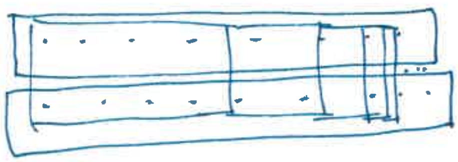
Notes from Lec 1's PDF

continued and refined

Algorithmic Gap: Does greedy do better than $\ln n$?

Fact: Greedy no better than $\ln n\,(1-\varepsilon)$. (even for unweighted)

Pf:



$OPT = 2$.

$Alg = \log_2(n/2)$ $\Rightarrow$ algo gap $= \dfrac{\log_2 n - 1}{2} \cong \dfrac{\ln n}{2\ln 2}$.

To get $\approx \ln n$, use setup where $OPT = k$ vertically

But each other set covers $1/k$ of remainder.

$$\Rightarrow \#\text{sets} = \log_{\frac{1}{(1-\frac{1}{k})}} n. \quad \Rightarrow \text{gap} = \frac{\log_{(1-1/k)} n}{k} \cong \frac{\ln n}{k \ln\left(\frac{1}{1-\frac{1}{k}}\right)}$$

but $\ln(1+\varepsilon) = \varepsilon + \Theta(\varepsilon^2)$ for ε small.

$$\cong \left(1-\Theta\left(\tfrac{1}{k}\right)\right)\cdot \ln n.$$

So another algorithm?

Before that: a greedy algo for weighted case.

at each step, pick set that $\max\left(\dfrac{\text{coverage}}{\text{cost}}\right)$.

Thm: Greedy is $\ln n$-apx for weighted set cover.

Pf: (sketch) Same idea as before. Show that if costs $c_1, c_2 \dots c_t$ for sets in steps $1,2..t$

$$n_t \le n\left(1-\frac{c_1}{OPT}\right)\left(1-\frac{c_2}{OPT}\right)\cdots\left(1-\frac{c_t}{OPT}\right)$$

$$\le n \exp\left(\frac{\sum_i c_i}{OPT}\right).$$

etc. ∎

A charging proof for weighted set cover.

Greedy: Pick set that ~~covers~~ has the max $\frac{\text{~~new~~ coverage}}{\text{cost}}$, until all elements covered.

Claim: Cost $\le H_n \cdot \text{cost}(OPT)$.

Pf: sps. S_t = t-th set picked, # elements uncovered before pick $S_t = n_t$
after pick $S_t = n_{t+1}$.

$\Rightarrow$ by greedy choice $\dfrac{n_t - n_{t+1}}{c(S_t)} \geq \dfrac{n_t}{OPT}$

$\Rightarrow c(S_t) \le OPT \cdot \dfrac{n_t - n_{t+1}}{n_t}$

$$\Rightarrow \underbrace{\sum_t c(S_t)}_{= ALG} \le OPT \left\{ \frac{n_0 - n_1}{n_0} + \frac{n_1 - n_2}{n_1} + \cdots \cdots \right\}$$

$$= OPT \left\{ \underbrace{\frac{1}{n} + \frac{1}{n} + \cdots \frac{1}{n}}_{n_0 - n_1} + \underbrace{\frac{1}{n_1} + \frac{1}{n_1} + \cdots}_{n_1 - n_2} + \cdots \cdots \right\}$$

$$\le OPT \left\{ \frac{1}{n} + \frac{1}{n-1} + \cdots \frac{1}{n_1 + 1} + \frac{1}{n_1} + \frac{1}{n_1 - 1} + \cdots \right\}$$

$$= OPT \cdot H_n.$$

A diff view: charge cost of $c(S_t)$ to all elements newly covered by it, equally.

Now: Consider set S^* in OPT, covers some elements $e_1, e_2 \dots e_t$ $\longrightarrow$ in order ~~picked~~ covered.

charge to $e_1 \le \dfrac{c(S^*)}{t}$ (cos we picked some set at least as good as S^*)

$e_2 \le \dfrac{c(S^*)}{t-1}$ ....

$\Rightarrow$ total charge $\le \sum_{S^*} c(S^*) \cdot H_n = OPT \cdot H_n$. $\Rightarrow$ total charge to elements covered by $S^* \le c(S^*) \cdot H_n$

Linear Program - based Algos:

Idea: Relax-and-Round

1. write an IP for Set Cover. (IP = Integer (Linear) Program).
2. "Relax" it to an LP (LP = Linear Program).
3. Solve this LP. Fact: can solve LP's in poly time.
4. "Round" the fractional solution to Integers.

Usually:
1. $IP(I) = Opt(I)$.
2. $LP(I) \le IP(I)$.
3. $Alg(I) \le \alpha . LP(I) \Rightarrow Alg(I) \le \alpha . OPT(I)$.

Set Cover: variable $x_S \in \{0,1\}$ for each set $S \in \{S_1, S_2 \ldots S_m\}$.

$\boxed{IP.}$

$$\min \sum c_S x_S$$

$$\text{st} \sum_{S: e \in S} x_S \ge 1 \quad \forall e \in U.$$

$$\cancel{x_S \in \{0,1\}}.$$

$\boxed{LP}$

$$x_S \ge 0$$

Round: Imagine each x_S as a prob. value. (Fact: $x_S \in [0,1]$, no reason for x_S to be larger).

Algo:

For $T =$ ______ times

  $\forall S \in \mathcal{F}$

    select S independently w.p. x_S.

T rounds of sampling

What if this is not a feasible solution?

Clean-up: $\forall$ element e, pick cheapest set covering e if e not covered by sampling.

Lemma: $E[\text{cost of solution}] \le T.\,LP(I) + \left[\sum_{e \in U} (\text{cheapest set covering } e)\right] e^{-T}$.

Pf: $E[\text{cost of each round}] = \sum_S c_S . \Pr[S \text{ picked}] = \sum c_S x_S = LP(I)$.

$\uparrow$ Lin of exp.

Now: $\text{prob}(e \text{ not covered in one round}) = \prod_{S: e \in S} (1 - x_S) \le e^{-\sum_{S: e \in S} x_S} \le e^{-1}$

$\Rightarrow \Pr(e \text{ not covered in } T \text{ rounds}) \le e^{-T}$.

Now use Linearity of expectation again. $\blacksquare$

Hence set $T = \ln n$.

$E[\text{cost}] \le (\ln n).LP + \frac{1}{n}.n.LP$ (b/c LP value $\ge$ cheapest set covering e for any e)

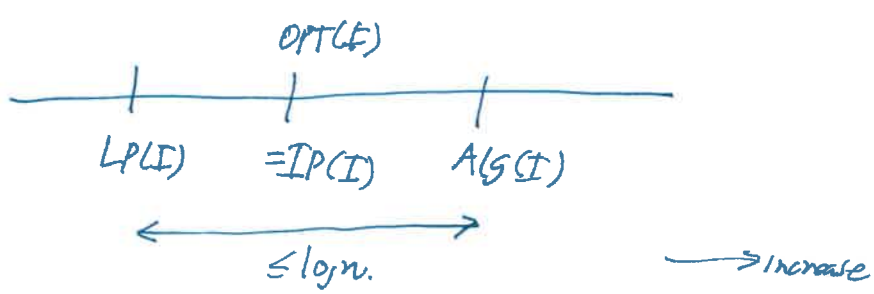
$\boxed{= (\ln n + 1)\,LP.}$

— x —

HW: Show that if sets are of size B, then ~~both~~ LP rounding gives $O(\ln B)$ apx.

Greedy too (but see more later).

Picture



Ask 2 questions:

① Algorithmic gap: does $\exists$ instance where $\frac{Alg(I)}{OPT(I)} = \Omega(\lg n)$. ↗ using this LP + rounding, cannot beat $\lg n$.

② Integrality gap: does $\exists$ instance s.t. $\frac{OPT(I)}{LP(I)} = \Omega(\lg n)$.

→ shows that using this ~~approach~~ LP cannot beat the log-apx. no matter what rounding we do.
[as long as we relate ourselves to the LP value, of course!].

Algo gap: $\boxed{\text{see in HW}}$.

Integrality gaps:

. Take $U = \{ x \in \{0,1\}^d \mid \|x\|_1 = d/2 \}$

$n = |U| = \binom{d}{d/2} \approx \Theta\left(\frac{2^d}{\sqrt{d}}\right)$. → $d \approx \log_2 n$.

• Sets: all "dictator" sets $S_i = \{ x \in U \mid x_i = 1 \}$. costs = 1.

• OPT $\geq d/2 + 1$ Else $\exists$ element not covered

• LP value: set $2/d$ on each set. (ie. $x_S = 1 \;\; \forall S$).

$\Rightarrow$ total LP value $= d \cdot 2/d = 2$.

$\Rightarrow$ Integrality gap $\geq \left(\frac{d/2+1}{2}\right) = \Omega(d) = \Omega(\log n)$.

Fact: Can do better, get $\ln n$ for integrality gap as well.

Derandomizing the Randomized Algo. (Pipeage Rounding)

9

Sps $\left\{\begin{array}{l} x^* \\ z^* \end{array}\right.$ = OPT fractional LP sol^n.

$LP = \sum_i z_e^*$.

$$\boxed{\begin{array}{l} \max \sum_i z_e \\ z_e \leq \sum_{S \ni e} x_S \\ z_e \leq 1 \\ \sum_S x_S \leq k \\ x_S, z_e \geq 0. \end{array}}$$ LP.

__Fact 1__ : $LP \geq OPT$ (relaxation).

An "idealized" Rounding

Pick each set S w.p. x_S, indep.

__Fact 2__ : $E[\text{coverage}] = \underbrace{\sum_e \underbrace{\left(1 - \prod_{S \ni e}(1-x_S)\right)}_{f_e(\bar{x})}}_{f(\bar{x}) = \sum_e f_e(\bar{x})} \geq \sum_e \left(1 - e^{-\sum x_S}\right) \geq \sum_e \left(1 - e^{-z_e}\right)$

$\geq \sum_e z_e (1 - 1/e) = LP(1-1/e)$.

__Fact 3__ : But only guarantee that we pick k elements in expectation.

__Fact 4__ : if solution x is integral, then coverage $= f(x)$.
and get $= k$ sets

__So idea__ :

$$\underset{LP}{x^*} \longrightarrow x_{(1)} \longrightarrow x_{(2)} \longrightarrow \cdots \longrightarrow \underset{\text{Integral.}}{x_{int}}$$

$$f(x^*) \leq f(x_{(1)}) \leq f(x_{(2)}) \quad \cdots \quad \leq f(x_{int})$$

$\geq (1-1/e) \cdot LP$

fact 2 $\geq (1-1/e) OPT$

fact 1.

but this is integral so actually tells you sol^n.

How to round (increase integer coordinates)?

Assume $\sum_i' x_S = K$

↑ else raise some x_S until sat.

if $\bar{x}$ fractional, has at least 2 fractional sets

say x_{S_1}, x_{S_2}

"Move along line $e_{S_1} - e_{S_2}$"

Consider solution $\bar{x}_\varepsilon \leftarrow \bar{x} + \varepsilon(e_{S_1} - e_{S_2})$

Let $g(\varepsilon) = f(\bar{x}_\varepsilon)$ be univariate function in ε.

- $g(0) = f(\bar{x}_0) = f(\bar{x})$
- $g(\varepsilon) = \sum_e' f_e(z_\varepsilon) = \sum'_{e \notin S_1 \cup S_2} f_e(z_\varepsilon) + \sum'_{e \in S_1 \setminus S_2} f_e(z_\varepsilon) + \sum'_{e \in S_2 \setminus S_1} f_e(z_\varepsilon) + \sum'_{e \in S_1 \cap S_2} f_e(z_\varepsilon)$

$f_e(z)$

$1 - \left[\prod_{\substack{S \neq S_1 \\ S \ni e}} (1 - x_S)\right](1 - x_{S_1} - \varepsilon)$

$=$ linear fn in ε.

also linear fn in ε

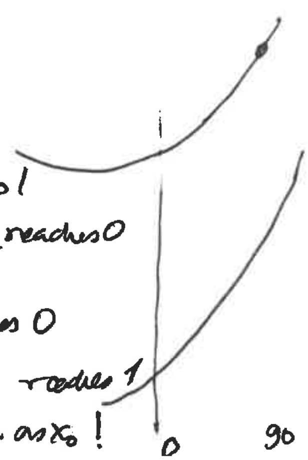
$1 - \prod_{\substack{S \neq S_1, S_2 \\ S \ni e}} (1 - x_S)(1 - x_{S_1} - \varepsilon)(1 - x_{S_2} + \varepsilon)$

$= c_0 + c_1 \varepsilon + c_2 \varepsilon^2$

↑ positive!

$\Rightarrow$ convex in ε.

Claim: $g(\varepsilon)$ is convex in ε.



$\Rightarrow$ consider $x_{+\varepsilon_0}$ st x_{S_1} reaches 1 or x_{S_2} reaches 0

$x_{-\varepsilon'}$ st x_{S_1} reaches 0 or x_{S_2} reaches 1

one must be at least as high as x_0! go to that one.