

Blind selection sort analysis

Christopher He

October 24, 2023

1 Exact running time

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

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

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

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

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

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

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 $|inv_i(a)| \geq a_i - i$. \square

Conjecture 1. *The number of swaps $s(a)$ performed by blind selection sort on permutation a is*

$$s(a) = \sum_{i < a_i} a_i - i + \sum_{i > a_i} |inv_i(a)| - \sum_i |ginv_i(a)| \quad (1)$$

Proof. In general, the number of swaps involving a_i (where a_i is the left element) is equal to $|inv_i(a)|$. 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, we subtract $\sum_i |ginv_i(a)|$ 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, $\sum_i |ginv_i(a)| = \sum_{i < a_i} |ginv_i(a)|$. \square

Conjecture 2. *The number of swaps $s(a)$ performed by blind selection sort on permutation a is*

$$s(a) = \sum_i \text{inv}_i(a) - 2 \sum_i |\text{ginv}_i(a)| \quad (2)$$

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

$$s(a) = \sum_{i < a_i} a_i - i + \sum_{i > a_i} |\text{inv}_i(a)| - \sum_i |\text{ginv}_i(a)| \quad (3)$$

If the following equality is proven (TODO: prove this by proving that good inversions balance the distance)

$$\sum_{i < a_i} a_i - i = \sum_{i \leq a_i} \text{inv}_i(a) - \sum_i \text{ginv}_i(a) \quad (4)$$

then this statement is proven.

$$\begin{aligned} s(a) &= \sum_{i < a_i} a_i - i + \sum_{i > a_i} |\text{inv}_i(a)| - \sum_i |\text{ginv}_i(a)| \\ &= \sum_{i \leq a_i} \text{inv}_i(a) - \sum_i |\text{ginv}_i(a)| + \sum_{i > a_i} |\text{inv}_i(a)| - \sum_i |\text{ginv}_i(a)| \\ &= \sum_i \text{inv}_i(a) - 2 \sum_i |\text{ginv}_i(a)| \end{aligned} \quad (5)$$

□

2 Structure of the worst case input

Theorem 1. *If a is the worst case permutation of $1, 2, \dots, N$ for blind selection sort, then $\text{ginv}_i(a) = \emptyset$.*

Proof. Assume $\text{ginv}_i(a) \neq \emptyset$, then pick $(i, j) \in \text{ginv}_i(a)$ such that $\forall k, i < k < j, a_k < k$. So we start with $a = [a_1, \dots, a_i, \dots, a_j, \dots, a_N]$. We can construct a new permutation by swapping a_i and a_j i.e $a' = [a_1, \dots, a_j, \dots, a_i, \dots, a_N]$. This effectively removes the good inversion (i, j) . Furthermore, let $k, i < k < j$. The number of inversions involving k remains the same after the swap. TODO: finish. By Conjecture 2, $s(a') = s(a) + 1$, thus contradicting that a is the worst case. □

This gives us the immediate result that in the worst input a , if an element starts before its position, all elements before it are smaller (otherwise $\text{ginv}(a) \neq \emptyset$).

Corollary 1. *In the worst case permutation of $1, 2, \dots, N$ for blind selection sort, if $a_j > j$ then $\forall i < j, a_i < a_j$.*

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

Lemma 3. *If a is the worst input for blind selection sort and $a_j > j$, then $\forall i < j, a_i > i$.*

Proof. We will prove the equivalent statement: if $j > 1, a_j > j$, then $a_{j-1} > j-1$. Assume for the sake of contradiction that $a_{j-1} < j-1$. By Corollary 1, we know $a_{j-1} < a_j$. So switch the position of a_{j-1} , and a_j . We created an inversion, and we know this is not a good inversion since $a_{j-1} < j-1 < j$ (in other words, after moving a_{j-1} to the right, it is still after its correct position.) So by Conjecture 2, we made an even worse input. \square

Definition 1. *If $a_k = x$ then $idx(x) = k$.*

Lemma 4. *If a is the worst input for blind selection sort and $\exists j > 1$ such that $a_j > j$, then $a_{j-1} = j-1$.*

Proof. Assume for the sake of contradiction that $a_{j-1} \neq j-1$ and let $k = idx(j-1)$. We know that $k > j-1$, otherwise $(k, j-1) \in ginv(a)$. Switch a_{j-1} with a_k . \square

Proof. Since $a_N < N$ (otherwise it is frozen), then by Lemma 2 and 3, the statement is proven. \square