IT584 Approximation Algorithms	January 9, 2024
Lecture 3	
Lecturer: Rachit Chhaya	*Scribe/s:* ***Roy Bhargav Ravi (202311038)***

1 The Set Cover Problem

For a weighted set cover problem, we have a set of elements $E = \{e_1, e_2, \ldots, e_n\}$, along with subsets $S_1, S_2, \ldots, S_m$ where each $S_j \subseteq E$, and each subset has a non-negative weight $w_j \geq 0$. Our aim is to find a collection of subsets that covers all elements in E with minimum total weight. In other words, we want to identify a subset $I \subseteq \{1, \ldots, m\}$ that minimizes the expression:

$$\sum_{j \in I} w_j$$

while ensuring that the union of the selected subsets covers all elements in E:

$$\bigcup_{j \in I} S_j = E$$

For example: Let E={A,B,C,D,E,F,G,H} and the subsets be $S_1 = \{A, B\}$, $S_2 = \{A, B, C, D, E, F, G, H\}$, $S_3 = \{C, D, E, F, G, H\}$ with $w_1 = 2$, $w_2 = 15$, $w_3 = 9$, then $I = \{1, 3\}$.

For the unweighted set cover problem, each subset's weight is 1.

1.1 Greedy Approach to the Set Cover Problem

To solve the set cover problem, we employ a greedy algorithm for the set cover problem. The algorithm is defined as:

$I \leftarrow \emptyset$
$\hat{S}_j \leftarrow S_j \quad \forall j$
while I is not a set cover **do**
  $l \leftarrow argmin_{j:\hat{S}_j \neq \emptyset} \frac{w_j}{|\hat{S}_j|}$

  $I \leftarrow I \bigcup \{l\}$

  $\hat{S}_j \leftarrow \hat{S}_j - S_l \quad \forall j$

where I is the index set for the indices of the subsets to be included in the solution. The algorithm will provide an approximate solution to the set cover problem.

Mathematical Terminologies(to be referred):

1. *The k^{th} Harmonic number is given by:*

$$H_k = 1 + \tfrac{1}{2} + \tfrac{1}{3} + \ldots + \tfrac{1}{k}$$

2. *For positive numbers $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$, the following holds:*

$$\min_{i=1,\ldots,k} \tfrac{a_i}{b_i} \leq \tfrac{\sum_{i=1}^{k} a_i}{\sum_{i=1}^{k} b_i} \leq \max_{i=1,\ldots,k} \tfrac{a_i}{b_i}$$

1.2 Approximation Analysis of Greedy Algorithm

Claim: *The greedy algorithm is an H_n-approximation algorithm for the set cover problem.*
Proof

Let OPT denote the value of an optimal solution to the set cover problem. Given that an optimal solution encompasses all n elements with a solution of weight OPT, it follows that there exists a subset covering its elements with an average weight of at most $\frac{OPT}{n}$. Similarly, once k elements have been covered, the optimal solution can address the remaining $n-k$ elements with a solution of weight OPT, indicating the existence of a subset covering its remaining uncovered elements with an average weight of at most $\frac{OPT}{n-k}$. So, in general, the greedy algorithm is about $OPT/(n-k+1)$ to cover the k^{th} uncovered element, giving a performance guarantee of $\sum_{k=1}^{n} \frac{1}{n-k+1} = H_n$.

Lemma 1 *In the k^{th} iteration,*

$$\min_{j:\hat{S}_j \neq 0} \frac{w_i}{\left|\hat{S}_j\right|} \leq \frac{OPT}{n_k}$$

Here, n_k represents the number of elements uncovered at the start of iteration k.

Proof
Let O denote the indices of sets in the optimal solution.

Claim:

$$\begin{aligned} \min_{j:\hat{S}_j \neq 0} \frac{w_i}{\left|\hat{S}_j\right|} &\leq \frac{\sum_{j \in O} w_j}{\sum_{j \in O}\left|\hat{S}_j\right|} \\ &= \frac{OPT}{\sum_{i \in O}\left|\hat{S}_j\right|} \leq \frac{OPT}{n_k} \end{aligned}$$

This implies $\sum_{j \in O}\left|\hat{S}_j\right| \geq n_k$, as the optimal solution covers at least as many elements as those uncovered by the greedy algorithm.

■

Now, formalizing the intuition, let n_k denote the number of elements that remain uncovered at the start of the k^{th} iteration. If the algorithm takes l iterations, then $n_1 = n$, and we set $n_{l+1} = 0$. Let I_k denote the indices of the sets chosen in iterations 1 through $k-1$, and for each $j = 1, \ldots, m$, let $\hat{S}_j$ denote the set of uncovered elements in S_j at the start of this iteration; that is, $\hat{S}_j = S_j - \bigcup_{p \in I_k} S_p$.

Lemma 2 *For the set j chosen in the k^{th} iteration,*

$$w_j \leq \frac{n_k - n_{k+1}}{n_k} \cdot OPT$$

Proof Consider the proof.

In the proof, since j minimizes the ratio, we have $\frac{w_j}{\left|\hat{S}_j\right|} \leq \frac{OPT}{n_k}$, leading to $w_j \leq \frac{\left|\hat{S}_j\right| OPT}{n_k} = \frac{n_k - n_{k+1}}{n_k} \cdot OPT$, where $n_{k+1} = n_k - \left|\hat{S}_j\right|$. ■

To conclude the proof, given the claimed inequality by lemma 2, we can prove the claim. Let I contain the indices of the sets in our final solution. Then

$$\begin{aligned} \sum_{j \in I} w_j &\leq \sum_{k=1}^{l} \frac{n_k - n_{k+1}}{n_k} \cdot OPT \\ &\leq OPT \cdot \sum_{k=1}^{l} \left(\frac{1}{n_k} + \frac{1}{n_k - 1} + \ldots + \frac{1}{n_{k+1}+1} \right) \\ &= OPT \cdot \sum_{k=1}^{l} \frac{1}{i} \\ &= H_n \cdot OPT \end{aligned}$$

where the inequality follows from the fact that $\frac{1}{n_k} \leq \frac{1}{n_k - i}$ for $0 \leq i < n_k$.

■