Analysis of Algorithms 2020/2021

Practice 2

S. Xiao Fernandez Marin

| Code | Plots | Memory | Total |
|------|-------|--------|-------|
| | | | |

1. Introduction.

In this practice, starting from the time functions I had coded in the previous one, I study the efficiency of two new algorithms which use divide and conquer approach. They are the functions: MergeSort and QuickSort. For each algorithm, it will be analyzed on tables of different sizes and also the results obtained will be compared with the theoretical study of the algorithm.

2. Objectives

2.1 Section 1

First of all, I studied the sorting algorithm corresponding to the method MergeSort. This algorithm works dividing each table in subtables until it has each element individually, when it starts sorting by comparing each element and combining them.

2.2 Section 2

In this part, I proved that the algorithm MergeSort works correctly, by adapting exercise5.c, given in the previous practice. It is expected that the average clock time it takes to sort is smaller than the one in InsertSort, as MergeSort does not work with the table at all once, but dividing and working with smaller elements.

2.3 Section 3

In this section I have implemented the sorting algorithm QuickSort. This algorithm consists of three functions: quicksort, split and median.

2.4 Section 4

The objective of this exercise is to study the execution of QuickSort algorithm by adapting exercise5.c, and then comparing it with the theoretical result of the algorithm.

2.5 Section 5

In this exercise, I had to add the function quicksort_ntr. This function is the quicksort function but without the tail recursion.

3 Tools and Methodology

Basically, the tool that I have been using during this task is Visual Studio Code, because it facilitates us the debugging of the code. I have also executed all the exercises with valgrind to prove that there were no dynamic memory allocation errors. To create the plots, I have used Gnuplot.

3.1 Section 1

To do this exercise, I have guided myself using the pseudocode we can find in the slides of the subject. First I understood how the MergeSort algorithm works writing it in a paper. Then, I started to code using the pseudocode. Lastly, to prove that it sorts correctly, I have used the exercise4.c given in the previous assignment and I have also executed it with valgrind to see there were no memory errors.

3.2 Section 2

Then, I have adapted the file exercise5.c of the previous practice for using MergeSort algorithm, and I have obtained the table with the average clock time, and the average, minimum and maximum number of times that the BO is executed depending on the size of the permutation. In this part I have also plotted the resulting values to compare them with the theoretical result.

3.3 Section 3

For this exercise, I have also used the slides and the pseudocode from the second unit of the subject to understand how the algorithm QuickSort works and then to code it. This algorithm requires two additional functions: split, that divides the elements of the table, and median, that determines the position of the pivot. Then, I have also used the exercise4.c adapted to QuickSort to check that it orders correctly.

3.4 Section 4

In this exercise, again adapted the file exercise5.c of the previous practice for using QuickSort algorithm, I have obtained the table with the average clock time, and the average, minimum and maximum number of times that the BO is executed depending on the size of the permutation. Then, as in exercise 2, I have plotted the resulting values and compared them with the theoretical result.

3.5 Section 5

We were asked to implement a new functions to remove the tail recursion. Using the file exercise5.c, I compared the average clock time, and the average, minimum and maximum number of times that the BO is executed with each function implemented.

4. Source code

4.1 Section 1

```
int mergesort(int* table, int ip, int iu){
  int middle = 0, i = 0;
  int aux1 = 0, aux2 = 0, ob = 0;
  if (table == NULL || ip < 0 || ip > iu)return ERR;
 if (ip == iu)return OK;
 else{
   middle = (int)(iu + ip) / 2;
    aux1 = mergesort(table, ip, middle);
    if (aux1 == ERR) return ERR;
    else ob = ob + aux1;
    aux2 = mergesort(table, middle + 1, iu);
    if (aux2 == ERR) return ERR;
   else ob = ob + aux2;
   i = merge(table, ip, iu, middle);
if (i == ERR) return ERR;
    i = i + ob;
  }
  return i;
}
```

```
int merge(int* table, int ip, int iu, int imiddle){
  int i = ip, j = imiddle + 1, k = 0, cont = 0;
  int *auxtable = NULL;
  if (table == NULL || ip < 0 || iu < ip || imiddle > iu || imiddle < ip) return ERR;
  auxtable = (int *)malloc(((iu - ip + 1) * sizeof(int)));
  if (auxtable == NULL) return ERR;
  while (i <= imiddle && j <= iu) {
    cont++;
    if (table[i] < table[j]){</pre>
      auxtable[k] = table[i];
    else{
     auxtable[k] = table[j];
      j++;
    k++;
  }
  if (i > imiddle){
    while (j <= iu){
     auxtable[k] = table[j];
      j++;
      k++;
   }
  else if (j > iu){
    while (i <= imiddle){</pre>
     auxtable[k] = table[i];
      j++;
      k++;
  for (i = ip, j = 0; i \le iu; i++, j++){
   table[i] = auxtable[j];
  free(auxtable);
  return cont;
}
```

```
int quicksort(int* table, int ip, int iu){
  int pos = 1, middle = 0, ob = 0, aux_left = 0, aux_right = 0;
  if (!table || ip < 0 || iu < 0 || iu < ip) return ERR;
  if (ip == iu) return ob;
  ob += split(table, ip, iu, &pos);
  if (ob == ERR)return ERR;
  middle = pos;
  if (ip < middle - 1) {
    aux_left = quicksort(table, ip, middle - 1);
    if (aux_left == ERR)return ERR;
    ob = ob + aux_left;
  }
  if (middle + 1 < iu) {
    aux_right = quicksort(table, middle + 1, iu);
    if (aux_right == ERR) return ERR;
    ob = ob + aux_right;
  }
  return ob;
int split(int* table, int ip, int iu,int *pos){
  int i = 0, k = 0, ob = 0, swap = 0, swap1 = 0, swap2 = 0;
  if (!table || ip < 0 || iu < 0 || iu < ip || !pos) return ERR;
  ob += median(table, ip, iu, pos);
  if (ob == ERR) return ERR;
  if ((*pos) > iu || (*pos) < ip) return ERR;
  k = table[*pos];
  swap = table[ip];
  table[ip] = table[*pos];
  table[*pos] = swap;
  *pos = ip;
  for (i = ip + 1; i \le iu; i++){
    if (table[i] < k){</pre>
      (*pos)++;
      swap1 = table[i];
     table[i] = table[*pos];
     table[*pos] = swap1;
    ob++;
  }
  swap2 = table[ip];
  table[ip] = table[*pos];
  table[*pos] = swap2;
  return ob;
}
```

```
int median(int *table, int ip, int iu,int *pos){
  if (!table || ip < 0 || iu < 0 || iu < ip || !pos) return ERR;
  *pos = ip;
  return 0;
}
4.5 Section 5
int quicksort_ntr (int* table, int ip, int iu){
  int pos = 1, middle = 0, ob = 0, aux_left = 0;
  if (!table || ip < 0 || iu < 0 || iu < ip) return ERR;</pre>
  while (ip < iu) {
    ob += split(table, ip, iu, &pos);
    if (ob == ERR)return ERR;
    middle = pos;
    aux_left = quicksort_ntr(table, ip, middle);
    if (aux_left == ERR)return ERR;
    ob = ob + aux_left;
    ip = middle + 1;
  }
  return ob;
}
```

5. Results, Plots

5.1 Section 1

I executed this exercise with exercise4.c and with valgrind to prove that the algorithm MergeSort works correctly and that there are not memory errors. We also checked that it orders correctly by entering a table of size 300.

```
| Archivo Editar Ver Buscar Terminal Ayuda | Archivo Editar Ver Bus
```

5.2 Section 2

In exercise 2, we were asked to prove the same algorithm as in one but obtaining the table with the average clock time, and the average, minimum and maximum times the BO is executed depending on the size of the permutation. Also I have checked it with valgrind. This is the result using min 1, max 100 and 100 permutations.

```
Archivo Editar Ver Buscar Terminal Ayuda

Archivo Editar Ver Buscar Terminal Ayuda

xiaogixlao-Lenovo-ideapad-320-15IXB:-(Escritorio/Ing. Inf II/AALG/P2 - FINALS valgrind --leak-check-full ./exercise5 -nun_min 1 -nun_max 100 -t

ncr 5 -nunp 100 -outputFile MergeSort

=15701a: Mencheck, a menory error detector

=15701a: Copyright (C) 2002-2017, and CNU GL'd, by Julian Seward et al.

=15701a- Using Valgrind-3.13.0 and LibVEX; rerun with -h for copyright info

=15701a- Command: ./exercise5 -nun_min 1 -nun_max 100 -tncr 5 -nunp 100 -outputFile MergeSort

Practice number 1, section 5

Droub: Your group

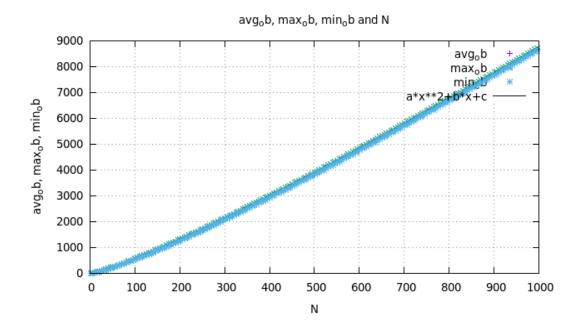
Correct output

=15701a-

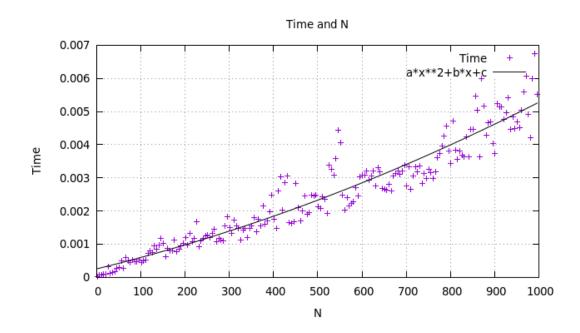
EARD SUMMARY: 0 errors from 0 contexts (suppressed: 0 from 0)

xlaogilao-Lenovo-ideapad-320-15IXB:-/Escritorio/Ing. Inf II/AALG/P2 - FINALS
```

Plot comparing the best, worst and average number of BOs for MergeSort of size from 0 to 1000. We can observe that the theoretical number and the practical one are almost the same. Also we can see that the best, worst and average number of BOs is similar, but not the same, as MergeSort depends on the size of the table. In the plot it is not so visible but as we have seen in the file generated before, there are differences between the three.



Plot comparing average clock time for MergeSort. In this plot I compare the average clock time with a table of size 1000. We can observe peaks, that is because this algorithm depends a lot in how sorted is the table at the beginning.



5.3 Section 3

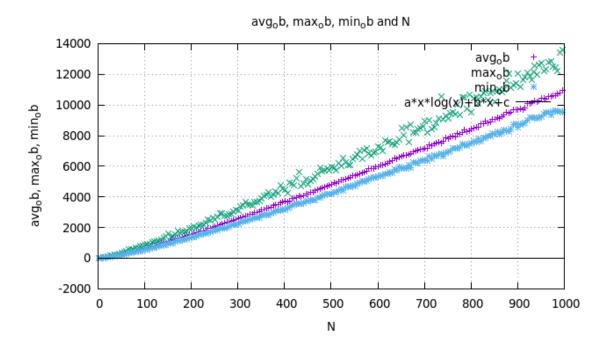
I have checked that the algorithm QuickSort works properly using also file exercise4.c adapted to QuickSort with valgrind checking. This is the result of entering a table of size 300 to be sorted.

5.4 Section 4

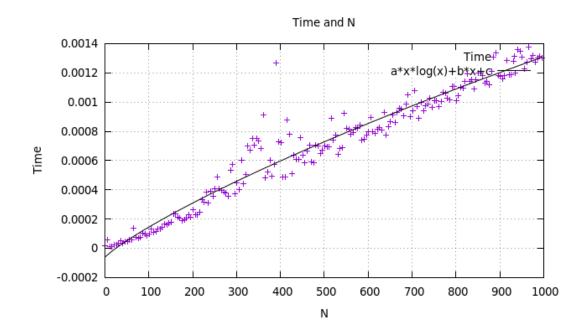
Now I checked algorithm QuickSort with file exercise5.c, which generated a file called "AAquicksort" with the average, minimum and maximum number of times the BO is executed in a table of 100 permutations.

```
xino@xiao-Lenovo-ideapad-320-15IKB:-/Escritorio/ing. Inf II/AALG/P2 - FINALS valgrind --leak-check=full ./exercise5 -num_min 1 -num_max 100 -t ncr 5 -numP 100 -outputFile AAquicksort ==17929== Rencheck, a nenory error detector ==17929== Copyright (C) 2002-2017, and CAU CRL'd, by Julian Seward et al. ==17929== Command: ./exercise5 -num_min 1 -num_max 100 -incr 5 -numP 100 -outputFile AAquicksort ==17929== Command: ./exercise5 -num_min 1 -num_max 100 -incr 5 -numP 100 -outputFile AAquicksort ==17929== Command: ./exercise5 -num_min 1 -num_max 100 -incr 5 -numP 100 -outputFile AAquicksort ==17929== Command: ./exercise5 -num_min 1 -num_max 100 -incr 5 -numP 100 -outputFile AAquicksort ==17929== Lan use at ext: 0 bytes in 0 blocks ==17929== Lan use at ext: 0 bytes in 0 blocks ==17929== Louse at ext: 0 bytes in 0 blocks ==17929== Louse at ext: 0 bytes in 0 blocks ==17929== Louse at ext: 0 bytes in 0 blocks ==17929== For counts of detected and suppressed errors, rerun with: -v ==17929== For Counts of detected and suppressed errors, rerun with: -v ==17929== For Counts of detected and suppressed errors, rerun with: -v ==17929== ERROR SUMMARY: 0 errors from 0 contexts (suppressed: 0 from 0) xlao@xlao-Lenovo-ldeapad-320-15IKB:-/Escritorlo/Ing. Inf II/AALG/P2 - FINALS []
```

Plot comparing the best, worst and average number of BOs for QuickSort. We can see that three parameters have different results, that is because the algorithm depending on how sorted is the table will choose a particular number, with whom it will do swap, and that number is not always the same.



Plot comparing average clock time for QuickSort. In this plot we have used a table also of size 1000. We can observe some peaks also, as in MergeSort, but it adjusts quite a lot to the theoretical average clock time.

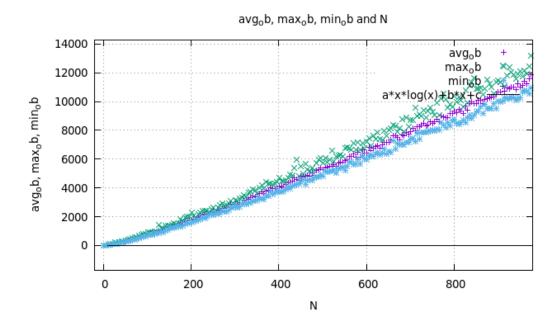


5.5 Section 5

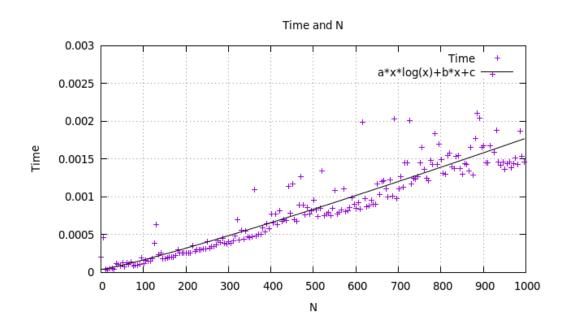
I have checked that the algorithm QuickSort_ntr works properly using also file exercise4.c adapted to QuickSort_ntr with valgrind checking. This is the result of entering a table of size 300 to be sorted.

Now I checked algorithm QuickSort_ntr with file exercise5.c, which generated a file called "AAquicksort_ntr" with the average, minimum and maximum number of times the BO is executed in a table of 10 permutations.

Plot comparing the best, worst and average number of BOs for QuickSort_ntr. We can see that three parameters have different results, that is because the algorithm depending on how sorted is the table will choose a particular number, with whom it will do swap, and that number is not always the same.



Plot comparing average clock time for QuickSort. In this plot we have used a table also of size 1000. We can observe some peaks also, as in MergeSort, but it adjusts quite a lot to the theoretical average clock time.



5. Answers to theoretical Questions.

5.1 Question 1: Compare the empirical performance of the algorithms with the theoretical average case for each case. If the traces of the performance graphs are very sharp, why do you think this happens?

- *MergeSort*: As we have seen in the time plots, the practical time is very similar to the theoretical average case (in the case of MergeSort is N(log(N)). We can also observe that in the plot there are some sharp peaks, this was expected because this algorithm depends a lot in how bad sorted is the table, as it does not go over the complete table. So due to the randomness of the tables we generate, is normal that the graphs are sharp.
- *QuickSort*: On the other hand, QuickSort plots are also sharp, but considerable less than in MergeSort. We can say that in terms of time, MergeSort is quite more efficient than QuickSort, specially when sorting big tables, where QuickSort is less efficient as its average case is 2N log (N) + O(N).

5.2 Question 2: Analyze the result obtained when comparing quicksort and quicksort_src both in the case you observe differences or not.

As we can see in the plots, the quicksort_ntr time does not have much differences between the quicksort with the tail recurssion. We see that the quicksort_ntr have bigger peaks of time than the one with the tail recurssion, but does not mean anything because it depends lot of how the initial table is sorted.

When talking about the obs, we can see that the quicksort_ntr is closer to the average ob than the quicksort one.

5.3 Question 3: What are the best and worst cases for each algorithm? What should be modified in practice to strictly calculate each case (include average case too)?

MergeSort:

- Best case = $\frac{1}{2}$ * (N log (N))
- Worst case = $N \log (N) + O(N)$
- Average case = $N \log (N)$

QuickSort:

- Best case = We have not seen this in theory class, but knowing how the algorithm works, we can conclude that the best case for QuickSort is when the pivot is in the middle of the table.
- Worst case = (N2/2) (N/2)
- Average case = $2N \log (N) + O(N)$

The modifications we should do to strictly calculate each case, are basically found in the functions generate_perms and generate_permutations, because for example to take the table sorted we have to eliminate the loop in charge of doing swap between each of the elements of the table randomly.

5.4 Question 4: Which of the two algorithms studied is empirically more efficient? Compare this result with the theoretical prediction. Which algorithm(s) is/are more efficient from the point of view of memory management? Justify your answer.

As seen before, the average case of both algorithms is similar, being MergeSort some better. However, the problem of QuickSort lays in the big size tables, because as we have seen, it is of order N2/2 + O(N), the same as the local sorting algorithms or even MergeSort.

However, if we have to decide between one of them, we will choose QuickSort, at it does not require additional memory, what makes it more efficient and avoids us possible memory errors such as segmentation fault, memory loss or problems allocating.

6. Final Conclusions.

As we have seen in this practice, algorithm MergeSort is the most effective compared to InsertSort and QuickSort, but QuickSort is the most efficient.

On the other hand, it is important for us that the algorithms are the less complex to understand, and QuickSort is not an easy algorithm to understand. Also in QuickSort you have to take into account the position of the pivot because this affects the result.

It has been interesting to create the function quicksort_ntr, that makes the algorithm have less basic operations than the quicksort. This has taught as that making a little modification on the code makes big changes on an algorithm.

Another way could have been modifying the split and median functions.

The first one by returning the average table position as the position of the pivot, and the number of basic operations that it performs is limited to just one.

The second one, choosing the position of the intermediate value between the first element of the table, the last and the one that it is in the intermediate position of the table. Obviously this routine takes a greater number of basic operations. So although some time is lost choosing the pivot, this at the end will result more efficient because having a better pivot will decrease the execution time as the algorithm will sort faster.