READ THIS NOW!

- Print your name in the space provided below.
- There are 9 short-answer questions, priced as marked. The maximum score is 100.
- When you have finished, sign the pledge at the bottom of this page and turn in the test and your fact sheet.
- Aside from the allowed one-page fact sheet, this is a closed-book, closed-notes examination.
- No laptops, calculators, cell phones or other electronic devices may be used during this examination.
- You may not discuss this examination with any student who has not taken it.
- Failure to adhere to any of these restrictions is an Honor Code violation.

Name (Last, First) ____

printed

Pledge: On my honor, I have neither given nor received unauthorized aid on this examination.

____

signed

1. Consider the following algorithm for computing the product of two N x N matrices:

```
for (row = 1; row <= N; row++) {                 // 1 before, 2 each
                                                 //   pass, 1 to exit
   for (col = 1; col <= N; col++) {              // 1 before, 2 each
                                                 //   pass, 1 to exit
      dotprod = 0;                               // 1
      for (pos = 1; pos <= N; pos++) {           // 1 before, 2 each
                                                 //   pass, 1 to exit
         dotprod = dotprod + a[row][pos]*b[pos][col]; // 7
      }
      product[row][col] = dotprod;               // 3
   }
}
```

a) [10 points] Using the rules given in class for counting operations, find a simplified function T(N) that counts the exact number of operations the algorithm would perform.

$$T(N)=1+\sum_{R=1}^{N}\left(2+1+\sum_{C=1}^{N}\left(2+1+1+\sum_{P=1}^{N}(2+7)+1+3\right)+1\right)+1$$

$$=2+\sum_{R=1}^{N}\left(4+\sum_{C=1}^{N}\left(8+\sum_{P=1}^{N}9\right)\right)$$

$$=2+\sum_{R=1}^{N}\left(4+\sum_{C=1}^{N}(8+9N)\right)$$

$$=2+\sum_{R=1}^{N}\left(4+8N+9N^{2}\right)$$

$$=2+4N+8N^{2}+9N^{3}$$

b) [5 points] To what big-Θ complexity class does your answer to the previous part belong? (No proof is necessary.)

Clearly, T(N) is Θ(N^3).

2. [15 points] Consider the following two functions:

$$f(n) = \sqrt{n} \text{ and } g(n) = \log n$$

Determine the complexity relationship between the two functions. That is, determine whether *f* is strictly Ω of *g*, or *f* is strictly O of *g*, or *f* is Θ of *g*. Whichever conclusion you reach, justify your conclusion completely by applying relevant theorems from the course notes.

Since f(n) does not occur in the list of terms for Theorem 5, we cannot apply that theorem. So, the only alternatives are to use the definitions or to take the appropriate limit and apply Theorem 8 or its Corollary.

$$\lim_{n\to\infty}\frac{f(n)}{g(n)} = \lim_{n\to\infty}\frac{\sqrt{n}}{\log n} \quad \text{which is of the form inf/inf, so apply l'Hopital's Rule}$$

$$= \lim_{n\to\infty}\frac{\frac{1}{2}n^{1/2}}{\frac{1}{n\ln 2}}$$

$$= \lim_{n\to\infty}\frac{n^{1/2}\ln 2}{2} = \infty$$

Since the limit is infinite, by the Corollary to Theorem 8, f is strictly Ω(g).

3. [12 points] Recall that the access time for a hard drive is the sum of three separate measures of performance. Suppose that an existing hard drive design was modified by making the following two changes:

- the rotational speed is increased by 10%
- the number of tracks on each surface is decreased by 20%, with a corresponding decrease in the radius of the platters

No other characteristics of the original drive design are changed, aside from the obvious change in the storage capacity. Describe what effect these changes, if any, will have on each of the three measures, and why.

The three factors are seek time, rotational latency, and transfer time.

The seek time depends solely on how far the head has to move when servicing a request. Since the second change reduces the radius of the platters but does not change the track spacing, the maximum distance the head could move has been reduced and therefore the average seek time will be reduced.

The rotational latency depends solely on how long it takes for the correct sector to rotate to the position of the head, and that depends on the rotational speed. Since the first change increases the speed at which the platters rotate, the average rotational latency will be decreased.

The transfer time depends solely on the rotational speed and the number of sectors to be read. The change in rotational speed will reduce the transfer time; the number of sectors to be read is independent of the drive parameters.

4. [12 points] In a B-tree of order 5, assume that the deletion of a particular value from a leaf node causes that leaf node to underflow, and that the underflow cannot be alleviated by "borrowing" (and therefore must be alleviated by "merging"). Is it necessary, or possible, or impossible that this "merging" will cause the parent of the leaf to underflow? If it is necessary, explain why. If it is possible, but not necessary, explain under what conditions the parent would underflow and under what conditions it would not. If it is impossible, explain why. Using examples to illustrate your argument, explain clearly.

Merging requires that we move the data values from one child node into the other child node, AND that we move a value down from the parent node (unless we're in the root). So, merging reduces the number of data values stored in the parent node.

Whether the parent node will underflow depends on how many data values the parent stored before the merge was carried out. For each node in the B-tree, there is a minimum number of values it must store. The value is 1 for the root node, and floor(m/2) for the other nodes.

So, the parent will underflow iff it was at its minimum fullness before the merging took place.

Therefore, it is possible but not necessary that the parent will underflow.

Supporting examples were presented in class.

5. Suppose you are designing a buffer pool for an application that repeatedly accesses a collection of 10000 records, labeled R0, R1, …, R9999.

   a) [8 points] Logically, what is a *replacement policy* and why is one necessary?

   A replacement policy is the set of rules used to determine which element, if any, of the buffer pool will be replaced when another element is added to the pool.

   A replacement policy is necessary because buffer pools must have a finite cap on their capacity and therefore may become full.

   b) [4 points] Give a specific example of a possible access pattern for which LRU replacement will produce better performance than FIFO replacement, assuming that the buffer pool has 4 slots.

   Here's one simple example. Access four distinct records to fill the pool, say:

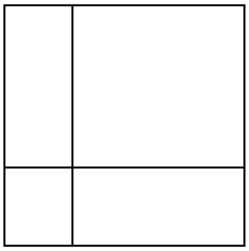
   R0, R1, R2, R3

   Now, access the first record again (so it's not the LRU entry), then access a new record to force a replacement. LRU will replace R1 and FIFO will replace R0:

   R0, R1, R2, R3, R0, R4

   Now access R0 again; LRU will score a hit but FIFO will score a miss: R0, R1, R2, R3, R0, R4, R0

6. [6 points] The PR-quadtree partitions a finite, square region into four identical sub-regions (quadrants). An alternative type of quadtrees, let's call it an R-quadtree, allows unequal partitioning of sub-regions, as shown below. Given the same set of data points, would you expect the PR-quadtree to have more or fewer levels than an R-quadtree? Why?



When a new element is added, the R-quadtree would never need to perform more than one split in order to separate it from any previous element.

The PR-quadtree might have to perform many splits in order to achieve the necessary separation.

So, the PR-quadtree might add more than 1 level on an insertion for which the R-quadtree would add a single level.

However, note that it is possible that an insertion into the PR-quadtree will require NO splits but the same insertion into an R-quadtree might require 1 split.

So, overall we'd expect the PR-quadtree to contain more levels (perhaps many more) than the R-quadtree given the same data.

7. [6 points] If a PR-quadtree stores points that fall into a world with side S, and the closest pair of points are a distance D apart, then the maximum height the PR-quadtree could have is given by

$$\left\lceil \log\left(\frac{S\sqrt{2}}{D}\right)\right\rceil$$

Explain why the bound involves the expression $\sqrt{2}$. Do not attempt to prove that the bound is correct.

We would guarantee that the closest pair of points is separated if the lowest level in the PR-quadtree contains leaf nodes whose regions are too small to contain both those points.

Given a square of side X, the maximum distance between any two points in the square will be the length of the diagonal, which is $X\sqrt{2}$.

8. Recall the B-tree data structure.

a) [4 points] What is the relationship between the number of values stored in an internal node of the B-tree and the number of children that node has?

The number of children an internal node has is 1 more than the number of data values in the node.

b) [9 points] B-trees and their relatives are generally used to index very large databases, in which even the index cannot be entirely stored in memory at once. Why is it better to use a B-tree, whose nodes store many values and have many children, rather than an AVL tree?

First, because B-tree nodes store many data values and AVL nodes store only one, a B-tree would have far fewer levels and therefore would require retrieving far fewer separate disk reads to retrieve nodes.

Second, reading a sector (or a few sectors) at once is almost as fast as reading just a few bytes of data from disk, so reading one of the larger B-tree nodes will cost only a little more than reading one of the AVL nodes.

9. [9 points] In C++, every container implementation that allocates memory dynamically should include a(n)

destructor , and a(n)

copy constructor , and a(n)

assignment operator overload .