Blind selection sort analysis

Christopher He

October 20, 2023

1 Exact running time

Let a be a permutation of $1, 2, ..., N$. Let $inv = \{(i, j) \mid 1 \leq i < j \leq N, a_i > a_j\}$ be the set of inversions of a. Let $ginv = \{(i, j) \in inv \mid a_j > j\}$ be the set of good inversions. Let $inv_k = \{(i, j) \in inv \mid i = k\}$ denote inversions where the left is k, and similarly define $ginv_k$.

First we proved that if $(i, j) \in inv$ and a_j starts before its correct position (j), then so does a_i.

Lemma 1. *If $(i, j) \in ginv$ then $a_i > i$.*

Proof. If $(i, j) \in ginv$ then $a_i > a_j > j > i$, so $a_i > i$. □

We then proved that if an element starts before its position i.e $a_i > i$, the distance $a_i - i$ is at most $|inv_i|$.

Lemma 2. *If $a_i > i$ then $a_i - i \leq |inv_i|$.*

Proof. There are $a_i - 1$ elements smaller than a_i, but only $i - 1$ available positions to its left. That means that at least $a_i - 1 - (i - 1) = a_i - i$ are on its right, forming that many inversions. So $|inv_i| \geq a_i - i$. □

Conjecture 1. *The number of swaps performed by blind selection sort is*

$$\sum_{i<a_i} a_i - i + \sum_{i>a_i} |inv_i| - \sum_i |ginv_i| \tag{1}$$

Proof. In general, the number of swaps involving a_i (where a_i is the left element) is equal to $|inv_i|$. For example, let $a = [6, 5, 4, 3, 2, 1]$. The number of swaps involving 3 is 2. When the algorithm searches for 1, it will generate one swap for 3. Another one will be generated when 2 is being searched.

But notice that a swap with a_i has a net effect of shifting a_i to the right. If an element starts before its position $(a_i > i)$, then by Lemma 2, the element shifts at most $a_i - i$ times before it is frozen.

Lastly, the subtracting the term $\sum_i |ginv_i|$ is to account for that fact that elements may be frozen "accidentally", saving one swap for every unfrozen element before it. Note that by Lemma 1, $\sum_i |ginv_i| = \sum_{i<a_i} |ginv_i|$. □

Conjecture 2. *The number of swaps performed by blind selection sort is*

$$\sum_i inv_i - 2\sum_i |ginv_i| \tag{2}$$

Proof. From Conjecture 1 we know the number of swaps is

$$\sum_{i<a_i} a_i - i + \sum_{i>a_i} |inv_i| - \sum_i |ginv_i| \tag{3}$$

If

$$\sum_{i<a_i} a_i - i = \sum_{i\leq a_i} inv_i - \sum_i ginv_i \tag{4}$$

then this statement is proven.

$$\begin{aligned} &\sum_{i<a_i} a_i - i + \sum_{i>a_i} |inv_i| - \sum_i |ginv_i| \\ &= \sum_{i\leq a_i} inv_i - \sum_i |ginv_i| + \sum_{i>a_i} |inv_i| - \sum_i |ginv_i| \\ &= \sum_i inv_i - 2\sum_i |ginv_i| \end{aligned} \tag{5}$$

□

2 Structure of the worst case input

Theorem 1. *The worst case permutation of* $1, 2, ..., N$ *for blind selection sort starts with an increasing, consecutive sequence*