

# CPXA: Exercises for Tutorial Classes

October 2024–January 2025. EPITA. ING1

Send any mistake to Alexandre Duret-Lutz <adl@lrde.epita.fr>.

## INEFFECTIVE SORTS

```
DEFINE HALFHEARTEDMERGESORT(LIST):  
  IF LENGTH(LIST) < 2:  
    RETURN LIST  
  PIVOT = INT(LENGTH(LIST) / 2)  
  A = HALFHEARTEDMERGESORT(LIST[:PIVOT])  
  B = HALFHEARTEDMERGESORT(LIST[PIVOT:])  
  // UMMMMM  
  RETURN [A, B] // HERE. SORRY.
```

```
DEFINE FASTBOGOSORT(LIST):  
  // AN OPTIMIZED BOGOSORT  
  // RUNS IN O(N LOG N)  
  FOR N FROM 1 TO LOG(LENGTH(LIST)):  
    SHUFFLE(LIST):  
    IF ISSORTED(LIST):  
      RETURN LIST  
  RETURN "KERNEL PAGE FAULT (ERRDR CODE: 2)"
```

```
DEFINE JOBINTEVIEWQUICKSORT(LIST):  
  OK SO YOU CHOOSE A PIVOT  
  THEN DIVIDE THE LIST IN HALF  
  FOR EACH HALF:  
    CHECK TO SEE IF IT'S SORTED  
    NO, WAIT, IT DOESN'T MATTER  
    COMPARE EACH ELEMENT TO THE PIVOT  
    THE BIGGER ONES GO IN A NEW LIST  
    THE EQUAL ONES GO INTO, UH  
    THE SECOND LIST FROM BEFORE  
  HANG ON, LET ME NAME THE LISTS  
  THIS IS LIST A  
  THE NEW ONE IS LIST B  
  PUT THE BIG ONES INTO LIST B  
  NOW TAKE THE SECOND LIST  
  CALL IT LIST, UH, A2  
  WHICH ONE WAS THE PIVOT IN?  
  SCRATCH ALL THAT  
  IT JUST RECURSIVELY CALLS ITSELF  
  UNTIL BOTH LISTS ARE EMPTY  
  RIGHT?  
  NOT EMPTY, BUT YOU KNOW WHAT I MEAN  
  AM I ALLOWED TO USE THE STANDARD LIBRARIES?
```

```
DEFINE PANICSORT(LIST):  
  IF ISSORTED(LIST):  
    RETURN LIST  
  FOR N FROM 1 TO 10000:  
    PIVOT = RANDOM(0, LENGTH(LIST))  
    LIST = LIST[PIVOT:] + LIST[:PIVOT]  
    IF ISSORTED(LIST):  
      RETURN LIST  
  IF ISSORTED(LIST):  
    RETURN LIST  
  IF ISSORTED(LIST): // THIS CAN'T BE HAPPENING  
    RETURN LIST  
  IF ISSORTED(LIST): // COME ON COME ON  
    RETURN LIST  
  // OH JEEZ  
  // I'M GONNA BE IN SO MUCH TROUBLE  
  LIST = [ ]  
  SYSTEM("SHUTDOWN -H +5")  
  SYSTEM("RM -RF ./")  
  SYSTEM("RM -RF ~/*")  
  SYSTEM("RM -RF /")  
  SYSTEM("RD /S /Q C:\*") // PORTABILITY  
  RETURN [1, 2, 3, 4, 5]
```

# Tutorial 1

## 1.1 Small Loops

How many lines are printed by the following loops? Give your answer as a function of  $N$ .

- ```
1. for (int i = 0; i < N; ++i)
    for (int j = 1; j <= N; ++j)
        puts("Loop 1");
```
- ```
2. for (int i = 0; i < N; ++i)
    for (int j = N/2; j > 0; --j)
        puts("Loop 2");
```
- ```
3. for (int i = 0; i < N; ++i)
    for (int j = i; j >= 0; j--)
        puts("Loop 3");
```
- ```
4. for (int i = 0; i < N; ++i)
    for (int j = 1; j < N; j *= 2)
        puts("Loop 4");
```

Hint: rewrite the inner loop so it increments by 1 while still performing the same number of iterations.

- ```
5. for (int i = 0; i < N; ++i)
    for (int j = i; j < N - 2; j++)
        puts("Loop 5");
```
- ```
6. for (int i = 0; i < N; ++i)
    for (int j = i; j >= 0; j -= 2)
        puts("Loop 6");
```

Hint: rewrite the inner loop so it increments by 1. Then distinguish between odd and even  $N$ .

## 1.2 House

How many '#' are printed by 'house( $n$ )'? Give your answer as a function of  $n$ .

```
1 void house(unsigned n)
2 {
3     for (unsigned y = 0; y < n; ++y)
4         {
5             for (unsigned x = 1; x < n - y; ++x)
6                 putchar(' ');
7             for (unsigned x = n - y; x <= n + y; ++x)
8                 putchar('#');
9             putchar('\n');
10        }
11    for (unsigned y = 1; y < n; ++y)
12        {
13            for (unsigned x = 1; x < 2 * n; ++x)
14                putchar('#');
15            putchar('\n');
16        }
17 }
```

## 1.3 Simple Computations

- Sum the first 1000 odd natural numbers.
- Compute  $\log_{10}(42)$  to two decimal places by hand. You may use a pocket calculator, but assume that its  $\boxed{\log}$  key is broken and cannot be used.
- \* Compute  $\log_2(42)$  similarly.

## 1.4 Binary Trees

Consider non-empty *full* binary trees (i.e., *internal nodes* always have two children, and *leaf nodes* have no children). Let  $h$  denote the *height* of the tree (a single-node tree has  $h = 0$ ),  $n_i$  the number of internal nodes,  $\ell$  the number of leaves, and  $n = n_i + \ell$  the total number of nodes.

- Express  $n_i$  as a function of  $\ell$ .
- Give lower and upper bounds for:
  - $\ell$  as a function of  $h$ ,
  - $n_i$  as a function of  $h$ ,
  - $n$  as a function of  $h$ .

Keep in mind that (even full) binary trees are not always *balanced*.

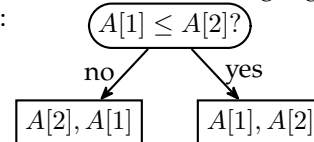
- Deduce some lower bounds for:
  - $h$  as a function of  $\ell$ ,
  - $h$  as a function of  $n_i$ ,
  - $h$  as a function of  $n$ .

Since  $h$  is an integer, use  $\lfloor \cdot \rfloor$  or  $\lceil \cdot \rceil$  when appropriate.

## 1.5 Decision Tree \*

Let's construct a tree that represents some sorting algorithm for an array of size  $n$ . Each internal node is labeled by a question of the form " $A[i] \leq A[j]$ ?" for some given  $i$  and  $j$ . The left child is used when the answer is negative, while the right child is used when the answer is positive. Each leaf is labeled by a possible answer of the sorting algorithm, i.e., the order in which values should be rearranged to be sorted. Several leaves may have identical labels.

For instance here is a sorting algorithm for an array of 2 values:



- Devise a decision tree for sorting an array of 3 values.
- What is the minimum number of leaves needed to sort an array of  $n$  values?
- Deduce the minimal height of a decision tree for sorting  $n$  values.
- What does this teach us about the complexity of *comparison sorting algorithms* (i.e., sorting algorithms based on comparisons of the values to sort)?

## 2.1 Binary Search

`BINARYSEARCH(A, b, e, v)` takes a sorted array  $A$ , and looks up a value  $v$  in the semi-open range of  $A$  delimited by indices  $b \geq 0$  (included if  $b < e$ ) and  $e$  (always excluded).

If there exists an index  $p$  such that  $b \leq p < e$  and  $A[p] = v$ , `BINARYSEARCH` returns  $p$ . If there is no such index, `BINARYSEARCH` returns the index  $i$  of the value before which  $v$  should be inserted if we wanted to add this value while keeping  $A$  ordered. In other words, when  $v$  is not found in  $A$  between  $b$  and  $e$ , the returned value  $i$  is such that:

$$\begin{cases} b = e \text{ or } v < A[b] & \text{if } i = b \\ A[i - 1] < v < A[i] & \text{if } b < i < e \\ b = e \text{ or } A[e - 1] < v & \text{if } i = e \end{cases}$$

`BINARYSEARCH(A, b, e, v)`

```

1  if b < e then
2      m ← ⌊(b + e)/2⌋
3      if v = A[m] then
4          return m
5      else
6          if v < A[m] then
7              return BINARYSEARCH(A, b, m, v)
8          else
9              return BINARYSEARCH(A, m + 1, e, v)
10 else
11     return b
    
```

1. Prove that  $\lfloor (b + e)/2 \rfloor = b + \lfloor (e - b)/2 \rfloor$  when  $b$  and  $e$  are natural numbers.
2. Although the previous equality is true from a mathematical standpoint, it does not hold anymore if the operations are implemented using a low level type such as `unsigned int`. Suggest two `unsigned int` values  $a$  and  $b$  verifying  $a < b$  but such that `assert((a+b)/2 == (a + (b-a)/2))` would signal a problem. Which one of these two expressions should you use when implementing this algorithm?
3. Give an **upper bound** of  $T(n)$ , the time complexity of `BINARYSEARCH` expressed as a function of the array size  $n = e - b$ . Your expression should be recursive, involving  $T(\lfloor n/2 \rfloor)$ .
4. Solve this recursive expression by substituting  $T(\dots)$  by its definition until you reach  $T(1)$ . You may assume that  $n$  is a power of 2.
5. Are the recursive calls to `BINARYSEARCH` in this function *tail calls*?
6. Perform a *tail call elimination* in `BINARYSEARCH` as if you were a compiler (i.e., using `gotos`), then refactor it to be more readable to the human (getting rid of the `gotos`).
7. What is the complexity of your non-recursive version of `BinarySearch`?

## 2.2 Insertion Sort + Binary Search

Let's consider again the insertion sort as seen in the lecture, where  $A$  is an array of  $n$  objects (with indices starting at 0) that can be compared.

`INSERTIONSORT(A, n)`

```

1  for i ← 1 to n - 1 do
2      key ← A[i]
3      j ← i - 1
4      while j ≥ 0 and A[j] > key do
5          A[j + 1] ← A[j]
6          j ← j - 1
7      A[j + 1] ← key
    
```

1. What is the number of comparisons performed in the worst and best cases? We only want to count comparisons between objects, i.e., the number of times  $A[j] > key$  is executed.
2. What is the total number of assignments of objects? (Lines 2, 5, and 7.)
3. Write a variant of `INSERTIONSORT` that uses `BINARYSEARCH` to find where to insert  $key$ , and then shift all objects to the right of this position to make some room for the new value.
4. \* What is the number of comparisons and assignments in this new algorithm? (Did you also account for the number of comparisons in `BINARYSEARCH`?)

## 2.3 The $\Theta$ Notation

By definition

$$\Theta(g(n)) = \left\{ f(n) \mid \begin{array}{l} \exists c_1 > 0, \exists c_2 > 0, \exists n_0 \in \mathbb{N}, \\ \forall n \geq n_0, c_1 g(n) \leq f(n) \leq c_2 g(n) \end{array} \right\}$$

For convenience, we often write  $f(n) = \Theta(g(n))$  instead of  $f(n) \in \Theta(g(n))$ . Similarly  $n^3 + \Theta(n^2)$  should be understood as some formula for the form " $n^3 + f(n)$ " where  $f(n) \in \Theta(n^2)$ .

1. Prove that for any two positive functions  $g_1$  and  $g_2$  we have:
  - a)  $\Theta(g_1(n)) + \Theta(g_2(n)) = \Theta(g_1(n) + g_2(n))$
  - b)  $\Theta(g_1(n)) + \Theta(g_2(n)) = \Theta(\max(g_1(n), g_2(n)))$
  - c)  $\Theta(g_1(n)) \cdot \Theta(g_2(n)) = \Theta(g_1(n) \cdot g_2(n))$
2. \* **Disprove** the following properties:
  - a)  $g_1(n) = \Theta(g_1(n/2))$
  - b)  $g_1(n) = \Theta(g_2(n)) \iff \log_2 g_1(n) = \Theta(\log_2 g_2(n))$

## 2.4 Ternary Search \*

Propose an implementation of `TERNARYSEARCH` and study its complexity.

## 🎀 Tutorial 3 🎀

### 3.1 Using the “master theorem”

Let us recall the *master theorem* from the lecture.

Given a recurrence equation such as  $T(n) = aT(n/b) + O(1) + f(n)$  with  $a \geq 1$ ,  $b > 1$ .

- If  $f(n) = O(n^{(\log_b a) - \epsilon})$  for some  $\epsilon > 0$ , then  $T(n) = \Theta(n^{\log_b a})$ .
- If  $f(n) = \Theta(n^{\log_b a})$ , then  $T(n) = \Theta(n^{\log_b a} \log n)$ .
- If  $f(n) = \Omega(n^{(\log_b a) + \epsilon})$  for some  $\epsilon > 0$ , and if  $af(n/b) \leq cf(n)$  for some  $c < 1$  and all large values of  $n$ , then  $T(n) = \Theta(f(n))$ .
- In other cases, the theorem does not apply.

Use this theorem to solve the following equations:

1.  $T(n) = 3T(n/3) + \Theta(1)$
2.  $U(n) = 2U(n/3) + \Theta(1)$
3.  $V(n) = V(n/2) + n + 2$
4.  $W(n) = W(\lfloor n/2 \rfloor) + W(\lceil n/2 \rceil) + \Theta(\log n)$
5.  $X(n) = 2X(n/2) + \Theta(n \log n)$

### 3.2 Matrix Addition

1. Write an iterative algorithm,  $\text{ADD}(A, B, n)$ , that sums two  $n \times n$  matrices.
2. What is the complexity of your algorithm? (The master theorem is useless here.)
3. Argue that it is impossible to do better.

### 3.3 Naive Block Matrix Multiplication

Input: two  $n \times n$  matrices  $A$  and  $B$  where  $n$  is a power of 2.

Output: the  $n \times n$  matrix  $C = A \times B$ .

```

BMUL(A, B, n)
0  if n = 1 then
1      C[0][0] ← A[0][0] × B[0][0]
2      return C
3      ( A11  A12 ) ← A           // slice A in four n/2 × n/2 blocks
   ( A21  A22 )
4      ( B11  B12 ) ← B
   ( B21  B22 )
5  C11 ← ADD(BMUL(A11, B11, n/2), BMUL(A12, B21, n/2), n/2)
6  C12 ← ADD(BMUL(A11, B12, n/2), BMUL(A12, B22, n/2), n/2)
7  C21 ← ADD(BMUL(A21, B11, n/2), BMUL(A22, B21, n/2), n/2)
8  C22 ← ADD(BMUL(A21, B12, n/2), BMUL(A22, B22, n/2), n/2)
9  return ( C11  C12 )
   ( C21  C22 )
    
```

1. Give a recursive equation for the complexity of this algorithm, in function of  $n$ . (Assume that slicing matrices and grouping sub-matrices is done using copies, so this is not constant time.)
2. Use the master theorem to find the complexity.

3. How do you think this algorithm compares to the naive multiplication (done with a triple loop)?

### 3.4 Strassen’s Block Matrix Multiplication

Same questions for the following algorithm, where SUB performs the subtraction of two matrices.

```

SMUL(A, B, n)
0  if n = 1 then
1      C[0][0] ← A[0][0] × B[0][0]
2      return C
3      ( A11  A12 ) ← A
   ( A21  A22 )
4      ( B11  B12 ) ← B
   ( B21  B22 )
5  M1 ← SMUL(ADD(A11, A22, n/2), ADD(B11, B22, n/2), n/2)
6  M2 ← SMUL(ADD(A21, A22, n/2), B11, n/2)
7  M3 ← SMUL(A11, SUB(B12, B22, n/2), n/2)
8  M4 ← SMUL(A22, SUB(B21, B11, n/2), n/2)
9  M5 ← SMUL(ADD(A11, A12, n/2), B22, n/2)
10 M6 ← SMUL(SUB(A21, A11, n/2), ADD(B11, B12, n/2), n/2)
11 M7 ← SMUL(SUB(A12, A22, n/2), ADD(B21, B22, n/2), n/2)
12 C11 ← SUB(ADD(M1, M4, n/2), ADD(M5, M7, n/2), n/2)
13 C12 ← ADD(M3, M5, n/2)
14 C21 ← ADD(M2, M4, n/2)
15 C22 ← ADD(SUB(M1, M2, n/2), ADD(M3, M6, n/2), n/2)
16 return ( C11  C12 )
   ( C21  C22 )
    
```

Note that you do not need to understand why the algorithm is correct to compute its complexity.

### 3.5 Fast Exponentiation

This computes  $x^p$  for  $p \in \mathbb{N}$ .

```

FASTPOWER(x, p)
1  if p = 0 then
2      return 1
3  if odd(p) then
4      return x × FASTPOWER(x × x, ⌊p/2⌋)
5  else
6      return FASTPOWER(x × x, p/2)
    
```

1. Assuming the multiplication used here is done in constant time, what are the worst-case and best-case complexities as a function of  $p$ ? Deduce the complexity in the general case.
2. This algorithm can be applied to any type that has an associative operation (here  $\times$ ) with a neutral element (here 1), i.e., any type that has a monoid structure. Suggest a variant of FASTPOWER for computing the  $p$ th power of a  $n \times n$  matrix using SMUL, and compute its complexity as a function a  $p$  and  $n$ .
3. Can you see how to use FASTPOWER to compute the product of two integers on a computer whose multiplication operation is broken?

## Tutorial 4

Today we compress files by just changing how each character is encoded at the bit level.

### 4.1 Histogram and Fixed-Size Encodings

Let us assume that the file to compress is represented as an array  $data[0..n-1]$  of  $n$  8-bit characters.

1. Suggest an algorithm to compute the histogram of characters in  $data$ . I.e., the function  $HIST(data, n)$  should return an array of length 256 giving the number of occurrences of each letter.
2. Give the complexity of  $HIST$  as a function of  $n$ .
3. What is the complexity of computing the number of different characters occurring in  $data$  once the histogram is built?
4. How many bits are necessary to distinguish  $p$  characters? Call this value  $b$ .
5. We decide that the compressed file we will output starts with 256 bits indicating for each possible character whether it was used in the input. (These 256 bits are enough to find  $p, b$ , and the way each character is encoded.) This header is followed by the  $n$  characters, each encoded on  $b$  bits. Compute the compression ratio as a function of  $n$  and  $p$ . What is its limit when  $n \rightarrow \infty$ ?

### 4.2 Prefix Code

We now consider a variable-length code: each character can be encoded using a different number of bits. Obviously we want frequent characters to use short codes, while long codes should be reserved to infrequent characters.

Also our encoding must have the *prefix property*: no code should be prefix of another code. If we do not respect this property, for instance encoding "a" with 11, and "b" with 111, then we cannot tell if 11111 encodes "ab" or "ba".

1. Consider the following encoding:  
letter a b c d e f  
code 0 101 100 111 1101 1100  
Decode 11011100110001001101.
2. Suggest an algorithm that takes one such "code table" and a string of  $n$  bits as input, and outputs the decoded string.
3. How many encodings with the prefix property exist for  $p$  letters?
4. Assuming  $p > 1$ , argue that if one letter is encoded with  $p$  bits or more, there must exist a better encoding.
5. Let's look for an *optimal* encoding. Assume the file to compress has 10000 characters chosen between abcdefg with the following frequencies:

letter	a	b	c	d	e	f	g
freq.	15%	42%	7%	11%	13%	9%	3%

Suggest an encoding of these letters such that the resulting size of all encoded letters is 3062.5 bytes (multiple solutions exist, but you cannot do less on this example).

### 4.3 Huffman

The HUFFMAN algorithm computes an optimal encoding by representing any encoding as a tree whose leaves are the letters to encode. The code of a letter can be deduced from the path from the root to the leaf labeled by this letter: each left branch represent a 0, and each right branch is a 1. Furthermore, each node of the tree has an attribute *freq* that is the sum of the frequencies of all the letters below.

Lines 1–5 the algorithm build a forest  $F$  in which each tree corresponds to a character occurring in the file to compress.

The two trees whose roots have the smallest frequencies are then combined, and the operation is repeated lines 6–11 until there is only one tree left in the forest.

```
HUFFMAN(letters, freqs, p)
1  for  $i \leftarrow 0$  to  $p - 1$  do
2       $z \leftarrow \text{NEWLEAF}(\text{letters}[i])$ 
3       $z.\text{freq} \leftarrow \text{freqs}[i]$ 
4      INSERT( $F, z$ )
5  for  $i \leftarrow 1$  to  $p - 1$  do
6       $z \leftarrow \text{NEWNODE}()$ 
7       $z.\text{left} \leftarrow \text{EXTRACTMIN}(F)$ 
8       $z.\text{right} \leftarrow \text{EXTRACTMIN}(F)$ 
9       $z.\text{freq} \leftarrow z.\text{left}.\text{freq} + z.\text{right}.\text{freq}$ 
10     INSERT( $F, z$ )
11  return EXTRACTMIN( $F$ )
```

1. Execute this algorithm with the frequencies from previous question.
2. Suggest a data structure for  $F$  so that EXTRACTMIN and INSERT will be efficient.
3. What is the complexity of HUFFMAN as a function of  $p$ ?
4. Write a recursive procedure that takes the tree built by HUFFMAN and prints the list of letters (in any order) with their associated code. What is the complexity of this procedure?
5. To decode a file, you need to map each code to its character. Suggest a way to save the HUFFMAN tree using at most  $10p - 1$  bits.
6. If the histogram is passed to HUFFMAN has a sorted list, show that  $F$  can be represented more efficiently using two sorted lists (one for trees that are leaves, and the other for larger trees). What does the complexity of HUFFMAN become?
7. Assume the frequencies of the  $p$  letters are the  $p$  first values of the Fibonacci sequence. What does the HUFFMAN encoding look like?

## 5.1 Frankenstein's QUICKSORT

We consider a variant of QUICKSORT where the pivot value chosen by the partition procedure is the median of the values.

Input: an array  $A[\ell..r-1]$  of integers

Output:  $A$  is sorted in place

FRANKENSORT( $A, \ell, r$ )

```

1  if  $r - \ell > 1$  then
2       $p \leftarrow$  FRANKENPART( $A, \ell, r$ )
3      FRANKENSORT( $A, \ell, p$ )
4      FRANKENSORT( $A, p, r$ )
    
```

Input: an array  $A[\ell..r-1]$  of integers

Output: an index  $p$ , and the array  $A$  is rearranged

so that  $A[\ell..p-1] \leq A[p..r-1]$

FRANKENPART( $A, \ell, r$ )

```

1   $x \leftarrow A[\text{MEDIAN}(A, \ell, r)]$ 
2   $i \leftarrow \ell - 1; j \leftarrow r$ 
3  repeat forever
4      do  $i \leftarrow i + 1$  until  $A[i] \geq x$ 
5      do  $j \leftarrow j - 1$  until  $A[j] \leq x$ 
6      if  $i < j$  then
7           $A[i] \leftrightarrow A[j]$ 
8      else
9          return  $i + (i = \ell)$ 
    
```

We assume that  $\text{MEDIAN}(A, \ell, r)$  returns the *index* of the median of  $A[\ell..r-1]$ , and that it is allowed to reorder the values of  $A[\ell..r-1]$ . For some given  $\ell$  and  $r$ , let  $n = r - \ell$  be the length of the considered sub-array.

1. Justify that a comparison-based MEDIAN needs at least  $n - 1$  comparisons regardless of the implementation.
2. Explain (in two lines) how to implement MEDIAN with  $\Theta(n \log n)$  comparisons.
3. Justify that with the following MEDIAN implementation,  $A[\text{MEDIAN}(A, \ell, r)]$  is the median of  $A[\ell..r-1]$ :

MEDIAN( $A, \ell, r$ )

```

1   $m \leftarrow \lfloor (\ell + r) / 2 \rfloor$ 
2  for  $i \leftarrow \ell$  to  $m$  do
3       $min \leftarrow i$ 
4      for  $j \leftarrow i + 1$  to  $r - 1$  do
5          if  $A[j] < A[min]$  then  $min \leftarrow j$ 
6           $A[i] \leftrightarrow A[min]$ 
7  return  $m$ 
    
```

4. Give, as a function of  $n$ , the complexity of MEDIAN above.
5. Give, as a function of  $n$ , the complexity of FRANKENPART when using this MEDIAN.
6. Give a recursive equation (of the form  $T(n) = aT(n/b) + f(n)$ ) for the complexity of FRANKENSORT.
7. Solve this equation using the master theorem. How does this algorithm compare to QUICKSORT?
8. Let us modify the main algorithm as follows, without changing the definition of MEDIAN:

FRANKENSORT2( $A, \ell, r$ )

```

1  if  $r - \ell > 1$  then
2       $p \leftarrow$  MEDIAN( $A, \ell, r$ )
3      FRANKENSORT2( $A, p + 1, r$ )
    
```

Explain why it is still correct, i.e., why the array returned by FRANKENSORT2 is sorted although the recursion is only done on the second half, and FRANKENPART is not even called...

9. What is the complexity of FRANKENSORT2.
10. Since the recursive call to FRANKENSORT2 is a tail call, perform a tail call elimination to rewrite FRANKENSORT2 as an iterative procedure. Then inline the code of MEDIAN, and finally, maybe give FRANKENSORT2 its real name.

## 5.2 A QUICKSORT-based SELECTION

This exercise is almost the converse of the previous one: instead of using MEDIAN in QUICKSORT, we derive a MEDIAN implementation from QUICKSORT... More generally we want to *select* the rank- $i$  value (i.e., the  $i$ th smallest value) in an array of size  $n$ . Rank 0 is the minimum, rank  $n - 1$  is the maximum, and rank  $\lfloor n/2 \rfloor$  is the median.

1. Suggest an algorithm to find the rank- $i$  value in a  $n$ -sized array, and give its complexity.
2. Execute the following algorithm on a few examples, and explain why it returns the rank- $i$  value.

Input: a non-empty array  $A[\ell..r-1]$  of integers, and a rank  $i < r - \ell$

Output: the rank- $i$  value

SELECTION( $A, \ell, r, i$ )

```

1  if  $r - \ell = 1$  then return  $A[\ell]$ 
2   $m \leftarrow$  RANDOMIZEDPARTITION( $A, \ell, r$ )
3   $k \leftarrow m - \ell$ 
4  if  $i < k$ 
5      then return SELECTION( $A, \ell, m, i$ )
6      else return SELECTION( $A, m, r, i - k$ )
    
```

Input: a non-empty array  $A[\ell..r-1]$  of integer

Output: an index  $p$ , and the array  $A$  is rearranged

so that  $A[\ell..p-1] \leq A[p..r-1]$

RANDOMIZEDPARTITION( $A, \ell, r$ )

```

1   $x \leftarrow A[\text{RANDOM}(\ell, r)]$ 
2   $i \leftarrow \ell - 1; j \leftarrow r$ 
3  repeat forever
4      do  $i \leftarrow i + 1$  until  $A[i] \geq x$ 
5      do  $j \leftarrow j - 1$  until  $A[j] \leq x$ 
6      if  $i < j$  then
7           $A[i] \leftrightarrow A[j]$ 
8      else
9          return  $i + (i = \ell)$ 
    
```

3. Assuming RANDOM runs in  $\Theta(1)$ , compute the complexity of SELECTION in the worst case.
4. This algorithm has an average-case complexity of  $\Theta(n)$ , we will prove it during the lecture.

## 🎨 Tutorial 6 🎨

### 6.1 Induction proof for big-O

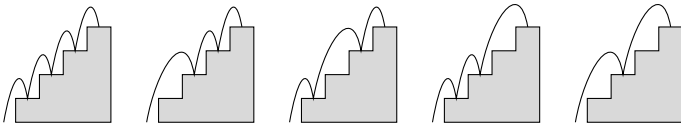
Assuming  $a > 0$  is a constant, and  $f(n) = O(n)$ , solve

$$T(n) = \begin{cases} a & \text{if } n < 2 \\ T(n-2) + f(n) & \text{if } n \geq 2 \end{cases}$$

and prove your result by induction.

### 6.2 Climbing stairs

Consider  $n \geq 1$  stairs, and a person that can climb them using steps of either 1 stair at a time, 2 stairs at a time, or any combination of those (and always going up). Let  $S_n$  be the number of ways to climb those  $n$  stairs. For instance, there are  $S_4 = 5$  ways to climb 4 stairs:



1. Find  $S_1$ ,  $S_2$ , and  $S_3$ .
2. Give a recursive definition of  $S_n$  for  $n > 0$ . Hint: you have two options for the first step.
3. Can we extend the definition for  $n = 0$ ?
4. Write a *recursive* algorithm that takes  $n$  and computes  $S_n$ , following the above definition.
5. Let  $T(n)$  be the time complexity for the above recursive algorithm. Can you guess an expression for  $T(n)$ , maybe involving  $S_n$ ? (Do not attempt to prove it yet.)
6. Propose a  $\Theta(n)$ -time iterative algorithm for computing  $S_n$ .
7. Let us now interpret this problem as a set of matrix operations. Propose a matrix  $M$  so that the following equality holds for  $n > 1$ :

$$\begin{bmatrix} S_{n-1} \\ S_n \end{bmatrix} = M \begin{bmatrix} S_{n-2} \\ S_{n-1} \end{bmatrix}$$

8. It follows that for  $n > 0$ :

$$\begin{bmatrix} S_{n-1} \\ S_n \end{bmatrix} = M^{n-1} \begin{bmatrix} S_0 \\ S_1 \end{bmatrix} = M^{n-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Can you suggest an algorithm for computing  $S_n$  with sub-linear complexity? (Hint: remember exercise 3.5.)

9. If you know how to solve linear recurrence equations, use the recursive definition of question 3 to find a closed form for  $S_n$ .

If you don't know how to do that, it's OK: there is more than one way to skin a cat. The previous equation shows that  $S_n$  is the sum of the two bottom elements of  $M^{n-1}$ . Therefore all you need to do is compute  $M^{n-1}$  for arbitrary  $n$ . (Hint: all real symmetric matrices are diagonalizable, and computing the power of a diagonal matrix is easy.)

10. Find a constant  $k$  such that  $S_n = \Theta(k^n)$ .
11. Prove that your guess from question 5 is correct, by induction.
12. Assuming all values are stored using 32-bit integers, estimate the  $n$  after which the algorithm of questions 6 and 8 will stop providing accurate answers.
13. If the algorithm from question 8 had to be implemented using integers with arbitrary precision (a.k.a. *bignums*), would it keep its complexity?

### 6.3 Barcodes \*

In a barcode, information is transmitted through a sequence of vertical lines of two different width (thin and thick), alternating two colors (black and white). The colors do not convey any information, only the widths of the lines do. We call "cell" the width of a thin line. A thick line is two cells.

A binary code can be transmitted by representing 1 with one cell (thin line), and 0 with two cells (thick lines), always alternating colors. For instance the code 10010 can be transmitted as BWBWBWB (where B and W represent black and white cells).

One can observe that encoding 0 is costlier than encoding 1, so codes containing many 0 will take more space.

Here are the list of codes one can transmit with a barcode of  $n > 0$  cells, for small  $n$ :

$$\begin{aligned} n = 1 & \quad \{1\} \\ n = 2 & \quad \{11, 0\} \\ n = 3 & \quad \{111, 01, 10\} \\ n = 4 & \quad \{1111, 101, 110, 011, 00\} \end{aligned}$$

Let  $G(n)$  be the number of codes that can be transmitted with a barcode of  $n$  cells.

1. Give a closed formula for  $G(n)$ , for  $n > 0$ .
2. Propose a bijection between the set of barcodes of  $n$  cells, and the set  $\{0, 1, \dots, G(n) - 1\}$ . This would allow encoding any number smaller than  $G(n)$  in a barcode of size  $n$ .

## 7.1 Balanced Partition

Consider a set of  $n$  people who want to board a unstable barge. For simplicity, we assume that people will sit either on the port (=left) side or starboard (=right) side. To balance the barge, we would like that the sum of the weights of people sitting on one side is as close as possible to the weights of people sitting on the other side.

The goal is to create an algorithm that will take the set of weights of all passengers as input, and assign these people to a side in order to balance the barge in the best way possible.

Let's number  $1, 2, \dots, n$  the passengers, and use  $w_1, w_2, \dots, w_n$  for their respective weights. Assume all weights are integers with  $w_i \geq 1$ .

Let  $S_p$  and  $S_s$  be the set of people assigned to Port or Starboard, and let  $W_p = \sum_{k \in S_p} w_k$  and  $W_s = \sum_{k \in S_s} w_k$  the weights of these two groups.

With these notations, our goal is to construct  $S_p$  and  $S_s$  such that  $S_p \cup S_s = \{1, \dots, n\}$  and  $|W_s - W_p|$  is minimal. We say that the barge is perfectly balanced if  $W_s - W_p = 0$ .

### 7.1.1 Greedy Attempts

- Let us consider the following procedure: people board the barge one after the other, and go sit on the side where the total weight is the smallest (this can be implemented by looking at a bubble level to see in which direction the barge is tilted).

Construct a sequence of weights (corresponding to the boarding of passengers) so that:

- when boarding has been completed according to the above procedure, the barge is *strongly* unbalanced (for some subjective definition of *strongly*)
- for the same set of passengers, there exists an assignment with perfect balance.

- Let us improve the algorithm and board people by decreasing weights. Propose a set of weights for which a perfect balance is possible, and still the new procedure will fail to balance the barge.

### 7.1.2 Dynamic Programming

- Give a formula for  $W_i$ , the *ideal* value of  $W_p$  and  $W_s$ , as a function of  $w_1, \dots, w_n$ . By ideal we mean that setting  $W_p$  and  $W_s$  to this value (if this is possible) would give us a boaring with perfect balance.

- If we wanted to evaluate all possible assignments of the passagers to one side of the boat, in order to use the better one, how many assignments would we need to evaluate? Write the answer as a function of  $n$ .

- Let  $P(i, m)$  be a Boolean function telling us if there is a subset of  $\{1, 2, \dots, i\}$  (the first  $i$  passengers) whose total weight is exactly  $m$ .

Please specify the domain of the values of  $i$  and  $m$ .

- Propose a recursive definition of  $P(i, m)$ .
- Compute  $P(i, m)$  for all values of  $i$  and  $m$  over the following sets of weights:  $\{19, 6, 14, 9, 17\}$ . Present your results as a table.
- What is the complexity of an algorithm that would compute this table?
- Where in this should we look for the weight  $p$  of a group that is closest to the ideal weight  $W_i$ ?
- Express the minimal value of  $|W_s - W_p|$  as a function of  $p$  and  $W_i$ .
- Compute the minimal value of  $|W_s - W_p|$  for the weights  $\{19, 6, 14, 9, 17\}$ .
- If  $P(i, m)$  is true, how can we find the set of people whose weight is  $m$ ?
- Write an algorithm that takes a list of weights as input, and compute table  $P$ .
- Write an algorithm that takes a list of weights as input and constructs the sets  $S_s$  and  $S_p$  such that  $|W_s - W_p|$  is minimal.

## 7.2 Knapsack

What is the connection between the previous problem, and the following picture?

