College of Computer & Information Science | Fall 2007
Northeastern University | 18 September 2007
CS G399: Algorithmic Power Tools I | Scribe: San Tan

Lecture Outline:

- Rounding of LP relaxations
- Randomized rounding for set cover

This lecture introduces the technique of rounding LP relaxations for designing approximation algorithms. We introduce the randomized rounding approach using the set cover problem. The rounding algorithm produces a set cover solution that, with probability almost 1, covers all elements with cost $O(\ln n)$ times that of optimal cost.

1 Rounding for set cover

Recall the definition of the set cover problem.

Problem 1. Given a universe $\mathcal{U} = \{e_1, \ldots, e_n\}$ of n elements, a collection $\mathcal{S}$ of m subsets of $\mathcal{U}$, and a cost function $c : \mathcal{S} \to Q^+$, the set cover problem seeks a minimum-cost subset of $\mathcal{S}$ that covers all elements of $\mathcal{U}$.

Last class we formulated a linear programming relaxation for the set cover problem.

$$\begin{aligned} min \quad & \sum_{S \in \mathcal{S}} c_S x_S \\ s.t. \quad & \sum_{e \in S} x_S \geq 1; \forall e \in \mathcal{U} \\ & x_S \geq 0; \forall S \in \mathcal{S} \\ & x_S \in [0,1], \forall S \in \mathcal{S} \end{aligned}$$

Suppose $\{x_S^*\}$ is an optimal solution for above LP problem. Notice x_S^* could be a fractional value between 0 and 1. In order to apply the LP solution back to the set cover problem, we need to round $\{x_s^*\}$. Let us consider two natural rounding methods.

- **Nearest rounding:**

$$x_S^* \geq \frac{1}{2} \to x_S = 1$$

$$x_S^* < \frac{1}{2} \to x_S = 1$$

- **Randomized rounding:** Interpret x_s as a probability that set s is selected.

To analyze the rounding approaches, we must ask two questions:

- **q1:** What's the cost of rounding algorithm (compared with optimal cost given by LP)?
- **q2:** Are all elements covered?

For nearest rounding, consider this scenario: for every set S that contains a given element e, we have $x_S^* < \frac{1}{2}$. In this case, e will definitely not be covered after rounding. It is easy to construct examples where this is the case for all elements. This scenario is surely undesirable.

Let's consider randomized rounding.

- **q1:** Suppose X = collection of selected sets after randomized rounding. Then the total cost is

  $$C(X) = \sum_{S \in X} c(S).$$

  The expected total cost is

  $$\begin{aligned} E[C(X)] &= \sum_{S \in X} \Pr[S \in X] \cdot c(S)] \\ &= \sum_{S \in X} x_S^* c(S) \\ &= OPT_{LP} \\ &\leq OPT \end{aligned}$$

  So the expected cost of randomized rounding is as good as the LP cost, which we know is a lower bound on the cost of an optimal set cover.

- **q2:** For an element e, assume that it occurs in sets $s_1, s_2, ..., s_k$. The probability that e is not covered is

  $$p_{\bar{e}} = \prod_{i=1,\cdots,k} (1 - x_{s_i}^*)$$

  Since $x_{s_1}^* + x_{s_2}^* + \cdots + x_{s_k}^* \geq 1$, we have

  $$\begin{aligned} p_{\bar{e}} &= (1 - x_{s_1})(1 - x_{s_2}) \cdots (1 - x_{s_k}) \\ &\leq \left(1 - \frac{1}{k}\right)^k \\ &\leq \frac{1}{e} \end{aligned}$$

  So the probability that e is covered by X is

  $$p_e = 1 - p_{\bar{e}} \geq 1 - \frac{1}{e}.$$

The expected cost of randomized rounding is good, but we have not guaranteed that all elements are covered. In fact, there is a very good chance that all the elements are not covered. On the other hand, any given element is covered with at least a constant non-zero probability. An improvement to the simple randomized rounding is to repeat randomized rounding for t times and combine the result

$$X = \bigcup_{i=1}^{t} X_i,$$

in which X_i is the result of i^{th} randomized rounding. The improved randomized rounding has

$$E[C(X)] \leq t \cdot OPT,$$

$$\Pr[\text{given element is not covered}] \leq \frac{1}{e^t}.$$

Then the probability that at least one element is not covered is

$$\begin{aligned} \Pr[\text{some element is not covered}] &\leq \sum_{e_i \in \mathcal{U}} \Pr[e_i \text{is not covered}] \\ &= \tfrac{n}{e^t} \end{aligned}$$

(Note that Pr[some element is not covered] is not equal to $1 - (1 - \frac{1}{e^t})^n$, since the events defined by the coverage of each element are not independent.)

Given a threshold ε, we now compute how many rounds of randomized rounding are needed to assure that the probability that some element is not covered is lower than ε.

$$\frac{n}{e^t} \leq \varepsilon \Rightarrow t \geq \ln \frac{n}{\varepsilon}.$$

So a solution that repeats randomized rounding for $\frac{n}{\varepsilon}$ times and return $X = \bigcup_{i=1}^{t} X_i$ satisfies

$$E[C(X)] \leq OPT \cdot \ln(\frac{n}{\varepsilon}),$$

$$p[\text{X is not a valid solution}] \leq \varepsilon.$$

From Markov's inequality $p_r[Y \geq \alpha] \leq \frac{E[Y]}{\alpha}$, we have

$$p_s[C(X) \geq 4 \cdot OPT \cdot \ln(\frac{n}{\varepsilon})] \leq \frac{E[C(X)]}{4 \cdot OPT \cdot \ln(\frac{n}{\varepsilon})} \leq \frac{1}{4}.$$

If we set $\varepsilon = \frac{1}{4}$,

$$p[X \text{ is valid and } C(X) \leq 4 \cdot OPT \cdot \ln(4n)] \geq \frac{1}{2},$$

and corresponding algorithm is

1. Solve LP;

2. Repeat randomized rounding for $\ln(4n)$ times and get a solution $X = \bigcup_{i=1}^{t} X_i$;

3. If $C(X) \geq 4 \cdot OPT \cdot \ln(4n)$ or X is not valid, go back to step 2; otherwise return X as a final solution.

The expected number of iterations of step 2 of the above algorithm before the algorithm terminates is at most 2. In fact, the probability that the algorithm terminates in T iterations is at least $1 - 1/2^T$.