Feb. 19, 2019

CS 5330: Randomized Algorithms

Optimization:

(1) Design ILP to solve problem ← NP-hard?

integer linear program

→ only integer solutions

(2) Relax ILP → LP ← linear program

→ real number solutions

black box

(3) Round solution to integers

← "Randomized Rounding"

Set Cover

→ n base items $P = \{p_1, p_2, \dots, p_n\}$

→ m sets $S = \{S_1, S_2, \dots, S_m\}$, $S_j \subseteq P$

→ Find a min-sized collection of sets

$$C = \{S_{i_1}, S_{i_2}, \dots, S_{i_k}\}$$

so that each p_j is in some set S_{i_ℓ}

Ex: $P = \{A, B, C\}$

$S = \{\{A, B\}, \{B, C\}, \{A, C\}\}$

$\Rightarrow C = \{\{A, B\}, \{B, C\}\}$

→ NP-Complete

→ Greedy Alg is $\Theta(\log n)$ approximation

Write problem as integer linear program:

(1) Variables $X_1, X_2, \dots, X_m$ ← include S_j in cover iff $X_j = 1$

(2) ILP:

$$\min \sum_{j=1}^{m} X_j$$ ← minimize number of sets in cover

$$\text{s.t.} \sum_{j : p_i \in S_j} X_j \geq 1 \quad \forall i \in [1, n]$$

$$X_j \in \{0, 1\} \quad \forall j \in [1, m]$$

Linear objective
Linear constraints

sum for all sets that contain p_i

force at least one set containing p_i

S_j is <u>in</u> cover or <u>not</u> in cover

<u>Claim:</u> if $X_1, X_2, \dots, X_m$ satisfy constraints, then $\{S_j : X_j = 1\}$ is a valid set cover.

<u>Claim:</u> if C is a valid set cover, then $\{X_j = 1 \text{ iff } S_j \in C\}$ satisfies the constraints

$\Rightarrow$ Solution to ILP is a minimum Set Cover

(Still NP-hard)

Relax!

Let $X_j \in [0,1]$ instead $X_j \in \{0,1\}$

$\Rightarrow$ **linear Program (LP)**

We can solve LP's in Poly-time!

Fast!

But: is solution still useful?

Ex $P = \{A, B, C\}$

$S = \{A, B\}, \{B, C\}, \{A, C\}$

$\downarrow$ $X_1 = 1/2$ $\quad$ $\downarrow$ $X_2 = 1/2$ $\quad$ $\downarrow$ $X_3 = 1/2$

Claim 1) $LP \leq ILP$

(every solution to ILP is also a solution to ILP, so $\min(LP) \leq \min(ILP)$)

What about rounding?

→ either all or none

→ not a good approximation as sets get large

Idea: Randomized Rounding

For each X_j, let X_j^* be solution to LP.

$$\boxed{\text{For each } X_j, \text{ add } S_j \text{ to } C \text{ with probability } X_j^*}$$

Let $Y_j = 1$ if add S_j $\quad Pr(Y_j) = X_j^*$
$\quad = 0$ otherwise

ILP solution

$$E[\#\text{ sets}] = E\left[\sum_{j=1}^{n} Y_j\right] = \sum_{j=1}^{n} E[Y_j] = \sum_{j=1}^{n} X_j^* \leq OPT$$

LP objective!!
minimum possible...

Is it a cover?

$$Pr[p_j \text{ not covered}] = \prod_{i: p_j \in S_i} (1 - Pr(\text{add } S_i))$$

$$= \prod_{i: p_j \in S_i} (1 - X_i^*)$$

$$\leq \prod_{i: p_j \in S_i} e^{-X_i^*} \quad \leftarrow e^{-x} \geq 1 - x$$

$$\leq e^{-\sum_{i: p_j \in S_i} X_i^*}$$

$$\leq e^{-1}$$

$\Rightarrow p_j$ is covered w.p. $\geq (1 - \frac{1}{e})$

Better idea: Repeat $\cancel{2}$ 3 $\ln(n)$ times.
Take union of all sets.

$$Pr(P_j \text{ not covered in } \underline{\underline{\text{all}}} \text{ iterations}) \leq \left(\frac{1}{e}\right)^{3\ln n}$$
$$\leq \frac{1}{n^3}$$

$$Pr(\text{Any } P_j \text{ not covered}) \leq \sum_{j=1}^{n} Pr(P_j \text{ not covered})$$
$$\leq \frac{n}{n^3} \leq \frac{1}{n^2}$$

$$\Rightarrow Pr(C \text{ is a set cover}) \geq \left(1 - \frac{1}{n^2}\right)$$

$$\text{Expected cost} \leq 3\ln(n) \cdot E\left[\sum x_j^*\right]$$
$$\leq 3\ln(n)\, OPT$$

Show: w.h.p, $\# \text{ sets} \leq 6 \ln(n)\, OPT$
$\Rightarrow 6\ln(n)$ approximation

Random vars:

$$\left.\begin{array}{l} y_1^1, y_2^1, \ldots, y_m^1 \\ y_1^2, y_2^2, \ldots, y_m^2 \\ \quad\vdots \\ y_1^{3\ln(n)}, y_2^{3\ln(n)}, \ldots, y_m^{3\ln n} \end{array}\right\}$$

$3m\ln(n)$ random variables

independent

$y_j^k \in [0,1]$

$$Y = \sum_{k=1}^{3\ln(n)} \sum_{j=1}^{m} y_j^k$$

$$\forall \ell : \sum_{j=1}^{m} \Pr(y_j^\ell = 1) = \sum_{j=1}^{n} x_j^* = LP$$

$$\Rightarrow E[Y] = \sum_{k=1}^{3\ln(n)} LP = 3\ln(n)\, LP$$

$\#$ sets chosen $\leq Y$

$$\Pr[Y \geq (2)(3\ln n)(LP)]$$

$$\leq \Pr[Y \geq (1+1)E[Y]] \leq e^{-E[Y]\, 1^2 \cdot \frac{1}{3}}$$

$$\leq e^{-3\ln(n) LP \cdot \frac{1}{3}}$$

$LP \geq 1$

$$\leq e^{-\ln n}$$

$$\leq \frac{1}{n}$$

Since $LP \leq OPT$,

$$\Pr[\#\text{ sets} \geq 6\ln(n)\, OPT] \leq 1/n$$

If you are more careful, can get a tighter approximation.