Greedy algorithm for partial set cover problem

In the set cover problem we are given a universe E of elements and a family $\{S1, S2, ..., Sm\}$ of E's subsets. The goal is to find a collection of subsets indexed by $I \subset \{1, ..., m\}$ that minimizes $\sum_{j \in I} w_j$ such that $\left|\bigcup_{j \in I} S_j\right| = |E|$. Consider the partial cover problem, in which one finds a collection of subsets indexed by I that minimizes $\sum_{j \in I} w_j$ such that $\left|\bigcup_{j \in I} S_j\right| \geq p|E|$ where $0 < p < 1$ is some constant. Consider the greedy algorithm for partial cover.

1. Prove that the greedy algorithm for partial cover constructs a solution of value no more than $\left(1 + \ln\left(\frac{1}{1-p}\right)\right) \cdot OPT^{SC}$, where OPT^{SC} is the value of the optimal solution to the set cover problem (note: real set cover, not partial set cover).
2. Prove that the greedy algorithm for the partial cover problem gives an $H_{\lceil |E| \cdot p \rceil}$-approximation algorithm, where $H_n = 1 + \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{n}$.

Solution 1.1.

Let be:

- n the number of elements;
- C_i the number of elements uncovered at iteration i.

In class we have seen that in the greedy algorithm when a set S_i is added, there is a set with cost less or equal to $\frac{OPT^{SC}}{C_i}$.

Follows that the element e_j covered by set S_i has cost:

$$cost(e_j) \leq \frac{OPT^{SC}}{C_i} \leq \frac{OPT_{SC}}{n - j + 1}$$

We count the cost of the last set separately and we call $n' < \lfloor np \rfloor$ the number of elements covered before that the last set is added:

$$\sum_{j=1}^{n'} cost(e_j) \leq \sum_{j=1}^{n'} \frac{OPT_{SC}}{n - j + 1} = \left(H_n - H_{n-n'}\right) \cdot OPT_{SC}$$

$$\Rightarrow \left(H_n - H_{n-n'}\right) \cdot OPT_{SC} \leq \left(H_n - H_{n-\lfloor np \rfloor}\right) \cdot OPT_{SC} \leq \left(1 + \ln(n) - \ln(n - np)\right) \cdot OPT_{SC} =$$
$$= \left(1 + \ln\left(\frac{1}{1-p}\right)\right) \cdot OPT_{SC}$$

The last set will cost at most OPT_{SC} thus:

$$C_{GREEDY} \leq \left(1 + \ln\left(\frac{1}{1-p}\right)\right) \cdot OPT_{SC} + OPT_{SC} = \left(2 + \ln\left(\frac{1}{1-p}\right)\right) \cdot OPT_{SC}$$

□

Solution 1.2.

Choose C_i to represent the currently uncovered number of items. Thus there are exactly $n-C_i$ covered items at each moment and another $C'_i = C_i - (n - \lceil pn \rceil)$ items to cover. In this algorithm instead of taking the set corresponding to the minimum cost among $\frac{w_i}{|S_i \cap C_i|}$, the new set's cost is calculated dividing by the minimum between $|S_i \cap C_i|$ and C'_i. Thus, we will take the set corresponding to the element e_j that minimizes

$$\frac{w_j}{min\{|S_j \cap C_i|, C'_i\}}$$

This change is needed to properly handle the partial cover problem so that the cost of solution is only evaluated within the first $\lceil np \rceil$ items. Let c_i be the cost charged to e_i, so if e_i belongs to S_j we have $c_i = \frac{w_j}{min\{|S_j \cap C_i|, C'_i\}}$.
Let SOL be an optimal solution. Using the algorithm just described above, we have

$$c_i = \frac{w_j}{min\{|S_j \cap C_i|, C'_i\}} \leq \frac{w_u}{min\{|S_u \cap C_i|, C'_i\}} \qquad \forall u \text{ with } |S_u \cap C_i| > 0$$

This implies

$$\frac{w_j}{min\{|S_j \cap C_i|, C'_i\}} \cdot min\{|S_u \cap C_i|, C'_i\} \leq w_u \quad \forall u$$

We have

$$OPT = \sum_{j \in SOL} w_u \geq \frac{w_j}{min\{|S_j \cap C_i|, C'_i\}} \cdot \sum_{j \in SOL} min\{|S_u \cap C_i|, C'_i\} = c_i \cdot \sum_{j \in SOL} min\{|S_u \cap C_i|, C'_i\} \geq c_i C'_i$$

remembering that $\bigcup_{u \in SOL} S_u \cap C_i$ covers at least C'_i items. It follows:

$$c_i \leq \frac{OPT}{C'_i} \leq \frac{OPT}{i}$$

The total cost of solution is

$$C^{GRE}_{PSC} = c_{\lceil pn \rceil} + ... + c_1 \leq OPT \cdot \left(\frac{1}{\lceil pn \rceil} + ... + 1 \right) = H_{\lceil pn \rceil} \cdot OPT$$