Delta\mathrm{Inv}(\pi, \rho(i, i{+}1)) = 1$, since the inversion (π_i, π_{i+1}) is removed. □

Theorem 3.5 *SbR and SbRT are 2-approximable and SbT is $\frac{4}{3}$-approximable.*

Observe that the distance is $O(n^2)$ because any permutation can have at most $\binom{n}{2}$ inversions. Besides that, the variation in the number of inversions caused by a short rearrangement can be calculated in constant time, since at most three elements are affected. As there are $O(n)$ possible short rearrangements, the algorithms spend $O(n)$ time to choose the best rearrangement to apply at each step. Thus, all three algorithms have time complexity $O(n^3)$.

We note that the exact algorithm for the problem of Sorting by Super Short Rearrangements has time complexity of $O(n^2)$ [8] and it is also a valid algorithm for SbR, SbT, and SbRT. Given a permutation π, at each step, this algorithm removes an inversion with a rearrangement of length 2. Therefore, a sorting sequence retrieved by this algorithm has a cost of exactly $2\,\mathrm{Inv}(\pi)$, which means that this algorithm is a 2-approximation for SbR and SbRT (Lemma 3.3) and a $\frac{4}{3}$-approximation for SbT (Lemma 3.2). However, the greedy algorithms presented above retrieve sorting sequences with a cost of at most $2\,\mathrm{Inv}(\pi)$ and, so, they have better practical results. Furthermore, a sorting sequence S retrieved by one of the greedy algorithms has a cost of exactly $2\,\mathrm{Inv}(\pi)$ if, and only if, it does not contain any rearrangement of length 3.

4 Improved Approximation for Transpositions

This section presents an improved approximation algorithm for permutations with many inversions. The algorithm is defined in Algorithm 1.

For each element $\pi_k = i$, with $1 \leq i \leq n$, Algorithm 1 applies a sequence of short transpositions to move π_k to its correct position. In this way, the permutation π is sorted at the end of this algorithm. This algorithm has time complexity of $O(n^2)$, since it has n iterations and, in the worst case of each iteration, $O(n)$ short transpositions are applied in π. Lemma 4.1 gives an upper bound on the cost of a sorting sequence used by Algorithm 1.

Algorithm 1 *Approximation Algorithm for SbT*

```
function sort ($\pi$)
  $c \leftarrow 0$
  for $i$ from 1 to $n$ do
    if $\pi_i \neq i$ then
      let $k$ be the position of element $i$
      if $k - i$ is even then
        $\pi \leftarrow \pi \cdot \tau(k-2, k, k+1) \cdot \tau(k-4, k-2, k-1) \cdot \ldots \cdot \tau(i, i+2, i+3)$
      otherwise
        $\pi \leftarrow \pi \cdot \tau(k-2, k, k+1) \cdot \tau(k-4, k-2, k-1) \cdot \ldots \cdot \tau(i, i+1, i+2)$
      $c \leftarrow c + 3\lfloor \frac{k-i}{2} \rfloor + 2\,((k-i) \mod 2)$
  return $c$
```

Lemma 4.1 *For any permutation π, Algorithm 1 finds a sorting sequence S with $f(S) \leq \frac{3}{2}\text{Inv}(\pi) + \frac{\gamma(\pi)}{2}$.*

Proof. We analyze the average cost to remove an inversion in each iteration of the algorithm.

Consider the i-th iteration. Let π be the permutation at the beginning of this iteration and π' be the permutation at the end of this iteration. Furthermore, let $\pi_k = i$. In this iteration, we know that $\pi = (1\ 2\ \ldots\ (i-1)\ \pi_i\ \ldots\ \pi_k\ \ldots\ \pi_n)$ and that there exists an inversion (π_j, π_k), for any $i \leq j < k$, since $j < k$ and $\pi_k < \pi_j$. After moving π_k to its correct position with the sequence S', the permutation $\pi' = (1\ 2\ \ldots\ (i-1)\ i\ \pi_i\ \ldots\ \pi_{k-1}\ \pi_{k+1}\ \ldots\ \pi_n)$.

If $k - i$ is even, then S' contains $\frac{k-i}{2}$ transpositions of length 3 and $f(S') = \frac{3}{2}(k-i)$. Otherwise, S' has $\lfloor \frac{k-i}{2} \rfloor$ transpositions of length 3 and one transposition of length 2, which means that $f(S') = 3\lfloor \frac{k-i}{2} \rfloor + 2 = \frac{3}{2}(k-i) + \frac{1}{2}$, because $k - i$ is odd. Since S' removes $k - i$ inversions and does not add any inversions in the permutation, we conclude that S' has an average cost of $\frac{3}{2}$ to remove one inversion and, if $k - i$ is odd, it has an additional cost of $\frac{1}{2}$. Therefore, $f(S) = \frac{3}{2}\text{Inv}(\pi) + \frac{1}{2}m$, where m is the number of iterations where a transposition of length 2 is used.

Since any transposition of S removes at least one inversion and no inversions are added, any isolated element of $G(\pi)$ is not affected by S. Thus, the permutation π is not changed in $n - \gamma(\pi)$ iterations of the algorithm. Consequently, $m \leq \gamma(\pi)$ and $f(S) \leq \frac{3}{2}\text{Inv}(\pi) + \frac{1}{2}\gamma(\pi)$. □

Theorem 4.2 *Algorithm 1 is a $(1 + \gamma(\pi)/(3\,\text{Inv}(\pi)))$-approximation for SbT.*

If $\text{Inv}(\pi) \geq cn^2$, where c is a constant, then the approximation factor of Algorithm 1 is

$$1 + \frac{\gamma(\pi)}{3\,\text{Inv}(\pi)} \leq 1 + \frac{n}{3\,cn^2} = 1 + \frac{1}{3\,cn}.$$

As an example, if $n = 200$ and $\text{Inv}(\pi) = 10000$, then the approximation factor is approximately 1.007.

Lemma 4.3 *The approximation factor of Algorithm 1 is at most $\frac{5}{3}$.*

Proof. The graph $G(\pi)$ has at least $\frac{\gamma(\pi)}{2}$ edges. Thus, the approximation factor is

$$1 + \frac{\gamma(\pi)}{3\,\mathrm{Inv}(\pi)} \leq 1 + \frac{\gamma(\pi)}{3\,\gamma(\pi)/2} = 1 + \frac{2}{3} = \frac{5}{3}.$$

□

5 Approximation Algorithms for Signed Permutations

In this section, we present greedy algorithms for Sb$\bar{\mathrm{R}}$ and Sb$\bar{\mathrm{R}}$T with approximation factor of 3, which are called 3-$\bar{\mathrm{R}}$ and 3-$\bar{\mathrm{R}}$T, respectively. Furthermore, we present a $\frac{7}{3}$-approximation algorithm for Sb$\bar{\mathrm{R}}$T called $\frac{7}{3}$-$\bar{\mathrm{R}}$T.

5.1 *The* 3*-Approximation Algorithms for Sb$\bar{R}$ and Sb$\bar{R}$T*

To show how these greedy algorithms work, we define a score function ϕ based on the inversion graph of π and on the sets $E_\pi^{\mathrm{even}^-}$ and $E_\pi^{\mathrm{odd}^+}$.

Given a permutation π and a rearrangement sequence S, let π' be the resulting permutation after applying S to π. The score function is defined as

$$\begin{aligned}\phi(\pi, S) &= \frac{\left(\sum_{v \in G(\pi)} d(v) + |E_\pi^{\mathrm{even}^-}| + |E_\pi^{\mathrm{odd}^+}|\right) - \left(\sum_{v \in G(\pi')} d(v) + |E_{\pi'}^{\mathrm{even}^-}| + |E_{\pi'}^{\mathrm{odd}^+}|\right)}{\sum_{\beta \in S} |\beta|} \\ &= \frac{(2\,\mathrm{Inv}(\pi) + |E_\pi^{\mathrm{even}^-}| + |E_\pi^{\mathrm{odd}^+}|) - (2\,\mathrm{Inv}(\pi') + |E_{\pi'}^{\mathrm{even}^-}| + |E_{\pi'}^{\mathrm{odd}^+}|)}{\sum_{\beta \in S} |\beta|},\end{aligned}$$

since $\sum_{v \in G(\pi)} d(v) = 2\,\mathrm{Inv}(\pi)$ for any permutation π.

The identity permutation is the only permutation with $\mathrm{Inv}(\pi) = |E_\pi^{\mathrm{even}^-}| = |E_\pi^{\mathrm{odd}^+}| = 0$. An optimal sorting sequence is the one with highest score function. So, the greedy choice of 3-$\bar{\mathrm{R}}$ and 3-$\bar{\mathrm{R}}$T is to always apply the short rearrangement with highest score function until the permutation is sorted. The following lemmas are used to prove the approximation factor of both algorithms.

Lemma 5.1 *(Galvão et al. [5]) For any signed permutation π and signed 2-reversal $\bar{\rho}$, we have that $|E_\pi^{\mathrm{even}^-}| + |E_\pi^{\mathrm{odd}^+}| = |E_{\pi'}^{\mathrm{even}^-}| + |E_{\pi'}^{\mathrm{odd}^+}|$, where $\pi' = \pi \cdot \bar{\rho}$.*

Lemma 5.2 *For any signed permutation π and short signed reversal $\bar{\rho}$, we have that $\phi(\pi, \bar{\rho}) \leq 3$.*

Proof. Let $\pi' = \pi \cdot \bar{\rho}$ and $\bar{\rho} = \bar{\rho}(i, j)$. Recall that $\Delta\mathrm{Inv}(\pi, \bar{\rho}) = \mathrm{Inv}(\pi) - \mathrm{Inv}(\pi')$. We can divide the proof into the following cases:

(i) *$\bar{\rho}$ has length 1.* As a 1-reversal only changes the sign of element π_i, the number of inversions in π' is the same as in π. Since the parity of $\mathrm{ent}(\pi_i)$ does not change, we have $(|E_{\pi}^{\text{even}^-}| + |E_{\pi}^{\text{odd}^+}|) - (|E_{\pi'}^{\text{even}^-}| + |E_{\pi'}^{\text{odd}^+}|) \leq 1$. Therefore, $\phi(\pi, \bar{\rho}) \leq 1$.

(ii) *$\bar{\rho}$ has length 2.* A signed 2-reversal can remove at most one inversion and, consequently, $2(\Delta\text{Inv}(\pi, \bar{\rho})) \leq 2$. Moreover, we have that $|E_{\pi}^{\text{even}^-}| + |E_{\pi}^{\text{odd}^+}| = |E_{\pi'}^{\text{even}^-}| + |E_{\pi'}^{\text{odd}^+}|$ (Lemma 5.1). Therefore, $\phi(\pi, \bar{\rho}) \leq 1$.

(iii) *$\bar{\rho}$ has length 3.* A signed 3-reversal can remove at most three inversions and, consequently, $2(\Delta\text{Inv}(\pi, \bar{\rho})) \leq 6$. Moreover, we have that $(|E_{\pi}^{\text{even}^-}| + |E_{\pi}^{\text{odd}^+}|) - (|E_{\pi'}^{\text{even}^-}| + |E_{\pi'}^{\text{odd}^+}|) \leq 3$, since only three elements are affected by $\bar{\rho}$ and $E_{\pi}^{\text{even}^-} \cap E_{\pi}^{\text{odd}^+} = \emptyset$. Therefore, $\phi(\pi, \bar{\rho}) \leq \frac{9}{3} = 3$.

□

Lemma 5.3 *For any signed permutation π and short transposition τ, we have that $\phi(\pi, \tau) \leq \frac{7}{3}$.*

Proof. Analogous to the proof of Lemma 5.2. □

Lemma 5.4 *For any signed permutation $\pi \neq \iota$, there exists a short signed reversal $\bar{\rho}$ with $\phi(\pi, \bar{\rho}) \geq 1$.*

Proof. Let $\pi' = \pi \cdot \bar{\rho}$. If $\text{Inv}(\pi) = 0$, then $\text{ent}(\pi_i) = 0$ for all elements π_i of π and $|E_{\pi}^{\text{odd}^+}| = 0$. As $\pi \neq \iota$, there exists an element $\pi_i < 0$ in $E_{\pi}^{\text{even}^-}$ and, after applying $\bar{\rho}(i, i)$ in π, this element becomes positive in π'. Therefore, we have that $|E_{\pi}^{\text{even}^-}| - |E_{\pi'}^{\text{even}^-}| = 1$ and $\phi(\pi, \bar{\rho}) = 1$.

If $\text{Inv}(\pi) > 0$, then there exists an inversion (π_i, π_{i+1}) in π (Lemma 2.1). Thus, the signed 2-reversal $\bar{\rho}(i, i+1)$ removes this inversion, resulting in $2(\Delta\text{Inv}(\pi, \bar{\rho})) = 2$ and $|E_{\pi}^{\text{even}^-}| + |E_{\pi}^{\text{odd}^+}| = |E_{\pi'}^{\text{even}^-}| + |E_{\pi'}^{\text{odd}^+}|$ (Lemma 5.1). Therefore, $\phi(\pi, \bar{\rho}) = 1$.□

Theorem 5.5 3-$\bar{R}$ *and* 3-$\bar{R}T$ *are* 3*-approximation algorithms for Sb$\bar{R}$ and Sb$\bar{R}T$, respectively.*

Proof. Since the algorithm 3-R̄ always searches for the short signed reversal with highest score function, at each step the algorithm finds a short signed reversal $\bar{\rho}$ with $\phi(\pi, \bar{\rho}) \geq 1$ (Lemma 5.4). The value of $\phi(\pi, S)$, for any sorting sequence S, is less than or equal to 3 (Lemma 5.2) and the algorithm ensures a score function greater than or equal to 1, which gives us the approximation factor of 3.

The proof for 3-R̄T is analogous. □

Algorithms 3-R̄ and 3-R̄T have time complexity $O(n^3)$. This analysis is similar to the one used for algoritms $\frac{4}{3}$-T, 2-R, and 2-RT. Note that the score function ϕ can be calculated in constant time for any short rearrangement, since at most three elements are affected.

5.2 The 7/3-Approximation Algorithm for Sb$\bar{R}T$

Given a permutation π and a rearrangement sequence S, let π' be the resulting permutation after applying S to π. We define a new score function ψ as

$$\psi(\pi, S) = \frac{\left(\sum_{v \in G(\pi)} d(v) + c_{\text{odd}}(\pi)\right) - \left(\sum_{v \in G(\pi')} d(v) + c_{\text{odd}}(\pi')\right)}{\sum_{\beta \in S} |\beta|}$$
$$= \frac{(2\,\text{Inv}(\pi) + c_{\text{odd}}(\pi)) - (2\,\text{Inv}(\pi') + c_{\text{odd}}(\pi'))}{\sum_{\beta \in S} |\beta|}.$$

The identity permutation is the only one with $\text{Inv}(\pi) + c_{\text{odd}}(\pi) = 0$. The greedy choice of the $\frac{7}{3}$-approximation algorithm is to always apply the short rearrangement with highest score function until the permutation is sorted. The following lemmas present bounds on the values of the score function.

Lemma 5.6 *For any signed permutation π and short rearrangement β, we have that $\psi(\pi, \beta) \leq \frac{7}{3}$.*

Proof. Let $\pi' = \pi \cdot \beta$. We can divide the proof into the following cases:

(i) β is a signed reversal of length 1. A signed 1-reversal affects only one element and does not change the number of inversions in the permutation. Besides that, since only the component containing the affected element may become even, we have that $c_{\text{odd}}(\pi) - c_{\text{odd}}(\pi') \leq 1$. Therefore, $\psi(\pi, \bar{\rho}) \leq 1$.

(ii) β is a rearrangement of length 2. Let π_i and π_{i+1} be the elements affected by β. We further divide our analysis into the following subcases:

(a) (π_i, π_{i+1}) is an inversion. The rearrangement β removes the inversion (π_i, π_{i+1}) and, consequently, $2(\Delta\text{Inv}(\pi, \beta)) = 2$. Besides that, note that π_i and π_{i+1} are in the same component of $G(\pi)$. So, β affects only one component and $c_{\text{odd}}(\pi) - c_{\text{odd}}(\pi') \leq 1$.

(b) (π_i, π_{i+1}) is not an inversion. The rearrangement β adds the inversion (π'_i, π'_{i+1}) and, consequently, $2(\Delta\text{Inv}(\pi, \beta)) = -2$. Furthermore, since π_i and π_{i+1} are in distinct components of $G(\pi)$, β affects only two components and $c_{\text{odd}}(\pi) - c_{\text{odd}}(\pi') \leq 2$.

In both cases, we have that $(2\,\text{Inv}(\pi) + c_{\text{odd}}(\pi)) - (2\,\text{Inv}(\pi') + c_{\text{odd}}(\pi')) \leq 3$ and, consequently, $\psi(\pi, \beta) \leq \frac{3}{2}$.

(iii) β is a rearrangement of length 3. Let π_i, π_{i+1}, and π_{i+2} be the elements affected by β. We further divide our analysis into the following subcases:

(a) π_i, π_{i+1}, and π_{i+2} are in the same component of $G(\pi)$. We have that $2(\Delta\text{Inv}(\pi, \beta)) \leq 6$ (Lemma 2.2) and $c_{\text{odd}}(\pi) - c_{\text{odd}}(\pi') \leq 1$.

(b) Two elements of $\{\pi_i, \pi_{i+1}, \pi_{i+2}\}$ are in a component C_1 and the remaining is in a component C_2 of $G(\pi)$. Since there is only one edge connecting the elements π_i, π_{i+1}, and π_{i+2}, we have that $2(\Delta\text{Inv}(\pi, \beta)) \leq 2$. Furthermore, β affects only two components of $G(\pi)$ and $c_{\text{odd}}(\pi) - c_{\text{odd}}(\pi') \leq 2$.

(c) π_i, π_{i+1}, and π_{i+2} are in distinct components of $G(\pi)$. In this case, we have that $2(\Delta\text{Inv}(\pi, \beta)) \leq -4$ and $c_{\text{odd}}(\pi) - c_{\text{odd}}(\pi') \leq 3$. Note that a short transposition $\tau(i, j, i+3)$, with $i < j < i+3$, adds two inversions and a signed short reversal $\bar{\rho}(i, i+2)$ adds three inversions in the permutation.

In all cases, we have that $(2\,\mathrm{Inv}(\pi) + c_{\mathrm{odd}}(\pi)) - (2\,\mathrm{Inv}(\pi') + c_{\mathrm{odd}}(\pi')) \leq 7$. Therefore, $\psi(\pi, \beta) \leq \frac{7}{3}$.

□

Lemma 5.7 *For any signed permutation $\pi \neq \iota$, there exists a short rearrangement β with $\psi(\pi, \beta) \geq 1$.*

Proof. Let $\pi' = \pi \cdot \beta$. If $\mathrm{Inv}(\pi) = 0$, then there are no edges in $G(\pi)$ and each component of $G(\pi)$ is a single vertex. As $\pi \neq \iota$, there exists an odd component $C = \{\pi_i\}$ and, after applying $\beta = \bar{\rho}(i, i)$ in π, the component C becomes even. Thus, we have that $(2\,\mathrm{Inv}(\pi) + c_{\mathrm{odd}}(\pi)) - (2\,\mathrm{Inv}(\pi') + c_{\mathrm{odd}}(\pi')) = 1$ and $\psi(\pi, \beta) \geq 1$.

If $\mathrm{Inv}(\pi) > 0$, then there exists an edge $e = (\pi_i, \pi_{i+1})$ in $G(\pi)$. Let C be the component that contains e. Suppose that e is not a cut-edge. A rearrangement $\beta = \tau(i, i+1, i+2)$ removes the inversion (π_i, π_{i+1}) and, since e is not a cut-edge, the components of $G(\pi')$ are the same components of $G(\pi)$. Moreover, the parity of the components remains the same, as a transposition does not change signs of elements. Thus, we have $(2\,\mathrm{Inv}(\pi) + c_{\mathrm{odd}}(\pi)) - (2\,\mathrm{Inv}(\pi') + c_{\mathrm{odd}}(\pi')) = 2$ and $\psi(\pi, \beta) \geq 1$.

Now, suppose that e is a cut-edge. Let C_1 and C_2 be the components of $C - e$. Since e is a cut-edge, the elements π_i and π_{i+1} are in distinct components of $C - e$. We further divide this proof into the following cases:

(i) C_1 and C_2 are both even. After applying $\beta = \tau(i, i+1, i+2)$ on π, we have that $2(\Delta\mathrm{Inv}(\pi, \beta)) = 2$ and $c_{\mathrm{odd}}(\pi) - c_{\mathrm{odd}}(\pi') = 0$. Note that C is also even and the parity of the components C_1 and C_2 remains the same.

(ii) C_1 and C_2 are both odd. After applying $\beta = \bar{\rho}(i, i+1)$ on π, we have that $2(\Delta\mathrm{Inv}(\pi, \beta)) = 2$ and $c_{\mathrm{odd}}(\pi) - c_{\mathrm{odd}}(\pi') = 0$. Note that C is even and the components C_1 and C_2 become even after applying the signed 2-reversal in π.

(iii) C_1 and C_2 have distinct parities. After applying $\beta = \tau(i, i+1, i+2)$ on π, we have that $2(\Delta\mathrm{Inv}(\pi, \beta)) = 2$ and $c_{\mathrm{odd}}(\pi) - c_{\mathrm{odd}}(\pi') = 0$. Note that C is odd and the parity of the components C_1 and C_2 remains the same.

In all cases, we have that $(2\,\mathrm{Inv}(\pi) + c_{\mathrm{odd}}(\pi)) - (2\,\mathrm{Inv}(\pi') + c_{\mathrm{odd}}(\pi')) = 2$ and $\psi(\pi, \beta) \geq 1$. □

Theorem 5.8 *$\frac{7}{3}$-$\bar{R}T$ is a $\frac{7}{3}$-approximation algorithm for Sb$\bar{R}T$.*

Proof. Since the algorithm $\frac{7}{3}$-R̄T always searches for the short rearrangement with highest score function ψ, at each step the algorithm finds a short rearrangement β with $\psi(\pi, \beta) \geq 1$ (Lemma 5.7). The value of $\psi(\pi, S)$, for any sorting sequence S, is less than or equal to $\frac{7}{3}$ (Lemma 5.6) and the algorithm ensures a score function greater than or equal to 1, which gives us the approximation factor of $\frac{7}{3}$. □

Algorithm $\frac{7}{3}$-R̄T has time complexity of $O(n^4)$. The analysis is similar to the previous greedy algorithms, except that it takes linear time to find the variation in the number of odd components caused by a short rearrangement [5].

6 Conclusion

In this work, we introduced the study of Sorting Permutations by Length-Weighted Short Rearrangements problems. We developed approximation-algorithms for five such problems. For future work, we aim at improving the approximation factors for problems involving signed permutations. Furthermore, we intend to study a new variation of this problem where the rearrangements' length-limit is equal to an integer λ, with $\lambda \leq n$.

References

[1] Berman, P., S. Hannenhalli and M. Karpinski, *1.375-Approximation Algorithm for Sorting by Reversals*, in: *Proceedings of the 10th Annual European Symposium on Algorithms (ESA'2002)*, Lecture Notes in Computer Science **2461**, Springer-Verlag Berlin Heidelberg New York, 2002 pp. 200–210.

[2] Bulteau, L., G. Fertin and I. Rusu, *Sorting by Transpositions is Difficult*, SIAM Journal on Computing **26** (2012), pp. 1148–1180.

[3] Caprara, A., *Sorting Permutations by Reversals and Eulerian Cycle Decompositions*, SIAM Journal on Discrete Mathematics **12** (1999), pp. 91–110.

[4] Elias, I. and T. Hartman, *A 1.375-Approximation Algorithm for Sorting by Transpositions*, IEEE/ACM Transactions on Computational Biology and Bioinformatics **3** (2006), pp. 369–379.

[5] Galvão, G. R., O. Lee and Z. Dias, *Sorting Signed Permutations by Short Operations*, Algorithms for Molecular Biology **10** (2015), pp. 1–17.

[6] Hannenhalli, S. and P. A. Pevzner, *Transforming Cabbage into Turnip: Polynomial Algorithm for Sorting Signed Permutations by Reversals*, Journal of the ACM **46** (1999), pp. 1–27.

[7] Heath, L. S. and J. P. C. Vergara, *Sorting by Short Swaps*, Journal of Computational Biology **10** (2003), pp. 775–789.

[8] Jerrum, M. R., *The Complexity of Finding Minimum-length Generator Sequences*, Theoretical Computer Science **36** (1985), pp. 265–289.

[9] Jiang, H., H. Feng and D. Zhu, *An 5/4-Approximation Algorithm for Sorting Permutations by Short Block Moves*, in: *Proceedings of the 25th International Symposium on Algorithms and Computation (ISAAC'2014)*, Lecture Notes in Computer Science **8889** (2014), pp. 491–503.

[10] Jiang, H., D. Zhu and B. Zhu, *A (1+ϵ)-Approximation Algorithm for Sorting by Short Block-Moves*, Theoretical Computer Science **437** (2012), pp. 1–8.

[11] Lefebvre, J.-F., N. El-Mabrouk, E. R. M. Tillier and D. Sankoff, *Detection and validation of single gene inversions*, Bioinformatics **19** (2003), pp. i190–i196.

[12] Nguyen, T. C., H. T. Ngo and N. B. Nguyen, *Sorting by Restricted-Length-Weighted Reversals*, Genomics Proteomics & Bioinformatics **3** (2005), pp. 120–127.

[13] Pinter, R. Y. and S. Skiena, *Genomic Sorting with Length-Weighted Reversals*, Genome Informatics **13** (2002), p. 2002.

[14] Rahman, A., S. Shatabda and M. Hasan, *An Approximation Algorithm for Sorting by Reversals and Transpositions*, Journal of Discrete Algorithms **6** (2008), pp. 449–457.

[15] Vergara, J. P. C., "Sorting by Bounded Permutations," Ph.D. thesis, Virginia Polytechnic Institute and State University (1998).

[16] Walter, M. E. M. T., Z. Dias and J. Meidanis, *Reversal and Transposition Distance of Linear Chromosomes*, in: *Proceedings of the 5th International Symposium on String Processing and Information Retrieval (SPIRE'1998)* (1998), pp. 96–102.