Sorting circular permutations by reversal

Andrew Solomon, Paul Sutcliffe and Raymond Lister

University of Technology, Sydney, Australia
`andrews,psutclif,raymond@it.uts.edu.au`

Abstract. Unsigned circular permutations are used to represent tours in the traveling salesman problem as well as the arrangement of gene loci in circular chromosomes. The minimum number of segment reversals required to transform one circular permutation into another gives some measure of distance between them which is useful when studying the 2-opt local search landscape for the traveling salesman problem, and, when determining the phylogeny of a group of related organisms. Computing this distance is equivalent to sorting by (a minimum number of) reversals. In this paper we show that sorting circular permutations by reversals can be reduced to the same problem for linear reversals, and that it is NP-hard. These results suggest that for most practical purposes any computational tools available for reversal sort of linear permutations will be sufficiently accurate.
These results entail the development of the algebraic machinery for dealing rigorously with circular permutations.

1 Introduction

A circular permutation can be thought of as a necklace with n distinct beads. Rotating and flipping the necklace do not change the object but one necklace may be transformed into any other by cutting it in two places, reversing one segment and rejoining the ends, or a composition of such operations. This paper addresses the problem of finding a minimum length sequence of segment reversals required to transform one circular permutation into another.

Tours in the symmetric traveling salesman problem are precisely circular permutations of the cities. In the context of the traveling salesman problem, segment reversal is called a *2-opt move* and is used to define a combinatorial landscape which is subjected to local search techniques in order to find local minima [11].

Among others, Boese [4] suggests a correlation between the values of local minima and their distance from other local minima – the so called "big valley" hypothesis which informs a number of successful heuristics for traversing the landscape. Boese uses the number of breakpoints (pairs which are adjacent in one permutation, but not the other) as an estimate of reversal distance. Our motivation for the present work is to have a more accurate measure of reversal distance for the purpose of investigating the big valley hypothesis.

Historically, the question of determining reversal distance between circular permutations was first posed in 1982 by Watterson et. al. [14] in the context of

computational biology. A circular permutation models the arrangement of gene loci around a circular chromosome such as is found in bacterial and mitochondrial DNA. While point mutations occur freqently, the order of gene loci is quite stable over the generations. When the gene order does change, it is usually by segment reversal [5]. Therefore, the reversal distance between two chromosomes is a measure of their evolutionary distance.

Watterson's paper gives the first rudimentary bounds on the reversal distance where it is noted that since each reversal can eliminate at most two breakpoints, half the number of breakpoints is a lower bound on the reversal distance. A simple ratchet algorithm is given to show that for circular permutations of n points, at most n reversals are required to transform one permutation into any other.

This paper sparked a great deal of interest but subsequent investigations focussed on the simpler case of linear chromosomes. In the remainder of this section we review the progress made in the linear case.

1.1   Sorting linear permutations by reversal

As we shall see, the problem of finding a minimum length sequence of reversals transforming one permutation into another is equivalent to sorting a permutation by (a minimum number of) segment reversals. Henceforth we use the initials SBR to refer to this problem.

Kececioglu and Sankoff [9, 10] give a greedy algorithm for sorting a permutation σ by reversal bounded above by the number $b(\sigma)$ of breakpoints which is therefore a 2-approximation algorithm. Computing upper and lower bounds on reversal distance, Kececioglu and Sankoff go on to give an algorithm for computing an exact reversal sort for linear permutations.

Bafna and Pevzner [2] improved on these results by formulating a $\frac{7}{4}$-approximation algorithm to sort a permutation by reversals. Along the way they defined the problem of sorting *signed* permutations by reversals, where each point has not only a position, but also an orientation. Signed permutations are arguably more significant biologically as genes have extension as well as position.

Using elaborate graph theoretic constructions, Caprara [5] solved in the affirmative a longstanding conjecture of Kececioglu and Sankoff that sorting by reversals is NP-hard. In contrast Hannenhalli and Pevzner [8] give a polynomial algorithm which sorts signed permutations.

David A. Christie [6] finds a polynomial time $\frac{3}{2}$-approximation algorithm for sorting unsigned permutations by reversals, and this remains the best known approximation factor.

Bounding approximability of SBR, Berman and Karpinski [3] show that it is NP-hard to approximate the reversal length of a linear permutation to a factor better than 1.0008.

1.2 Notational Preliminaries

Formalizing the notion of a circular permutation is a delicate matter and clarity is well served by taking some pains to carefully define the notion of a permutation.

Linear Permutations Rather than regarding a permutation as a function from a set to itself, we distinguish the set Σ of n objects being permuted, from the ordered set $[n] = \{0, 1, \ldots n-1\}$ of positions in which we place each object. Then a permutation is a bijection $\pi : \Sigma \to [n]$ such that $a\pi$ denotes the position in which we place object $a \in \Sigma$, that is, permutations act on the right. Occasionally, it will be convenient to visualize all the objects of Σ in their positions under π as $(\pi_0 \pi_1 \ldots \pi_{n-1})$, which is to say $\pi_i = i\pi^{-1}$.

Fix some permutation $\iota : \Sigma \to [n]$. Then ι defines a canonical ordering of the elements of Σ and ι will be called the *identity* permutation. Then $\Sigma = \{\iota_0, \ldots, \iota_{n-1}\}$. (Identifying the sets $[n]$ and Σ would enable us to revert to the usual notions of permutation and identity.)

A *reversal* $\rho(i, j)$ (with $i < j$ in $[n]$) of a linear permutation is a bijection on the set of positions whose effect can be visualized as reversing the segment from position i to position j, transforming $(\pi_0 \ldots \pi_i \pi_{i+1} \ldots \pi_j \ldots \pi_{n-1})$ into $(\pi_0 \ldots \pi_j \pi_{j-1} \ldots \pi_{i+1} \pi_i \pi_{j+1} \ldots \pi_{n-1})$. Precisely, for $x \in \Sigma$ define

$$\rho(i, j) : [n] \longrightarrow [n]$$
$$x \longmapsto \begin{cases} i + j - x & \text{if } i \leq x \leq j \\ x & \text{otherwise.} \end{cases}$$

then it is easy to see that $\pi\rho(i, j) = (\pi_0 \ldots \pi_j \pi_{j-1} \ldots \pi_i \pi_{j+1} \ldots \pi_{n-1})$ as required.

Circular Permutations The notion of circular permutation we are trying to capture is an arrangement of the elements of Σ around the vertices of a regular n-gon subject to the condition that, like a necklace, rotating the n-gon, or flipping it over does not change the circular permutation that it represents.

Arbitrarily, we label the vertices of the n-gon by the elements of $\mathbb{Z}_n$ from 0 at the twelve o'clock position and proceed clockwise up to $n - 1$. A *circular arrangement* of the elements of Σ around the vertices of the n-gon is then a bijection $\pi : \Sigma \to \mathbb{Z}_n$. In a similar way to the treatment of linear permutations, fix an arbitrary circular arrangement $\iota : \Sigma \to \mathbb{Z}_n$ and refer to ι as the *identity* arrangement. For $i \in \mathbb{Z}_n$ define the *elementary rotation* $r : \mathbb{Z}_n \to \mathbb{Z}_n$ by $ir = i \oplus 1$ and *canonical reflection* $s : \mathbb{Z}_n \to \mathbb{Z}_n$ by $is = \ominus i$, where $\oplus$ and $\ominus$ denote addition and negation (or subtraction) in $\mathbb{Z}_n$. For example, $(\pi_0 \pi_1 \pi_2 \pi_3 \pi_4)r = (\pi_4 \pi_0 \pi_1 \pi_2 \pi_3)$ and $(\pi_0 \pi_1 \pi_2 \pi_3 \pi_4)s = (\pi_0 \pi_4 \pi_3 \pi_2 \pi_1)$.

The maps r and s generate all $2n$ rotations and reflections of the regular n-gon. Together, these form the *dihedral group* D_n, which has presentation

$$\langle r, s \mid s^2, r^n, rs = sr^{n-1} \rangle. \tag{1}$$

To capture the idea that π, πr and πs all represent the same circular permutation, define a *circular permutation* to be the set πD_n for some circular arrangement π. It is then clear that $\pi D_n = \pi r D_n = \pi s D_n$ as required. Any circular arrangement in πD_n defines a linear permutation by identifying $\mathbb{Z}_n$ with $[n]$. Call such a permutation a *linearization* of πD_n and denote the set of all $2n$ linearizations of πD_n by $\mathrm{lin}(\pi D_n)$.

For $i, j \in \mathbb{Z}_n$, define the *interval* $[i,j]$ to be the set $\{i, i \oplus 1, \ldots, j \ominus 1, j\}$. For example if n is 6 then $[3,1] = \{3,4,5,0,1\}$ while $[1,3] = \{1,2,3\}$.

Let $x \in \Sigma$. Then a *circular reversal* $\rho^c(i,j)$ is defined by

$$\rho^c(i,j) : \mathbb{Z}_n \;\to\; \mathbb{Z}_n$$
$$x \longmapsto \begin{cases} i \oplus j \ominus x & \text{if } x \in [i,j] \\ x & \text{otherwise.} \end{cases}$$

As an example of the way a circular reversal acts on a circular arrangement, notice that when $n = 6$,

$$(\pi_0 \pi_1 \pi_2 \pi_3 \pi_4 \pi_5)\rho^c(1,3) = (\pi_0 \pi_3 \pi_2 \pi_1 \pi_4 \pi_5)$$

and

$$(\pi_0 \pi_1 \pi_2 \pi_3 \pi_4 \pi_5)\rho^c(4,1) = (\pi_5 \pi_4 \pi_2 \pi_3 \pi_1 \pi_0).$$

The technical report [12] inspired a number of notational decisions in this section. In particular, the symbols used to denote reversal, arithmetic in $\mathbb{Z}_n$ and intervals appear also in [12].

1.3 Mathematical Preliminaries

In the linear case, the problem one attempts to solve is to find, for two permutations σ and τ a minimum length sequence of reversals $\alpha_1, \ldots, \alpha_k$ such that $\sigma \alpha_1 \ldots \alpha_k = \tau$, however

$$\sigma \alpha_1 \ldots \alpha_k = \tau \text{ if and only if}$$
$$\iota \tau^{-1} \sigma \alpha_1 \ldots \alpha_k = \iota$$

and since ι is the identity permutation, we see that a minimum length sequence of reversals transforming σ into τ is equivalent to a minimum length sequence of reversals which transforms $\iota \tau^{-1} \sigma$ into ι, which is to say, the sequence of reversals *sorts* $\iota \tau^{-1} \sigma$. The reversal distance between a permutation π and the identity will be called the *reversal length* of the permutation and denoted $l(\pi)$.

In the circular case, the primary problem is to find, given two circular arrangements σ and τ, a minimum length sequence of circular reversals $\alpha_1^c, \ldots, \alpha_k^c$ such that $\sigma \alpha_1^c \ldots \alpha_k^c \in \tau D_n$. Once again, notice that $\sigma \alpha_1^c \ldots \alpha_k^c \in \tau D_n$ if and only if $\iota \tau^{-1} \sigma \alpha_1^c \ldots \alpha_k^c \in \iota D_n$. Regarding ιD_n as the *identity* circular permutation, we see that the sequence $\alpha_1^c, \ldots, \alpha_k^c$ *sorts* the circular arrangement $\iota \tau^{-1} \sigma$.

The reversal distance between a circular arrangement and the identity will be called the *reversal length* of the arrangement.

For the remainder of the paper, fix some n as the size of Σ and let r denote the elementary rotation and s the canonical reflection in D_n. We give some useful facts describing the interaction of circular reversals with the elements of the dihedral group D_n.

Lemma 1.
$$s\rho^c(i,j) = r^{i\oplus j}\rho^c(j\oplus 1, i\ominus 1).$$

Proof. Noting that for any $x \in \mathbb{Z}_n$, $x \in [i,j]$ if and only if $\ominus x \in [\ominus j, \ominus i]$ we have

$$x s\rho^c(i,j) = \begin{cases} i\oplus j\oplus x & \text{if } x\in[\ominus j, \ominus i] \\ \ominus x & \text{otherwise} \end{cases}$$

while

$$x r^{i\oplus j}\rho^c(j\oplus 1, i\ominus 1) = \begin{cases} (j\oplus 1)\oplus(i\ominus 1)\ominus(x\oplus i\oplus j) & \text{if } x\oplus i\oplus j\in[j\oplus 1, i\ominus 1] \\ x\oplus i\oplus j & \text{otherwise} \end{cases}$$

$$= \begin{cases} \ominus x & \text{if } x\in[1\ominus i, \ominus 1\ominus j], \text{ subtracting } i\oplus j \text{ everywhere} \\ x\oplus i\oplus j & \text{otherwise} \end{cases}$$

$$= \begin{cases} x\oplus i\oplus j & \text{if } x\in[\ominus j, \ominus i] \\ \ominus x & \text{otherwise} \end{cases}$$

as required.

The reader may easily verify the following equations.

$Eqn-I$	$r\rho^c(i,j) = \rho^c(i\ominus 1, j\ominus 1)r$
$Eqn-II$	$s\rho^c(i,j) = \rho^c(\ominus j, \ominus i)s$
$Eqn-III$	$\rho^c(i,j) = sr^{i\oplus j}\rho^c(j\oplus 1, i\ominus 1) = \rho^c(j\oplus 1, i\ominus 1)sr^{i\oplus j}$
$Eqn-IV$	$\rho^c(i\oplus 1, i) = sr^{2i\oplus 1}$

Eqn-I and Eqn-II ensure that for any $\rho^c(i,j)$ and any $d \in D_n$, $d\rho^c(i,j) = \rho^c(i',j')d$ for some $i',j' \in \mathbb{Z}_n$. Suppose there is a sequence $\alpha_1, \ldots, \alpha_k$ of reversals such that $\sigma\alpha_1 \ldots \alpha_k \in \iota D_n$. Then for any $\tau \in \sigma D_n$, $\tau = \sigma d$ so that $\sigma = \tau d^{-1}$ and

$$\sigma\alpha_1 \ldots \alpha_k = \tau d^{-1}\alpha_1 \ldots \alpha_k$$
$$= \tau\beta_1 \ldots \beta_k d^{-1} \in \iota D_n$$

for some reversals $\beta_1, \ldots, \beta_k$, so that $\tau\beta_1 \ldots \beta_k \in \iota D_n$. Consequently, τ has length at most k. By symmetry, this shows that any two circular arrangements in the same circular permutation have the same length, so we may speak of the *reversal length of a circular permutation* πD_n and denote it by $l^c(\pi D_n)$.

Proposition 2. *The following table expresses each non-identity element of D_n as a minimum length product of linear reversals.*

Element of D_n	As reversals
Orientation preserving elements	
$r^i, i \in \{1, 2\}$	$\rho(0, n - i)\rho(1, n - 1)$
$r^i, 2 < i < n - 2$	$\rho(0, n - i - 1)\rho(n - i, n - 1)\rho(0, n - 1)$
$r^i, i \in \{n - 2, n - 1\}$	$\rho(1, n - 1)\rho(0, i)$
Orientation reversing elements	
s	$\rho(1, n - 1)$
$sr^i, 0 < i < n - 2$	$\rho(0, i)\rho(i + 1, n - 1)$
$sr^i, i \in \{n - 2, n - 1\}$	$\rho(0, i)$

Proof. To verify the equality of the expressions on the left and right is an easy exercise. The proof that the expressions on the right hand side are of minimum length is tedious and inessential to the development of the remainder of the paper, so we omit it.

2 Reducing circular SBR to linear SBR

It is clear that if a sequence $\alpha_1, \ldots, \alpha_k$ sorts a linearization of πD_n then it certainly sorts some circular arrangement of πD_n so that the reversal length of πD_n is bounded above by the minimum reversal length amongst its linearizations.

Theorem 3. *If πD_n can be sorted in m circular reversals, then there is some linearization $\sigma \in \mathrm{lin}(\pi D_n)$ which can be sorted in at most m linear reversals.*

A direct result is that $l^c(\pi D_n)$ is bounded below by the minimum length amongst its reversals so that together with the observation above, we have

Corollary 4. *$l^c(\pi D_n)$ is precisely the minimum value of $l(\sigma)$ for any linearization σ of πD_n.*

Proof (of theorem). By way of a basis for an induction on m, suppose πD_n has reversal length 0. Then $\pi \in \iota D_n$, whence $\iota = \pi t$ for some $t \in D_n$. Consequently, the linearization πt of πD_n is sorted and has a reversal length of 0 as required.

Now suppose πD_n has reversal length m. That is, there is a sequence of circular reversals $\alpha_1^c, \ldots, \alpha_m^c$ such that $\pi \alpha_1^c \ldots \alpha_m^c \in \iota D_n$. Put $\pi_0 = \pi$ and for $1 \le i \le m$, set $\pi_i = \pi \alpha_1^c \ldots \alpha_i^c$.

By the inductive hypothesis, there is some linearization $\sigma_1 \in \mathrm{lin}(\pi_1 D_n)$ which is sortable in $m - 1$ linear reversals. Say $\gamma_2, \ldots, \gamma_m$ is a sequence of linear reversals sorting σ_1.

We now focus on the relationship between the linear permutation σ_1 and the circular arrangement $\pi_1 = \pi_0 \alpha_1 = \pi_0 \rho^c(i, j)$ for some $i, j \in \mathbb{Z}_n$. The presentation at (1) shows that an element of the dihedral group may always be written as a rotation, or as a reflection followed by a rotation, giving us only two cases to consider: Case (i) $\sigma_1 = \pi_1 r^k$; Case (ii) $\sigma_1 = \pi_1 sr^k$.

In Case (i) $\sigma_1 = \pi_0 \rho^c(i,j) r^k$, and by Eqn-I $\sigma_1 = \pi_0 r^k \rho^c(i \oplus k, j \oplus k)$. There are three subcases to consider: as elements of $\mathbb{Z}$ either (a) $i \oplus k \le j \oplus k$, (b) $i \oplus k = j \oplus k \oplus 1$, or (c) $i \oplus k > j \oplus k \oplus 1$. In case (a), set $\sigma_0 = \pi_0 r^k$ and $\gamma_1 = \rho(i \oplus k, j \oplus k)$. This gives $\sigma_0 \gamma_1 = \sigma_1$ and the sequence $\gamma_1, \gamma_2, \ldots, \gamma_m$ linearly sorts $\sigma_0 = \pi_0 r^k$ as required. In case (b), Eqn-IV gives $\rho^c(i \oplus k, j \oplus k) = sr^{2j \ominus 2k \oplus 1}$ so that

$$\sigma_1 = \pi_0 r^k sr^{2j \ominus 2k \oplus 1}$$
$$= \pi_0 sr^{2j \ominus k \oplus 1}$$

Putting $\sigma_0 = \pi_0 sr^{2j \ominus k \oplus 1}$ and $\gamma_1 = 1_{\mathbb{Z}_n}$, gives the required sequence of linear reversals. In case (c) Eqn-III gives $\rho^c(i \oplus k, j \oplus k) = sr^{i \oplus j \ominus 2k} \rho^c(j \oplus k \oplus 1, i \oplus k \ominus 1)$ so that

$$\sigma_1 = \pi_0 r^k \rho^c(i \oplus k, j \oplus k)$$
$$= \pi_0 r^k sr^{i \oplus j \ominus 2k} \rho^c(j \oplus k \oplus 1, i \oplus k \ominus 1)$$
$$= \pi_0 sr^{i \oplus j \ominus k} \rho^c(j \oplus k \oplus 1, i \oplus k \ominus 1)$$

Since $i \oplus k > j \oplus k \oplus 1$, $j \oplus k \oplus 1 \le i \oplus k \ominus 1$ so that $\rho(j \oplus k \oplus 1, i \oplus k \ominus 1)$ is a linear reversal. Putting $\sigma_0 = \pi_0 sr^{i \oplus j \ominus k}$, and $\gamma_1 = \rho(j \oplus k \oplus 1, i \oplus k \ominus 1)$ then ensures that the sequence $\gamma_1, \ldots, \gamma_m$ sorts σ_0 linearly as required.

In Case (ii)

$$\sigma_1 = \pi_1 sr^k$$
$$= \pi_0 \rho^c(i,j) sr^k$$
$$= \pi_0 s\rho^c(\ominus j, \ominus i) r^k$$
$$= \pi_0 sr^k \rho^c(k \ominus j, k \ominus i).$$

As above, there are three subcases to consider: as elements of $\mathbb{Z}$ either (a) $k \ominus j \le k \ominus i$, (b) $k \ominus j = k \ominus i \oplus 1$, or (c) $k \ominus j > k \ominus i \oplus 1$. In case (a) put $\sigma_0 = \pi_0 sr^k$ and $\gamma_1 = \rho(k \ominus j, k \ominus i)$ and $\gamma_1, \ldots, \gamma_m$ is the required sequence of linear reversals which sorts σ_0. In case (b), Eqn-IV gives $\rho^c(k \ominus j, k \ominus i) = sr^{2k \ominus 2i \oplus 1}$ so that

$$\sigma_1 = \pi_0 sr^k sr^{2k \ominus 2i \oplus 1}$$
$$= \pi_0 r^{k \ominus 2i \oplus 1}$$

Putting $\sigma_0 = \pi_0 r^{k \ominus 2i \oplus 1}$ and $\gamma_1 = 1_{\mathbb{Z}_n}$, gives the required sequence of linear reversals. Finally, in case (c) Eqn-III gives $\rho^c(k \ominus j, k \ominus i) = sr^{2k \ominus i \ominus j} \rho^c(k \ominus i \oplus 1, k \ominus j \ominus 1)$ so that

$$\sigma_1 = \pi_0 sr^k \rho^c(i \oplus k, j \oplus k)$$
$$= \pi_0 sr^k sr^{2k \ominus i \ominus j} \rho^c(k \ominus i \oplus 1, k \ominus j \ominus 1)$$
$$= \pi_0 r^{k \ominus i \ominus j} \rho^c(k \ominus i \oplus 1, k \ominus j \ominus 1)$$

Since $k \ominus j > k \ominus i \oplus 1$, $k \ominus i \oplus 1 \le k \ominus j \ominus 1$ so that $\rho(k \ominus i \oplus 1, k \ominus j \ominus 1)$ is a linear reversal. Putting $\sigma_0 = \pi_0 r^{k \ominus i \ominus j}$, and $\gamma_1 = \rho(k \ominus i \oplus 1, k \ominus j \ominus 1)$ then ensures that the sequence $\gamma_1, \ldots, \gamma_m$ sorts σ_0 linearly as required. $\square$

In summary, we see that given an algorithm $\mathcal{L}$ to solve the minimum length SBR problem for linear permutations, in order to solve SBR for some circular permutation πD_n, we need only apply $\mathcal{L}$ to each of the $2n$ linearizations of πD_n, and take the shortest solution. Calling a subroutine a polynomial number of times in this way is called a *Turing reduction* [7].

3   Circular Sort by Reversals is NP-hard

By recourse to a result of Berman and Karpinski [3] on the inapproximability of linear SBR, we show that circular SBR is NP-hard. The core of our proof is the following

Lemma 5. *Let $\alpha_1, \ldots, \alpha_m$ be a sequence of circular reversals taking σ to an element of ιD_n. Then there is a sequence $\beta_1, \ldots, \beta_k$ of linear reversals such that $k \le m$ and $\sigma \beta_1 \ldots \beta_k \in \iota D_n$.*

Proof. Proceed by induction on m. The $m = 0$ case is trivial. If α_1 is a linear reversal, put $\beta_1 = \alpha_1$ and appeal to the inductive hypothesis with the permutation $\sigma \beta_1$.

Therefore we may assume that α_1 is not a linear reversal. That is: $\alpha_1 = \rho^c(x, y)$ with $x > y$. There are two cases: (i) $\alpha_1 = \rho^c(i+1, i)$; and (ii) $\alpha_1 = \rho^c(i, j)$ with $i > j + 1$.

In case (i) Eqn-IV gives $\alpha_1^c = sr^{2i \oplus 1}$. By use of Eqn-I and Eqn-II

$$
\begin{aligned}
\sigma \alpha_1 \ldots \alpha_m &= \sigma s r^{2i \oplus 1} \alpha_2 \ldots \alpha_m \\
&= \sigma \alpha_2' \ldots \alpha_m' s r^{2i \oplus 1}
\end{aligned}
$$

so that $\sigma \alpha_2' . \alpha_m' \in \iota D_n$ and we are finished by appeal to the inductive hypothesis.

In case (ii) Eqn-III gives $\alpha_1 = sr^{i \oplus j} \rho(j \oplus 1, i \ominus 1)$ and $i > j + 1$ ensures $j \oplus 1 \le i \ominus 1$. By Eqn-III we have $\alpha_1 = \rho(j \oplus 1, i \ominus 1) sr^{i \oplus j}$.

Therefore

$$
\begin{aligned}
\sigma \alpha_1 &= \sigma \rho(j \oplus 1, i \ominus 1) sr^{i \oplus j} \alpha_2 \ldots \alpha_m \\
&= \sigma \rho(j \oplus 1, i \ominus 1) \alpha_2' \ldots \alpha_m' s r^{i \oplus j} \in \iota D_n
\end{aligned}
$$

so that setting $\beta_1 = \rho(j \oplus 1, i \ominus 1)$, $\sigma \beta_1$ is circularly sorted in $m - 1$ circular reversals, which completes the proof by appeal to the inductive hypothesis.

As an immediate consequence of Lemma 5 and Proposition 2 we have

Proposition 6. *For any linear permutation σ,*

$$
l^c(\sigma D_n) \le l(\sigma) \le l^c(\sigma D_n) + 3. \square
$$

Theorem 7 (Restatement of Theorem 6 in [3]). *For any positive ϵ_1, ϵ_2 it is NP-hard to distinguish linear permutations with $2240k$ breakpoints that have length below $(1236 + \epsilon_1)k$ from those whose length is above $(1237 - \epsilon_2)k$.*

In particular, setting $k = 4m$ and bounding ϵ_1, ϵ_2 we have

Corollary 8. *For $0 < \epsilon_1, \epsilon_2 < \frac{1}{10}$ it is NP-hard to distinguish between linear permutations with $2440 \times 4m$ breakpoints that have length below $l = (1236 + \epsilon_1)4m$ and those with length above $u = (1237 - \epsilon_2)4m$. Note that*

$$u - l = 4m - (\epsilon_1 + \epsilon_2)4m > 3\frac{1}{5}m > 3.$$

Finally, we are in a position to prove

Theorem 9. *The problem of computing the reversal length of a circular permutation is NP-hard.*

Proof. We show that the problem of estimating the length of a linear permutation with precision determined by Corollary 8 can be reduced in constant time to the problem of computing the reversal length of the associated circular permutation. Consequently the latter problem must be NP-hard.

To estimate the length of a linear permutation σ, compute the reversal length $l^c(\sigma D_n)$ of the corresponding circular permutation. The reversal length of σ is then approximated by Proposition 6.

With l and u defined as in Corollary 8, let σ be a permutation whose reversal length $l(\sigma)$ is either below l or above u. We show that $l(\sigma) < l$ if and only if $l^c(\sigma D_n) < l$. The forward direction is immediate from the statement of Proposition 6. For the reverse direction, if $l^c(\sigma D_n) < l$ then $l^c(\sigma D_n) + 3 < u$ since we defined l and u to be at least 3 apart. Since $l^c(\sigma D_n) + 3$ is an upper bound on $l(\sigma)$, we have that $l(\sigma) < u$, whence by definition of σ, $l(\sigma) < l$. $\qquad\square$

4   Conclusion

We showed that determining a reversal sort for circular permutations can be reduced to finding a minimum length sort amongst its $2n$ linearizations (Theorem 3).

Using an inapproximability result on linear SBR, it is shown that determining reversal distance between circular permutations is NP-hard (Theorem 9).

In practical terms, to approximate reversal length for a circular permutation it is sufficient to compute it for one of its linearizations using any of the programs already developed for this purpose (for example [10], [2]). This estimate will be accurate to within three reversals (Proposition 6) and NP-hardness of SBR for circular permutations assures us that using tools for linear permutations is likely to be as efficient as developing specific algorithms for circular permutations.

In case reversal lengths in a given situation are so small that an error margin of three is significant, Bafna and Pevzner's theorem [2, Theorem 5] concerning

the expected reversal length of a random permutation suggests that n will also be small. Therefore it may well be feasible to compute the length of the $2n$ linearizations for an exact result. This will be the subject of a future experimental investigation.

References

1. David A. Bader, Bernard M. E. Moret, Mi Yan, *A linear time algorithm for computing inversion distance between signed permutations with an experimental study*, Journal of Computational Biology, Volume 8, Number 5, 2001 pp483–491.
2. V. Bafna and P. A. Pevzner, *Genome rearrangements and sorting by reversals. SIAM Journal on Computing*, **25** (1996), 272-289.
3. P. Berman, M. Karpinski, *On some tighter inapproximability results (extended abstract)*, in "Automata, languages and programming (Prague, 1999)", Lecture Notes in Comput. Sci., **1644**, pp. 200–209, Springer, Berlin, 1999.
4. K. D. Boese, *Cost Versus Distance In the Traveling Salesman Problem*, Technical Report CSD-950018, UCLA Computer Science Department, May 1995.
5. Alberto Caprara, *Sorting Permutations by Reversals and Eulerian Cycle Decompositions* SIAM Journal on Discrete Mathematics, Volume 12, Number 1 (1999) pp. 91-110.
6. David A. Christie, *A 3/2-approximation algorithm for sorting by reversals*, in Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 244-252, San Francisco, California, 25-27 January 1998.
7. Michael R. Garey and David S. Johnson, "Computers and Intractability", W. H. Freeman, New York, 1979.
8. S. Hannenhalli, P. A. Pevzner, *Transforming cabbage into turnip: a polynomial algorithm for sorting signed permutations by reversals*, Journal of ACM, 46, 1-27, 1999.
9. John Kececioglu and David Sankoff, *Efficient bounds for oriented chromosome-inversion distance* Proceedings of the 5th Symposium on Combinatorial Pattern Matching, Springer-Verlag Lecture Notes in Computer Science 807, 307-325, 1994.
10. John Kececioglu and David Sankoff, *Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement* Algorithmica **13**, 180-210, 1995.
11. S. Lin and B. Kernighan, *An efficient heuristic for the traveling salesman problem*, Operations Research, 21(2):498–516, 1973.
12. J. Meidanis, M. E. M. T. Walter and Z. Dias, *Reversal distance of signed circular chromosomes*, Technical Report IC-00-23 (December 2000), Instituto de Computação, Universidade Estadual de Campinas, `http://www.ic.unicamp.br/ic-tr-ftp/2000/Abstracts.html`
13. S. Micali and V. Vazirani, *An $O(\sqrt{|V|}|E|)$ algorithm for finding maximum matchings in general gaphs*, Proceedings of the 21st Symposium on Foundations of Computer Science, 17–27, 1980, (cited in [10]).
14. G. Watterson, W. Ewens, T. Hall and A. Morgan, *The chromosome inversion problem*, J. Theor. Biol. **99** (1982), 1-7.