

Q1. On pages 55 and 56 of the textbook there is a proof for the correctness of the program `avg`. The author used the invariant

$$Inv(i, \text{sum}): 0 \leq i < \text{len}(A) \wedge \text{sum} = \sum_{k=0}^{i-1} A[k].$$

As you can see, the invariant is a predicate in two variables: `i`, `sum`. These two variables are used in the program `avg`, but neither is the variable on which the induction proof is based. The author is using simple induction, but it is not very clear what the induction variable is (also the induction predicate itself is ambiguous). We want to make sure you understand what is going on there by having you re-write the proof by yourself in a style similar to the one we use in lectures. Here is the predicate you should be proving:

$$Q(j) : \text{At the beginning of the } j^{\text{th}} \text{ iteration, } \text{sum} = \sum_{k=0}^{i-1} A[k].$$

Remark 1: By the program's design, the variables `i`, `sum` may change with each iteration (in other words, both are functions of j). This is why it might be more appropriate to write

$$Q(j) : \text{sum}_j = \sum_{k=0}^{i_j-1} A[k] \quad \text{or} \quad Q(j) : \text{sum}(j) = \sum_{k=0}^{i(j)-1} A[k].$$

where i_j (or $i(j)$) means the value of program variable `i` at the beginning of the j^{th} iteration (the same with `sum`). That said, we believe that the version above Remark 1 is the best to work with as long as one understands that `i`, `sum` are iteration dependent.

Remark 2: Using j as an index in Remark 1 has a different meaning from that which is intended by the author. The author is using indices to differentiate between the values of `i`, `sum` at the beginning of an (arbitrary) iteration and their values after the iteration is run.

The end goal is to prove $Q(\text{len}(A))$. A proof by induction will show that

$$\forall j \in \{1, \dots, \text{len}(A)\}, Q(j).$$

Your proof must follow the style used in lectures.

Proof. It will be shown using simple induction that $\forall j \in \{1, \dots, \text{len}(A)\}, Q(j)$.

Base Case. Let $j = 1$. At the beginning of the first iteration of the loop, $i = 0$ and $\text{sum} = 0$. Notice that in the expression $\sum_{k=0}^{-1} A[k]$, the lower bound is greater than the upper bound. This results in an empty sum that evaluates to 0. Therefore the base case holds.

Induction Hypothesis. Suppose that for some $j \in \{1, \dots, \text{len}(A) - 1\}$, $Q(j)$ is true.

Induction Step. i increases by 1 at the end of each iteration, so at the beginning of the j^{th} iteration, $i = j - 1 < \text{len}(A)$, which means that the loop body is entered. From the induction hypothesis,

$$\text{sum} = \sum_{k=0}^{j-2} A[k]$$

Running through the loop, on line 9, $A[j - 1]$ is added to sum . After this line is executed, the new value of sum becomes

$$sum = \sum_{k=0}^{j-2} A[k] + A[j - 1] = \sum_{k=0}^{j-1} A[k]$$

Then, line 10 executes and the value of i is incremented and becomes j . After the line is executed, the program returns to the top of loop, signifying the beginning of the $j + 1$ th iteration. sum remains unchanged since the execution of line 9, and is equal to

$$\sum_{k=0}^{j-1} A[k] = \sum_{k=0}^{i-1} A[k]$$

as desired.

By the principle of simple induction, at the beginning of the j th iteration, $Q(j)$ holds for any $j \in \{1, 2, \dots, \text{len}(A)\}$. In particular, $Q(\text{len}(A))$, and the proof is done.

□

Q2. –Following up on the Q1– We mentioned this in class, but it is good to remind you again of the fact that, proving $\forall j \in \{1, \dots, \text{len}(A)\}, Q(j)$ is equivalent to proving $\forall n \in \mathbb{N}, Q'(n)$ where

$$Q'(n) : 0 \leq n < \text{len}(A) \Rightarrow Q(n+1).$$

Explain the equivalence.

Proof. It will be proven that the two statements imply each other, that is,

$$\forall j \in \{1, \dots, \text{len}(A)\}, Q(j) \iff \forall n \in \mathbb{N}, 0 \leq n < \text{len}(A) \implies Q(n+1)$$

To prove the forward direction, suppose that $Q(j)$ holds true for all $j \in \{1, \dots, \text{len}(A)\}$. Let n be a natural number such that $0 \leq n < \text{len}(A)$. It follows that $n \in \{0, 1, \dots, \text{len}(A) - 1\}$, which implies that $n+1 \in \{1, \dots, \text{len}(A)\}$. By the original assumption, $Q(n+1)$ is true.

Conversely, suppose that for all natural n , if $0 \leq n < \text{len}(A)$ then $Q(n+1)$ is true. Let $j \in \{1, \dots, \text{len}(A)\}$. Then $j-1 \in \{0, \dots, \text{len}(A) - 1\}$. It is clear that $j-1$ is a natural number and that $0 \leq j-1 < \text{len}(A)$. By the assumption in the beginning, $Q(j)$ is true.

Thus both statements are equivalent to each other.

□

Q3. Solve questions 6 to 10 on pages 64 to 66 of the textbook. There is a typo in Q8 line 12 (it should be $c = 1$ instead of $c == 1$). Proofs must follow the lecture's format and level of clarity. Only **one** of the five questions will be selected for marking. A serious attempt of all the mentioned questions is a necessary condition for receiving more than 0 points.

Question 6.

1. Give a loop invariant that characterizes the values of a and y .

Proof. Let x be a natural number and i is a positive natural number. Define the loop invariant to be

$$Inv(i) : y = 11 - i \text{ and } a = x + \frac{y(y+1)}{2} - 55$$

It will be proven that at the beginning of the i th iteration of the loop, $Inv(i)$ is true.

Base Case. At the beginning of the first iteration, $a = x$ and $y = 10$. Thus

$$y = 11 - 1 = 11 - i$$

As well,

$$x + \frac{y(y+1)}{2} - 55 = x - \frac{10(11)}{2} - 55 = x - 55 + 55 = x = a$$

The base case has been proven to be true.

Induction Hypothesis. Now suppose that for some positive natural j , $Inv(i)$ holds true at the beginning of the j th iteration. Then the loop invariant states that

$$y = 11 - j \text{ and } a = x + \frac{(11-j)(12-j)}{2} - 55$$

Induction Step. If $a \leq 0$, the loop terminates and never reaches the beginning of the $j+1$ th iteration. Otherwise, on line 6 in the j th iteration, a is subtracted by $y = 11 - j$. Using the induction hypothesis, the value of a is now

$$\begin{aligned} a &= x + \frac{(11-j)(12-j)}{2} - 55 - (11-j) = x + \left(\frac{(11-j)(12-j)}{2} - (11-j) \right) - 55 \\ &= x + \frac{(11-j)(12-j-2)}{2} - 55 = x + \frac{(10-j)(11-j)}{2} - 55 \end{aligned}$$

Finally, on line 7, the value of y is decremented by 1 and becomes $y = 11 - j - 1$. The program then arrives at the beginning of the $j+1$ th iteration. Indeed, it is true that

$$y = 11 - (j+1) \text{ and } a = x + \frac{(10-j)(11-j)}{2} - 55 = x + \frac{y(y+1)}{2} - 55$$

Thus $Inv(i)$ is true at the beginning of the $j+1$ th iteration, verifying the validity of the loop invariant.

□

2. Show that sometimes this code fails to terminate

Proof. Let $x = 56$. It will be shown directly that at the beginning of every iteration, $a > 0$, which implies that the loop can never terminate.

Consider the beginning of an arbitrary j th iteration. From the loop invariant found in part (a), it is true that $y = 11 - j$ and $Inv(i)$, so

$$a = x + \frac{y(y+1)}{2} - 55 = 56 + \frac{(11-j)(12-j)}{2} - 55 = 1 + \frac{(11-j)(12-j)}{2}$$

Note that $\frac{(11-j)(12-j)}{2} < 0$ if and only if $11 < j < 12$, which is impossible because j is a natural number. Thus $\frac{(11-j)(12-j)}{2} \geq 0$ so

$$a = 1 + \frac{(11-j)(12-j)}{2} \geq 1 > 0$$

Therefore, $a > 0$ at the beginning of all iterations, which means that the loop never terminates. □

Question 7. State the pre- and post-conditions that the algorithms must satisfy, then prove that each algorithm is correct.

- (a) *Proof.* Pre-condition: a is a non-zero number and b is a natural number.

Post-condition: Returns the value of a^b .

The correctness of the program will be proved using complete induction on b . For the entirety of the proof, a is fixed to be an arbitrary non-zero number.

Base Case. When $b = 0$, the method returns 1, which is exactly a^b .

Induction Hypothesis. Suppose that the program returns the correct value of a^b for all $b \leq n$, for some $n \in \mathbb{N}$.

Induction Step. Let $b = n + 1$. Then $b \geq 1$, so line 3 is not reached. Consider two possible cases:

If b is even, line 5 and 6 will be executed. A new variable x is initialized to be equal to `exp_rec(a, b / 2)`. Since $\frac{b}{2}$ is natural and less than or equal to n , By the induction hypothesis, the value of x is $a^{\frac{b}{2}}$. The program returns $x * x$, which is equal to $a^{\frac{b}{2}} \cdot a^{\frac{b}{2}} = a^b$ on line 6.

If b is odd, line 8 and 9 will be executed. By a similar argument as the case above, a variable x is initialized to the value $a^{\frac{b-1}{2}}$ and `x * x * a` is returned, which has the value of $a^{\frac{b-1}{2}} \cdot a^{\frac{b-1}{2}} \cdot a = a^b$

Therefore, it can be concluded that the program satisfies the postconditions for all natural b . □

(b) *Proof.* Let a be a non-zero number and b be a natural number.

Define the loop variant to be exp , which is initialized to be the natural number b . Every iteration, a floor division of 2 is applied to it, so it is always natural and decreasing. Thus the program will eventually terminate.

For a positive natural i , define the loop invariant

$$P(i) : \text{mult}^{\text{exp}} \cdot \text{ans} = a^b$$

It will be shown using induction on i that for $i \in \{1, 2, \dots, \lfloor \log_2(b) \rfloor + 2\}$, $P(i)$ is true at the beginning of the i th iteration.

Base Case. Let $i = 1$. At the start of the very first iteration, $\text{mult} = a$, $\text{ans} = 1$, and $\text{exp} = b$, so it follows that

$$\text{mult}^{\text{exp}} \cdot \text{ans} = a^b \cdot 1 = a^b$$

Thus the base case holds.

Induction Hypothesis. Let $j \in \{1, 2, \dots, \lfloor \log_2(b) \rfloor + 1\}$. Suppose that at the beginning of the j th iteration, $P(j)$ is true. That is,

$$\text{mult}^{\text{exp}} \cdot \text{ans} = a^b$$

At this point, exp has been halved $j - 1$ times, so along with the fact that exp is initially equal to b ,

$$\left\lfloor \frac{b}{2^{j-1}} \right\rfloor \geq \left\lfloor \frac{b}{2^{\lfloor \log_2(b) \rfloor}} \right\rfloor \geq \left\lfloor \frac{b}{2^{\log_2(b)}} \right\rfloor = \left\lfloor \frac{b}{b} \right\rfloor = 1$$

Which means that the loop condition holds and the loop body is entered.

Induction Step. Let $\text{mult}_0, \text{exp}_0, \text{ans}_0$ be the value of $\text{mult}, \text{exp}, \text{ans}$ at the beginning of the j th iteration and let $\text{mult}_1, \text{exp}_1, \text{ans}_1$ denote the value of the variables at the beginning of the $j + 1$ th iteration.

Now, two cases will be considered for the analysis of the loop execution.

If exp is odd, then line 7 will run and multiply ans by mult , which implies that $\text{ans}_1 = \text{ans}_0 \cdot \text{mult}_0$.

Line 8 will run and will multiply mult by itself, so $\text{mult}_1 = \text{mult}_0^2$.

As well, since exp is odd, line 9 will result in $\text{exp}_1 = \text{exp}_0 // 2 = \frac{\text{exp}_0 - 1}{2}$

After line 9, the current iteration ends and the $j + 1$ th iteration begins. Using algebraic manipulation and the induction hypothesis, $P(j + 1)$ can be verified to be true:

$$\begin{aligned} \text{mult}_1^{\text{exp}_1} \cdot \text{ans}_1 &= (\text{mult}_0^2)^{\frac{\text{exp}_0 - 1}{2}} \cdot (\text{ans}_0 \cdot \text{mult}_0) = (\text{mult}_0)^{\text{exp}_0 - 1} (\text{mult}_0 \cdot \text{ans}_0) \\ &= \text{mult}_0^{\text{exp}_0} \cdot \text{ans}_0 = a^b \end{aligned}$$

Otherwise if exp is even, then line 7 will not run, so $\text{ans}_1 = \text{ans}_0$.

From line 8, it is seen that $\text{mult}_1 = \text{mult}_0^2$

exp is even, so line 9 produces $\text{exp}_1 = \text{exp}_0 / 2 = \frac{\text{exp}_0}{2}$.

The execution of line 9 denotes the end of the j th iteration and the start of the $j + 1$ th iteration. Similar to the previous case, it is seen that

$$\text{mult}_1^{\text{exp}_1} \cdot \text{ans}_1 = (\text{mult}_0^2)^{\frac{\text{exp}_0}{2}} \cdot (\text{ans}_0) = \text{mult}_0^{\text{exp}_0} \cdot \text{ans}_0 = a^b$$

Thus, in either case, $P(j + 1)$ has been shown to hold.

By the principle of simple induction, it has been shown that $P(\lfloor \log_2(b) \rfloor + 2)$ is true at the beginning of iteration number $\lfloor \log_2(b) \rfloor + 2$. Since exp was initialized as b and has been halved $\lfloor \log_2(b) \rfloor + 1$ times, the value of exp at the beginning of the iteration is 0, which means that the loop terminates. Using the fact that $P(\lfloor \log_2(b) \rfloor + 2)$ holds,

$$\text{mult}^{\text{exp}} \cdot \text{ans} = \text{mult}^0 \cdot \text{ans} = \text{ans} = a^b$$

Finally, the program returns ans on line 10, which satisfies its postconditions, the program is correct.

□

Question 8. Prove that `majority()` is correct.

Proof. Define the loop variant to be $\text{len}(A) - i$. Notice that it is a natural number and decreases every iteration, but is bounded below by 0. Thus the program will terminate eventually.

Before continuing to prove correctness, a new definition will be defined to make the proof more concise and clear to read:

A list of integers that has more than half of its entries equal to some integer x

is called a **majority list** for x .

Let A be a majority list for x . As well, let n_i represent the absolute difference between the number of occurrences of x and the other values in $A[0 : i]$. For $i \in \{1, 2, \dots, \text{len}(A)\}$, define the loop invariant

$P(i)$: if $A[0 : i]$ is a majority list for x , then $m = x$ and $c \geq n_i$. As well,

if $A[0 : i]$ is not a majority list for x , then $c \leq n_i$ or $m = x$

It will be shown using simple induction that for $i \in \{1, 2, \dots, \text{len}(A)\}$, $P(i)$ is true at the beginning of the i th loop iteration.

Base Case. Let $i = 1$. Suppose that $A[0 : 1]$ is a majority list for x . Then it must be that $A[0] = x$ and $n_1 = 1$. Upon entering the loop for the first time, $m = A[0]$ and $c = 1 \geq n_1$, which satisfies the conclusion of the first implication.

Secondly, suppose that $A[0 : 1]$ is not a majority list for x . Then $A[0] \neq x$ and $n_1 = 1$. The first time entering the loop, $c = 1 \leq n_1$, which is sufficient to prove the second statement to be true.

Thus at the beginning of the first iteration, $P(1)$ is true.

Induction Hypothesis. Let $i = k$, for some $k \in \{1, 2, \dots, \text{len}(A) - 1\}$. Suppose that at the start of the k th iteration, $P(k)$ is true. Since $k \leq \text{len}(A) - 1$, the loop body will enter. The end goal is to show that $P(k + 1)$ is true at the beginning of the $k + 1$ th iteration.

Induction Step. Let m_j, c_j denote the values of m and c at the start of the j th iteration. To prove the first implication, suppose that $A[0 : k + 1]$ is a majority list for x . Consider the following two cases:

1. $A[0 : k]$ is a majority list for x :

By the induction hypothesis, $m_k = x$ and $c_k \geq n_k$. Executing the k th iteration, it is clear that $c_k \neq 0$, as $c_k \geq n_k > 0$. Thus lines 11 and 12 are skipped. Then, if $A[k] = m_k = x$, $n_{k+1} = n_k + 1$. By line 14, $c_{k+1} = c_k + 1 \geq n_k + 1 = n_{k+1}$. If $A[k] \neq x$, then $n_{k+1} = n_k - 1$. Line 16 instead yields $c_{k+1} = c_k - 1 \geq n_k - 1 = n_{k+1}$. Also, since m is unchanged, $x = m_k = m_{k+1}$. Therefore the first implication holds.

2. $A[0 : k]$ is not a majority list for x :

In this case, the only way for $A[0 : k]$ to not be a majority list for x is if $A[k] = x$ and $A[0 : k]$ has half its elements equal to x . This means that $n_k = 0$, and by the induction hypothesis, either $c_k = 0$ or $m_k = x$. Executing the k th iteration, if $c_k = 0$, running lines 11 and 12 result in $m_{k+1} = A[k] = x$ and $c_{k+1} = 1$. Otherwise, if $c_k \neq 0$, then it must be true that $m_k = x$. $A[k] = x$, so the condition on line 13 is satisfied, running line 14, so $c_{k+1} = c_k + 1 \geq n_k + 1 = n_{k+1}$. m remains unchanged so $x = m_k = m_{k+1}$. In either case, the first implication is still shown to be true.

Therefore, the first implication holds true.

Now, proving the second implication, suppose that $A[0 : k + 1]$ is not a majority list for x . Again, consider cases.

1. $A[0 : k]$ is not a majority list for x :

By the induction hypothesis, either $c_k \leq n_k$ or $m_k = x$. Consider what happens if $A[k] = x$. Executing the k th iteration, if $c_k = 0$, line 11 results in $m_{k+1} = A[k] = x$. If not, then the value of m is unchanged so $m_{k+1} = m_k$. Next, if the condition on line 13 is true, then $x = A[k] = m_k = m_{k+1}$. Finally, if line 16 runs, it means that $m_k = x \neq A[k]$. From the induction hypothesis, $c_k \leq n_k$ must be true. Then $c_{k+1} = c_k - 1 \leq n_k - 1 = n_{k+1}$. In all situations, either $m_{k+1} = x$ or $c_{k+1} \leq n_{k+1}$.

Now, consider what happens when $A[k] \neq x$. Again, executing the k th iteration, if $c_k = 0$, line 12 results in $c_{k+1} = 1$, but notice that since $A[0 : k]$ is not a majority list, appending a value not equal to x will increase the value of n_{k+1} , so $n_{k+1} = n_k + 1$. Then $c_{k+1} = 1 \leq n_k + 1 = n_{k+1}$. If $c_k \neq 0$, then similar to the previous case, $m_{k+1} = m_k$. If $m_k = A[k] \neq x$, it must be true that $c_k \leq n_k$. Then, line 14 will add 1 to c_k , so $c_{k+1} = c_k + 1 < n_k + 1 = n_{k+1}$. Otherwise, if $m_k \neq A[k]$, then m_k can either be equal or not equal to x . If $m_k = x$, the conclusion follows immediately. If not, $c_k \leq n_k$ and from line 16, c_k is decremented so $c_{k+1} = c_k - 1 \leq n_k - 1 = n_{k+1}$.

Therefore the second implication holds true at the beginning of the $k + 1$ th iteration.

2. $A[0 : k]$ is a majority list for x .

The only situation where this can be true is if $n_k = 1$ and $A[k] \neq x$. If this condition is not true, then $A[0 : k + 1]$ is a majority list for x . Continuing on, by the induction hypothesis, $c_k \geq n_k = 1$ and $m_k = x$. Iterating through the k th iteration, notice that c_k is non-zero. Thus lines 11 and 12 are skipped. The value of m_k remains unchanged, so $m_{k+1} = m_k = x$, which makes the second implication true.

Thus it has been shown that both the implications are true, so $P(k + 1)$ is true at the beginning of the $k + 1$ th iteration.

By the principle of simple induction, for all $i \in \{1, \dots, \text{len}(A)\}$, $P(i)$ holds true at the beginning of the i th iteration. When $i = \text{len}(A)$, the loop condition is false so the loop terminates. Notice that $A[0 : \text{len}(A)] = A$ is a majority list for x , so by $P(i)$, the program correctly returns $m = x$, satisfying the postconditions of the program. □

Question 9. Studying bubblesort. (Note: Inner loop condition should be $i < \text{len}(L) - k - 1$)

- (a) State and prove an invariant for the inner loop

Proof. Let L be a list of numbers. Define the inner loop invariant to be

$$Q(n) : \text{For all } j \in \mathbb{N}, j < i, L[j] \leq L[i]$$

Let $k \in \{0, 1, \dots, \text{len}(L) - 1\}$. It will be proven using simple induction that for all $n \in \{1, \dots, \text{len}(L) - k\}$, at the start of the n th iteration, $Q(n)$ is true.

Base Case. Let $n = 1$. At the beginning of the first iteration, $i = 0$. It follows that $Q(1)$ is vacuously true because there is no natural number $j < 0$.

Induction Hypothesis. Suppose that for some $n \in \{1, 2, \dots, \text{len}(L) - k - 1\}$, at the beginning of the n th iteration, $Q(n)$ is true. Notice that at the end of every iteration, i is incremented once. Thus $i = n - 1 < \text{len}(L) - k - 1$, so the loop body will be entered. It will be shown that at the start of the $n + 1$ th iteration, $Q(n + 1)$ is true.

Induction Step. Iterating through the loop, if $L[n - 1] > L[n]$, their positions in the list will be swapped. This guarantees that $L[n - 1] \leq L[n]$ after line 11. Next, i is incremented by 1, so its new value is n . This marks the end of the n th iteration and the start of the $n + 1$ th iteration.

$L[n - 1] \leq L[n]$ is true based on the previous iteration of the loop. Additionally, by the induction hypothesis, for all natural $j < n - 1$, $L[j] \leq L[n - 1] \leq L[n]$ if line 11 did not run, or $L[j] \leq L[n]$ if it did. Regardless, $L[j] \leq L[n]$. It can be concluded that for $j < n = i$, $L[j] \leq L[n] = L[i]$, so $Q(n + 1)$ is true.

By the principle of simple induction, for all $n \in \{1, 2, \dots, \text{len}(L) - k\}$, at the start of the n th iteration, $Q(n)$ is true. □

- (b) State and prove an invariant for the outer loop

Proof. For a positive natural m , define the outer loop invariant to be

$$P(m) : \text{The last } m \text{ elements of } L \text{ are sorted.}$$

In other words, it is true that $L[\text{len}(L) - m] \leq L[\text{len}(L) - m + 1] \leq \dots \leq L[\text{len}(L) - 1]$. It will be proven using simple induction that for $m \in \{1, 2, \dots, \text{len}(L)\}$, $P(m)$ is true at the start of the m th iteration.

Base Case. Let $m = 1$. When the loop is first entered, the last element of L is obviously sorted relative to itself. Thus $P(1)$ holds true.

Induction Hypothesis. Suppose that for some $m \in \{1, \dots, \text{len}(L) - 1\}$, $P(m)$ is true at the beginning of the m th iteration. Notice that k is initialized to be 0 and is incremented every outer loop iteration. Thus $k = m - 1 < \text{len}(L)$, so the loop body is entered.

Induction Step. Running the m th iteration, i is initialized to be 0 and the inner loop runs. First, it will be shown that the inner loop terminates.

Let the inner loop variant be $\text{len}(L) - k - i$. This value is natural and decreasing, so the inner loop will terminate.

Now, by the previous part, $Q(\text{len}(L) - m + 1)$ is true at the beginning of the $(\text{len}(L) - m + 1)$ th iteration. This can be seen by taking $k = m - 1$. Recall that at the beginning of this iteration, $i = \text{len}(L) - m = \text{len}(L) - k - 1$, so the loop terminates. Thus before line 13, $L[\text{len}(L) - m - 1] \leq L[\text{len}(L) - m]$, as $\text{len}(L) - m - 1 < i$. After line 13, the $m + 1$ th iteration begins.

By the induction hypothesis, there is the following chain of inequalities:

$$L[\text{len}(L) - m - 1] \leq L[\text{len}(L) - m] \leq L[\text{len}(L) - m + 1] \leq \dots \leq L[\text{len}(L) - 1]$$

which means the last $m + 1$ elements are sorted, so $P(m + 1)$ is true.

By the principle of simple induction, for all $m \in \{1, 2, \dots, \text{len}(L)\}$, $P(m)$ is true at the beginning of the m th iteration. \square

- (c) Prove that `bubblesort` is correct, according to its specifications

Proof. Let the outer loop variant be $\text{len}(L) - k$. It is easy to see that this value is natural and decreasing. Thus the program will eventually terminate.

From the previous part, it is true that $P(\text{len}(L))$ holds at the beginning of the $\text{len}(L)$ th iteration. At this iteration, $k = \text{len}(L) - 1$, so the loop condition is true and the loop body enters. When this iteration reaches the inner loop, notice that $i = 0 = \text{len}(L) - k - 1$, so the inner loop does not run. Then k is incremented and the loop returns to the top of the loop, where the loop condition fails and terminates the loop. Since L remained unchanged throughout the iteration, $P(\text{len}(L))$ is still true, so the last $\text{len}(L)$ elements of the list are sorted, which is the entire list. Thus the program satisfies its postconditions. \square

Question 10. Consider the following generalization of the `min` function.

- (a) Prove that this algorithm is correct

Proof. This program, given that A is a non-empty list of numbers and k is a positive natural less than or equal to $\text{len}(A)$, returns the k th smallest element. It will be shown using complete induction on $\text{len}(A)$ that this program is correct.

Base Case. Let A be a list of numbers such that $\text{len}(A) = 1$. Then k must be equal to 1. Taking $A[0]$ as the pivot, line 4 will partition an empty list, so $\text{len}(L) = 0$. Then the condition on line 5 is true, so the program correctly returns the pivot as the smallest element on line 6.

Induction Hypothesis. Suppose that for any list B such that $\text{len}(B) \leq n$, `extract(B, k)` correctly returns the k th smallest element.

Induction Step. Let A be a list such that $\text{len}(A) = n + 1$. Running through the program, choosing $A[0]$ as the pivot, line 4 separates A into two lists L, G , where L contains all elements in A such that they are less than `pivot` and G contains all elements in A such that they are greater than or equal to `pivot`. Continuing on to line 5, if $\text{len}(L) = k - 1$, then L contains $k - 1$ elements that are smaller than the pivot, so the pivot is the k th smallest element, which the program returns on line 6.

Otherwise, notice that $\text{len}(L), \text{len}(G) < n + 1 \implies \text{len}(L), \text{len}(G) \leq n$, which means that L and G are covered with the induction hypothesis.

If $\text{len}(L) \geq k$, the k th smallest element of A is contained in L . The program returns `extract(L, k)`, which by the induction hypothesis, returns the k th smallest element of L , which is equal to the k th smallest element of A .

If not, then it must be that the k th smallest element of A is inside of G , and is equal to the $(k - \text{len}(A) - 1)$ th smallest element of G .

Indeed, the program returns `extract(G, k - len(L) - 1)`, which by the induction hypothesis returns the $(k - \text{len}(A) - 1)$ th smallest element of G , which is the k th smallest element of A .

Thus the program returns the correct output for lists of size $n + 1$, and by the principle of complete induction, the program is correct.

□

- (b) Analyze the worse-case running time of this algorithm.

Let n denote number of elements in the list A . Define $T(n)$ to be the worst-case runtime of `extract`. The base case is when $n = 1$. As outlined in the correctness proof above, an empty list is partitioned, an if statement is evaluated, and a number is returned; all of which is constant time.

When $n > 1$, only one recursive call is made. In particular, either `extract(L, k)` or `extract(G, len(L) - k - 1)` is called. Other than that, the program also partitions A , which

takes linear time, and evaluates two if statements, and performs subtraction, which is done in practically constant time. The worst-case runtime will occur when the program makes a recursive call to the bigger list between L and G . Thus the recurrence for the worst-case runtime is

$$T(n) = \begin{cases} c, & \text{if } n = 1; \\ T(\max\{|L|, |G|\}) + dn + f, & \text{if } n > 1. \end{cases}, \text{ where } c, d, f \text{ are constants.}$$

In order to find an asymptotic upper bound for $T(n)$, each recursive call must be as big as possible. This can be done by making the list partition in a way such that either L or G is empty. In this case, $\max\{|L|, |G|\} = n - 1$. Rewrite $T(n)$ as

$$T(n) = \begin{cases} c, & \text{if } n = 1; \\ T(n - 1) + dn + f, & \text{if } n > 1. \end{cases}$$

Now, proceed with finding the closed form for $T(n)$ using repeated substitution. For some natural $n > 0$,

$$\begin{aligned} T(n) &= T(n - 1) + dn + f \\ &= T(n - 2) + d(n + (n - 1)) + 2f \\ &= T(n - 3) + d(n + (n - 1) + (n - 2)) + 3f \\ &\quad \vdots \\ &= T(1) + d \left(\frac{n(n + 1)}{2} - 1 \right) + (n - 1)f \end{aligned}$$

It will be proven by induction on n that $T(n) = \frac{d}{2}n^2 + (\frac{d}{2} + f)n + c - d - f$.

Base Case. Let $n = 1$. Then

$$\frac{d}{2}n^2 + \left(\frac{d}{2} + f \right) n + c - d - f = \frac{d}{2} + \left(\frac{d}{2} + f \right) + c - d - f = c = T(1)$$

Thus the closed form is valid for $n = 1$.

Induction Hypothesis. Suppose that the closed form holds for some $n = k$. That is,

$$T(k) = \frac{d}{2}k^2 + \left(\frac{d}{2} + f \right) k + c - d - f$$

The goal is to show that

$$T(k + 1) = \frac{d}{2}(k + 1)^2 + \left(\frac{d}{2} + f \right) (k + 1) + c - d - f$$

Induction Step. Using the recursive definition of $T(k + 1)$,

$$T(k + 1) = T(k) + d(k + 1) + f = \left(\frac{d}{2}k^2 + \left(\frac{d}{2} + f \right) k + c - d - f \right) + d(k + 1) + f$$

$$= \frac{d}{2}k^2 + \left(\frac{d}{2} + f\right)k + c - d - f + dk + \frac{d}{2} + \frac{d}{2} + f$$

Reordering terms yields

$$\begin{aligned} &= \left(\frac{d}{2}k^2 + dk + \frac{d}{2}\right) + \left(\left(\frac{d}{2} + f\right)k + \frac{d}{2} + f\right) + c - d - f \\ &= \frac{d}{2}(k+1)^2 + \left(\frac{d}{2} + f\right)(k+1) + c - d - f \end{aligned}$$

which is equal to the claimed closed form for $T(k+1)$.

By the principle of induction, it is true that for all naturals $n > 0$,

$$T(n) = \frac{d}{2}n^2 + \left(\frac{d}{2} + f\right)n + c - d - f$$

It will be shown that $T(n) \in \mathcal{O}(n^2)$.

Let $c_0 = c$, $n_0 = 1$. Notice that for all $n > 1$,

$$\begin{aligned} T(n) &= \frac{d}{2}n^2 + \left(\frac{d}{2} + f\right)n + c - d - f \leq \frac{d}{2}n^2 + \left(\frac{d}{2} + f\right)n^2 + (c - d - f)n^2 \\ &= \left(\frac{d}{2} + \frac{d}{2} + f + c - d - f\right)n^2 = cn^2 = c_0n^2 \end{aligned}$$

Thus the worst-case time complexity of this algorithm is $\mathcal{O}(n^2)$.

Q4. Solve questions 6,7,10,12, and 14 on pages 46 to 48 of the textbook. Only **two** of the five questions will be selected for marking. A serious attempt of all the mentioned questions is a necessary condition for receiving more than 0 points.

Question 6. Let $T(n)$ be the number of binary strings of length n where every 1 is immediately preceded by a 0.

(a) Develop a recurrence for $T(n)$

If the binary string is of length 1, it must be '0', since every binary string that starts with '1' must be at least length 2.

If the binary string is of length 2, it can either be '00' or '10'. Any other string would violate the condition.

Notice that for any binary string that satisfies the condition above and is length n , where $n \geq 3$, it can begin with either a 1 or a 0.

If it begins with a 1, then the subsequent character must be a 0. If the first '10' is taken out of the string, there remains a binary string with length $n - 2$ that satisfies the condition, of which there are $T(n - 2)$ unique possibilities.

If it begins with a 0, take out the first character. There remains a binary string that follows the conditions above with length $n - 1$, which there are $T(n - 1)$ unique possibilities for.

Thus the total number of these particular binary strings is $T(n - 1) + T(n - 2)$, and the recurrence for $T(n)$ is

$$T(n) = \begin{cases} 1, & \text{if } n = 1; \\ 2, & \text{if } n = 2; \\ T(n - 1) + T(n - 2), & \text{if } n \geq 3. \end{cases}$$

(b) Find a closed form for $T(n)$

The closed form for $T(n)$ is $T(n) = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{n+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{n+1} \right)$, which will be proven to be correct in the next part.

(c) Prove that the closed form is correct using induction

Proof. It will be proven that the closed form is correct using complete induction on $n \in \mathbb{N}^+$.

Base Case. Let $n = 1$. Then $T(n) = 1$. As well,

$$\begin{aligned} \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^2 - \left(\frac{1-\sqrt{5}}{2} \right)^2 \right) &= \frac{1}{4\sqrt{5}} ((1 + 2\sqrt{5} + 5) - (1 - 2\sqrt{5} + 5)) \\ &= \frac{1}{4\sqrt{5}} (4\sqrt{5}) = 1 \end{aligned}$$

Thus the closed form is valid for $n = 1$.

Now, let $n = 2$. Then $T(n) = 2$. As well,

$$\begin{aligned} \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^3 - \left(\frac{1-\sqrt{5}}{2} \right)^3 \right) &= \frac{1}{8\sqrt{5}} ((1+3\sqrt{5}+15+5\sqrt{5}) - (1-3\sqrt{5}+15-5\sqrt{5})) \\ &= \frac{1}{8\sqrt{5}} (16\sqrt{5}) = 2 \end{aligned}$$

Therefore the closed form holds true for $n = 1$ and $n = 2$.

Induction Hypothesis. Let $k \in \mathbb{N}^+$. Suppose that for all $i \leq k$, the closed form holds true. The goal is to show that the closed form is valid for $n = k + 1$.

Induction Step. Using the recursive definition for $T(n)$,

$$T(k+1) = T(k) + T(k-1)$$

The values of k and $k-1$ are covered within the induction hypothesis, so

$$\begin{aligned} T(k)+T(k-1) &= \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{k+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{k+1} \right) + \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^k - \left(\frac{1-\sqrt{5}}{2} \right)^k \right) \\ &= \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{k+1} \left(1 + \frac{2}{1+\sqrt{5}} \right) - \left(\frac{1-\sqrt{5}}{2} \right)^{k+1} \left(1 + \frac{2}{1-\sqrt{5}} \right) \right) \\ &= \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{k+1} \left(1 + \frac{2(1-\sqrt{5})}{-4} \right) - \left(\frac{1-\sqrt{5}}{2} \right)^{k+1} \left(1 + \frac{2(1+\sqrt{5})}{-4} \right) \right) \\ &= \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{k+1} \left(\frac{1+\sqrt{5}}{2} \right) - \left(\frac{1-\sqrt{5}}{2} \right)^{k+1} \left(\frac{1-\sqrt{5}}{2} \right) \right) \\ &\quad + \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{k+2} - \left(\frac{1-\sqrt{5}}{2} \right)^{k+2} \right) \end{aligned}$$

Which is the desired form for $T(k+1)$.

By the principle of induction, the closed form is correct for all $k \in \mathbb{N}^+$.

□

Question 7. Let $T(n)$ denote the number of distinct full binary trees with n nodes. Give a recurrence for $T(n)$. Then use induction to prove that $T(n) \geq \frac{1}{n} 2^{(n-1)/2}$.

Proof. Full binary trees can be defined recursively in the following way:

1. A leaf node is a full binary tree
2. If T_1, T_2 are full binary trees, then the binary tree obtained from attaching T_1, T_2 to the root node is a full binary tree.

There are no full binary trees that have an even number of nodes, so when n is even, $T(n) = 0$.

When $n = 1$, The only tree is the single root node, which is a full binary tree, so $T(1) = 1$.

When n is odd and greater than 1, the root node must have two children. Notice that each of the children must be full binary trees as well. Additionally, the sum of the number of nodes of the left child and the right child must add up to $n - 1$. Suppose that the left child has k nodes and the right child has $n - k - 1$ nodes. Thus the number of full binary trees with this characteristic is $T(k)T(n - k - 1)$.

It can be concluded that the number of binary trees with n nodes can be found by considering all the combinations of smaller full binary trees. The recurrence for $T(n)$ is

$$T(n) = \begin{cases} 0, & \text{if } n \text{ is even;} \\ 1, & \text{if } n = 1; \\ \sum_{i=1}^{\frac{n-1}{2}} T(n-2i) \cdot T(2i-1), & \text{otherwise.} \end{cases}$$

Now, it will be shown using complete induction that $T(n) \geq \frac{1}{n}2^{(n-1)/2}$ for all odd naturals n .

Base Case. Let $n = 1$. Then $T(n) = 1$ and $\frac{1}{n}2^{(n-1)/2} = 1$, so the base case holds.

Induction Hypothesis. Let k be an odd natural number. Suppose that for all $i \leq k$, $T(i) \geq \frac{1}{i}2^{(i-1)/2}$. The goal is to show that $T(k+2) \geq \frac{1}{k+2}2^{(k+1)/2}$.

Induction Step. Using the recursive definition for $T(n)$, as well as the induction hypothesis,

$$\begin{aligned} T(k+2) &= \sum_{i=1}^{\frac{k+1}{2}} T(k+2-2i) \cdot T(2i-1) \geq \sum_{i=1}^{\frac{k+1}{2}} \frac{2^{(k+1-2i)/2}}{k+2-2i} \cdot \frac{2^{(2i-2)/2}}{2i-1} = \sum_{i=1}^{\frac{k+1}{2}} \frac{2^{(k-1)/2}}{(k+2-2i)(2i-1)} \\ &= 2^{(k-1)/2} \sum_{i=1}^{\frac{k+1}{2}} \frac{1}{(k+2-2i)(2i-1)} \end{aligned}$$

Since i is between 1 and $\frac{k+1}{2}$, $k+2-2i > 0$ and $2i-1 > 0$. By the AGM inequality,

$$(k+2-2i)(2i-1) \leq \left(\frac{k+1}{2}\right)^2 \leq \frac{(k+1)(k+2)}{4}$$

Thus

$$\frac{1}{(k+2-2i)(2i-1)} \geq \frac{4}{(k+2)(k+1)}$$

Returning back to the sum, it follows that

$$T(k+2) \geq 2^{(k-1)/2} \sum_{i=1}^{\frac{k+1}{2}} \frac{1}{(k+2-2i)(2i-1)} \geq 2^{(k-1)/2} \sum_{i=1}^{\frac{k+1}{2}} \frac{4}{(k+2)(k+1)}$$

$$= 2^{(k-1)/2} \left(\frac{k+1}{2} \right) \left(\frac{4}{(k+2)(k+1)} \right) = \frac{2^{(k-1)/2} \cdot 2}{k+2} = \frac{2^{(k+1)/2}}{k+2}$$

as desired.

Therefore by the principle of induction, $T(n) \geq \frac{1}{n} 2^{(n-1)/2}$ for all odd naturals n .

□

Question 12. Analyze the worse-case runtime of `fast_rec_mult`.

Let $T(n)$ denote the worst-case runtime of `fast_rec_mult` on the inputs x and y , where n represents the number of digits of x and y .

The base case of this method is when $n = 1$. In this case the program returns $\mathbf{x} \star \mathbf{y}$ on line 4, which runs in constant time because x and y are small enough numbers.

If $n > 1$, lines 6 and below run. Assuming nice inputs, it can be ensured that a, b, c, d are $\frac{n}{2}$ digit numbers. The non-recursive steps are just a combination of basic arithmetic, which takes at most linear time.

Three recursive calls are made on $\frac{n}{2}$ digit numbers, which in total, has worst-case runtime of $3T(\frac{n}{2})$.

Therefore the full definition of T is

$$T(n) = \begin{cases} c, & \text{if } n = 1; \\ 3T\left(\frac{n}{2}\right) + dn, & \text{if } n > 1. \end{cases}$$

The asymptotic upper bound for the extra steps is $\mathcal{O}(n)$. Since $1 < \log_2(3)$, by Master Theorem, $T(n) = \Theta(n^{\log_2 3})$.

Question 14. Recall the recurrence for the worst-case runtime of quicksort:

$$T(n) = \begin{cases} c, & \text{if } n \leq 1; \\ T(|L|) + T(|G|) + dn, & \text{if } n > 1; \end{cases}$$

where L and G are the partitions of the list.

1. Suppose the lists are always evenly split; that is, $|L| = |G| = \frac{n}{2}$ at each recursive call. Find a tight asymptotic bound on the runtime of quicksort using this assumption.

Proof. If the lists are evenly split, the recurrence for the runtime of quicksort is

$$T(n) = \begin{cases} c, & \text{if } n \leq 1; \\ 2T\left(\frac{n}{2}\right) + dn, & \text{if } n > 1. \end{cases}$$

Notice that the non-recurrence part of the function is $\Theta(n)$. Since $1 = \log_2(2)$, by the Master Theorem,

$$T(n) \in \Theta(n \log n)$$

□

2. Now suppose that the lists are always very unevenly split: $|L| = n - 2$ and $|G| = 1$ at each recursive call. Find a tight asymptotic bound on the runtime of quicksort using this assumption.

Proof. When $|L| = n - 2$, $|G| = 1$, the recursive runtime of quicksort is

$$T(n) = \begin{cases} c, & \text{if } n \leq 1; \\ T(n-2) + dn + c, & \text{if } n > 1. \end{cases}$$

A closed form for $T(n)$ will be found using repeated substitution.

$$\begin{aligned} T(n) &= T(n-2) + dn + c = T(n-4) + d(n-2) + dn + 2c \\ &= T(n-6) + d(n-4) + d(n-2) + dn + 3c = \dots \end{aligned}$$

It can be conjectured that a possible closed form for $T(n)$ is

$$T(n) = \left(\sum_{i=0}^{\lfloor \frac{n}{2} \rfloor - 1} (n - 2i) \right) d + \left(\left\lfloor \frac{n}{2} \right\rfloor + 1 \right) c$$

It will be proven that this is indeed true by induction on n .

Base Case. Let $n = 0$. Then $T(n) = c$ but also

$$\left(\sum_{i=0}^{\lfloor \frac{0}{2} \rfloor - 1} (0 - 2i) \right) d + \left(\left\lfloor \frac{0}{2} \right\rfloor + 1 \right) c = 0d + c = c$$

The exact same argument can be used for $n = 1$, so it can be reasoned that the claim holds for both $n = 0$ and $n = 1$.

Induction Hypothesis. Let k be a natural number. Suppose that the closed form for $T(k)$ holds. The goal is to show that the closed form works for $T(k+2)$.

Induction Step. This can be proven with algebraic manipulations. Utilizing the induction hypothesis,

$$T(k+2) = T(k) + dn + c = \left(\sum_{i=0}^{\lfloor \frac{k}{2} \rfloor - 1} (k - 2i) \right) d + \left(\left\lfloor \frac{k}{2} \right\rfloor + 1 \right) c + d(k+2) + c$$

Reindexing the sum,

$$= \left(\sum_{i=1}^{\lfloor \frac{k+2}{2} \rfloor - 1} (k+2 - 2i) \right) d + \left(\left\lfloor \frac{k}{2} \right\rfloor + 1 \right) c + d(k+2) + c$$

Assimilating the terms together yields

$$= \left(\sum_{i=0}^{\lfloor \frac{k+2}{2} \rfloor - 1} (k+2-2i) \right) d + \left(\left\lfloor \frac{k+2}{2} \right\rfloor + 1 \right) c$$

Therefore this expression is a valid closed form for $T(k+2)$.

By the principle of induction, for all natural n , the closed form for $T(n)$ is

$$T(n) = \left(\sum_{i=0}^{\lfloor \frac{n}{2} \rfloor - 1} (n-2i) \right) d + \left(\left\lfloor \frac{n}{2} \right\rfloor + 1 \right) c$$

By manipulating the expression, it can be reduced to a more explicit form:

$$\begin{aligned} \left(\sum_{i=0}^{\lfloor \frac{n}{2} \rfloor - 1} (n-2i) \right) d + \left(\left\lfloor \frac{n}{2} \right\rfloor + 1 \right) c &= \left(\sum_{i=0}^{\lfloor \frac{n}{2} \rfloor - 1} n - \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor - 1} 2i \right) d + \left(\left\lfloor \frac{n}{2} \right\rfloor + 1 \right) c \\ &= \left(n \left(\left\lfloor \frac{n}{2} \right\rfloor - 1 \right) - \left\lfloor \frac{n}{2} \right\rfloor \left(\left\lfloor \frac{n}{2} \right\rfloor - 1 \right) \right) d + \left(\left\lfloor \frac{n}{2} \right\rfloor + 1 \right) c \\ &\implies T(n) = \left(\left\lfloor \frac{n}{2} \right\rfloor - 1 \right) \left(n - \left\lfloor \frac{n}{2} \right\rfloor \right) d + \left(\left\lfloor \frac{n}{2} \right\rfloor + 1 \right) c \end{aligned}$$

It will be shown that $T(n) \in \Theta(n^2)$. Since $\frac{n}{2} - 1 \leq \left\lfloor \frac{n}{2} \right\rfloor \leq \frac{n}{2}$, then

$$\begin{aligned} \left(\frac{n}{2} - 2 \right) \left(\frac{n}{2} \right) d + \frac{n}{2} c &\leq T(n) \leq \left(\frac{n}{2} - 1 \right) \left(\frac{n}{2} + 1 \right) d + \left(\frac{n}{2} + 1 \right) c \\ &\implies \frac{d}{4} n^2 + \left(\frac{c}{2} - d \right) n \leq T(n) \leq \frac{d}{4} n^2 + \frac{c}{2} n + c - d \end{aligned}$$

From this, it can be seen that $T(n) \in \Theta(n^2)$.

□