Discrete Applied Mathematics 235 (2018) 59–80

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

On the average number of reversals needed to sort signed permutations

Thaynara Arielly de Lima[a,*], Mauricio Ayala-Rincón[b,*]

[a] Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus Samambaia, 74690-900, Goiânia, GO, Brazil
[b] Departamentos de Matemática e Computação, Universidade de Brasília, Campus Darcy Ribeiro, Edifício CIC/EST 70910-900, Brasília, DF, Brazil

ARTICLE INFO

Article history:
Received 4 June 2015
Received in revised form 12 September 2017
Accepted 23 September 2017
Available online 1 November 2017

Keywords:
Combinatorics of permutations
Sorting signed permutations by reversal
Average number of reversals

ABSTRACT

It is well-known that signed permutations can be sorted in polynomial time, while sorting unsigned ones is an $\mathcal{NP}$-hard problem.

Although sorting signed permutations by reversals is an "easy" computational problem, the study of the average reversal distance is justified because it gives rise to interesting combinatorial questions and also because in applications related with genome similarity analysis, exact solutions for this problem have been used to design approximate algorithms for its unsigned $\mathcal{NP}$-hard version. Thus, the average number of reversals needed to sort signed permutations represents a good control mechanism for the quality of approximate solutions for the unsigned case. This paper analyzes the average number of reversals needed to sort signed permutations by exploring combinatorial properties of structures related with signed permutations, such as breakpoint graphs, and provides a recurrence formula for this average number.

The recurrence formula is built based on the analysis of the probability that two nodes of the breakpoint graph belong to the same alternating cycle among all the breakpoint graphs related with permutations of length n. Through this analysis it is possible to compute the average reversal distance for signed variations of the identity permutation. Also, lower and upper bounds of the average reversal distance for signed permutations are provided. Additionally, based on computational data, it is shown how these bounds can be used in order to propose concrete upper and lower bounds.

1. Introduction

A permutation of length $n \in \mathbb{N}$ can be seen as a bijective function in the set $[n] := \{1, \ldots, n\}$. The symmetric group, denoted as usual as S_n, is the well-known structure of all permutations of length n with the operator function composition. To obtain a signed permutation, a sign in $\{+, -\}$ is attributed to each image. Thus, a signed permutation on $[\pm n] = \{\pm 1, \ldots, \pm n\}$ can be represented as a string $\pi = \pi(1) \ldots \pi(n)$, where $\pi(k) \in [\pm n]$ and $|\pi(i)| \neq |\pi(j)|$ if $i \neq j$.

For $i \in \mathbb{N}$, the unary operator "$-$" switches the sign: $-(+i) = -i$, $-(-i) = +i$. Signed permutations are used to represent the linear structure of genes inside genomes, in which genes appear with different orientations. A reversal ρ_{ij}, where $1 \leq i \leq j \leq n$, is an operation that for each π reverses the positions of each x from $\pi(i)$ until $\pi(j)$ and switches its sign: $\rho_{i,j}\pi = \pi(1) \ldots - \pi(j) \ldots - \pi(i) \ldots \pi(n)$. The genome rearrangement problem restricted to reversals consists in finding the minimum number of reversals needed to transform a permutation π into the identity permutation, that is,

* Corresponding authors.
E-mail addresses: thaynaradelima@ufg.br (T.A. de Lima), ayala@unb.br (M. Ayala-Rincón).

https://doi.org/10.1016/j.dam.2017.09.014
0166-218X/

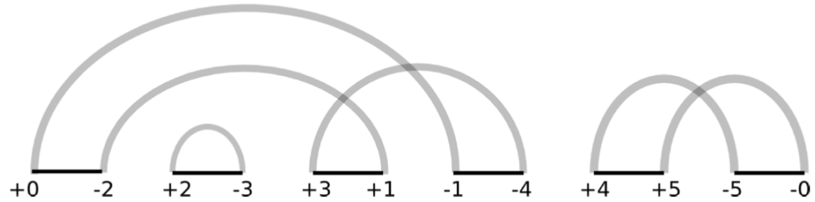

Fig. 1. $G(\pi)$ for $\pi = +2\ +3\ -1\ +4\ -5$.

$id = +1\ +2 \ldots + n$. The minimum number of reversals necessary to sort π is called the reversal distance. It is well-known that sorting signed permutations by reversals is polynomial [9], but in contrast sorting unsigned permutations by reversals is an $\mathcal{NP}$-hard problem [5]. Thus, the development of approximate solutions for the latter problem is of great interest and some approaches apply solutions to the former problem in order to build approximate solutions for the latter one [1,13] and [12]. A reasonable control mechanism for estimating the quality of approximate solutions is the average number of reversals needed to sort permutations of length n, both for the unsigned and the signed cases.

Here we focus on the signed case and our main goal is to obtain an expression for the average number of reversals. We reach this goal by analyzing combinatorial properties of the breakpoint graph associated to each permutation. Breakpoint graphs are essential for studying the reversal distance [2]. The relation between the number of alternating cycles and black edges in the breakpoint graph of a permutation has been proved crucial for the combinatorial analysis and even for the development of algorithms for the reversal and other distance problems. For our development, the key property is the probability that two nodes are in the same alternating cycle of the breakpoint graph.

In addition to reversals other operations are of practical and theoretical interest. For example, another interesting operation is block interchange. A block interchange is an operation that interchanges two blocks without changing the order of entries within each block. The two blocks do not need to be adjacent. Sorting unsigned permutations by block interchange is known to be polynomial [6]. Miklós Bóna showed that the average number of block interchanges needed to sort permutations of length n is closer to the difference between $\frac{n}{2}$ and a logarithmic factor [3]. Based on Bóna's strategy, in this work we propose a recurrence relation for the average number of reversals needed to sort signed permutations.

Observation 1.1. *We start with the observation that the set of signed permutations of length n is a group, indeed, a subgroup of S_{2n}. Consider the function*

$$\begin{array}{lll} \varphi \ : \ \{\pi \mid \pi \text{ is a signed permutation on } [\pm n]\} & \to & S_{2n} \\ \text{such that } \pi = \pi(1)\ldots\pi(n) & \mapsto & \pi' = \pi(1)'\pi(2)'\ldots\pi(2n-1)'\pi(2n)' \end{array}$$

where, for $i = 1, \ldots, n$,

(i) $\pi(2i-1)' = 2\pi(i) - 1$ *and* $\pi(2i)' = 2\pi(i)$, *if* $\pi(i) \in \mathbb{Z}_+$;
(ii) $\pi(2i-1)' = -(2\pi(i))$ *and* $\pi(2i)' = -(2\pi(i)) - 1$, *if* $\pi(i) \in \mathbb{Z}_-$.

Notice that the image of φ is a subgroup of S_{2n}. Also, since φ is an injective function, we can identify the set of all signed permutations on $[\pm n]$ with this subgroup.

In the sequel, the group of signed permutations over $[\pm n]$ will be denoted as $S_n^{\pm}$.

For defining the breakpoint graph of a signed permutation $\pi = \pi(1)\cdots\pi(n)$, first we extend π by adding $\pi(0) = +0$ and $\pi(n+1) = -0$[1] and associate to each $\pi(i)$, $1 \le i \le n$, the ordered pair $-\pi(i)\ \pi(i)$.

Definition 1.1 (*Breakpoint Graph*)**.** The *breakpoint graph* of a given $\pi \in S_n^{\pm}$, is a bi-colored graph $G = \langle V, E\rangle$, denoted as $G(\pi)$, such that the set of vertices is given as $V = [\pm n] \cup \{\pm 0\}$ and the set of edges E as:

(i) there is a gray edge between vertices with labels $+i$ and $-(i+1)$, $0 \le i < n$ and $+n$ and -0;
(ii) there is a black edge between vertices with labels $\pi(i)$ and $-\pi(i+1)$, $0 \le i < n$ and $\pi(n)$ and $\pi(n+1)$.

A simple example is depicted in Fig. 1.

Notice that the breakpoint graph of a signed permutation can be decomposed in alternating cycles, where a cycle is called alternating if the colors of every two consecutive edges are distinct.

Denote by $d(\pi)$, $b(\pi)$ and $c(\pi)$ the reversal distance, the number of black edges and the number of alternating cycles in the breakpoint graph of π, respectively.

[1] The choice of ± 0 should not create confusion: syntactically $+0 \neq -0$. Also, it is possible to use $\pm(n+1)$ instead ± 0, but for brevity, we prefer the use of ± 0.