

Homework 2

September 25, 2007

Due Date: October 2

1. (15 points) In this problem we consider when a buffer is going to be maintained under a variety of situations. For each, explain if you would select an array, a circular array, a dynamic array, a dynamic circular array, a singly-linked list, or a doubly-linked list to implement the buffer, and very briefly explain why.
 - (a) (5 points) The buffer is a bounded buffer that is to hold no more than 1000 elements. The only methods that are commonly used are to insert and remove elements from both the front and back of the buffer, and these methods must be as efficient as possible. The only other method needed is the ability to iterate over the buffer. The buffer is expected to hold between 500 and 1000 elements at any given time.
 - (b) (5 points) We consider the variation of part (a) in which the buffer is unbounded (i.e., there is no a prior bound on number of elements that will be placed in the buffer).
 - (c) (5 points) We next consider a different variation from part (a) in which efficient support must be provided to iterate over the buffer removing any elements that satisfy a given property (which can be checked in constant time for each element).
2. (15 points) In order to have a *dynamic array* (such as Java's `ArrayList`) that has no upper limit on the size, the standard implementation is to double the array size copying over the elements in use when the array is full. In this problem you explore the overhead in time complexity per insertion caused by the resizing¹. For this analysis, we'll assume there is a cost of 1 to insert an element at the end of the positional collection when it is not full. We treat the cost to resize a full array of x elements to a half-full array of $2x$ elements as x since x elements must be copied into the new array. (If desired the cost of allocating and initializing the array could also be included. However, copying the elements in the dominant cost.)
 - (a) (8 points) First consider the situation in which the dynamic array is initialized to length 1. Consider the program that inserts 100 elements at the end of the dynamic array. Compute the cost of the overhead associated with resizing (i.e., the cost of allocating new arrays and copying over the elements).
 - (b) (1 points) What is the cost per element inserted into the dynamic array (by the user program) of the overhead associated with resizing the dynamic array?
 - (c) (6 points) Now consider when the dynamic array is initialized to length 32. Now what is the cost per element inserted into the dynamic array of the overhead associated with the dynamic array?
3. (15 points) Suppose the partitioning procedure presented in class for randomized quicksort was applied to partition an n element array around a randomly selected pivot element. What is the expected number of elements to the right of the pivot after the partitioning is complete?

Note: While you could use your intuition to guess the answer for this problem, you are expected to use the formal definition of expected value. Be sure to show your work starting from the definition of expected value!

¹Appendix B.5 discusses *amortized analysis*. Informally, the cost of reorganization is amortized over the public methods executed that caused the reorganization to be performed.

4. (15 points) Below is code for binary search which assumes that A is sorted such that $A[i] \leq A[i + 1]$. The initial call made is `binarySearch(x, A, 0, 14)`.

```
boolean binarySearch(x, A, p, r)
1  if (p == r)
2    return (x == A[p])
3  q = ⌊(p + r)/2⌋
4  if (x == A[q])
5    return true
6  if x < A[q]
7    binarySearch(x, A, p, q - 1)
8  else
9    binarySearch(x, A, q + 1, r)
```

You are to exactly compute the expected number of times that x is compared to an element of A (in line 2, 4 or 6) when searching $A[0..14]$ where for $0 \leq i \leq 14$, x is $A[i]$ with a probability of $1/30$ and with probability $1/2$, x is not in A .

5. (15 points) Prove a lower bound for the problem of computing which elements in a sorted array A are less than a given element x under the comparison-based model in which information about the relationship between x and the elements of A can only be learned by asking if $x < A[i]$ for any $i \in \{0, 1, \dots, n - 1\}$.
6. (25 points) The *nuts and bolts* problem is defined as follows. You are given a collection of n bolts of different diameters and n corresponding nuts. You are allowed a **test operation** in which you try a nut and bolt together from which you can determine whether the nut is too large, too small, or an exact match for the bolt. You can not directly compare two nuts or directly compare two bolts. Each test operation takes constant time, and the problem is to match with each bolt to its nut.
- (a) (15 points) Give an algorithm to correctly pair up the n nuts and bolts in expected $O(n \log n)$ time. *You can use the bound we derived on the expected cost of randomized quicksort.*
- (b) (10 points) Show that any algorithm for the nuts and bolts problem must use the test operation $\Omega(n \lg n)$ times in the worst case.

Extra Credit Problems:

- 7* (2 points) The following program determines the maximum value in an unordered array $A[0..n - 1]$.

```
1  max = -∞
2  (int i = 0; i < n; i++)
3      if A[i] > max
4          then max = A[i]
```

What is the expected number of times the assignment in line 4 is executed? (Assume that all numbers in A are drawn randomly from the interval $[0,1]$.) You may find it useful to let s_1, s_2, \dots, s_n be n random variables, where s_i represents the number of times (0 or 1) that line 4 is executed during the i th iteration of the **for** loop. Since $s = s_1 + s_2 + \dots + s_n$, by linearity of expectations $E(s) = E(s_1) + \dots + E(s_n)$.

- 8* (3 points) Generalize your work from Problem 4 to compute the expected number of elements examined when the array has $n = 2^r - 1$ elements (for r an integer) and with probability $1/2$, x is not in the array, and with probability $1/(2n)$, $x = A[i]$ for $i = 0, 1, \dots, n - 1$.