

Discussion Group Problems for Week 3

For: January 27–January 31

Problem 1. Java Review

At this point, most of you should be comfortable enough to work with Java. Let's take some time to review a few concepts in Java so that we can limit our Java-related issues and, hence, focus on the algorithms when solving future Problem Sets.

- (a) What is the difference between a class and an object? Illustrate with an example.

Solution: A class can be seen as a 'template', or a 'blueprint', specifying what kind of methods/operations should be supported and its behaviour. An object is an instance of a class.

- (b) Why does the `main` method come with a `static` modifier?

Solution: Java program's `main` method has to be declared `static` because the keyword `static` allows `main` to be called without creating an object of the class in which the `main` method is defined. If we omit the `static` keyword before `main`, Java program will successfully compile but it won't execute.

- (c) Give an example class (or classes) that uses the modifier `private` incorrectly (i.e., the program will not compile as it is, but would compile if `private` was changed to `public`).

Solution:

Below is a possible solution:

```
class SecretHolder {
    private int secret;

    public SecretHolder(int value) {
        this.secret = value;
    }
}

class Test {
    public static void main(String[] args) {
        SecretHolder holder = new SecretHolder(5);
        holder.secret = 6; // Compile-time error!
    }
}
```

- (d) The following question is about Interfaces.

- (d)(i) Why do we use interfaces?

Solution: An interface can be seen as a ‘contract’ that is signed by a class whenever it **implements** the interface. The reason for using interfaces is that whenever we see a class that implements that interface, we know for certain that it supports the operations specified by that interface.

(d)(ii) Give an example of using an interface.

Solution: See Problem Set 1.

(d)(iii) Can a method return an object with an interface type?

Solution: Yes, see Problem Set 1: `ShiftRegisterTest.java`. It was also covered in Lecture 2.

(e) Refer to `IntegerExamination.java`, which can be found in the same folder as this PDF. Without running the code, predict the output of the `main` method. Can you explain the outputs?

Solution:

The expected output consists of the following 6 strings:

I am in `addOne`. The value of `i` is 8

I am in `myIntAddOne`. The value of `j` is 8

I am in `myOtherIntAddOne`. The value of `k` is 8

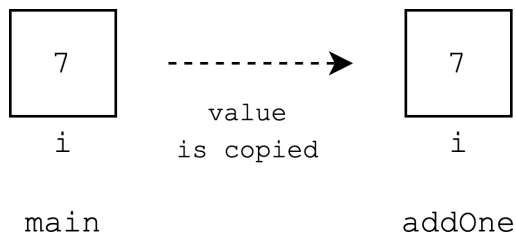
The final value of `i` back in `main` is 7

The final value of `j` back in `main` is 8

The final value of `k` back in `main` is 7

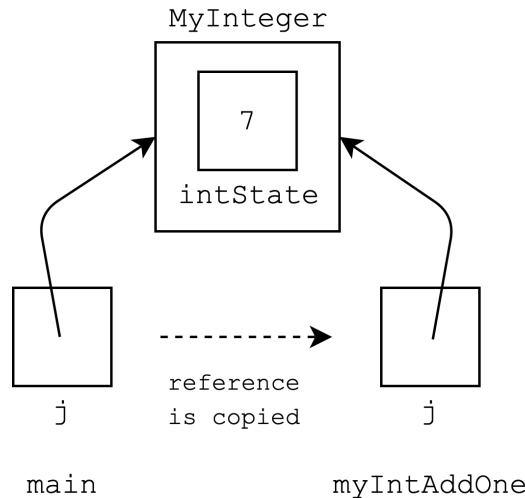
Most of the intuition is already mentioned in the code itself as comments. If you understand that, then you should understand the answer.

- Within the `addOne` method, the variable `i` refers to the argument of the method and not the variable `i` in the `main` method. The value of `i` in `addOne` is copied over from `i` in `main`. As such, the variable `i` in the `main` method remains unchanged as 7.



Solution: *(Continued)*

- Because `MyInteger` is a class (as opposed to a primitive data type like `int`), `j` is an object. That means that when it is passed as an argument to the `myIntAddOne` method, the variable `j` in the method is pointing to the same object `j` in the `main` method. As a result, when the `intState` of `j` in `myIntAddOne` is increased, that change is reflected in `j` in `main`, thus the final value of `j` in `main` is 8.



- We see something similar happening for variable `k` and `myOtherIntAddOne`. The variable `k` in `myOtherIntAddOne` initially points to the same object that `k` in `main` is pointing to. However, a new `MyInteger` object is subsequently constructed and assigned to `k` in `myOtherIntAddOne`. The original `MyInteger` object, which `k` in `main` is still pointing to, remains unchanged.
- In the above diagram, we have visualised the object reference as a box with an arrow pointing to the object instance. More precisely, this pointer is actually an address value. For example, `k` may contain something like `0x00AB38A0`, which is the value of the memory address where the referenced `MyInteger` object is stored.

When `k` is passed as an argument to the `myOtherIntAddOne` method, what's really happening is that you're passing a copy of `0x00AB38A0` over to the variable `k` in `myOtherIntAddOne`. In other words, aside from the contents being the same, there is no link between `k` in the `myOtherIntAddOne` method and `k` in the `main` method.

When a new `MyInteger` object is constructed and assigned to `k`, `k` in `myOtherIntAddOne` will now contain the address of a brand new `MyInteger` object. However, `k` in `main` will still be containing `0x00AB38A0`, the address of the original `MyInteger` object, which still stores 7.

This is why it's incorrect to call Java either 'pass-by-reference' or 'pass-by-value'; it's more accurate to call it 'pass-reference-by-value'.

(f) Can a variable in a parameter list for a method have the same name as a member (or static)

variable in the class? If yes, how is the conflict of names resolved?

Solution: Use `this`.

```
class Example {
    int value = 5;

    public void clash(int value) {
        // Refers to the 'value' argument.
        System.out.println(value);
        // Refers to the 'value' member in the class.
        System.out.println(this.value);
    }
}
```

Problem 2. Asymptotic Analysis

This is a good time for a quick review of asymptotic big-O notation. For each of the expressions below, what is the best (i.e. tightest) asymptotic upper bound (in terms of n)?

(a) $f_1(n) = 7.2 + 34n^3 + 3254n$

Solution: $f_1(n) = O(n^3)$

The general strategy is to take the most dominant term and drop the coefficients. Here, that refers to the term with the largest degree, $34n^3$.

(b) $f_2(n) = n^2 \log n + 25n \log^2 n$

Solution: $f_2(n) = O(n^2 \log n)$

In general, a term of the form n^a for any positive constant a will be dominant over a term of the form $\log^b n$ for any positive constant b . For example, $n^{0.02}$ is dominant over $\log^{2040} n$. Hence, $n^2 \log n$ is the dominant term in $f_2(n)$.

(c) $f_3(n) = 2^{4 \log n} + 5n^5$

Solution: $f_3(n) = O(n^5)$

Here, we use the following two properties to simplify the $2^{4 \log n}$ term:

$$a^{mn} = (a^m)^n \qquad a^{\log_a b} = b$$

Using these properties,

$$\begin{aligned} 2^{4 \log n} &= (2^{\log n})^4 \\ &= n^4 \end{aligned}$$

This makes $5n^5$ the dominant term, resulting in the final answer.

(d) $f_4(n) = 2^{2n^2+4n+7}$

Solution: $f_4(n) = O(2^{2n^2+4n})$

We can move the 7 down from the exponent by doing the following:

$$2^{2n^2+4n+7} = 2^{2n^2+4n} \cdot 2^7$$

We can then remove the 2^7 coefficient. However, we cannot remove any of the other terms, neither can we remove any of the coefficients from the exponents.

For example, $2^{2n} \neq O(2^n)$. One possible intuition is that $2^{2n} = (2^n)^2$.

Problem 3. More Asymptotic Analysis!

Let f and g be functions of n where $f(n) = O(n)$ and $g(n) = O(\log n)$. Find the best asymptotic bound (if possible) of the following functions.

(a) $h_1(n) = f(n) + g(n)$

Solution:
$$\begin{aligned} h_1(n) &\leq c_1 n + c_2 \log n \\ &= O(n) \end{aligned}$$

(b) $h_2(n) = f(n) \times g(n)$

Solution:
$$\begin{aligned} h_2(n) &\leq c_1 n \cdot c_2 \log n \\ &= c_1 c_2 n \log n \\ &= O(n \log n) \end{aligned}$$

(c) $h_3(n) = \max(f(n), g(n))$

Solution:
$$\begin{aligned} h_3(n) &= \max(f(n), g(n)) \\ &= O(f(n) + g(n)) \\ &= O(n) \end{aligned}$$

(d) $h_4(n) = f(g(n))$

Solution:
$$\begin{aligned} h_4(n) &\leq c_1 g(n) \\ &\leq c_1 c_2 \log n \\ &= O(\log n) \end{aligned}$$

Note that this explanation is not very rigorous, but the intuitive idea should be there.

(e) $h_5(n) = f(n)^{g(n)}$

Solution: Trick question! The constant matters in the exponent here.

For example, if $f(n) = n$ and $g(n) = \log n$, then $h_5(n) = O(n^{\log n})$. However, if $g(n) = 2 \log n$, then $h_5(n) = O(n^{2 \log n}) = O((n^{\log n})^2)$.

Problem 4. Application of Binary Search

Given a sorted array of $n - 1$ unique integers in the range $[1, n]$, how would you find the missing element? Discuss possible naive solutions and possibly faster solutions.

Solution: The idea is to check the element in the middle of the the array first (at index $\lfloor \frac{n}{2} \rfloor$). If the element is $\lfloor \frac{n}{2} \rfloor$, then we know that all the elements in the left half of the array are present, so the missing element is in the right half, so we should recurse on that half instead, else the element we are looking at should be $\lfloor \frac{n}{2} \rfloor + 1$, which would mean the missing element should be in the left half of the array.

Problem 5. Another Application of Binary Search

You have n piles of homework and the i^{th} pile has $\text{piles}[i]$ pieces of homework. Unfortunately, you realised you have h hours left before all your homework is due. In a moment of panic, you try to figure out the rate k (that is, the “*pieces of homework*”-*per-hour*) at which you need to do your homework at in order to finish everything on time. Assume we only want integer rates. ($k \in \mathbb{Z}$, nothing fractional.)

Here’s how you plan to go about it: At every hour, you choose a pile of homework and start clearing pieces of homework from that pile. If the pile has less than k pieces of homework remaining, you decide to just finish that pile itself, and not start on the next pile yet during the same hour. We all need to take breaks after all, *right*?

To maintain your sanity, you want to minimise the number of pieces of homework you do per hour, i.e. k , while still finishing all piles of homework in time. You can assume that there exists a k that allows you to complete all piles of homework within h hours given the above constraints. Discuss how you can find the minimum integer k such that you can finish all your homework within h hours.

Here’s an example: Let’s say you had 10 hours, the piles are of size $[1, 7, 2, 3, 5]$. If we set $k = 7$, then we would take 5 hours. Each pile takes 1 hour to finish (since we finish 7 pieces per hour and there are 5 piles). 5 is less than our hour limit of 10. So $k = 7$ is viable.

If we set $k = 1$, then it would take $1 + 7 + 2 + 3 + 5 = 18$ hours (the first pile takes an hour, but the second pile takes 7, and so on), which is more than our allowance of 5 hours.

If we set $k = 3$, it would take $1 + 3 + 1 + 1 + 2 = 8$ hours. (1 hour to solve the first pile, 3 hours to solve the second pile, and so on)

If we set $k = 2$, it would take $1 + 4 + 1 + 2 + 3 = 11$ hours. (1 hour to solve the first pile, 2 hours to solve the second pile, and so on)

So minimum possible integer k in this example is 3, not 2.

A naive solution would be to try values k from 1 onwards, and checking if for that value k whether it is viable or not. If it is, we stop. Otherwise, we try for k being one larger.

Question: Is there a more efficient way to find k ? What is the algorithm?

Solution: Let's begin by identifying the possible range of values for k . We know the minimum is 1 (we have to do at least 1 piece of homework per hour in order to progress towards our goal of finishing all our homework), and the maximum is the number of pieces of homework in the largest pile (e.g. if the largest pile contains 5 pieces of homework, any k greater than 5 would not make sense as we are trying to minimize k while being subjected to the constraint of not being able to move on to a *new pile* within the same hour).

Now, with this range of possible values of k , we can turn the question into a binary search problem. Let's perform a binary search on this range of possible values for k . The idea is to check whether the value in the middle of the current range of possible values is *valid* (i.e. you can complete all piles of homework within h hours when working at this rate k , that is

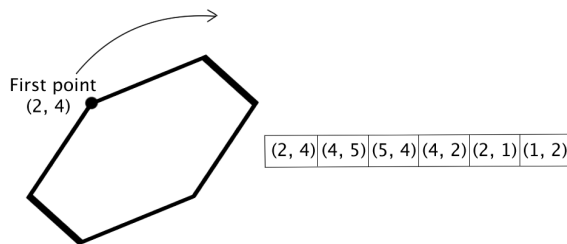
$$\sum_{i=1}^n \lceil \frac{piles[i]}{k} \rceil \leq h$$

). If the middle value is *valid*, we can attempt to recurse on the range of values in the *left half* to find a possible smaller k that is valid. If the middle value is *invalid*, then we can recurse on the range of values in the *right half* to find a bigger k that satisfies the constraints.

Problem 6. Yet Another Application of Binary Search

(Optional) Given an array of n x and y -coordinates of an n -sided convex polygon in clockwise order, find a bounding box around the polygon. Discuss possible naive solutions and possibly faster solutions. A convex polygon is a polygon where all interior angles are less than 180 degrees.

An example of such an array is shown below:



A *bounding box* is an axis-aligned rectangle (defined by 4 corner points) such that the entire polygon is contained within the box.

For the example above, the bounding box is defined by the 4 corner points: (1, 1), (1, 5), (5, 1), (5, 5).

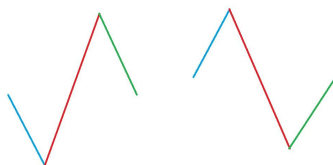
Can you come up with a solution that takes $\Theta(n)$ time? Can you come up with something faster?

Solution: The problem is equivalent to finding the minimum and maximum x and y -coordinates of the polygon. We can start off by assuming that no two consecutive points will have the same x or y -coordinate value.

The key idea to note is that a **convex** polygon's x and y -coordinates will have **at most** two “changes in direction”, i.e. be either increasing-decreasing-increasing, or decreasing-increasing-decreasing. If you start at the minimum or maximum point, then it is possible to have only a single “change in direction”, i.e. it would be either increasing-decreasing or decreasing-increasing, or have none at all, i.e. simply increasing or decreasing. This question then becomes a problem of peak-finding, similar to Q1 of Problem Set 2.

The complexity of this question is greatly reduced by the fact that the points are in a clockwise order, and that the end points are adjacent points of a closed convex polygon. This lets us skip over a lot of the edge cases that come with binary searching an array with two “changes in directions”.

Let us illustrate this with some diagrams. Suppose we plot out x -coordinates against index of the points for an arbitrary polygon. We might see a curve that's similar to the two curves below:



One important insight is that the blue and green parts of the curve do not overlap, since it is a closed convex polygon. We are thus able to verify which part of the array we are in easily as long as we take note of the first element's value, the middle element's value and the direction that the array **starts with**. For example, in an array that goes increasing-decreasing-increasing, if our binary search lands on a decreasing portion of the array, we know that we are between the global maximum and minimum. If we land on an increasing portion of the array, we know that we are on the first increasing section if the value of the middle element is greater than that of the first element, and we know that we are on the second increasing section if the value of the middle element is smaller than that of the first element. A similar argument can be made for the decreasing-increasing-decreasing case. Knowing this, we can easily binary search in the right direction to find a peak/valley. You can then find the other extreme end with another simple binary search.

As for the scenario where two consecutive points have the same x or y -coordinates—in a convex polygon, if two adjacent points have the same x -coordinates, then those two points must be the top and bottom points of a fully vertical side, and this side would be the leftmost or rightmost side of the polygon. Similarly, same y -coordinates means they are the left and right points of a fully horizontal side that is at the top or the bottom of the polygon. As such, even if the array is not composed of unique values for consecutive points, you can be sure that the binary search won't be stuck at a plateau because any plateau is guaranteed to be a peak.