MIDTERM

Please answer all questions. Begin each question on a new page. Reserve some space for scratch work, which you should hand in. All the best!

1. 15 points.
   For any constant $d > 0, c > 0$, show that

   $$\sum_{i=1}^{n} i^d \log^c i = \Theta(n^{d+1} \log^c n).$$

   ANSWER: This was discussed in class, and is standard. Let S be the above sum. Clearly,

   $$S \le \sum_{i=1}^{n} n^d \log^c n = n^{d+1} \log^c n.$$

   For lower bound, you discard the smallest $n/2$ terms. If $m = \lceil n/2 \rceil$, we get

   $$S \ge \sum_{i=m}^{n} m^d \log^c m = m^{d+1} \log^c m = \Omega(n^{d+1} \log^c n).$$

2. 10+15 points
   Prove tight upper and lower bounds on $T(n)$ where:
   (a) $T(n) = n^3 \log^3 n + 9T(n/3)$.
   (b) $T(n) = n^2 \log^3 n + 9T(n/3)$.

   ANSWER: (a) We claim this is case (+) of the Master theorem, so $T(n) = \Theta(n^3 \log^3 n)$. We need to show the regularity condition, that $f(n) = n^3 \log^3 n \ge 9f(n/3)/c$ for some $0 < c < 1$.

   (b) The master theorem does not apply even though this looks almost like case (+). Try it! Upper bound: Two answers would be acceptable: $T(n) = O(n^{2+\epsilon})$ for any $\epsilon > 0$ or $T(n) = O(n^2 \log^4 n)$. The latter, of course is better, in fact optimal. You can show this by induction.

3. 20 points
   In the non-randomized linear time median algorithm, what running time would we get if we divide the n input elements into groups of size 3 each? Show your analysis.

   ANSWER: If you use groups of size 3, then you can be sure that the median of medians has at least $2(n/6) = n/3$ elements above it, and at least the same number below it. Hence, in the recursive call, we may have to search in a set of size $2n/3$. In the first recursive call, you had to find the median of $n/3$ medians. Hence the recurrence is

   $$T(n) = n + T(n/3) + T(2n/3).$$

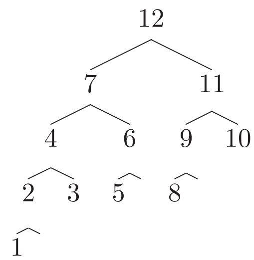
   But from our homework problem, we know that this recurrence has solution $T(n) = \Theta(n \log n)$.

4. 10+20 points
   (a) Draw two AVL trees, both of height 4. One has maximum size and the other has mininum size.
   (b) Starting with an empty AVL tree, insert the following set of keys, in this order:

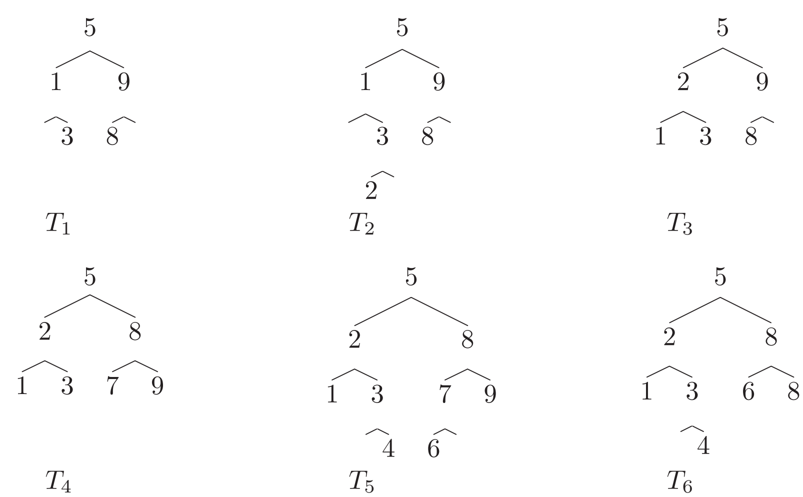
   $$5, 9, 1, 3, 8, 2, 7, 6, 4.$$

Now delete key 9. Please show the tree at the end of each operation. (You can show intermediate trees if you like).

ANSWER:
(a) The largest AVL tree of height 4 is just the complete binary tree of height 4, and this has 31 vertices. Here is the smallest size AVL tree of height 4: note that the software we use for drawing binary tree does not draw the child with outdegree 1 correctly.



(b) The first 5 insertions produce the tree T_1 with no rotations. The next insertion of key 2 produces T_2 which is double-rotated into T_3. Inserting key 7 causes a rotation to produce T_4. Inserting keys 6 and 4 causes no rotation, yielding T_5. Finally, deleting 9 from T_5 causes a rotation, yielding T_6.



5. 10+20 points
(a) Show a counter example to Markov's inequality in case the random variable X is not non-negative. HINT: you can use a sample space with only 3 points.
(b) Professor Yap likes to play a game where you roll a pair of dice. If dice turns up a pair of numbers $(X, Y) \in \{1, 2, \ldots, 6\}^2$ then he pays you XY dollars. E.g., if you roll $(X, Y) = (3, 6)$, you win 18 dollars. He charges you 13 dollars to play this game. Would you play? Give mathematical reasons only.

ANSWER:
(a) Markov's inequality says that $\Pr\{X > 2E[X]\} \le 1/2$ for non-negative X. Suppose $S = \{a, b, c\}$ and $\Pr(w) = 1/3$ for each $w \in S$. If $X(a) = -9$, $X(b) = 6$, $X(c) = 6$, then $E[X] = 1$ but $\Pr(X > 2) > 1/2$.
(b) There is a slow way and a fast way to compute the expected value of ij. It is easy to compute $E[X] = \texttt{E}[[]Y] = 7/2$. Then using the fact that X, Y are independent, you get $E[XY] = 49/4 = 12.25$. So, no, you shouldn't play.

6. 15+15 points
(a) Let p be a prime, and $K \subseteq \mathbb{Z}_p$ be a set of n keys to be stored using the "perfect hashing scheme" in the text (and homework). Assume a conventional computer memory

divided into words, each word large enough to store any element of $\mathbb{Z}_p$. How many words of memory do you need for implementing the perfect hashing scheme?
(b) In part (a), we assume the primary table has n slots. Suppose we now use m slots where $m < n$, in order to save space. What is the expected value of the sum $X = \sum_{i=0}^{m-1} n_i^2$ where n_i is the size of the ith bucket. In particular, show that $E[X] \le 3n$ when $m = \lfloor n/2 \rfloor$. HINT: Recall the proof that $E[X] < 2n$ when $m = n$.

ANSWER: (a) Suppose the ith slot in the primary table $T[i]$ constains a_i, b_i, m_i and an index into the secondary table S_i. Thus the size of T is $4n$. Suppose we assume, as in the homework, that $X = \sum_{i=0}^{m-1} n_i^2$ is at most $4n$. Then the total size of the secondary tables is $4n$. The overall space is therefore at most $8n$ words.
(b) Follow the same analysis as the text. Let C be the number of conflicts. It was shown that $E[X] = n + 2E[C]$, and almost immediately from the properties of universal hash functions, we get $E[C] = n(n-1)/(2m)$. If $m = \lfloor n/2 \rfloor$, then $E[C] \le n$ so $E[X] = n + 2E[C] \le 3n$.