NP-completeness of sorting unsigned permutations by reversals

Thaynara Arielly de Lima and Mauricio Ayala-Rincón
Departamento de Matemática, Universidade de Brasília

September 17, 2013

Abstract

Since it is well-known that the problem of sorting unsigned permutations by bounded reversals is $\mathcal{NP}$-hard but no proof of the fact that this problem is in $\mathcal{NP}$ is available in the literature, here a standard proof of this fact is given.

MIN-SBP is in $\mathcal{NP}$

It is well-known that the problem of genome rearrangement by reversals corresponds to sorting permutations by bounded reversals in the symmetric group S_n, denoted as *MIN-SBR*. This problem consists in finding the minimum number of reversals needed to transform a permutation into the identity permutation. Almost twenty years ago A. Caprara showed that *MIN-SBR* is an $\mathcal{NP}$-hard problem ([2]), but to the best of our knowledge there is no explicit proof in the literature of the fact that the associated decision problem belongs to $\mathcal{NP}$, despite the fact that this is mentioned in several papers. Here, we present a standard proof of this fact following standard techniques as those presented in textbooks on complexity and combinatorics as [1].

Optimization Problems

In general, an *Optimization Problem* is the problem of finding the *best* solution from all feasible solutions.

Definition 1 ([1]). *An optimization problem* $\mathfrak{P}$ *is characterized by the quadruple of objects* $\langle I_{\mathfrak{P}}, SOL_{\mathfrak{P}}, m_{\mathfrak{P}}, goal_{\mathfrak{P}}\rangle$*, where:*

1. *$I_{\mathfrak{P}}$ is the set of instances of $\mathfrak{P}$;*

2. *$SOL_{\mathfrak{P}}$ is a function that associates to any input instance $x \in I_{\mathfrak{P}}$ the set of feasible solutions of x;*

3. *$m_{\mathfrak{P}}$ is a measure function that associates to any pair (x, y) such that $x \in I_{\mathfrak{P}}$ and $y \in SOL_{\mathfrak{P}}$ a positive number called the value of the feasible solution y;*

4. *$goal_{\mathfrak{P}} \in \{MIN, MAX\}$ specifies whether $\mathfrak{P}$ is a maximization or a minimization problem.*

Example 1. *Given a permutation $\pi \in S_n$,* MIN-SBR *consists in finding the minimum number of reversals needed to transform π into the identity permutation. Formally, this problem is defined as follows:*

1. *$I_{MIN-SBR} = \{\pi \mid \pi \in S_n\}$;*

2. *$SOL_{MIN-SBR}(\pi) = \{\rho = \rho_1, \rho_2, \ldots, \rho_k \mid k \in \mathbb{N}, \rho_i$ reversal in S_n for $1 \leq i \leq k$ and $\pi\rho_1 \ldots \rho_k = id\}$;*

3. *$m_{MIN-SBR}(\pi, \rho) =\mid \rho \mid$;*

4. *$goal_{MIN-SBR} = MIN$.*

One can note that any optimization problem $\mathfrak{P}$ has an associated decision problem $\mathfrak{P}_D$. So, given an optimization problem $\mathfrak{P}$ and an instance x, denote by :

1. $SOL^*_{\mathfrak{P}}(x)$ the set of optimal solutions of x relative to goal (MIN or MAX);

2. for every $y^*(x) \in SOL^*_{\mathfrak{P}}(x)$, $m_{\mathfrak{P}}(x, y^*(x)) = goal_{\mathfrak{P}}\{v \mid v = m_{\mathfrak{P}}(x, z)$ and $z \in SOL_{\mathfrak{P}}(x)\}$;

3. $m^*_{\mathfrak{P}}(x)$ the value of any optimal solution $y^*(x)$ of x.

The decision problem $\mathfrak{P}_D$ associated to an optimization problem $\mathfrak{P}$ is formally defined as below.

Definition 2 (Decision Problem ($\mathfrak{P}_D$) associated to $\mathfrak{P}$ [1])**.** *Given an optimization problem $\mathfrak{P}$ and an instance $x \in I_{\mathfrak{P}}$ and $k \in \mathbb{N}$, decide whether $m^*_{\mathfrak{P}}(x) \geq k$ (if $goal_{\mathfrak{P}} = MAX$) or whether $m^*_{\mathfrak{P}}(x) \leq k$ (if $goal_{\mathfrak{P}} = MIN$).*

By analogy with the class $\mathcal{NP}$ of decision problems, one can define the class $\mathcal{NPO}$ of optimization problems.

Definition 3 (The class $\mathcal{NPO}$ of decision problems [1])**.** *An optimization problem $\mathfrak{P} = \langle I_{\mathfrak{P}}, SOL_{\mathfrak{P}}, m_{\mathfrak{P}}, goal_{\mathfrak{P}} \rangle$ belongs to the class $\mathcal{NPO}$ iff*

1. *the set of instances $I_{\mathfrak{P}}$ is recognizable in polynomial time;*
2. *there exists a polynomial q such that, given an instance $x \in I_{\mathfrak{P}}$, for any $y \in SOL_{\mathfrak{P}}(x)$, $|y| \leq q(|x|)$ and, besides, for any y such that $|y| \leq q(|x|)$, it is decidable in polynomial time whether $y \in SOL_{\mathfrak{P}}(x)$;*
3. *the measure function $m_{\mathfrak{P}}$ is computable in polynomial time.*

It is well-known that given any permutation $\pi \in S_n$, the minimum number of reversals needed to sort π is at most $n-1$. So, in the following, for any $\pi \in S_n$, it will be required that $|\rho| \leq n-1$ for every $\rho \in SOL_{MIN-SBR}(\pi)$.

Example 2. *MIN-SBR belongs to $\mathcal{NPO}$. Indeed,*

1. *given a permutation π, one can recognize in linear time whether $\pi \in S_n$ (e.g., check wheter π is to a bijective function on $\{1, \ldots, n\}$);*
2. *consider the polynomial $q(n) = n^2$. If $\rho \in SOL_{\mathfrak{P}}(\pi)$ then $|\rho| \leq n-1 < q(n)$. Moreover, for any sequence $\rho = \rho_1, \ldots, \rho_k, 1 \leq k < n$, testing if each ρ_i is a reversion takes linear time and if ρ is a feasible solution requires testing if $\pi\rho_1 \ldots \rho_k = id$, which can be clearly performed bounded by quadratic time (e.g., explicitly building $\pi\rho_1$ takes linear time at most, and so on);*
3. *finally, given a feasible solution ρ, the measure function (size of ρ), is trivially computable in linear time.*

The following theorem is the key result applied to conclude that *MIN-SBR$_D$* belongs to $\mathcal{NP}$.

Theorem 1 ([1])**.** *For any optimization problem $\mathfrak{P}$ in $\mathcal{NPO}$, the corresponding decision problem $\mathfrak{P}_D$ belongs to $\mathcal{NP}$.*

Proof. Assume q is the polynomial for $\mathfrak{P}_D$ as in Def. 3. Given an instance $x \in I_{\mathfrak{P}}$ and a natural k, we can solve $\mathfrak{P}_D$ by performing the following non-deterministic algorithm:

1. guess any string y such that $|y| \leq q(|x|)$ in time $q(|x|)$;
2. check if y belongs to $SOL_{\mathfrak{P}}(x)$ in polynomial time;
3. if the previous test is positive, compute $m_{\mathfrak{P}}(x, y)$, in polynomial time;

4. answer YES if $m_{\mathfrak{P}}(x, y) \leq k$;

5. answer NO otherwise.

$\square$

Corollary 1. *MIN-SBR$_D$ belongs to $\mathcal{NP}$.*

Proof. The statement follows from Theorem 1 and Example 2. $\square$

References

[1] G. Ausiello, M. Protasi, A. Marchetti-Spaccamela, G. Gambosi, P. Crescenzi, and V. Kann. *Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1st edition, 1999.

[2] A. Caprara. Sorting by reversals is difficult. In *RECOMB*, pages 75–83, 1997.