A lower bound on the reversal and transposition diameter

JOÃO MEIDANIS
Universidade de Campinas
Instituto de Computação
Cx.Postal 6176
13083-971 Campinas, São Paulo
Brazil

MARIA EMÍLIA M. T. WALTER
Universidade de Brasília
Departamento de Ciência da Computação
Campus Universitário – Asa Norte
70910-900 Brasília, Distrito Federal
Brazil

ZANONI DIAS
Universidade de Campinas
Instituto de Computação
Cx.Postal 6176
13083-971 Campinas, São Paulo
Brazil

August 22, 2004

Abstract

One possible model to study genome evolution is to represent genomes as permutations of genes and compute distances based on the minimum number of certain operations (rearrangements) needed to transform one permutation into another. Under this model, the shorter the distance, the closer the genomes are. Two operations that have been extensively studied are the reversal and the transposition. A reversal is an operation that reverses the order of the genes on a certain portion of the permutation. A transposition is an operation that "cuts" a certain portion of the permutation and "pastes" it elsewhere in the same permutation. In this note we show that the reversal and transposition distance of the signed permutation $\pi_n = (-1\ -2\ \ldots\ -(n-1)\ -n)$ with respect to the identity is $\lfloor n/2 \rfloor + 2$ for all $n \geq 3$. We conjecture that this value is the diameter of the permutation group under these operations.

1 Introduction

One possible model to study genome evolution is to represent genomes as permutations of genes and compute distances based on the minimum number of certain operations (rearrangements) needed to transform one permutation into another. Under this model, the shorter the distance, the closer the genomes are.

In this note we are interested in the diameter of permutation groups, that is, the maximum distance possible between two permutations of size n, under several operation choices. Two operations that have been extensively studied are the reversal and the transposition. A reversal is an operation that reverses the order of the genes on a certain portion of the permutation. A transposition is an operation that moves a certain portion of the permutation to another location in the same permutation.

Table 1 shows what is currently known about the diameter for signed and unsigned permutations under various combinations of the above operations. In this note we provide a lower bound for the diameter in the case of signed permutations evolving by transpositions and reversals. Full proofs of all theorems can be found on the technical report [7].

2 Definitions

Given a permutations π, we want to compute a shortest series of reversals and transpositions that transforms π into $\iota_n = (+1\ +2\ \ldots +n)$, that is, we want to find $\varrho_1, \varrho_2, \ldots, \varrho_k$, where ϱ_i is a reversal or a transposition, such that $\varrho_k \cdot \varrho_{k-1} \cdot \ldots \cdot \varrho_2 \cdot \varrho_1 \cdot \pi = \iota_n$ and k is minimum. We call k the reversal and transposition distance between π and ι_n and denote it by $d(\pi)$. Finding this minimum length series is also called sorting π. In the following an operation can be a reversal or a transposition.

A powerful tool for studying the reversal and transposition distance is the *Breakpoint Graph* of two permutations [1, 4, 5]. The diagram has exactly $n+1$ reality edges and the same number of desire edges. The idea is that reality edges indicate the situation as it is now, and desire edges indicate the situation sought. When reality equals desire in all edges, we have $\pi = \iota_n$ and $d(\pi) = 0$. Therefore, our goal is to apply reversals and transpositions so that reality becomes desire.

Note that the Breakpoint Graph of ι_n is the only one having $n+1$ cycles. So, the sequence of reversals and transpositions transforming π into ι_n must take the number of cycles from $c(\pi)$ to $n+1$. It is known that transpositions can change the number of cycles by at most two [1], and the reversals can change it by at most one [5].

For a permutation π, and an operation or series of permutations ϱ, denote by $\Delta c(\pi, \varrho)$ the difference $c(\varrho \cdot \pi) - c(\pi)$. Therefore, $\Delta c(\pi, \varrho) \in \{-2, -1, 0, 1, 2\}$. In the rest of the note we concentrate on sorting $\pi_n = (-1\ -2\ \ldots\ -(n-1)\ -n)$. Note that $c(\pi_n) = 1$.

3 Main Results

Theorem 1 *We have $d(\pi_n) \leq \lfloor \frac{n}{2} \rfloor + 2$ for $n \geq 3$.*

Proof: First we reverse the entire permutation π_n, obtaining

$$\pi'_n = (+n\ +(n-1)\ \ldots\ +2\ +1),$$

After that we recall the result from [2, 6], proving that the transposition distance $d_t(\pi'_n)$ is $\lfloor \frac{n}{2} \rfloor + 1$, for $n > 2$. The total number of operations is then $\lfloor \frac{n}{2} \rfloor + 2$, which is an upper bound on the distance $d(\pi_n)$ for $n \geq 3$. □

Our strategy is to show that this upper bound is also a lower bound. It is based on an analysis of the first two possible moves, reaching the following conclusion:

Theorem 2 *If $\varrho_k \cdot \varrho_{k-1} \cdot \ldots \cdot \varrho_2 \cdot \varrho_1 \cdot \pi_n = \iota_n$, for $n \geq 3$, then $\Delta c(\pi_n, \varrho_k \cdot \varrho_{k-1} \cdot \ldots \cdot \varrho_2 \cdot \varrho_1) \leq 2k - 3$.*

The proof appears in the technical report [7]. It follows from Theorem 2 that, if $\varrho_k \cdot \varrho_{k-1} \cdot \ldots \cdot \varrho_2 \cdot \varrho_1 \cdot \pi_n = \iota_n$ is an optimal sorting then $k = d(\pi_n)$ and:

$$n \leq 2d(\pi_n) - 3$$

or:

$$d(\pi_n) \geq \frac{n+3}{2}$$

or, after rounding up both sides,

$$d(\pi_n) \geq \left\lceil \frac{n+3}{2} \right\rceil = \left\lfloor \frac{n}{2} \right\rfloor + 2$$

From Theorems 1 and 2, follows the result:

Theorem 3 *We have* $d(\pi_n) = \lfloor \frac{n}{2} \rfloor + 2$ *for* $n \geq 3$*.*

4 Future Work

Apart from the open question about the value of the reversal and transposition diameter, two other issues merit further research. An interesting point to be studied is the diameter of signed permutations under reversals, transpositions, and transversals. A transversal acts by moving a block of genes to another place on the permutation, but with the genes reversed. This operation is biologically as natural as the transposition.

Another line of study is to consider different weights for transpositions and reversals. With equal weights, as was the case here, the minimum path consists predominantly of transpositions. It would be interesting to use weights suggested by what has been observed in practice. Apparently, transpositions should weigh about twice as much as reversals. Eriksen [3] has developed an $(1+\epsilon)$-approximation algorithm for the problem with this cost function.

Acknowledgments

This work was partially funded by Brazilian agencies CAPES and FAPESP.

References

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

[2] D. A. Christie. *Genome Rearrangement Problems*. PhD thesis, Glasgow University, 1998.

[3] N. Eriksen. (1+eps)-approximation of sorting by reversals and transpositions. *Theoretical Computer Science*, 2002. To appear.

[4] S. Hannenhalli and P. Pevzner. To cut... or not to cut (applications of comparative physical maps in molecular evolution). In *6th ACM-SIAM Sumposium on Discrete Algorithms*, 1996.

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

[6] J. Meidanis, M. E. Walter, and Z. Dias. Transposition distance between a permutation and its reverse. In R. Baeza-Yates, editor, *Proceedings of the 4th South American Workshop on String Processing (WSP'97)*, pages 70–79, Valparaiso, Chile, 1997. Carleton University Press.

[7] J. Meidanis, M. E. M. T. Walter, and Z. Dias. A lower bound on the reversal and transposition diameter. Technical Report IC-00-16, IC - Unicamp, October 2000. www.ic.unicamp.br/ic-tr/.

Operations	Diameter
Reversals (unsigned)	$D = n - 1$
Reversals (signed)	$D = n + 1$
Transpositions	$\lfloor n/2 \rfloor + 1 \leq D \leq \lfloor 3n/4 \rfloor$
Reversals, transpositions	$\lfloor n/2 \rfloor + 2 \leq D \leq \lfloor 3n/4 \rfloor$

Table 1: Results known about the diameter of permutation groups under genome rearrangement operations.