

# Data structures and algorithms

## DAT037, DAT038, TDA416, TDA417

### Re-examination

Thursday, 2020-08-20, 8:30–12:30, in Canvas and Zoom

---

|                        |   |
|------------------------|---|
| <b>Examiner(s)</b>     | Peter Ljunglöf, Nick Smallbone, Pelle Evensen.  |
| <b>Allowed aids</b>    | Course literature, other books, the internet.<br><i>You are not allowed to discuss the problems with anyone!</i>  |
| <b>Submitting</b>      | Submit your answers in <i>one single</i> PDF or Word or OpenOffice document.<br><i>You can write your answers directly into the red boxes in this file.</i>                                   |
| <b>Pen &amp; paper</b> | Some questions are easier to answer using pen and paper, which is fine!<br>In that case, take a photo of your answer and paste into your answer document.<br>Make sure the photo is readable! |
| <b>Assorted notes</b>  | You may answer in English or Swedish.<br>Excessively complicated answers might be rejected.<br>Write legibly – we need to be able to read and understand your answer!                         |
| <b>Exam review</b>     | If you want to discuss the grading, please contact Peter via email.   |

There are 6 basic, and 3 advanced questions, and two points per question. So, the highest possible mark is 18 in total (12 from the basic and 6 from the advanced). Here is what you need for each grade:

- To **pass** the exam, you must get 8 out of 12 on the basic questions.
- To get a **four**, you must get 9 out of 12 on the basic questions, plus 2 out of 6 on the advanced.
- To get a **five**, you must get 10 out of 12 on the basic questions, plus 4 out of 6 on the advanced.

| Grade | Total points | Basic points | Advanced |
|-------|--------------|--------------|----------|
| 3     | $\geq 8$     | $\geq 8$     | —        |
| 4     | $\geq 11$    | $\geq 9$     | $\geq 2$ |
| 5     | $\geq 14$    | $\geq 10$    | $\geq 4$ |

## During the exam (8:30–12:30)

ID checking:

- In the beginning you will be invited to a "breakout room" in Zoom, where your ID will be checked.
- If you are kicked out of the Zoom meeting, then enter again as quickly as possible by using the same link as earlier. You will be ID checked again before being let in.

Do's and don'ts:

- You are allowed to use books or the internet.
- Don't communicate with anyone else during the exam, neither orally nor in any other format. This includes posting questions in chat forums, etc.
- Don't wear earbuds or earphones.
- Don't leave your seat.

Contacting the guard or the examiner:

- If you need to go to the toilet, inform the exam guard via the Zoom chat. Write "Toabesök" / "Toilet visit" before you go to the toilet, and "Toabesök SLUT" / "Toilet visit ENDED" when you're back.
- If you need to contact the exam guard, write "Kontakt" / "Contact" in the Zoom chat. You will be invited to a breakout Zoom room, but it might take some time.
- If you have a question for the teacher, write an email to Peter ([peter.ljunglof@cse.gu.se](mailto:peter.ljunglof@cse.gu.se)), with the subject "Question regarding reexam for ABC123" (use your own course code). Write your name and personnummer in the email.

If you finish earlier:

- You have to stay logged in to Zoom until you are permitted to leave.
- If you finish earlier, write "Scannar lösningar" / "Scanning solutions" in the chat before you scan your solutions and merge them into one single document. When your solution is submitted, write "Lämnat in i Canvas" / "Submitted to Canvas". When the exam guard has ticked you off and added ##KLAR## before your Zoom name, you can leave the Zoom meeting. But not before that!

## When the exam finishes (12:30–12:50)

- The exam finishes at 12:30 sharp. After that you are not allowed to continue working on your solutions.
- You have 20 more minutes to scan your answers and submit them to Canvas.
  - You have to submit before 12:50, otherwise you automatically fail the exam.
- Submit your answers in Canvas as a single file upload to the examination assignment.

Make sure that you:

- Submit one single document in PDF, Word or OpenOffice/LibreOffice format.
- Scanned answers should be inserted as images in the document.
- Name the document "reexam-YYYYMMDD-NNNN.pdf", with YMDN replaced by your personnummer (and the correct suffix instead of .pdf, if you use Word or OpenOffice).
- After submitting, write "Lämnat in i Canvas" / "Submitted to Canvas" in the chat.
- If you have problems submitting to Canvas or creating a single document in the correct format, you may send your solutions in an email to Peter ([peter.ljunglof@cse.gu.se](mailto:peter.ljunglof@cse.gu.se)). Use the subject "Reexam solutions for ABC123" (use your own course code), and write your name and personnummer in the email.
  - You must still do this before 12:50!

# Question 1: Lagrange's four square theorem

In mathematics, *Lagrange's four-square theorem* states that every natural number  $n$  can be written as the sum of four squares. That is,  $n = a^2 + b^2 + c^2 + d^2$  where  $a, b, c$  and  $d$  are integers  $\geq 0$ . For example,  $42 = 6^2 + 2^2 + 1^2 + 1^2$ , and  $89 = 7^2 + 6^2 + 2^2 + 0^2$ .

[https://en.wikipedia.org/wiki/Lagrange%27s\\_four-square\\_theorem](https://en.wikipedia.org/wiki/Lagrange%27s_four-square_theorem)

The following program takes as input a natural number  $n$ , and checks that Lagrange's four-square theorem holds for all numbers up to  $n$ . It first calculates which numbers can be written as the sum of *two* squares  $a^2 + b^2$ , and then checks that all numbers up to  $n$  can be written as the sum of two numbers of the form  $a^2 + b^2$ .

```
boolean checkLagrangeTheorem(int n):
    boolean[] isSumOfTwoSquares = new boolean[n]

    for i in 0..n-1:
        isSumOfTwoSquares[i] = false

    // which numbers can be written as the sum of two squares
    for i in 0..floor(sqrt(n)):
        // note: floor(x) rounds x down to the nearest integer
        // e.g. floor(3.7) = 3
        for j in 0..floor(sqrt(n)):
            if i*i + j*j < n:
                isSumOfTwoSquares[i*i + j*j] = true

    for i in 0..n-1:
        boolean ok = false
        // check if i can be written as j+k,
        // where j and k are the sum of two squares
        for j in 0..i:
            k = i-j
            if isSumOfTwoSquares[j] and isSumOfTwoSquares[k]:
                ok = true
        if not ok:
            // i cannot be written as a sum of four
squares
            return false

    // all numbers in the range 0..n-1 can be written
    // as the sum of two squares
    return true
```

### Part A

What is the worst-case time complexity of this program, as a function of  $n$ ? Give your answer either using  $O$ -notation or using order of growth. You may assume that all arithmetic operations (including square root) take a constant amount of time.

(Replace this text with your answer)

### Part B

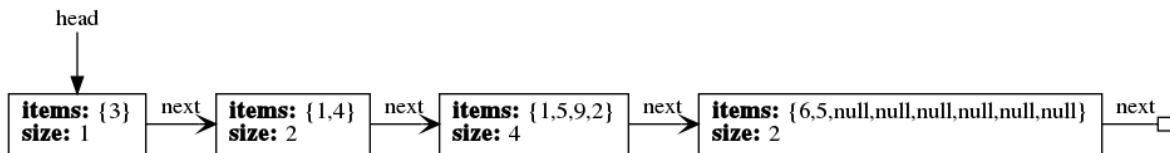
When you run the program for  $n = 1,000,000$ , it takes one minute. If you plan to run the program for one day, roughly how big a value of  $n$  will you be able to check? Explain your answer, with reference to the complexity you calculated in part A.

(Replace this text with your answer)

## Question 2: Linked dynamic arrays

A *linked dynamic array* is a data structure that stores a sequence of items, and supports getting and setting the  $i$ th item, and adding an item to the end of the sequence. Its design combines ideas from *linked lists* and *dynamic arrays*.

The data structure consists of a *linked list* of arrays. The last array in the list behaves specially: like a dynamic array, it can have empty space at the end. The other arrays in the list have no empty space. For example, here is a linked dynamic array containing the items 3, 1, 4, 1, 5, 9, 2, 6, 5:



Each node in the linked list has two data fields: *items*, an array of items, and *size*, which records how many items are stored in the array. In this example, the last array has six empty spaces since its *size* field is 2 but its array has a capacity of 8.

Notice that the arrays do not all have the same capacity. In a linked dynamic array, the first array always has a capacity of 1, and each array in the rest of the list is exactly twice as long as the previous array. In the example above, the array capacities go 1, 2, 4, 8.

Here is a partial implementation of linked dynamic arrays in a Java-like pseudocode. The class `LinkedDynamicArray` has just one field, `head`, which stores the first node in the linked list.

```
class LinkedDynamicArray<Elem>:
    // A class representing a linked list node
    class Node:
        Elem[] items // The array of items stored in this node
        int size      // The portion of the array that is used
        Node next    // The next node in the list

        // The constructor: returns a node containing
        // an array of size 'capacity'.
        Node(int capacity):
            items = new Elem[capacity]
            next = null
            size = 0

    // The first node in the list
    Node head

    // The constructor: returns a new empty linked dynamic array
    LinkedDynamicArray():
        head = new Node(1)

    // Return the number of items in the sequence
    int size():
        // To be implemented in part A of this question
```

```
// Return the item stored at index i
Elem get(int i):
    // To be implemented in question 7

// Add an item to the end of the sequence
void add(Elem item):
    // To be implemented in question 7
```

## Part A

Starting from the code above, implement the method `size()` which returns the number of items in the sequence (e.g. for the example above of 3,1,4,1,5,9,2,6,5, it should return 9).

Write your answer either as pseudocode or code in a programming language of your choice. Pseudocode should be in **full detail**, like the pseudocode above. You may only use functionality that exists in the code above or that you implement yourself. In particular, you may not assume that an iterator has been defined for `Node`.

```
int size():
    // Replace this comment with your answer
```

## Part B

What is the worst-case time complexity of `size` in terms of  $n$ , the number of items in the sequence?

**Hint:** Think about the relationship between  $n$  and the number of nodes in the list.

(Replace this text with your answer)

## Question 3: An unknown sorting algorithm

Consider the following in-place sorting algorithm:

```
void sortInPlace(int[] A):  
    for i in 1 ... A.length-1:  
        int x = A[i]  
        int j = i - 1  
        while A[j] > x:  
            A[j+1] = A[j]  
            j = j - 1  
            if j < 0:  
                break  
        A[j+1] = x
```

### Part A

Which well-known sorting algorithm is this? Explain why.

(Replace this text with your answer)

### Part B

Get an individualised list of numbers from here: <http://www.cse.chalmers.se/~evensen/exam20200820.html>

Arrange the numbers you got in an order that yields the greatest number of comparisons when running the above algorithm. That is, you want the condition  $A[j] > x$  to be executed as many times as possible.

(Replace this text with your list of numbers)

Motivate why the order you have given is the worst possible.

(Replace this text with your answer)

## Question 4: Hash tables

Get an individualised list of elements and hash values from here:

<http://www.cse.chalmers.se/~evensen/exam20200820.html>

Insert the elements you got in the order given and with the hash value given after the colon, into an initially empty hash table.

Example: [B:1, C:2, A:0] means that the following elements should be inserted: B (with hash value 1), C (with hash value 2), and A (with hash value 0), in that order.

### Part A

Insert the elements in a *separate chaining* hash table (“open hashing” / “linear chaining”).

Let the size of the hash table array be  $m = 5$ .

(Replace this text with the final hash table)

### Part B

Insert the elements in an *open addressing* hash table (“closed hashing” / “linear probing”).

Let the size of the hash table array be  $m = 10$ , and assume linear probing with **probe interval 7**, i.e.:

$$h' = (h + 7) \% m$$

(Replace this text with the final hash table)

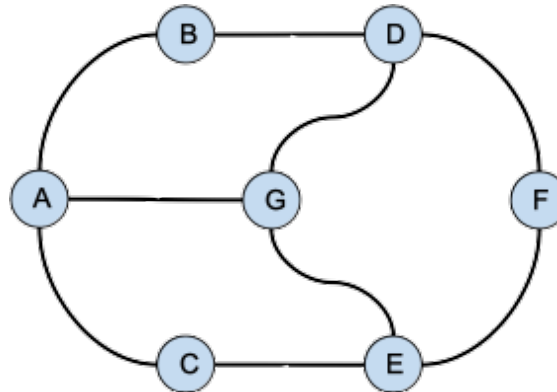
Explain what collisions you had and how each collision was resolved:

(Replace this text with your answer)



## Question 5: Make a graph directed

Assume the following undirected and unweighted graph:



Get a set of possible weights that you are allowed to use:

<http://www.cse.chalmers.se/~evensen/exam20200820.html>

Write down the set of possible weights that you got:

(Replace this text with set of weights)

### Part A

Add directions and assign weights from the possible weights (but no new edges) to the graph so that the following holds:

- all edges must be directed, and all weights must be unique (different from each other)
- there are an infinite number of possible paths from A to G
- the cheapest path from A to G (i.e., the path with the smallest total weight) is not the shortest path (i.e., the path passing through the fewest edges)
- if you remove the directions, the minimal spanning tree (MST) should have **total cost at least 50**

**Important:** all weights must come from the set of weights you got from the link above.

Each weight must only be used once, but you do not have to use all weights.

(Insert your graph here. It can be an image (e.g., a scanned/photographed hand-drawn graph, or created on the computer), or it can be drawn using ASCII art (then please change to a fixed-width font), or as a list of weighted edges (on the form “A→B[ 42 ]”).

### Part B

Run Dijkstra’s algorithm on your final graph, starting at node A. In which order are the nodes visited, and what is the calculated cost to reach them? (We say that a node is visited when it is removed from the priority queue). Also, show the contents of the priority queue, after the visited node has been processed.

Fill this table with the missing information:

[illegible]

## Question 6: AVL trees or red-black trees

**Important:** There are two questions n:o 6 – one about AVL trees and one about red-black trees. You can choose which one you want to answer.

**Grading of this question:** If you answer both questions, your points on question 6 will be the *maximum*, not the sum! This means that if you get 1 point on the AVL question and 1 point on the red-black question, you will get 1 point in total.

**In short:** Select one of the two to answer. Skip the other one.

## Question 6, alternative 1: AVL trees

Get an individualised binary search tree from here:

<http://www.cse.chalmers.se/~evensen/exam20200820.html>

### Part A

Annotate the nodes in the tree with their AVL balance factor. The balance factor of a node is defined as the height of the right subtree minus the height of the left subtree.

(Insert your AVL tree here. It can be an image (e.g., a scanned/photographed hand-drawn graph, or created on the computer), or it can be drawn using ASCII art (then please change to a fixed-width font).

### Part B

Take the AVL tree you created earlier and *transform* the tree using left and/or right *tree rotations* such that we get an unbalanced tree containing a so-called *left-left* case. You should show the resulting tree with updated balance factors, and explain which rotations you have used.

Insert your unbalanced tree here:

(Replace this with your unbalanced tree)

Explain which rotations you used and in which order:

(Write your explanation)

## Question 6, alternative 2: Red-black trees

Get an individualised binary search tree from here:

<http://www.cse.chalmers.se/~evensen/exam20200820.html>

### Part A

Colour the nodes in the tree red or black, so that the tree becomes a correctly balanced red-black tree, where no nodes should have two red children, and there is at least one red leaf node. *Do not transform the tree*, just colour the nodes! (Of course, this is not possible for all BSTs, but it is for the one you get)

**Important:** Don't forget the two additional restrictions: (1) no nodes should have two red children, and (2) at least one leaf node must be red. (This is to be able to do part B below.)

(Insert your red-black tree here. It can be an image (e.g., a scanned/photographed hand-drawn graph, or created on the computer), or it can be drawn using ASCII art (then please change to a fixed-width font).

### Part B

Insert a new red node into the tree, using the normal BST insertion algorithm, so that the final red-black tree is unbalanced, and the new red node is a **right child of another red node**.

**Important:** The new red node must be a right child of another red node.

(Insert your unbalanced red-black tree here)

Then rebalance the tree, so that you have a correct red-black tree again.

(Insert your rebalanced red-black tree here)

### Part C

Look at the tree you got from part B. If the tree has no red leaf nodes, you should skip part C.

Otherwise, repeat the exact same process as in part B. Insert a new red node, using the normal BST insertion algorithm, so that the final red-black tree is unbalanced, and the new red node is a right child of another red node.

(Insert your unbalanced red-black tree here)

Then rebalance the tree, so that you have a correct red-black tree again.

(Insert your rebalanced red-black tree here)

## Question 7 (advanced): Linked dynamic arrays, part II

This question builds on the linked dynamic array data structure defined in question 2. Please read that question before you attempt this one.

In this question you will implement two operations on linked dynamic arrays: getting the item at index  $i$ , and adding an item to the end of the sequence.

Write your answer either as pseudocode or code in a programming language of your choice. Pseudocode should be in **full detail**. You may only use functionality that exists in the `LinkedDynamicArray` class or that you implement yourself.

### Part A

Implement a method `get ( )` which returns the  $i$ th item in the sequence, counting from 0.

For example, if `x` is the sequence 3,1,4,1,5,9,2,6,5 from question 2, then `x.get ( 2 )` should return 4.

```
Elem get(int i):  
    // Replace this comment with your answer
```

What is the worst-case complexity of `get ( )` in terms of  $n$ , the total number of items in the sequence?

(Replace this text with your answer)

### Part B

To add an item to the end of a linked dynamic array, we first check if there is free space in the last array of the list. If there is, we store the new item in that space. Otherwise, we add a new array of *twice* the capacity to the linked list, and store the item there.

Implement a method `add ( )` which adds a new item to the end of the sequence.

Your method should take (where  $n$  is the length of the sequence):

- $O(\log n)$  time when the last array of the list is not full
- $O(n)$  time when the last array is full  
(assume that creating an array of size  $n$  takes  $O(n)$  time).

Recall that you can use `arr.length` to get the capacity of an array `arr` in Java.

```
void add(Elem item):  
    // Replace this comment with your answer
```

Explain why your method has this complexity.

(For the interested reader: this implies that the method takes amortised logarithmic time overall.)

(Replace this text with your answer)

### **Part C**

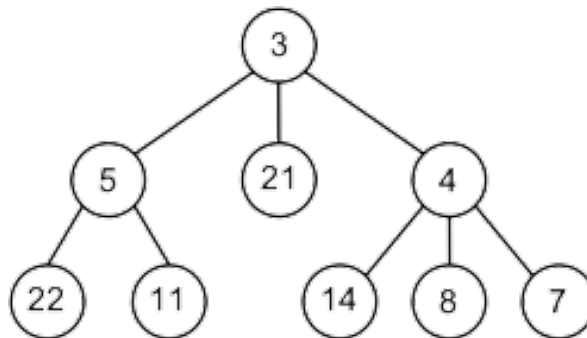
Is there a way to augment the class to get constant time complexity (amortised) for the `add ( )` method? You don't have to write any code, just explain what changes/additions you would do.

(Replace this text with your answer)

## Question 8 (advanced): Rose heaps

A “rose tree” is a tree where the nodes can have any number of children. Since we don’t know how many children a given node has, we cannot use a fixed number of children pointers (such as “left” and “right” for binary trees). One way to implement the children of a node is as a linked list. (Another possibility is as an array, but that’s not what we’re going to do here).

This question is about defining a new kind of heap that is based on a rose tree. The idea behind this heap is to have multiple children instead of just two (as is the case in a binary heap). A rose heap must satisfy the heap property: the value of each node is less than (or equal to) the value of its children. The following is an example of a rose heap:



An implementation for such a heap might look like this (we use a linked list to implement the children):

```
class RoseHeap<Elem>:
    private class Node:
        Elem value
        LinkedList<Node> children
        Node(Elem val): // The constructor for Node
            value = val
            children = new LinkedList<Node>()
    private Node root

    public Elem findMin():
        return root.value

    public void insert(Elem value):
        insertNode(new Node(value))

    public void insertNode(Node node):
        // To be implemented in part A
        if root == null:
            // What to do if the root is null?
        else if node.value < root.value:
            // What to do if the node is smaller than the root?
        else:
            // What to do if the root is smaller (or equal)?

    public Elem deleteMin():
        // To be implemented in part C
```



## Part A

Complete the implementation of `insertNode()`:

```
void insertNode(Node node):  
    // Replace this comment with your answer
```

## Part B

Get an individualised list of numbers from here: <http://www.cse.chalmers.se/~evensen/exam20200820.html>

(Replace this text with the list of numbers, in the order they arrived)

Build a new rose heap by hand, from the list of numbers you got.  
Add all elements to an initially empty heap, in the same order as they are given.

(Insert your rose heap here. It can be an image (e.g., a scanned/photographed hand-drawn graph, or created on the computer), or it can be drawn using ASCII art (then please change to a fixed-width font).

## Part C

Complete the implementation of `deleteMin()`:

```
Elem deleteMin():  
    // Replace this comment with your answer
```

## Part D

What is the worst-case complexity of your implementation of `deleteMin()`?

(Replace this text with your answer)

How does a rose heap compare to other heap implementations, such a binary heap or a skew heap? Are there any advantages or disadvantages?

(Replace this text with your answer)

## Question 9 (advanced): Paths of a given length

The following class can be used to implement binary search trees:

```
class BST<Elem>:
    private class Node:
        Node left, right
        Elem value
    Node root
```

Your task is to write a method that returns *all paths from the root to any leaf, of a given length*.

A path is here a list of values, i.e., the values that you encounter when you move along the path.

```
class BST<Elem>:
    ...
    List<Elem[]> pathsOfLength(int len):
        // Returns all paths from the root to a leaf,
        // of the given length 'len',
        // where a path is an array of values
```

Your solution should only visit a node at most once. Also, you should not visit more nodes than necessary (i.e., don't visit nodes that cannot possibly lead to a solution).

```
List<Elem[]> pathsOfLength(int len):
    // Replace this comment with your answer
```

Explain why your solution doesn't visit any node more than once:

(Replace this text with your answer)

What unnecessary nodes are not visited by `pathsOfLength()`?

(Replace this text with your answer)

You can assume that the class `List` has the following methods, and that all of them are efficient (i.e., constant time complexity, except `toArray()` which is linear).

```
class List<Elem>:
    List()    // The constructor

    Elem[] toArray()    // Returns a copy of the list as an array
    boolean isEmpty()
    int size()

    void addFirst(Elem e)
    Elem getFirst()
    Elem deleteFirst()

    void addLast(Elem e)
    Elem getLast()
    Elem deleteLast()
```

(The alert reader notices that this is more like Java's `Deque` interface than the `List` interface, but that's not really important for the task at hand).