

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 $j \in \{1, \dots, \text{len}(A)\}$, $Q(j)$ 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 $j \in \{1, 2, \dots, \text{len}(A), \text{len}(A) + 1\}$.

At the beginning of the $\text{len}(A) + 1$ th iteration, $i = \text{len}(A)$, so the loop terminates. The program returns $sum/\text{len}(A)$ on line 11, which is exactly the average of the numbers in A .

□

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. Define the loop invariant to be

$$Inv(x, y) : a = x + \frac{y(y+1)}{2} - 55$$

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

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

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

The base case has been proven to be true.

Induction Hypothesis. Now suppose that for some positive natural j , $Inv(a, y)$ holds true at the beginning of the j th iteration. y is initially 10 and is decremented at the end of every iteration, so at the beginning of the j th iteration, $y = 10 - (j - 1) = 11 - j$. Then the loop invariant states that

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

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

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

Finally, on line 7, the value of y is decremented by 1 and becomes $y = 10 - j$. The program then arrives at the beginning of the $j + 1$ th iteration. Indeed, it is true that

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

Thus $Inv(a, y)$ 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 using simple induction that at the beginning of every iteration, $a > 0$, which implies that the loop can never terminate.

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

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

Note that $\frac{(11-j)(12-j)}{2} < 0$ if and only if $11 < j$ and $j < 12$ at the same time, which is impossible for $j \in \mathbb{N}$. 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. When $b = n + 1$, $b \geq 1$, 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 $i \in \{1, 2, \dots, \lfloor \log_2(b) \rfloor + 1\}$, $P(i)$ is true at the beginning of the i th iteration.

Base Case. Let $i = 1$. At the start of the very first iteration, **mult** = a , **ans** = 1, and **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 $j \in \{1, 2, \dots, \lfloor \log_2(b) \rfloor\}$. Suppose that at the beginning of the j th iteration, $P(j)$ is true. That is,

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

Induction Step. Let **mult**₀, **exp**₀, **ans**₀ be the value of **mult**, **exp**, **ans** at the beginning of the j th iteration and let **mult**₁, **exp**₁, **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 **ans**₁ = **ans**₀ · **mult**₀.

Line 8 will run and will multiply **mult** by itself, so **mult**₁ = **mult**₀².

As well, **exp** can be rewritten as $2n + 1$, for some natural n , so line 9 will result in **exp**₁ = **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 **ans**₁ = **ans**₀.

Finally, from line 8, it is seen that **mult**₁ = **mult**₀²

exp can be written as $2k$ for some natural k , so line 9 produces **exp**₁ = **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 $P(\lfloor \log_2(b) \rfloor + 1)$ is true at the beginning of iteration number $\lfloor \log_2(b) \rfloor + 1$. 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 $P(\lfloor \log_2(b) \rfloor + 1)$ holds,

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

The program returns `ans` on line 10, which satisfies its postconditions. \square

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. 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 list with more than half of its entries equal to each other, and let x be that value. As well, let n_i represent the difference between the number of occurrences of x and the other values in $A[0 : i]$. For a positive natural i , 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 a list with half its elements equal to x , then $c = 0$ or $m = x$

It will be shown using complete 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. Consider two base cases.

Let $i = 1$. Suppose that $A[0 : 1]$ is a majority list for x . Then it must be that $A[0] = x$. Upon entering the loop for the first time, $m = A[0]$ and $c = 1$, which satisfies the conclusion of the first implication. Secondly, notice that the second statement is vacuously true because a list with one element cannot have half of its elements equal to x . Thus at the beginning of the first iteration, $P(1)$ holds true.

Induction Hypothesis. Let $i = k$, for some $k \in \{1, 2, \dots, \text{len}(A) - 1\}$. Suppose that for all $j \leq k$, at the start of the j th iteration, $P(j)$ is true. 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_i, c_i denote the values of m and c at the start of the i 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, $P(k)$ is true at the start of the k th iteration. That is, $m_k = x$ and $c_k \geq n_i$. Executing the k th iteration, it is clear that $c_k \neq 0$, as $c_k \geq n_i > 0$. Thus lines 11 and 12 are skipped. Then, if $A[k] = m_k = x$, by line 14, $c_{k+1} = c_k + 1 \geq n_k + 1 = n_{k+1}$. Also, $x = m_k = m_{k+1}$. Therefore the first implication holds.

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

The only way for $A[0 : k]$ to not be a majority list for x at the same time as $A[0 : k + 1]$ is if $A[k] = x$ and $A[0 : k]$ has half its elements equal to x . By the induction hypothesis, $c_k = 0$ 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 $m_k = x$. $A[k] = x$, 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 a list with half its elements equal to x . This means that $A[0 : k + 1]$ is a list with at least 2 elements. Again, consider cases.

1. Neither of the last 2 elements are equal to x :

In this case, $A[0 : k - 1]$ is a majority list for x . By the induction hypothesis, at the beginning of the $k - 1$ th iteration, $m_{k-1} = x$ and $c_{k-1} \geq n_{k-1} = 2$. Iterating through the $k - 1$ th iteration, only lines 16 and 17 are ran, which means that $m_k = m_{k-1}$ and $c_k = c_{k-1} - 1 \geq 1$. Again, iterating through the k th iteration of the loop, only lines 16 and 17 are ran, so $m_{k+1} = m_k = m_{k-1} = x$ and $c_{k+1} = c_k \geq 1$, which satisfies the conclusion.

2. Exactly 1 of the last 2 elements are equal to x :

In this case, $A[0 : k - 1]$ is a list with half of its elements equal to x . By the induction hypothesis, at the start of the $k - 1$ th iteration, either $c_{k-1} = 0$ or $m_{k-1} = x$.

In the $k - 1$ th iteration, if $c_{k-1} = 0$, then the iteration will run lines 11 and 12, resulting in $c_k = 1$ and $m_k = A[k - 1]$. The k th iteration will then run line 16, as $c_k \neq 0$ and $m_k = A[k - 1] \neq A[k]$. Thus $c_{k+1} = c_k - 1 = 0$.

Otherwise, $m_{k-1} = x$ is true, so either line 14 or 16 will run, depending on whether $A[k - 1] = x = m_{k-1}$. If the equality is true, then $A[k - 1] = x$, and $A[k] \neq x$. Line 14 will run in the $k - 1$ th iteration, making $c_k = c_{k-1} + 1$ and $m_k = m_{k-1}$. Then, since $c_k > 0$, $m_k = x$, and $x = A[k - 1] \neq A[k]$, line 16 will run, so the variable m remains unchanged, meaning $m_{k+1} = m_k = x$. On the other hand, if the equality is not true, $A[k - 1] \neq x$ and $A[k] = x$. Returning to the $k - 1$ th iteration, line 16 will run, so $m_k = m_{k-1} = x$ and $c_k = c_{k-1} - 1$. Next, in the k th iteration, if $c_k = 0$, $m_{k+1} = A[k] = x$. If not, line 14 will run, keeping $m_{k+1} = m_k = x$.

In all cases, either $c_{k+1} = 0$ or $m_{k+1} = x$.

3. The last 2 elements are equal to x :

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 complete 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.

- (a) State and prove an invariant for the inner loop
- (b) State and prove an invariant for the outer loop
- (c) Prove that `bubblesort` is correct, according to its specifications

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

- (a) Prove that this algorithm is correct
- (b) Analyze the worse-case running time of this algorithm.

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 $T(n)$
- (b) Find a closed form for $T(n)$
- (c) Prove that the closed form is correct using induction

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 $T(n) \geq \frac{1}{n}2^{(n-1)/2}$.

Question 10. Let $H(n)$ denote the number of binary strings of length n that have no odd length blocks of 1's.

Develop a recursive definition for $H(n)$, and justify why it is correct. Then find a closed form for H using repeated substitution.

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

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.
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.