Massachusetts Institute of Technology
6.854J/18.415J: Advanced Algorithms
David Karger

Handout 23
Wednesday, November 16, 2005

Problem Set 10 Solutions

Problem 1.

(a) We create an integer variable f_{ij}^k for each edge from i to j and for each pair k, and we state the following constraints. First we bound the range of each variable

$$\forall k\ \forall (i,j) \in E \qquad 0 \le f_{ij}^k \le 1,$$

then for each demand pair (v_k, w_k) we model the flow of unit value from the one of vertices to the other

$$\forall k \forall i \qquad \sum_j (f_{ij}^k - f_{ji}^k) = \begin{cases} 1, & \text{if } i = v_k, \\ -1, & \text{if } i = w_k, \\ 0, & \text{otherwise,} \end{cases}$$

and eventually, bound the flow along each edge by w:

$$\forall (i,j) \in E \qquad \sum_k (f_{ij}^k + f_{ji}^k) \le w.$$

Obviously, for each k the flow from v_k to w_k of unit value corresponds to the path between them, and any feasible solution to the integer linear program is a solution to our original problem, and the goal function is 0.

(b) In the relaxation we can get many fractional paths between each demand pair of total unit value. These paths can be read out of the solution in polynomial time in the following way. We find an edge e of the smallest positive flow value c, and subtract flow of this value from a cycle or a path going through e over edges of positive flow value. Repeating this procedure we will decompose the flow into at most m cycles and paths, where m is the number of edges. Furthermore, we can assume that all paths are simple paths, and no vertex appears twice on a path.

(c) For each demand pair we take paths into which we decomposed flow. Since the sum of the capacities of the paths equals 1, we randomly choose one of them with the probability that equals the capacity of the path.

Let (i,j) be an arbitrary edge. Let X_k be the Bernoulli variable equal to 1, if the path connecting the k^{th} demand pair goes over (i,j), and 0, otherwise. We know that

$$\mathbb{E}[\sum_k X_k] = \sum_k \mathbb{P}[X_k = 1] \le w = 1.$$

By the Chernoff bound, for n large enough, it holds that

$$\mathbb{P}[\sum_k X_k \geq 1 + 2\log n] \leq 2^{-(1+2\log n)} = \frac{1}{2n^2},$$

and by the union bound, the probability that over some edge goes at least $1 + 2\log n$ paths can be upper bounded by

$$n^2 \cdot \frac{1}{2n^2} = \frac{1}{2},$$

and after an expected constant number of attempts we get a $O(\log n)$-approximation.

(d) Once again we will make use of the Chernoff bound. As before, for a single edge

$$\mathbb{E}[\sum_k X_k] = \sum_k \mathbb{P}[X_k = 1] \leq w.$$

We can assume without loss of generality that $\mathbb{E}[\sum_k X_k]$ equals exactly w, since it is in our interest that $\sum_k X_k$ is as small as possible, and we can always add some virtual Bernoulli variables increasing the expectation. Slightly modifying,

$$\mathbb{P}[\sum_k X_k \geq w + 3\sqrt{w \log n}] = \mathbb{P}[\sum_k X_k \geq w(1 + 3\sqrt{\frac{\log n}{w}})],$$

and under the assumption that $w > \log n$, which yields $3\sqrt{\log n/w} < 2e - 1$, we get

$$\mathbb{P}[\sum_k X_k \geq w + 3\sqrt{w \log n}] \leq e^{-\left(3\sqrt{\log n/w}\right)^2 w/4} = e^{-9\log n/4}.$$

The probability that some edges carries at least $w + 3\sqrt{w\log n}$ paths can be upper-bounded by

$$n^2 \cdot e^{-9\log n/4} = e^{-\log n/4} = \frac{1}{\sqrt[4]{n}} \leq \frac{1}{\sqrt[4]{2}} < 1,$$

for $n \geq 2$, and after an expected constant number of attempts we get a 4-approximation.

Problem 2.

(a) The number of sets in C is not greater than 2^k, since there are only 2^k different subsets of B and therefore, at most 2^k different unions of subsets of B.

(b) This is true. Suppose that there exists a set B for C of the required property. If in some set $D \in B$ exactly one of x and y belongs to D, we can add to D the other element, not changing the union of each subset of B that equals some set in C. Therefore, we do not have to distinguish x and y, and we can safely remove one of them.

Such a simplification of the problem for all pairs x and y can be done in polynomial time.

(c) If $|C| > 2^k$, the answer is simply "no". We simplify the sets as in part (b), and we only need to show that the number of items that can appear in sets after the simplification is bounded by a function of k. It turns out that there can be at most $2^{2^k} - 1$ of them, since each two have to differ on belonging to at least one set. For fixed k we find a solution in constant time, for example checking all possible sets B.

Problem 3.

We will remove vertices from a given graph G of treewidth k in a good elimination ordering, and we can assume w.l.o.g. that G is connected. Some cliques in a current graph will be *active cliques*. For each active clique C and for each assignment of values to variables that are represented by vertices that belong to C, we store some specific assignment to some variables that have been already removed. Note that the size of any clique during the elimination is bounded by $k + 1$ (actually active cliques are always of size at most k).

The algorithm is following. In the beginning there is no active clique. When we remove a vertex v from the graph, we connect all its neighbors, and make them an active clique C. Note that if v is a member of some active clique K, K disappears from the graph, and all the other members of K become members of C. For each assignment ϕ of values to variables in C, we consider two possible assignments to the variable represented by v, and joining assignments from active cliques K_i (K_i is the i-th active clique to which v belongs) that are consistent on vertices in C and v, we choose the one that satisfies more clauses. This assignment becomes an assignment associated with ϕ in C. Note that the joining of assignments is possible, since no variable except v and these in C, appears in two assignments we are joining. This is a consequence of the fact that a variable that is already removed is set in exactly one active clique.

Eventually, we have only one vertex left, which constitutes an active clique on its own, and we simply choose a better of two assignments.

To show that the runtime of the algorithm is polynomial, it is enough to say that the single removal of a vertex takes polynomial time. This is true, since the number of active cliques is less than the number of variables, and we consider only a constant number of assignments.

The algorithm returns an optimal solution, since it can be proven by induction, that for any assignment ϕ to variables of an active clique C, we compute an assignment of some set of variables that in the original graph was separated from the rest of the graph by vertices in C, and therefore we maximize the number of satisfied clauses, that contained only these removed variables and the variables in C, for a given assignment of the variables in C. Eventually, we have two possibilities for the last variable, and one of them must be optimal.

Problem 4.

The optimum load OPT is always greater than the optimal maximum load at any given time. It suffices to show that scheduling a job according to Graham's rule we never exceed $(2 - 1/m) \cdot \text{OPT}$. Let p be the value of the job we are adding. We know that $p \leq \text{OPT}$, and that the average load can never exceed OPT. By the latter the average load right before the scheduling of the job has to be at most $\text{OPT} - p/m$, otherwise we would exceed the maximum possible average. The job is placed on the least loaded machine, and the load of this machine is not greater than $\text{OPT} - p/m$, so it will not be greater than $\text{OPT} + (1 - 1/m)p$, what can be in turn bounded by the required $(2 - 1/m)\text{OPT}$.

Problem 5.

(a) In any deterministic strategy we choose a next possible partner to date based only on ordering of the ones dated so far and k, the number of potential partners. We build a malicious set of partners in the following way. We observe a behavior of the strategy, when it is given worse and worse partners, until it eventually accepts one of them, and we make all partners that have not been dated the best ones (in any order). The strategy chooses on this set of partners the worst one.

(b) We split all possible partners into two sets: a sample set S, and a set T, from which we will choose a companion. Each possible partner goes to each of sets with the same probability, i.e. 1/2. We observe first all partners in S, by dating them, and then date partners in T, until we find a one better than the best in S. If one of sets is empty, or we do not find a required partner in T, we take as a companion the first or the last partner, respectively.

The probability that the best possible partner is in T and the second best in S equals by independence $1/2 \cdot 1/2 = 1/4$, and with probability at least 1/4 we will choose the best one.

(c) As before we split possible partners, and dating these in S we enquire information about the population at MIT. Then we date potential partners in T, until we find a one that is better than the $2\log k$-th in S.

We consider 3 possible cases for k large enough. Suppose first that in S there is no person better than $2\log k$-th best in S, and we need to take the last person. This implies that the best $2\log k$ persons are in set S, and the probability that it happens equals

$$\left(\frac{1}{2}\right)^{2\log k} = \frac{1}{k^2}.$$

The second case that we consider occurs when the $2\log k$-th best in S has rank greater than $12\log k$, what can happen only if among the best $12\log k$ elements in the whole set, there is at most $2\log k$ that belong to S, and we can bound the probability of this event, making use of independence of indicators X_i whether the i-th best belongs to S. Taking

$$Y = \sum_{i=1}^{12\log k} X_i,$$

by the Chernoff bound, we get

$$\mathbb{P}[Y \leq 2\log k] = P[Y \leq \left(1 - \frac{2}{3}\right) 6\log k] \leq e^{-\frac{1}{2}\cdot\frac{4}{9}\cdot 6\log k} = e^{-\frac{4}{3}\log k} \leq e^{-\ln k} = \frac{1}{k}.$$

The third and last case is when we find in T a companion better than the $2\log k$-th best in S, and the $2\log k$-th best in S has a rank smaller than $12\log k$.

Let R be a random variable equal to the rank of a companion that we choose. Eventually,

$$\mathbb{E}[R] = \sum_{i=1}^{3} \mathbb{P}[\text{case } i] \cdot \mathbb{E}[R|\text{case } i] \leq \frac{1}{k^2} \cdot k + \frac{1}{k} \cdot k + 1 \cdot 12\log k = O(\log k).$$

(d) If someone wants to maximize their chances, sometimes, it may be difficult to believe that a potential partner they are dating should be only a sample, and therefore they can fall into a trauma caused by the conflict between the belief that one should stick to the strategy and volcanic feelings that arise to this potential partner.