Lecture 7: Approximation via Randomized Rounding

Often LPs return a fractional solution where the solution x, which is supposed to be in $\{0,1\}^n$, is in $[0,1]^n$ instead. There is a generic way of obtaining an $\{0,1\}^n$ vector from a $x \in [0,1]^n$ one: let the ith variable be 1 with probability x_i, 0 with probability $(1-x_i)$. The resulting random vector $x^{\texttt{int}}$ has the desirable property that for any $i, \mathbf{E}[x^{\texttt{int}}{}_i] = x_i$. In particular, if x minimized a certain *linear* function $c \cdot x$, then by linearity of expectation, the expected cost of the random $\{0,1\}^n$ vector is also equal to $c \cdot x$. Of course, often the resulting $x^{\texttt{int}}$ won't satisfy the *constraints*, and therein lies the non-triviality.

We will need some facts in probability theory, mostly in the concentration of random variables around their means. I will collect these facts down in the appendix; the proofs of which can be found in texts on randomized algorithms or the web.

The Set Cover Problem. Recall the LP relaxation for the set cover problem.

$$\begin{aligned}
\min \quad & \sum_{j=1}^{m} c(S_j)x_j & & (1)\\
\text{subject to} \quad & \sum_{j:i\in S_j} x_j \geq 1 & \forall i \in U \quad & (2)\\
& x_j \in [0,1] & \forall j = 1...m \quad & (3)
\end{aligned}$$

Let x be a solution to the LP. Consider the experiment where we randomly select set S_j with probability x_j, *independently* of each other. Let X_j be the indicator random variable which is 1 if we picked S_j and 0 otherwise. The random variable $\sum_{j=1}^{m} c(S_j)X_j$ indicates the cost of the sets picked, and thus the expected cost of our solution is precisely $\sum_{j=1}^{m} c(S_j)\mathbf{E}[X_j]$ which is the LP solution. Are all elements covered? Fix an element i. The following claim shows that each element is covered with at least a constant probability.

Claim 1. *The probability that an element i is covered is at least $(1-1/e)$.*

Proof. Element i is covered iff a set containing i is picked. The probability that i is not covered, thus, is at most

$$\prod_{j:i\in S_j} (1-x_j)$$

due to independence. But this is at most

$$\prod_{j:i\in S_j} e^{-x_j} = e^{-\sum_{j:i\in S_j} x_j} \leq \frac{1}{e}$$

where the inequality follows from (2). □

So our randomized algorithm returns a collection of sets whose expected cost is the LP cost, and covers each element with probability at least $(1 - 1/e)$. Suppose we repeated this $2\ln n$ times and returned the union of the sets picked in each round. The expected cost of our solution is going to be $2\ln n$ times the LP solution. What's the probability that an element i is not covered in any of the rounds? Since the rounds are independent, this probability is at most $(\frac{1}{e})^{2\ln n} = 1/n^2$. By the union bound (Fact 2), the probability that there exists some element i which is not covered is at most $n \cdot \frac{1}{n^2} = 1/n$. Thus, we get the following theorem.

Theorem 1. *The algorithm described above returns a set cover with probability at least* $(1 - \frac{1}{n})$ *whose expected cost is* $2\ln n$ *times the LP solution.*

Randomized approximation algorithms often look like the theorem above. They "succeed" with high probability and return a solution whose expected cost is at most (or at least) a factor of the optimum solution. By taking a slight hit at the approximation factor, the guarantee on the expected cost can often be converted to a guarantee on the cost with high probability. For instance, Markov's inequality (Fact 3) tells us that the probability that the cost of the sets picked exceeds $4\ln n$ times the optimum is at most $1/2$. Thus, the algorithm returns a set cover of cost $4\ln n$ times the optimum with probability at least $\frac{1}{2} - \frac{1}{n} \geq 1/3$. If we repeat this algorithm t times, then with probability $1 - \frac{1}{3^t}$ (really, really high probability if t is some polynomial in n) we will encounter such a set cover.

Suppose we look at the set multicover problem where each element i needs to be covered at least b_i times for some positive integer b_i? The above analysis needs to be modified a bit; in particular, the probability that an element i is not covered in any iteration goes down from $1/e$ to $1/e^{b_i}$, although, it now needs to be covered b_i times. This is true, and I'll leave this calculation checking as an exercise. I now show a slightly different (but similar in spirit) randomized algorithm. It runs in a single round by sampling sets with higher probabilities.

Let $\hat{x}_j = \min(1, 6\ln n \cdot x_j)$. Sample each set S_j independently with probability $\hat{x}_j$. That's it. The expected cost now is $\sum_j c(S_j)\hat{x}_j \leq 6\ln n \cdot \mathtt{lp}$. To argue that we cover each element sufficient number of times, we use the Chernoff bound (Fact 4). Let X_j be the indicator random variable of the event whether S_j is picked or not. For an element i, define $Z_i = \sum_{j: i \in S_j} X_j$. If $Z_i \geq b_i$, then i is covered, otherwise not. The following claim, using a union bound, gives that with probability $(1 - 1/n)$, the sets picked form a set cover.

Claim 2. $\mathbf{Pr}[Z_i < b_i] \leq \frac{1}{n^2}$.

Proof. By re-ordering suppose $i \in S_j$ for $1 \leq j \leq k$. We may assume $\hat{x}_j < 1$ for all such j; since all others (say ℓ of them) are picked with probability 1, we can move to the residual problem where the element i needs to be covered $b_i - \ell$ times. Now, $\mathbf{E}[Z_i] = \sum_{j=1}^{k} \hat{x}_j = 6\ln n \cdot b_i$. Thus, Fact 4 gives us (with $\delta = 1 - \frac{1}{6\ln n} \geq 5/6$)

$$\mathbf{Pr}[Z_i < b_i] \leq \exp(-6\ln n \cdot \frac{25}{72}) \leq \frac{1}{n^2}$$

□

Facility Location Problem

Let's recall the LP relaxation for the facility location problem and it's dual. Let (x, y) and (α, β) be a pair of optimal primal-dual solutions. Recall the following claim from last class

$$\begin{aligned} \min \quad & \sum_{i\in F} f_i y_i + \sum_{i\in F, j\in C} c(i,j)x_{ij} && (4)\\ & \sum_{i\in F} x_{ij} = 1, \qquad \forall j \in C && (5)\\ & y_i \geq x_{ij}, \qquad \forall i\in F,\ \forall j \in C && (6)\\ & x, y \geq 0 \end{aligned}$$

$$\begin{aligned} \max \quad & \sum_{j\in C} \alpha_j && (7)\\ & \sum_{j\in C} \beta_{ij} \leq f_i, \quad \forall i \in F && (8)\\ & \alpha_j - \beta_{ij} \leq c(i,j), \quad \forall i\in F,\ \forall j\in C && (9)\\ & \beta \geq 0 \end{aligned}$$

Claim 3. *If $x_{ij} > 0$, then $c(i,j) \leq \alpha_j$.*

Before we describe the algorithm, we are going to make a structural assumption about (x, y). I will leave it as an exercise to prove that the assumption is without loss of generality.

Assumption: If $x_{ij} > 0$, then $y_i = x_{ij}$.

Algorithm. (In class, we looked at the clustering algorithm which ordered clients in increasing order of α_j, and gave us a $(2+2/e)$ approximation. Below we give a refined argument which attains a $(1 + 2/e)$-approximation.)

The algorithm proceeds as follows. Let $C^*_j := \sum_i c_{ij}x_{ij}$. Order the clients in increasing order of $(\alpha_j + C^*_j)$. Let U denote the set of unassigned clients initialized to C. Let j be the first client in U in the order. Form the cluster $N_2(j)$ consisting of j, all facilities $\{i : x_{ij} > 0\}$, and all clients $\{j' \in U : x_{ij'} > 0 \text{ for some } i \in N_2(j)\}$. j is called the cluster center of $N_2(j)$. Remove all clients of $N_2(j)$ from U and repeat till U is empty. At the end of this step all clients are assigned some cluster $N_2(j)$, for some client j.

Note that in any cluster $N_2(j)$, $\sum_{i\in N_2(j)} y_i = \sum_{i\in N_2(j)} x_{ij} = 1$. In each such cluster, open exactly one facility with probability y_i - this is valid since $\sum_i y_i = 1$. Let X be the set of facilities opened. The clients are assigned their nearest facility.

Claim 4. *The expected facility opening cost is precisely $\sum_{i\in F} f_i y_i$.*

Proof. The probability a facility i is in X is precisely y_i. □

Let us now calculate the expected connection cost for a client j. Note that a cluster center j pays expected connection cost $\sum_{i\in N_2(j)} c(i,j)x_{ij} = C^*_j$. Let's take a non-center client $j \in N_2(j')$. Let $\Gamma(j) := \{i : x_{ij} > 0\}$, and let $\{j_1, \ldots, j_r\}$ be the centers such that $N_2(j_\ell) \cap \Gamma(j)$ is non-empty. For $\ell = 1, \ldots, r$, let $p_\ell := \sum_{i\in N_2(j_\ell)} x_{ij}$ be the probability a facility in $N_2(j_\ell) \cap \Gamma(j)$ is opened. Note that $\sum_\ell p_\ell = 1$, and that these r events are independent.

Let $C^*_j(\ell)$ be the expected cost of connecting to a client in $N_2(j_\ell) \cap \Gamma(j)$ conditioned on the event that such a facility is open. Thus,

$$C^*_j(\ell) := \sum_{i\in N_2(j_\ell)} c(i,j) \cdot \frac{x_{ij}}{p_\ell}$$

Let's order $j_1, \ldots, j_\ell$ so that $C^*_j(1) \leq \cdots \leq C^*_j(\ell)$. Lastly, let's define $C^*_j(\infty)$ to be the expected cost of connecting j conditioned on the event that no client in $\Gamma(j)$ is opened. We'll bound this shortly. Now we are ready to bound the expected connection cost of client j.

Claim 5. *The expected connection cost of client j is at most*

$$p_1 C_j^*(1) + (1-p_1)p_2 C_j^*(2) + \cdots + (1-p_1)(1-p_2)\cdots(1-p_{r-1})p_r C_j^*(r) + C_j^*(\infty)\prod_{\ell=1}^{r}(1-p_\ell)$$

Proof. We are conditioning on mutually exclusive events: event ℓ is when no facility in $N_2(j_{\ell'}) \cap \Gamma(j)$ has been opened for $\ell' < \ell$, and a facility in $N_2(j_\ell) \cap \Gamma(j)$ has been opened; and the last event when no facility in $\Gamma(j)$ has been opened. □

Claim 6. $C_j^*(\infty) \le C_j^* + 2\alpha_j$.

Proof. If no facility in $\Gamma(j)$ is opened, then connect j to the facility i' opened in $N_2(j')$: recall $j \in N_2(j')$ and thus there exists $i \in N_2(j')$ with $x_{ij} > 0$. The connection cost to i' is $c(i', j) \le c(i, j) + c(i, j') + c(i', j') \le 2\alpha_j + c(i', j')$, where we have used Claim 3 and the fact that $\alpha_{j'} \le \alpha_j$. Taking expectations, we get

$$C_j^*(\infty) \le 2\alpha_j + \sum_{i \in N_2(j)} c(i', j')x_{i'j'}.$$

□

Putting it all together, we get the total connection cost of j is at most

$$\sum_{\ell=1}^{r}\left[p_\ell(1-p_{\ell-1})\cdots(1-p_1)C_j^*(\ell)\right] \; + \; (C_j^* + 2\alpha_j)\prod_{\ell=1}^{r}(1-p_\ell) \tag{10}$$

Since $\sum_{\ell=1}^{r} p_\ell = 1$, we have the product $\prod_{\ell=1}^{r}(1-p_\ell) \le 1/e$. As a first approximation, note that $(1-p_\ell) \le 1$ and $\sum_{\ell=1}^{r} p_\ell C_j^*(\ell) = C_j^*$. This gives us total expected facility + connection cost of at most

$$\sum_{i \in F} f_i y_i + \sum_{j \in C} C_j^* + \frac{1}{e}\sum_{j \in C}(C_j^* + 2\alpha_j) \le \mathtt{opt}(1 + 3/e)$$

giving a 2.104...-approximation.

One can analyze better. Now we state the following inequality:

Lemma 1. *Let $A_1 \le A_2 \cdots \le A_r$, and let $\sum_{i=1}^{r} p_i = 1$. Then,*

$$p_1 A_1 + p_2(1-p_1)A_2 + \cdots + p_r(1-p_1)(1-p_2)\cdots(1-p_{r-1})A_r \le \left(\sum_{i=1}^{r} p_i A_i\right)\cdot\left(1 - \prod_{i=1}^{r}(1-p_i)\right)$$

Proof. Define $q_i := \frac{p_i(1-p_1)\ldots(1-p_{i-1})}{1-\prod_{i=1}^{r}(1-p_i)}$. We need to prove that $\sum_{i=1}^{r} q_i A_i \le \sum_{i=1}^{r} p_i A_i$. Note that $\sum_{i=1}^{r} q_i = 1 = \sum_{i=1}^{r} p_i$. Furthermore, there exists a $1 < k < r$ such that $q_i \ge p_i$ for all $i \le k$ and $q_i \le p_i$ for all $i > k$; it's the largest i for which $\prod_{\ell=1}^{i}(1-p_\ell)$ exceeds $1 - \prod_{i=1}^{r}(1-p_i)$. Let $\delta_i = q_i - p_i$. Thus, $\delta_i \ge 0$ for $i \le k$, $\delta_i \le 0$ for $i > k$ and $\sum_i \delta_i = 0$.

$$\sum_{i=1}^{r}(q_i - p_i)A_i = \sum_{i=1}^{r}\delta_i A_i \le A_k\left(\sum_{i=1}^{k}\delta_i\right) + A_{k+1}\left(\sum_{i=(k+1)}^{r}\delta_i\right) = \left(\sum_{i=1}^{k}\delta_i\right)(A_k - A_{k+1}) \le 0$$

□

Given the above fact, we can get a stronger approximation factor. Let $q := \prod_{\ell=1}^{r}(1-p_\ell)$. Then, the expected connection cost is at most

$$(1-q)C_j^* + qC_j^*(\infty) \leq C_j^* + 2\alpha_j/e$$

Putting it together gives a $(1+2/e)$-approximation.

The Group Steiner Tree Problem (on a tree)

In this problem, the input is a rooted tree T on n vertices. There are k disjoint groups $g_1, \ldots, g_k$, where each group is a subset of leaves. Each edge of the tree has a cost, and the goal is to find the cheapest rooted sub-tree which contains a vertex from each group.

LP Relaxation.

$$\begin{aligned} \min \quad & \sum_{e \in E(T)} c_e x_e & x \geq 0 \qquad (11) \\ \text{subject to} \quad & x(\delta(S)) \geq 1 & \forall S \subseteq V : g_i \subseteq S \text{ for some } i, r \notin S \qquad (12) \end{aligned}$$

Algorithm. Let the root be r. Given a non-root vertex v in the tree, let e_v denote the edge connecting v to its parent. We'll abuse notation and let x_v denote x_{e_v}. We call an edge f an ancestor of edge e if the path from r to the end points of e contains f. f is the parent of e if its an ancestor and shares an end point with e.

Preprocessing x. We process the solution x as follows. Firstly, remove all vertices v with $x_v \leq 1/2g$. For any group g_i, we still have $\sum_{v \in g_i} x_v \geq 1/2$. Secondly, increase x_e for every edge to the nearest power of $(1/2)$. Note that the new LP cost at most doubles. Thirdly, for an edge e with parent f, if $x_e = x_f$, then merge the two edges into a single edge with the same x_e value. This doesn't change the feasibility of the solution. Note that the height of the tree now is at most $2\log g$: any path from root to a vertex v has strictly decreasing x_e in powers of $1/2$ ranging from at most 1 to at least $1/2g$. Finally, decrease all the x_{e_v}'s such that for any group g_i, $\sum_{v \in g_i} x_v = 1$. Since these are leaf edges and groups are disjoint, this can be done without changing the feasibility of the solution.

Independently, sample all edges e incident on the root with probability x_e. For every other edge, e with parent f, sample e independently with probability $\frac{x_e}{x_f}$. The resulting graph can have many components - throw away all except the component containing the root. The above step is repeated $(16 \ln n \cdot \log g)$ times independently and the final solution is the union of all edges picked; g is the maximum size of a group.

Theorem 2. *The above algorithm has an expected cost of $O(\log n \log g) \cdot \texttt{opt}$ and is a feasible group Steiner tree with high probability.*

Proof. In each iteration the probability that an edge e is sampled is exactly x_e. This is because, let the path from root to e be $e_p, e_{p-1}, \ldots, e_1, e$. The probability e survives this iteration is at most

$$\frac{x_e}{x_{e_1}} \frac{x_{e_1}}{x_{e_2}} \cdots \frac{x_{e_{p-1}}}{x_{e_p}} \cdot x_{e_p} = x_e$$

By linearity of expectation, the expected cost of the edges in any iteration is `opt`, and the first claim in the theorem follows.

Fix a group g_i and a iteration. We claim that the probability that the sub-tree sampled contains a vertex from g is at least $\frac{1}{8 \log g}$. This will prove the second claim of the theorem since the probability group g_i is not covered in $16 \ln n \log g$ iterations is at most $\left(1 - \frac{1}{8 \log g}\right)^{16 \ln n \log g} \leq \frac{1}{n^2}$. Union bound over all groups (there are n many at most) ends the argument.

For a vertex $u \in g_i$, let $\mathcal{E}_u$ be the event that u is contained in the subtree. We are trying to lower bound the event $\mathbf{Pr}[\bigcup_{u \in g_i} \mathcal{E}_u]$. Let's calculate $\mathbf{Pr}[\mathcal{E}_u]$. u is contained in the tree if and only if all the edges on the path from u to r are sampled. This probability is x_u. Furthermore, the constraint in the LP gives us $\sum_{u \in g_i} x_u = 1$. So, we have

$$\mu := \sum_{u \in g_i} \mathbf{Pr}[\mathcal{E}_u] = 1.$$

(Thus, if all the $\mathcal{E}_u$'s were independent in a group g_i, then in fact the probability that none of the vertices are contained in the tree would have been at most $1/e$.) Let $u \sim v$ if the events $\mathcal{E}_u$ and $\mathcal{E}_v$ are not independent. Define $\Delta := \sum_{u \sim v} \mathbf{Pr}[\mathcal{E}_u \wedge \mathcal{E}_v]$.

Claim 7. $\Delta \leq 2 \log g$

Proof. For u and v in the tree, let w be the least common ancestor. Note that conditioned on the event $\mathcal{E}_w$, the event $\mathcal{E}_u$ and $\mathcal{E}_v$ are independent. Thus the probability

$$\mathbf{Pr}[\mathcal{E}_u \wedge \mathcal{E}_v] = \mathbf{Pr}[\mathcal{E}_u \wedge \mathcal{E}_v | \mathcal{E}_w] \cdot \mathbf{Pr}[\mathcal{E}_w] = \mathbf{Pr}[\mathcal{E}_u | \mathcal{E}_w] \cdot \mathbf{Pr}[\mathcal{E}_v | \mathcal{E}_w] \cdot \mathbf{Pr}[E_w] = \frac{x_u x_v}{x_w}$$

Now for any interior node w, let T_w be the sub-tree rooted at w. Note that $x_w \geq \sum_{u \in g_i \cap T_w} x_u$. This is because of the feasibility of the LP constraint with S containing all vertices in T_w and all vertices in $g_i \setminus T_w$:

$$1 \leq x(\delta(S)) = x_w + \sum_{u \in g_i \setminus T_w} x_u = x_w + 1 - \sum_{u \in g_i \cap T_w} x_u$$

Let A_u denote the set of ancestors of u in the tree. Since the height of the tree is at most $2 \log g$, $|A_u| \leq 2 \log g$. Thus,

$$\Delta = \sum_{u \sim v} \frac{x_u x_v}{x_w} \leq \sum_{u \in g_i} \sum_{w \in A_u} \sum_{v \in g_i \cap T_w} \frac{x_u x_v}{x_w} \leq \sum_{u \in g_i} x_u \sum_{w \in A_u} \frac{1}{x_w} \sum_{v \in g_i \cap T_w} x_v \leq \sum_{u \in g_i} x_u |A_u| \leq 2 \log g$$

□

Now by Janson's inequality (Fact 5), we get that

$$\mathbf{Pr}[\bigwedge_{u \in g_i} \overline{\mathcal{E}_u}] \leq \exp\left(-\frac{\mu^2}{\mu + \Delta}\right) \leq \exp\left(-\frac{1}{4 \log g}\right) \leq 1 - \frac{1}{8 \log g}$$

since $e^{-x} \leq 1 - x + \frac{x^2}{2} \leq 1 - x/2$ when $0 \leq x \leq 1$. So we get that in any iteration and any group g_i, the probability we pick a vertex from the group is at least $\frac{1}{8 \log g}$. This completes the proof. □

Some Probabilistic Facts.

Fact 1 (Linearity of Expectation)**.** *Let* $X_1, \ldots, X_n$ *be random variables and let* $X := \sum_{i=1}^n X_i$*. Then,* $\mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[X_i]$*.*

Fact 2 (The Union Bound)**.** *Given* n *events* $\mathcal{E}_1, \ldots, \mathcal{E}_n$*, the probability that one of them occurs is at most the sum of their probabilities.*

$$\mathbf{Pr}[\mathcal{E}_1 \vee \mathcal{E}_2 \vee \cdots \vee \mathcal{E}_n] \leq \sum_{i=1}^n \mathbf{Pr}[\mathcal{E}_i]$$

Fact 3 (Markov's Inequality)**.** *Let* X *be a* non-negative *random variable. Then*

$$\mathbf{Pr}[X \geq t\mathbf{E}[X]\ ] \leq \frac{1}{t}$$

Fact 4 (Chernoff Bound)**.** $X_1, \cdots, X_n$ *be* n independent $\{0,1\}$ *random variables with* $\mathbf{Pr}[X_i = 1] = p_i$*. Let* $X := \sum_{i=1}^n X_i$ *and let* $\mu = \sum_{i=1}^n p_i = \mathbf{E}[X]$*. Then*

$$\mathbf{Pr}[X > (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$$

$$\mathbf{Pr}[X < (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^\mu$$

Useful versions of the above:

- *For any* $\delta > 0$*,*

$$\mathbf{Pr}[X > (1+\delta)\mu] \leq \exp(-\mu\delta^2/3)$$
$$\mathbf{Pr}[X < (1-\delta)\mu] \leq \exp(-\mu\delta^2/2)$$

- *For* $t > 0$*,*

$$\mathbf{Pr}[|X-\mu| > t] \leq 2\exp(-2t^2/n)$$

- *For* $t > 4\mu$*,*

$$\mathbf{Pr}[X > t] \leq 2^{-t}$$

Fact 5 (Janson's Inequality)**.** *Let* $X \subset U$ *obtained by sampling element* $i \in U$ *with probability* p_i *independently.* $A_1, \ldots, A_t$ *be subsets of* U *and let* $\mathcal{E}_i$ *be the event that* $A_i \subseteq X$*. Then, if* $\mu = \sum_{i=1}^t \mathbf{Pr}[\mathcal{E}_i]$ *and* $\Delta = \sum_{i,j:A_i \cap A_j \neq \emptyset} \mathbf{Pr}[\mathcal{E}_i \wedge \mathcal{E}_j]$*, then*

$$\mathbf{Pr}[\bigwedge_{i=1}^t \overline{\mathcal{E}_i}] \leq \exp\left(-\frac{\mu^2}{\mu+\Delta}\right).$$