Randomized Rounding and Set Cover

Based on Vazirani: Chapter 14

Last Revision: Nov 25, 2006

Recall the problem. Let $U = \{u_1, \ldots, u_n\}$ be a set and $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ be a family of subsets of X such that $X = \cup_i F_i$.

The *Set Cover* problem is to find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$ that covers X, i.e., $X = \cup_{F \in \mathcal{C}} F$. In what follows OPT is the size of a minumum set cover.

In this lecture we will see how to use Randomized Rounding to find a "random" subset $\mathcal{C} \in \mathcal{F}$ that has the following properties.

- $Pr(|\mathcal{C}] \geq 4 \ln n \cdot OPT) \leq \frac{1}{2} + o(1)$
- $Pr(\mathcal{C}$ is not a set cover$) \leq \frac{1}{n}$.

Combining the two gives that

$$Pr\left([|\mathcal{C}] \leq 4 \ln n \cdot OPT] \text{ and } [\mathcal{C} \text{ is a set cover}]\right) \geq \frac{1}{2} + o(1).$$

This implies that by repeating this construction many times and choosing the smallest cover $\mathcal{C}$ constructed, we have probability arbitrarily close to one of not having constructed a $4 \ln n$ approximate cover.

We start by constructing the integer linear program

Minimize $\Sigma_i \, x_i$
subject to conditions
$\forall j,\ \Sigma_{u_j \in F_i} \, x_i \geq 1$
$\forall i,\ x_i \in \{0, 1\}$

Given a feasible solution x to the linear program define

$$\mathcal{C}(x) = \{F_i \,:\, x_i = 1\}.$$

The condition $\forall j,\ \Sigma_{u_j \in F_i} \, x_i \geq 1$ means that $\mathcal{C}$ is a cover, so every x corresponds to a cover.

Working backwards, given cover $\mathcal{C}$, we can define $x_i(\mathcal{C})) = 1$ if $F_i \in \mathcal{C}$ and 0 otherwise. $x(\mathcal{C})$ is a feasible solution. This gives us a one-one correspondence between feasible solutions and set covers.

Since $|\mathcal{C}| = \Sigma_i \, x_i$, minimizing the objective function is equivalent to solving minimal set cover.

The original integer LP was

Minimize $\Sigma_i x_i$
subject to conditions
$\forall j,\ \Sigma_{u_j \in F_i} x_i \geq 1$
$\forall i,\ x_i \in \{0, 1\}$

We now relax the problem to a standard LP to find

Minimize $\Sigma_i x_i$
subject to conditions
$\forall j,\ \Sigma_{u_j \in F_i} x_i \geq 1$
$\forall i,\ 0 \leq x_i \leq 1$

(We can replace the last constraints by $\forall i,\ 0 \leq x_i$; why?)

Randomized Rounding for Set Cover
Solve the Relaxed Linear Program
    Calculate (x^*)
Set $\mathcal{C} = \emptyset$.
for $i = 1$ to n do
    Flip a x_i^*-biased coin
        If Heads put F_i into $\mathcal{C}$.

Note that

$$E(|\mathcal{C}|) = \sum_i x_i^* \leq OPT.$$

Our major problem is that nothing says that $\mathcal{C}$ must be a set cover.

Lemma: Let $u_j \in U$ and $\mathcal{C}$ as constructed by the randomized rounding algorithm. Then

$$Pr(\mathcal{C} \text{ does not cover } u_j) \leq \frac{1}{e}.$$

Proof: Suppose that u_j is contained in k sets 0f $\mathcal{F}$. W.L.O.G. we may assume that these are $F_1, F_2, \ldots, F_k$. From the linear program we have that

$$\sum_{i=1}^{k} x_i^* \geq 1$$

and each $x_i \leq 1$. Now

$$\begin{aligned} Pr(\mathcal{C} \text{ does not cover } u_j) &= Pr(\text{None of } F_1, F_2, \ldots, F_k \text{ are in } \mathcal{C}\ ) \\ &= \prod_{i=1}^{k} (1 - x_i^*) \\ &\leq \left(\frac{k - \sum_{i=1}^{k} x_i^*}{k} \right)^k \\ &\leq \left(1 - \frac{1}{k}\right)^k \leq \frac{1}{e}. \end{aligned}$$

Now *Run the Randomized Rounding algorithm* $K = \lceil 2 \ln n \rceil$ *times,* constructing n collections $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_K$ and then setting

$$\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \ldots \cup \mathcal{C}_K.$$

Note that $E(|\mathcal{C}|) = \Sigma_{i=1}^{K} E(|\mathcal{C}_i|) \le \lceil 2 \ln n \rceil OPT$.

Recall *Markov's inequality* that, for any random variable X and $\alpha > 0$,

$$Pr(|X| \ge \alpha) \le \frac{E|X|}{\alpha}.$$

Setting $X = |\mathcal{C}|$ and $\alpha = 4 \ln n \, OPT$ gives

$$Pr(|\mathcal{C}| \ge 4 \ln n \, OPT) \le \frac{\lceil 2 \ln n \rceil OPT}{4 \ln n \, OPT} = \frac{1}{2} + o(1).$$

Now, for any given j

$$Pr(\mathcal{C} \text{ does not cover } u_j) \le \left(\frac{1}{e}\right)^K \le \left(\frac{1}{e}\right)^{2 \ln n} = \frac{1}{n^2}.$$

The only way that $\mathcal{C}$ can not be a cover is if one of the u_j is not covered by $\mathcal{C}$ so

$$Pr(\mathcal{C} \text{ is not a set cover}) \le \sum_j Pr(\mathcal{C} \text{ does not cover } u_j) \le n \frac{1}{n^2} = \frac{1}{n}.$$