Princeton University
COS 521: Advanced Algorithms
Final Exam Fall 2005
Sanjeev Arora
Due Jan 13 4 pm at the latest in my office, Room 307

Instructions: The test has eight questions. The first question ("quickies") has 5 parts and is worth 30 points. Each of the remaining seven questions is worth 10 points each.

Finish the test within **48 hours** after first reading it. You can consult any notes/handouts from this class and feel free to quote, without proof, any results from there. You cannot consult any other source or person in any way.

DO NOT READ THE TEST BEFORE YOU ARE READY TO WORK ON IT.

Write and sign the honor code pledge on your exam (The pledge is "I pledge my honor that I have not violated the honor code during this exam and followed all instructions.")

I will be available Jan 3-13 by email to answer any questions. I will also offer to call you if your confusion does not clear up. In case of unresolved doubt, try to explain your confusion as part of the answer and maybe you will receive partial credit.

1. (30 points) The following questions are meant to be "quickies."

   (a) Show that every d-regular bipartite graph can be decomposed into d perfect matchings.

   (b) Suppose n balls are thrown randomly into n bins. Show that there with probability at least 0.99, at least $n^{0.5}$ bins have at least $\log n/(\log\log n)^2$ balls.

   (c) Suppose somebody gives you an oracle that, given any instance of TSP, finds the cost of the optimum tour. Show how to use the oracle to generate an optimum tour.

   (d) Suppose Harry and Sally play the following game. Each starts with a quarter and a dime. Each secretly selects a coin and places it under a cup. The players then raise their cups. If the coins are equal, Harry gets both. If they are different, Sally gets both. Write the payoff matrix for this game and compute the optimal strategies for Harry and Sally in the sense of the minimax theorem.

   (e) Suppose we are given n numbers $c_1, c_2, \ldots, c_n \in Q$ and wish to solve the following exponential size program:

$$\min \sum_i c_i x_i \tag{1}$$

$$\sum_{i \in S} x_i \geq 1 \qquad \forall S \subseteq \{1 \ldots, n\}, |S| = n/2 \tag{2}$$

   Give as fast an algorithm as you can.

2. (10 points) Consider a matrix of numeric data where each entry is fractional, but each row and column is an integer. Prove that you can "round off" this matrix, rounding each entry to the next integer above or below, *without* changing the row or column sums. (Hint: Flows.)

3. (10 points) There is a data stream of items each labeled with a value in $\{1, \ldots, N\}$. We wish to compute the following quantity by making one pass over the data:

$$\sum_i i^4 f_i,$$

   where f_i is the number of items that are labeled i. Sketch an algorithm that approximates this within factor $1 + \epsilon$ using space that is logarithmic in n (or close to that).

4. (10 points) This question concerns the problem of hiring out one room to tourists who desire it. The ith tourist can arrive at or after a certain date s_i and has to leave by a certain *deadline* date d_i. It is only worthwhile for him to visit if he can stay for t_i consecutive days in the interval $[s_i, d_i]$. Only one tourist can stay in the room at a time (they hate roommates).

   Note that in general not all tourists will be accomodated since we only have one room. Describe an algorithm that finds the largest set of tourists who can be accomodated.

Modify your algorithm to find the revenue-maximizing schedule (i.e. the one that rents out the room for the longest period of time).

The inputs are s_i, t_i, d_i and the running time can depend on $D = \sum_i d_i$.

5. (10 points) The following is one of the many hard problems that arise in genome mapping. Recall that a chromosome can be viewed as a string. We are given a set of n *markers* $\{\mu_1, \mu_2, \ldots, \mu_n\}$, which are positions on the chromosome. The goal is to output a linear ordering of these markers. We are given K "constraints," where a constraint is of the form $\{\mu_i, \mu_j \mu_k\}$. The constraint is *satisfied* in the final ordering if μ_j appears between μ_i, μ_k in that ordering. Note that other markers could also lie in between μ_i, μ_k, and it is also OK for μ_k to come before μ_i so long as μ_j lies between them.

   Let OPT denote the maximum number of constraints that can be satisfied in any ordering. Describe a polynomial-time algorithm to produce an ordering that satisfies at least $OPT/3$.

6. (10 points) Let d_{ij} be a distance function on n point, where $d_{ij} \geq 0$ denotes the distance between i, j. A *Euclidean embedding* with *distortion* C is a set of vectors $u_1, u_2, \ldots, u_n \in \Re^n$ such that for all i, j:

$$|u_i - u_j|_2^2 \leq d_{ij} \leq C \cdot |u_i - u_j|_2^2.$$

   Describe a polynomial-time algorithm to compute the minimum C for which such an embedding exists. (It is also OK to give an algorithm that closely approximates C.)

7. (10 points) Let $u_1, u_2, \ldots, u_n$ be unit vectors in $\Re^m$. Then show that there exists $\epsilon_1, \epsilon_2, \ldots, \epsilon_n = \pm 1$ such that

$$|\epsilon_1 u_1 + \epsilon_2 u_2 + \cdots \epsilon_n u_n|_2 \leq \sqrt{n}.$$

   Show that there also exist ϵ_i's so that the previous inequality is reversed.

8. (10 points) For any weighted graph on n vertices with edge weights w_{ij}, let L be the $n \times n$ matrix where $L_{ij} = -w_{ij}$ if $i \neq j$ and $L_{ii} = \sum_k w_{ik}$. As usual, $\lambda_{max}(\cdot)$ denotes the largest eigenvalue.

   (a) Show that L is positive semidefinite.
   (b) Show that the value of the maximum cut in the graph is upperbounded by $\frac{n}{4}\lambda_{max}(L)$.
   (c) Show that if $d_1, d_2, \ldots, d_n$ are any numbers satisfying $\sum_i d_i = 0$ then $\frac{n}{4}\lambda_{max}(L + diag(d))$ is also an upperbound on the value of the maximum cut. Here $diag(d)$ is the matrix in which all off-diagonal entries are zero and the diagonal entries are $d_1, d_2, \ldots, d_n$ respectively.