Correctness of Selection Sort

Wolfgang Jeltsch

6 February 2017

1 Introduction

This document contains a correctness proof for the selection sort implementation presented in the lecture. Note that this implementation uses an efficient approach based on swapping elements and therefore does not exactly correspond to the traditional presentation of selection sort.

2 Preservation of Elements

The `Selection_Sort` procedure changes the `Data` array only by calling the `Swap` procedure. It is trivial to see that `Swap` neither adds nor removes elements.

3 Sortedness of the Result

To proof that the `Data` array is finally sorted, we use loop invariants.

3.1 Outer Loop

We define the outer loop invariant $\mathcal{O}$ as follows:

$$\mathcal{O}(F, L, a, d) := \forall i \in [F, d) \,.\, \forall j \in (i, L] \,.\, a(i) \leq a(j) \tag{1}$$

Let F and L be `Data'First` and `Data'Last`. At the beginning of each loop iteration, the condition $\mathcal{O}(F, L, a, d)$ is supposed to hold, where a and d are the then current values of `Data` and `Destination_Index`. At the end of the loop, the condition $\mathcal{O}(F, L, a', L + 1)$ is supposed to hold, where a' is the then current value of `Data`.

3.1.1 Basis

Let a be the value of `Data` at the beginning of the loop. We have to prove that $\mathcal{O}(F, L, a, F)$ is true. We do this as follows:

$$\begin{aligned}
\top &\to \forall i \in \emptyset \,.\, \forall j \in (i, L] \,.\, a(i) \le a(j) \\
&\to \forall i \in [F, F) \,.\, \forall j \in (i, L] \,.\, a(i) \le a(j) \\
&\to \mathcal{O}(F, L, a, F) && \text{(using (1))}
\end{aligned}$$

3.1.2 Step

Let a and d be the values of `Data` and `Destination_Index` at the beginning of the loop iteration, and let a' be the value of `Data` at the end of the loop iteration. We have to prove that $\mathcal{O}(F, L, a, d)$ implies $\mathcal{O}(F, L, a', d + 1)$.

For our proof, we use the following propositions:

$$\forall i \in [F, d) \,.\, a(i) = a'(i) \tag{2}$$

$$\{\, a(i) \mid i \in [d, L] \,\} = \{\, a'(i) \mid i \in [d, L] \,\} \tag{3}$$

$$\forall i \in (d, L] \,.\, a'(d) \le a'(i) \tag{4}$$

The proofs of (2) and (3) are trivial and therefore not shown here; (4) is proved in Subsection 3.2.

We conduct the actual step as follows:

$$\begin{aligned}
\mathcal{O}(F, L, a, d) &\to \forall i \in [F, d) \,.\, \forall j \in (i, L] \,.\, a(i) \le a(j) && \text{(using (1))} \\
&\to \forall i \in [F, d) \,.\, \forall j \in (i, L] \,.\, a'(i) \le a(j) && \text{(using (2))} \\
&\to \forall i \in [F, d) \,.\, \forall j \in (i, L] \,.\, a'(i) \le a'(j) && \text{(using (2) and (3))} \\
&\to \big(\forall i \in [F, d) \,.\, \forall j \in (i, L] \,.\, a'(i) \le a'(j)\big) \land \forall j \in (d, L] \,.\, a'(d) \le a'(j) \\
& && \text{(using (4))} \\
&\to \forall i \in [F, d + 1) \,.\, \forall j \in (i, L] \,.\, a'(i) \le a'(j) \\
&\to \mathcal{O}(F, L, a', d + 1) && \text{(using (1))}
\end{aligned}$$

3.1.3 Conclusion

Let a' be the value of `Data` at the end of the loop. We have to prove that $\mathcal{O}(F, L, a', L + 1)$ implies that a' is sorted. We do this as follows:

$$\begin{aligned}
\mathcal{O}(F, L, a', L + 1) &\to \forall i \in [F, L + 1) \,.\, \forall j \in (i, L] \,.\, a'(i) \le a'(j) && \text{(using (1))} \\
&\to \forall i \in [F, L] \,.\, \forall j \in (i, L] \,.\, a'(i) \le a'(j)
\end{aligned}$$

3.2 Inner Loop

We define the inner loop invariant $\mathcal{I}$ as follows:

$$\mathcal{I}(a, d, s) := \forall i \in (d, s) \,.\, a(d) \le a(i) \tag{5}$$

At the beginning of each loop iteration, the condition $\mathscr{I}(a,d,s)$ is supposed to hold, where a, d, and s are the then current values of `Data`, `Destination_Index`, and `Source_Index`. At the end of the loop, the condition $\mathscr{I}(a',d,L+1)$ is supposed to hold, where a' and d are the then current values of `Data` and `Destination_Index`.

3.2.1 Basis

Let a and d be the values of `Data` and `Destination_Index` at the beginning of the loop. We have to prove that $\mathscr{I}(a,d,d+1)$ is true. We do this as follows:

$$
\begin{aligned}
\top &\rightarrow \forall i \in \emptyset \,.\, a(d) \leq a(i) \\
&\rightarrow \forall i \in (d,d+1) \,.\, a(d) \leq a(i) \\
&\rightarrow \mathscr{I}(a,d,d+1) && \text{(using (5))}
\end{aligned}
$$

3.2.2 Step

Let a, d, and s be the values of `Data`, `Destination_Index`, and `Source_Index` at the beginning of the loop iteration, and let a' be the value of `Data` at the end of the loop iteration. We have to prove that $\mathscr{I}(a,d,s)$ implies $\mathscr{I}(a',d,s+1)$.

Case 1: $a(s) < a(d)$. The elements at s and d are swapped. Therefore, the following properties hold:

$$\forall i \in (d,s) \,.\, a(i) = a'(i) \tag{6}$$

$$a(d) = a'(s) \tag{7}$$

$$a(s) = a'(d) \tag{8}$$

We reason as follows:

$$
\begin{aligned}
\mathscr{I}(a,d,s) &\rightarrow \forall i \in (d,s) \,.\, a(d) \leq a(i) && \text{(using (5))} \\
&\rightarrow \forall i \in (d,s) \,.\, a(d) \leq a'(i) && \text{(using (6))} \\
&\rightarrow \big(\forall i \in (d,s) \,.\, a(d) \leq a'(i)\big) \land a(d) = a'(s) && \text{(using (7))} \\
&\rightarrow \big(\forall i \in (d,s) \,.\, a(d) \leq a'(i)\big) \land a(d) \leq a'(s) \\
&\rightarrow \forall i \in (d,s+1) \,.\, a(d) \leq a'(i) \\
&\rightarrow \forall i \in (d,s+1) \,.\, a(s) \leq a'(i) && \text{(using the case condition)} \\
&\rightarrow \forall i \in (d,s+1) \,.\, a'(d) \leq a'(i) && \text{(using (8))} \\
&\rightarrow \mathscr{I}(a'd,s+1) && \text{(using (5))}
\end{aligned}
$$

Case 2: $a(s) \geq a(d)$. No swapping is done. Therefore, the following property holds:

$$\forall i \in [d,s] \,.\, a(i) = a'(i) \tag{9}$$

We reason as follows:

$$\begin{aligned}
\mathscr{I}(a,d,s) &\rightarrow \forall i \in (d,s)\,.\; a(d) \le a(i) && \text{(using (5))}\\
&\rightarrow \big(\forall i \in (d,s)\,.\; a(d) \le a(i)\big) \wedge a(d) \le a(s) && \text{(using the case condition)}\\
&\rightarrow \forall i \in (d,s+1)\,.\; a(d) \le a(i)\\
&\rightarrow \forall i \in (d,s+1)\,.\; a'(d) \le a(i) && \text{(using (9))}\\
&\rightarrow \forall i \in (d,s+1)\,.\; a'(d) \le a'(i) && \text{(using (9))}\\
&\rightarrow \mathscr{I}(a',d,s+1) && \text{(using (5))}
\end{aligned}$$

3.2.3 Conclusion

Let a' and d be the values of `Data` and `Destination_Index` at the end of the loop. We have to prove that $\mathscr{I}(a', d, L+1)$ implies (4). We do this as follows:

$$\begin{aligned}
\mathscr{I}(a',d,L+1) &\rightarrow \forall i \in (d,L+1)\,.\; a'(d) \le a'(i) && \text{(using (5))}\\
&\rightarrow \forall i \in (d,L]\,.\; a'(d) \le a'(i)
\end{aligned}$$