

ter 1

Correctness - Recitation 2

Algorithms correctness is necessary to guarantee that our code computes its goal for every given input. In that recitation, we will examine several algorithms, their running time and memory consumption, and prove they are correct.

1.0.1 (Leading Example.). Consider n numbers $a_1, a_2, \dots, a_n \in \mathbb{R}$. Given set Q of $|Q|$ queries, such each query $q \in Q$ is a tuple $(i, j) \in [n] \times [n]$ that calculates the $\max_{i \leq k \leq j} a_k$.

Correctness And Loop Invariant.

Definition. We will say that an algorithm \mathcal{A} (an ordered set of operations) computes $f : D_1 \rightarrow D_2$ if for every $x \in D_1$ the following equality holds: $f(x) = \mathcal{A}(x)$. When it's obvious what is the goal function f , we will abbreviate and say that \mathcal{A} is correct. Examples of functions f might be: file saving, summing numbers, or posting a message in the forum.

Definition. Loop Invariant is a property that characterizes a loop segment code and satisfy the following conditions:

1. Initialization. The property holds (even) before the first iteration of the loop.

2. Preservation. As long as one performs the loop iterations, the property still holds.

3. (Optional) Termination. Exiting from the loop carrying information.

1.1.1. Before dealing with the hard problem, let us face the naive algorithm to find the maximum of a given array.

The algorithm returns the maximum of $a_1 \dots a_n \in \mathbb{R}^n$

function $\text{max}(a_1 \dots a_n)$

for $i \in [n]$ **do**

$j \leftarrow 1$

while $j \leq [n]$ **and** $a_i \geq a_j$ **do**

$j \leftarrow j + 1$

end

if $j = n + 1$ **then**

return a_i

end

return Δ

Algorithm 1: naive maximum alg.

1. Consider the while loop. The property: "for every $j' < j \leq n + 1 \Rightarrow a_{j'} \leq a_i$ " is a loop invariant that is associated with it.

initialization condition holds, as the at the first iteration $j = 1$ and therefore the property is trivial. Assume by induction, that for every $j < n$, the property holds. Now, we consider the j -th iteration. If back again to line (5), then it means that $(j - 1) < n$ and $a_{j-1} \leq a_i$. Combining the above with the induction hypothesis, we get that $a_i \geq a_{j-1}, a_{j-2}, \dots, a_1$.

Proof. Split into cases, First if the algorithm return result at line (9), then due to the loop invariant, combining the fact that $j = n + 1$, it holds that $a_i \geq a_{j'}$ i.e a_i is the maximum of a_1, \dots, a_n . The second case, in which the algorithm returns Δ at line number (10) contradicts the fact that a_i is the maximum. This is an exercise. the running time is $O(n^2)$ and the space consumption is $O(n)$.

In The Cleverer Alg. Consider now the linear time algorithm:

Find the maximum of $a_1 \dots a_n \in \mathbb{R}^n$

$\max(a_i \dots a_j)$

1

2, $n]$ **do**

$\max(b, a_i)$

Algorithm 2: maximum alg.

Loop Invariant here? "at the i -th iteration, $b = \max\{a_1 \dots a_{i-1}\}$ ". The proof is almost identical to the naive case.

Linear Space Complexity Algorithms.

Naive. Consider the leading example; It's easy to write an algorithm that answers the queries at a total time of a $O(|Q| \cdot n)$ by answering each query individually. Can we achieve a better upper bound?

Find the $\max\{a_i \dots a_j\}$ for each query $(i, j) \in Q$

$\max(a_i \dots a_j)$

$\mathbb{M}^{n \times n}$

$[1, n]$ **do**

$\leftarrow a_i$

$[1, n]$ **do**

$\in [n]$ **do**

if $i + k \leq n$ **then**

$A_{i,i+k} \leftarrow \max(A_{i,i+k-1}, a_{i+k})$

end

Q **do**

$\leftarrow q$

$A_{i,j}$

Algorithm 3: Sub-Array. $O(n^2)$ space alg.

Consider the outer loop at the k -th step. The following is a loop invariant:

$$\text{for every } k' < k, \text{ s.t. } i + k' \leq n \Rightarrow A_{i,i+k'} = \max \{a_i, a_{i+1}, \dots, a_{i+k'}\}$$

Initialization condition trivially holds, assume by induction that $A_{i,i+k-1} = \max \{a_i \dots a_{i+k-1}\}$ at beginning of k iteration. By the fact that $(x, y), z)$ we get that

$$\max \{a_1 \dots a_{i+k-1}, a_{i+k}\} = \max \{\max \{a_1 \dots a_{i+k-1}\}, a_{i+k}\} = \max \{A_{i,i+k-1}, a_{i+k}\}$$

Right term is exactly the value which assigned to $A_{i,i+k}$ in the end of the k -th iteration. Thus in the beginning of $k + 1$ iteration the

Space Solution. Example for $O(n \log n + |Q| \log n)$ time and $O(n \log n)$ space algorithm. Instead of storing the whole matrix we store only the number of rows.

print the $\max \{a_i \dots a_j\}$ for each query $(i, j) \in Q$

on $\max(a_i \dots a_j)$

$A \leftarrow \mathbb{M}^{n \times \log n}$

for $i \in [n]$ do

$A_{i,1} \leftarrow a_i$

end

for $k \in [2, 4, \dots, 2^m, \dots, n]$ do

 for $i \in [n]$ do

 if $i + k \leq n$ then

$A_{i,k} \leftarrow \max \left(A_{i, \frac{k}{2}}, A_{i + \frac{k}{2}, \frac{k}{2}} \right)$

 end

 end

end

for $q \in Q$ do

$i, j \leftarrow q$

 decompose $j - i$ into binary representation $2^{t_1} + 2^{t_2} + \dots + 2^{t_l}$

 print $\max \{A_{i, 2^{t_1}}, A_{i+2^{t_1}, 2^{t_2}}, \dots, A_{i+2^{t_1}+2^{t_2}+\dots+2^{t_{l-1}}, 2^{t_l}}\}$

end

Algorithm 4: Sub-Array. $O(n \log n)$ space alg.