Set Multicover

The *Set Multicover* problem is the same as set cover except that every element needs to be covered a specified number of times. That is we want to pick a minimum-cost set cover such that element $e \in U$ must be covered at least a given r_e times.

We will require that each set be picked at most once.

The corresponding integer LP will be

Minimize $\Sigma_{F\in\mathcal{F}} C(F)x_F$
subject to conditions
$\forall e \in U, \qquad \Sigma_{e\in F}\, x_F \geq r_e$
$\forall F \in \mathcal{F} \qquad x_F \in \{0,1\}$

When we relax the LP we get $0 \leq x_F \leq 1$. This is not in *primal* form so we replace the equations by $0 \leq x_F$ *and* $-x_F \leq -1$.

The relaxation of the LP is now

Minimize $\Sigma_{F\in\mathcal{F}} C(F)x_F$
subject to conditions
$\forall e \in U, \qquad \Sigma_{e\in F}\, x_F \geq r_e$
$\forall F \in \mathcal{F}, \qquad -x_F \geq -1$
$\forall F \in \mathcal{F}, \qquad x_F \geq 0$

The primal was

Minimize $\Sigma_{F\in\mathcal{F}} C(F)x_F$
subject to conditions
$\forall e \in U, \qquad \Sigma_{e\in F}\, x_F \geq r_e$
$\forall F \in \mathcal{F}, \qquad -x_F \geq -1$
$\forall F \in \mathcal{F}, \qquad x_F \geq 0$

The dual is

Maximize $\Sigma_{e\in U}\, r_e y_e - \Sigma_{F\in\mathcal{F}}\, z_F$
subject to conditions
$\forall F \in \mathcal{F}, \qquad (\Sigma_{e\in F}\, y_e) - z_F \leq C(F)$
$\forall e \in U, \qquad y_e \geq 0$
$\forall F \in \mathcal{F}, \qquad z_F \geq 0$

We now generalize the greedy set cover algorithm to multicover

A Greedy MultiCover algorithm
$U \leftarrow X$
$\mathcal{C} \leftarrow \emptyset$
  select a $F \in \mathcal{F}$ that minimizes $\frac{Cost(F)}{|F\cap U|}$
  $\mathcal{C} \leftarrow \mathcal{C} \cup \{F\}$
  For each $e \in F$
    Let j be the number of times e appears in $\{F \in \mathcal{C}\}$
    Set $price(e,j) = \frac{Cost(F)}{|F\cap U|}$
    If $j = r_e$ remove e from U.
return($\mathcal{C}$)

This algorithm is very similar to the previous one. At each step it examines each remaining set and determines how many "live" elements it can cover and at what average cost per element. It then chooses the set that covers elements at minimum cost during that step.

Notice that the average cost can not decrease. Then $\forall e \in U$

$$price(e,1) \leq price(e,2) \leq \ldots \leq price(e, r_e).$$

Also, the total weight of the sets chosen is

$$\sum_{e\in U}\left(\sum_{j=1}^{r_e} price(e,j)\right).$$

Consider the setting

$\forall e \in U, \quad y_e = \frac{1}{H_n} price(e, r_e)$
$\forall F \in \mathcal{C}, \quad z_F = \frac{1}{H_n}\left(\Sigma_{\{e \text{ covered by } F\}}(price(e,r_e) - price(e,j_e))\right)$
$\forall F \in \mathcal{F} - \mathcal{C}, \quad z_F = 0$

where j_e is the copy of e that was covered by the inclusion of F.

We claim that (y,z) is a feasible solution to the dual.

We say that an item $e \in U$ is *alive* if e has not yet been covered r_e times.

Consider set F. Let $k = |F|$. Number (order) the elements of F in the order in which they are fully covered, i.e., the time at which they stop being alive.

If F is *not picked* by the algorithm then at the time the algorithm covers the last copy of e_i, F contains *at least* $k - i + 1$ alive items so

$$price(e_i, r_{e_i}) \leq \frac{C(F)}{k-i+1}.$$

Since $z_F = 0$ we get

$$\begin{aligned}\left(\sum_{i=1}^{k} y_{e_i}\right) - z_F &= \frac{1}{H_n}\sum_{i=1}^{k} price(e_i, r_{e_i}) \\ &\leq \frac{C(F)}{H_n}\left(\frac{1}{k} + \frac{1}{k-1} + \cdots + \frac{1}{1}\right) \leq C(F)\end{aligned}$$

so those equations are satisfied.

Now suppose $F \in \mathcal{C}$.
Suppose too, that right before F was chosen k' elements in F were already fully covered (not alive). Then

$$\left(\sum_{i=1}^{k} y_{e_i}\right) - z_F = \frac{1}{H_n}\left(\sum_{i=1}^{k'} price(e_i, r_{e_i}) + \sum_{i=k'+1}^{k} price(e_i, j_{e_i})\right)$$

Now notice that

$$\sum_{i=k'+1}^{k} price(e_i, j_i) = C(F).$$

Also, by the same reasoning as we used before, when item e_i, $i \le k'$ was chosen for the last time F had not yet been picked and therefore contained at least $k-i+1$ alive elements. This means $price(e_i, r_{e_i}) \le C(F)/(k-i+1)$ so

$$\sum_{i=1}^{k'} price(e_i, r_{e_i}) \le C(F) \sum_{i=1}^{k'} \frac{1}{k-1-i'} \le C(F) \sum_{j=k-k'+1}^{k} \frac{1}{j}.$$

Therefore

$$\left(\sum_{i=1}^{k} y_{e_i}\right) - z_F \le \frac{C(F)}{H_n}\left[(H_k - 1) + 1\right] \le C(F)$$

satisfying the other constraints.

Since $\forall F$, $z_F \ge 0$ (why) the setting (y, z) is a feasible solution to the dual.

Theorem: The greedy algorithm gives a H_n-approximation algorithm to set-multicover.

We first calculate the value of the dual objective function for our feasible solution.
By weak duality this gives a lower bound on OPT.
We then use this lower bound to prove the theorem.

The value of the dual objective function for (y, z) is

$$\begin{aligned}
\sum_{e \in U} r_e y_e \;-\; \sum_{F \in \mathcal{F}} z_F
&= \sum_{e \in U} r_e \frac{price(e, r_e)}{H_n} \\
&\quad - \sum_{F \in \mathcal{C}} \frac{1}{H_n}\left[\sum_{F \text{ covers } e} (price(e_i, r_{e_i}) - price(e_i, j_{e_i}))\right] \\
&= \frac{1}{H_n} \sum_{e \in U} \sum_{j=1}^{r_e} price(e, j) \\
&= \frac{1}{H_n} \sum_{F \in \mathcal{C}} C(F)
\end{aligned}$$

By the weak duality theorem

$$\frac{1}{H_n} \sum_{F \in \mathcal{C}} C(F) \le OPT$$

or

$$\sum_{F \in \mathcal{C}} C(F) \le H_n \cdot OPT$$

where OPT is the optimum cost for a solution to set-multicover. Since $\Sigma_{F \in \mathcal{C}}\, C(F)$ is the cost of the cover produced by the algorithm we are finished.