CS787: Advanced Algorithms	**Scribe:** Yinglun Zhu
Lecture 7: Randomized Rounding	**Date:** Feb. 14, 2019

7.1 The Set Cover Problem

Given a set of elements $E = \{e_1, e_2, \ldots, e_n\}$ and a collection of sebsets $S = \{s_1, s_2, \ldots, s_m\}, S_j \subseteq E$ for $1 \leq j \leq m$, the set cover problem is to identify the smallest sub-collection of S whose union equals E. It should be noticed that the vertex cover problem is a special case of the set cover problem where E equals the set of all edges and $S = \{s_1, s_2, \ldots, s_n\}$ with s_i be the set of edges incident on vertex i. It is special in the sense that each element appears in exactly two subsets.

7.1.1 ILP Formulation

Before formalizing the integer linear programming problem, we first introduce a slgihtly generalized version of the set cover problem where each subset s_i is associated with cost c_i (which used to be 1), and the goal is thus changed to picking a minimum cost collection of subsets that covers E. Let $x_i = \mathbb{1}\{s_i \text{ is in the set cover}\}$, then the problem can be formulated as the following integer linear program

$$\begin{aligned} \min_{x_i} \quad & \sum_{i=1}^{m} c_i x_i \\ \text{s.t.} \quad & \sum_{i: e \in s_i} x_i \geq 1, \quad \forall e \in E \\ & x_i \in \{0, 1\} \end{aligned} \tag{7.1.1}$$

Since ILP is NP-Hard, we relax the ILP to the following LP

$$\begin{aligned} \min_{x_i} \quad & \sum_{i=1}^{m} c_i x_i \\ \text{s.t.} \quad & \sum_{i: e \in s_i} x_i \geq 1, \quad \forall e \in E \\ & x_i \in [0, 1] \end{aligned} \tag{7.1.2}$$

Solving the linear program gives us a fractional solution which serves as a lower bound of the optimal solution. Rounding the fractional x_i to integers will then give us a valid solution to the original ILP problem.

7.2 Rounding Algorithms

We will first provide an deterministic rounding algorithm which generates an F-approximation to the solution, and then give a randomized rounding algorithm which outputs an $O(\log n)$-approximation solution.

7.2.1 Deterministic Rounding

Definition 7.2.1 *Let F be the maximum number of sets that an element belongs to.*

Algorithm 1 Deterministic rounding

Step 1: solve the LP and obtain solution $\{x_i\}$
Step 2: Let $x_i' = \mathbb{1}\left(x_i \geq \frac{1}{F}\right)$ and output $\{x_i'\}$ as the solution to the ILP

Theorem 7.2.2 *Algorithm 1 gives an F-approximation to the ILP 7.1.1.*

Proof: We are first going to prove that $\{x_i'\}$ is a feasible solution to the ILP. Since $\{x_i\}$ is a solution to the LP, for any $e \in E$, we have $\sum_{i:e\in s_i} x_i \geq 1$; and this implies at least one of $\{x_i, e \in s_i\}$ should be greater or equal to $\frac{1}{F}$ (pigeonhole pricipal). Let $x_i' = \mathbb{1}\left(x_i \geq \frac{1}{F}\right)$, we thus know that $\sum_{i:e\in s_i} x_i \geq 1, \ \forall e \in E$.

Moreover, since $x_i' \leq F x_i$ and $\{x_i\}$ serves as a lower bound of the ILP, we have

$$\sum_{i=1}^{m} c_i x_i' \leq F \cdot \sum_{i=1}^{m} c_i x_i \leq F \cdot OPT \tag{7.2.3}$$

■

Remark 7.2.3 *In the worst case, F may be of order $\Omega(n)$; thus, the randomized rounding algorithm that we are going to discuss later will generally give us better gaurantees.*

7.2.2 Concentration Inequalities

Before getting into the randomized rounding algorithms, we first introduce two concentration inequalities, namely, Markov's inequality and Chebyshev's inequality.

7.2.2.1 Markov's inequality

Theorem 7.2.4 *If X is a nonnegative random variable and $\epsilon > 0$, then we have*

$$\mathbb{P}\left(X > \epsilon\right) \leq \frac{\mathbb{E}(X)}{\epsilon} \tag{7.2.4}$$

Proof:

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}\left(X|X \geq \epsilon\right) \cdot \mathbb{P}\left(X \geq \epsilon\right) + \mathbb{E}\left(X|X < \epsilon\right) \cdot \mathbb{P}\left(X < \epsilon\right) \\ &\geq \mathbb{E}\left(X|X \geq \epsilon\right) \cdot \mathbb{P}\left(X \geq \epsilon\right) \\ &\geq \epsilon \cdot \mathbb{P}\left(X \geq \epsilon\right)\end{aligned} \tag{7.2.5}$$

Rearranging the inequality above gives us the desired result. ■

7.2.2.2 Chebyshev's inequality

Theorem 7.2.5 *Let X be a random variable with finite mean $\mathbb{E}(X)$ and finite variance $\mathbb{V}(X)$. Then for any $\epsilon > 0$, we have*

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq \epsilon\right) \leq \frac{\mathbb{V}(X)}{\epsilon^2} \tag{7.2.6}$$

Proof:

$$\begin{aligned}\mathbb{P}\left(|X - \mathbb{E}(X)| \geq \epsilon\right) &= \mathbb{P}\left(\left(X - \mathbb{E}(X)\right)^2 \geq \epsilon^2\right) \\ &\leq \frac{\mathbb{E}\left(\left(X - \mathbb{E}(X)\right)^2\right)}{\epsilon^2} \\ &= \frac{\mathbb{V}(X)}{\epsilon^2}\end{aligned} \tag{7.2.7}$$

where the inequality comes from the Markov's inequality. ■

7.2.3 Randomized Rounding

Algorithm 2 Randomized rounding

Step 1: solve the LP and obtain solution $\{x_i\}$
Step 2: $\forall i \in [m]$, pick s_i independently with probability x_i
Step 3: repeat step 2 until all elements are covered

In order to analyze the algorithm, we consider the following version of the algorithm.

Algorithm 3 Randomized rounding 2

Step 1: solve the LP and obtain solution $\{x_i\}$
Step 2: repeat the following procedure for $(\log n + 2)$ times
$(\star)$: $\forall i \in [m]$, pick s_i independently with probability x_i
Step 3: repeat step 2 until the following two conditions are satisfied:
(1): all elements are coverd
(2): the total cost is smaller or equal to $(4 \log n + 8)$ times the cost of the LP solution

It's easy to see that when the two conditons in step 3 are satisfied, we obtain a $O(\log n)$-approximation to the solution. Next, we will show that only two iterations in expectation are needed to satisfy both conditions in step 3.

We begin by analyzing the cost of the rounding scheme.

Lemma 7.2.6 *The expected cost of $(\star)$ is equal to the LP solution.*

Proof: Let Y_i be the indicator random variable with $Y_i = \mathbb{1}(s_i \text{ is picked})$, then the cost of $(\star)$ is just $Y = \sum_{i=1}^m c_i Y_i$. By the linearity of expectation, we have

$$\mathbb{E}(Y) = \sum_{i=1}^{m} c_i \mathbb{E}(Y_i) = \sum_{i=1}^{m} c_i x_i \tag{7.2.8}$$

■

Corollary 7.2.7 *The expected cost of the collection of subsets picked in any iteration of step 2 is* $(\log n + 2)$ *times the LP cost. The probability that this cost is greater than* $(4 \log n + 8)$ *times the LP cost is at most* $\frac{1}{4}$.

Proof: Follows from linearity of expectation and Markov's inequality. ■

Next, we consider the probabilty that elements are covered after one iteration of step 2.

Lemma 7.2.8 *After one iteration of step 2, condtion (1) in step 3 of Algorithm 3 fails with probability at most* $\frac{1}{4}$.

Proof: For any fixed $e \in E$, we have

$$\begin{aligned} \mathbb{P}(e \text{ is not covered in a step of } (\star)) &= \prod_{i:e \in s_i} (1 - x_i) \\ &\leq \prod_{i:e \in s_i} \exp(-x_i) \\ &= \exp\left(-\sum_{i:e \in s_i} x_i\right) \\ &\leq \exp(-1) \end{aligned} \tag{7.2.9}$$

by independence, we thus have

$$\begin{aligned} \mathbb{P}(e \text{ is not covered in step 2}) &\leq \exp(-\log n - 2) \\ &< \frac{1}{4n} \end{aligned} \tag{7.2.10}$$

Since there are n elements in total, by a simple union bound, we have

$$\mathbb{P}(\text{condion (1) fails}) \leq \frac{1}{4} \tag{7.2.11}$$

■

Combine Corollary 7.2.7 and Lemma 7.2.8 with a simple union bound, we have

$$\mathbb{P}(\text{Conditions in Step 3 are not satisfied}) \leq \frac{1}{2} \tag{7.2.12}$$

Thus, we only need at most two iterations of Step 2 in expectation.