Proof lemma 17.3

Show that every $e_i \in E$ is covered

x^* feasible $\Rightarrow \sum_{j : e_i \in S_j} x_j \geq 1$

f_i variables have coefficient 1 (the others 0)

$\Rightarrow$ one of these $x_j \geq \frac{1}{f_i} \geq \frac{1}{f}$

$\Rightarrow$ for every e_i $\exists$ variable x_j that is rounded to 1 $\square$

Proof Theorem 17.4

Algo runs in poly-time (solve LP, all data known)

Let $\hat{I} := \{ j \mid \hat{x}_j = 1 \} = \{ j \mid x_j^* \geq \frac{1}{f} \}$

$\Rightarrow f \cdot x_j^* \geq 1 = \hat{x}_j \quad \forall j$ with $\hat{x}_j = 1$

$$\left.\begin{array}{ll} \Rightarrow f x_j^* \geq \hat{x}_j & \\ \hat{x}_j = 0 \Rightarrow & f x_j^* \geq \hat{x}_j \end{array}\right\} f \cdot x_j^* \geq \hat{x}_j \quad \forall j$$

$$w(\hat{I}) = \sum_{j=1}^{m} w_j \hat{x}_j \leq \sum_j w_j \cdot f x_j^* = f \sum_j w_j x_j^*$$

$$= f \cdot OPT(LP) \leq f \cdot OPT(IP)$$

$\Rightarrow$ Algo is an f-approximation algorithm $\square$

Rounding a Dual Solution

Dual Rounding Algorithm for Set Cover:

1. Compute an optimal solution y^* to the dual of the set-cover-LP (17.2).
2. Let $I^* := \{j \mid \sum_{i:e_i \in S_j} y_i = w_j\}$; $\leftarrow$ take sets S_j with ~~dual slack~~ $= 0$

Lemma 17.5.
The collection of subsets S_j with $j \in I^*$ is a set cover.

Proof: … □

Theorem 17.6.
The dual rounding algorithm is an f-approximation algorithm for the set cover problem.

Proof: … □

Dual LP of the set cover LP

$$\max \sum_{i=1}^{n} y_i \qquad \leftarrow \text{dual variable for each } e_i$$

$$\text{s.t.} \sum_{i:\, e_i \in S_j} y_i \leq w_j \qquad \forall\ S_j,\ j = 1, \ldots, m$$

$$y_i \geq 0$$

<u>Proof Lemma 17.5</u>

Show $I^* = \{j \mid \sum_{i:\, e_i \in S_j} y_i^* = w_j\}$ is a set cover

Suppose e_k is not covered

$\Rightarrow$ every S_j containing e_k has $\sum_{e_i:\, e_i \in S_j} y_i^* < w_j$

Let $\varepsilon := \min_{j:\, e_k \in S_j} \{w_j - \sum_{i:\, e_i \in S_j} y_i^*\} > 0$

Consider vector $\bar{y}$ with $\bar{y}_i := y_i \quad \forall i \neq k$

$\bar{y}_k := y_k + \varepsilon$

$\Rightarrow$ $\bar{y}$ is feasible in the dual

$\Rightarrow \sum_i \bar{y}_i = \sum_i y_i^* + \varepsilon$, contradiction to y^* being optimal $\square$

<u>Proof</u> <u>Thm. 17.6</u>

Dual rounding gives an f-approximation

charging argument : charge y_i^* to each e_i in S_j if S_j is selected

this is what e_i must pay for being covered

Then $w(I^*) = \sum_{j \in I^*} w_j = \sum_{j \in I^*} \sum_{i : e_i \in S_j} y_i^*$

$= \sum_{i=1}^{n} \underbrace{|\{j \in I^* \mid e_i \in S_j\}|}_{\leq f_i \leq f} y_i^*$ ← what all e_i pay

$\leq f \cdot \sum_{i=1}^{n} y_i^* = f(\text{cost of } y^*) \leq \text{OPT(primal)}$

weak duality

$\leq f \cdot \text{OPT(IP)}$ $\square$

Primal-Dual Algorithm

Note: The two previous algorithms require solving a linear program. Special purpose algorithms are often much faster!

Idea: Construct a feasible dual solution that is "good enough".

Primal-dual algorithm for the set cover problem: need not solve an LP

1. Initialize: $y \equiv 0$ and $I = \emptyset$;
2. WHILE $\exists e_k \notin \bigcup_{j \in I} S_j$ DO
3. Increase y_k until $\exists j$ with $e_k \in S_j$ such that $\sum_{i: e_i \in S_j} y_i = w_j$;
4. Set $I = I \cup \{j\}$;

I is a set cover at termination

Theorem 17.7.

The primal-dual algorithm is an f-approximation algorithm for the set cover problem.

Proof: as before. □

we need only that $w_j = \sum_{i: e_i \in S_j} y_i$ when S_j is selected

and that y is feasible in the dual

need LP-theory only in the proof □

Greedy Algorithm

Idea: Iteratively, until all elements are covered, select a set that minimizes the ratio of its weight to the number of currently uncovered elements it contains.

Greedy algorithm for the set cover problem:

1. Initialize: $\mathcal{I} = \emptyset$ and $\hat{S}_j = S_j$ for all j;
2. WHILE $\mathcal{I}$ is not a cover DO
3. $\ell = \operatorname{argmin}\{\frac{w_j}{|\hat{S}_j|} \mid \hat{S}_j \neq \emptyset\}$:
4. Set $\mathcal{I} = \mathcal{I} \cup \{\ell\}$;
5. Set $\hat{S}_j = \hat{S}_j \setminus S_\ell$ for all j;

best bang for the buck

Theorem 17.8.

The greedy algorithm returns a cover $\mathcal{I}$ with $w(\mathcal{I}) \leq H_g \cdot z^*_{LP}$, where $g = \max_j |S_j|$ and $H_g = \sum_{k=1}^{g} \frac{1}{k} \approx \ln\ g$.

Proof: ... □

No performance guarantee better than H_n possible (unless $\mathcal{P} = \mathcal{NP}$)!

Proof Thm. 17.8

uses dual fitting technique

Construct an "infeasible dual solution" y such that

$$y \geq 0$$

$$\boxed{w(I) = \sum_{i=1}^{n} y_i} \quad (1)$$

I constructed by the greedy algorithm

Show that $\boxed{\bar{y} := \frac{1}{H_g} y \quad \text{is dual feasible}} \quad (2)$

Then $w(I) \overset{(1)}{=} \sum_{i}^{n} y_i = H_g \cdot \sum_{i=1}^{n} \bar{y}_i \underset{\text{weak duality}}{\overset{(2)}{\leq}} H_g \cdot OPT$

Construct y as follows:

Let S_ℓ be selected in iteration k

Set $y_i := \frac{w_\ell}{|\hat{S}_\ell|}$ for each $e_i \in \hat{S}_\ell$

$\Downarrow$

$\Rightarrow$ $e_i \in \hat{S}_\ell$ were uncovered before iteration k and are now covered

$\Rightarrow$ y_i is only set in this iteration

$\Rightarrow$ $w_\ell = \sum_{i : e_i \in \hat{S}_\ell} y_i \underset{\text{at the end}}{\Rightarrow} \sum_{j \in I} w_j = \sum_{i=1}^{n} y_i$

$\Rightarrow$ (1)

Show that $\bar{y}_i = \frac{1}{H_g} y_i$ is dual feasible

i.e. $\bar{y} \geq 0$ $\leftarrow$ clear

$$\sum_{i:\, e_i \in S_j} \bar{y}_i \leq w_j \quad \forall S_j$$

Consider S_j

Let u_k be the number of uncovered elements at the beginning of iteration k in S_j

Let $U_k := \{$ uncovered elements of S_j covered in iteration k $\}$

$\Rightarrow |U_k| = u_k - u_{k+1}$

Let S_e be chosen in iteration k

$\Rightarrow$ for all $e_i \in A_k$ $\quad \bar{y}_i = \frac{w_e}{H_g |\hat{S}_e|} \leq \frac{w_j}{H_g u_k}$

$\uparrow$ greedy choice rule

$\Rightarrow$ over all iterations k

$$\sum_{i:\, e_i \in S_j} \bar{y}_i = \sum_k \sum_{\substack{i:\, e_i \in U_k \\ e_i \in S_j}} \bar{y}_i \leq \sum_k (u_k - u_{k+1}) \frac{w_j}{H_g u_k}$$

$$= \frac{w_j}{H_g} \sum_k \frac{u_k - u_{k+1}}{u_k}$$

Now $\displaystyle \sum_k \frac{u_k - u_{k+1}}{u_k} = \sum_k \sum_{j = u_{k+1}+1}^{u_k} \frac{1}{u_k}$

$$\leq \sum_k \sum_{j = u_{k+1}+1}^{u_k} \frac{1}{j} \qquad j \leq u_k$$

$$= \sum_k \left(\sum_{j=1}^{u_k} \frac{1}{j} - \sum_{j=1}^{u_{k+1}+1} \frac{1}{j} \right)$$

$$= \sum_k \left(H_{u_k} - H_{u_{k+1}+1} \right)$$

telescopic sum

$$= H(u_1) - H(u_{\text{last iteration}}) \leq H(u_1) \leq H(g)$$

that covers elements from S_j

Definition of u_k $\Rightarrow$ $u_1 = |S_j| \leq g$

So we have

$$\sum_{i:\, e_i \in S_j} \bar{y}_i \leq \frac{w_j}{H_g} H_g = w_j \Rightarrow (2) \qquad \square$$