

Lecture # 20

Lecturer: Debmalya Panigrahi

Scribe: Reem Mokhtar

1 Overview

Here we cover the analysis technique of dual-fitting on combinatorial algorithms, specifically on the greedy approximation algorithm problem of SET-COVER. In addition, we discuss how the integrality gap establishes an upper bound on the approximation factor, especially as it relates to SET-COVER.

2 SET COVER via Dual Fitting

2.1 Dual Fitting

Given a basic combinatorial algorithm, use the LP-relaxation of the problem and its dual to show that the objective function value of the primal solution found is at most the objective function value of the dual computed, and thus show that the primal integral solution provided by the algorithm is *fully paid for* by the computed dual.

The main steps are as follows:

1. **Formulate the dual:** Obtain an LP-relaxation of the IP, and find the dual of the relaxed LP.
2. **Shrink the dual:** divide the dual by a suitable factor f .
3. **Show that it is is feasible:** show that the shrunk dual fits into the given instance.
4. **Lower bound on OPT:** the shrunk dual becomes a lower bound on OPT
5. **Approximation guarantee:** the factor (f) is the approximation guarantee of the algorithm.

2.2 Greedy SET COVER

In greedy set cover, we choose set that minimizes the cost per number of elements covered.

Algorithm 2.1: GREEDYSETCOVER(C, \mathbb{U})

```

 $C \leftarrow \emptyset$ 
for each  $C \neq \mathbb{U}$ 
  do  $\left\{ \begin{array}{l} \text{Find } S \text{ with minimum cost-effectiveness } \alpha \leftarrow \frac{\text{cost}(S)}{|S - C|} \\ \text{For each } e \in S - C, \text{ price}(e) \leftarrow \alpha \\ C \leftarrow C \cup S \end{array} \right.$ 
return ( $C$ )

```

2.3 Analysis of Greedy SET COVER via Dual-Fitting

Step 1: Formulate the Dual

We first formulate the IP of set cover. Given sets $S \in \mathbb{S}$, assign a 0-1 variable x_S .

$$\text{minimize } \sum_{S \in \mathbb{S}} c(S)x_S \quad (1)$$

$$\text{subject to } \sum_{S: e \in S} x_S \geq 1 \quad e \in \mathbb{U} \quad (2)$$

$$x_S \in \{0, 1\} \quad S \in \mathbb{S} \quad (3)$$

We then obtain the relaxed LP:

$$\text{minimize } \sum_{S \in \mathbb{S}} c(S)x_S \quad (4)$$

$$\text{subject to } \sum_{S: e \in S} x_S \geq 1 \quad e \in \mathbb{U} \quad (5)$$

$$x_S \geq 0 \quad S \in \mathbb{S} \quad (6)$$

Finally, we formulate the dual LP:

$$\text{maximize } \sum_{e \in \mathbb{U}} y_e \quad (7)$$

$$\text{subject to } \sum_{e: e \in S} y_e \leq c(S) \quad S \in \mathbb{S} \quad (8)$$

$$y_e \geq 0 \quad e \in \mathbb{U} \quad (9)$$

We can think of the dual intuitively as charging each element e some nonnegative prices $y_e \geq 0$ for its coverage by a set cover. In that case, we might think of charging lower-cost sets with low prices, and higher-cost sets with high prices. The sum of the prices in some set S , or $\sum_{e: e \in S} y_e$ cannot be more than the cost of the set, which gives the limit described in the inequality in (8). If we set the dual variable to $\text{price}(e)$, this dual is not feasible.

Step 2: Shrink the Dual

We need to find a suitable factor by which to divide the dual variables in order to shrink the dual.

Lemma 1. For each $e \in \{1, \dots, n\}$, $\text{price}(e) \leq \frac{OPT}{n-k+1}$.

Proof. In any iteration, the cost of covering the remaining elements by the optimal solution is at most OPT , which means that there exists a set with $\alpha = \frac{OPT}{|\bar{C}|}$. When element e is being covered, the size of \bar{C} is $n-k+1 \leq |\bar{C}|$. Since e is being covered by the most cost-effective set in this iteration, then:

$$\text{price}(e) \leq \frac{\text{OPT}}{|\bar{C}|} \leq \frac{\text{OPT}}{n-k+1} \quad (10)$$

□

Theorem 2. *The greedy algorithm is an H_n factor approximation algorithm for the minimum set cover problem, where $H_n = 1 + \frac{1}{2} + \dots + \frac{1}{n}$.*

Proof. Since the cost of each set chosen is divided amongst the new elements being covered, the total cost of the set cover chosen is $\sum_{e \in S} \text{price}(e)$. By Lemma 1,

$$\sum_{e \in S} \text{price}(e) \leq \left(1 + \frac{1}{2} + \dots + \frac{1}{n}\right) \cdot \text{OPT} = H_n \cdot \text{OPT} \quad (11)$$

□

We can show that if this dual is shrunk by a factor of H_n , then it fits into the given set cover instance. For each element e , let $y_e = \frac{\text{price}(e)}{H_n}$.

Lemma 3. *The vector $\mathbf{y} = (y_{e_1}, \dots, y_{e_n})$ is a feasible solution for the dual program (8).*

Proof. We need to show the sum of the prices in some set S is at most the cost of the set. Consider a set $S \in \text{mathbb{S}}$ consisting of k elements. If the elements are numbered in the order in which they are covered (e_1, \dots, e_n), at the iteration where element e_i is being covered, then S contains at least $k - i + 1$ uncovered elements. S has a cost-effectiveness of at most $\frac{c(S)}{k - i + 1}$. Since S is the most cost-effective set, then the price of the element $\text{price}(e_i) \leq \frac{c(S)}{k - i + 1}$, which means:

$$y_e \leq \frac{1}{H_n} \cdot \frac{c(S)}{k - i + 1} \quad (12)$$

Summing over all elements in S :

$$\sum_{e \in S} y_e \leq \frac{c(S)}{H_n} \cdot \left(\frac{1}{k} + \frac{1}{k-1} + \dots + 1\right) = \frac{H_k}{H_n} \cdot c(S) \leq c(S) \quad (13)$$

□

Step 3: Lower bound on OPT

Let OPT_f be the cost of an optimal fractional set cover, which is the optimal solution to the LP in (1). Let OPT be the optimal integral set cover. The cost of an optimal integral set cover is at least the cost of an optimal fractional set cover, or $\text{OPT}_f \leq \text{OPT}$. The cost of any feasible dual solution is a lower bound on the cost of the fractional set cover.

The lower bounding scheme of the Algorithm 2.1 is thus the cost of any dual feasible solution.

Step 4: Approximation Guarantee

Theorem 4. *The approximation guarantee of the greedy set cover algorithm is H_n .*

Proof. Since y is dual feasible, the cost of the set cover picked:

$$\sum_{e \in \mathbb{U}} \text{price}(e) = H_n \left(\sum_{e \in \mathbb{U}} y_e \right) \leq H_n \cdot \text{OPT} \quad (14)$$

where OPT is the cost of the optimal fractional set cover. □

3 Integrality Gap

In order to explain the why an optimal solution to the primal is not necessary a better solution than a feasible solution to the dual, we need to understand what the integrality gap (or integrality ratio) of an LP-relaxation is.

Definition 1. *The integrality gap of an integer program is the worst-case ratio over all instances of the problem of the value of an optimal solution to the integer programming formulation $\text{OPT}(I)$ to the value of an optimal solution to its linear programming relaxation $\text{OPT}_f(I)$.*

$$\sup_{I \in \mathcal{J}} \left(\frac{\text{OPT}(I)}{\text{OPT}_f(I)} \right) \quad (15)$$

In the case of a maximization problem, the integrality gap will be the infimum (instead of the supremum) of the ratio.

The best approximation factor we can yield is the integrality gap of the relaxation. An *exact relaxation* is an LP-relaxation with an integrality gap of 1.

Also, this is closely related to complementary slackness conditions, since we can show that if a condition holds within a factor of f , we can get an f -approximation algorithm.

3.1 Integrality Gap of SET-COVER

Claim 1. *The SET-COVER LP has integrality gap $\Omega(\log(n))$*

Proof. The integrality gap of set cover can be shown using a Galois (finite) field of two elements $GF^k(2)$.

Let each element $e \in \mathbb{U}$ be a binary string of length k . This field has number of ground elements $n = 2^k - 1$. Each set of elements $S_e = \{e' : e' \cdot e = 1\}$.

$$|S_e| = \frac{n}{2}$$

$$|S_{e'} : e' \in S_e| = \frac{n}{2} \text{ number of sets each element belongs to.}$$

There are $T = 2^k$ sets. Each ground element is contained in exactly half of the sets. Which means that if we fractionally choose a set at most $\frac{2}{T}$, then each element is fractionally covered by at most 1. Hence

$$\text{OPT}(I) \leq T \frac{2}{T} = 2.$$

Second, $\text{OPT}_f(I) \geq q$.

What is the best fractional solution? $x_S = \frac{2}{n}$ is feasible. Fractional objective = 2. x must show that each objective has $\log(n)$. Show that every integer solution has at least k sets.

Supposed not.

$\{S_{e_1}, \dots, S_{e_q}\}, q \leq k$. Suppose q does not span all elements, that is $\{e_1, \dots, e_q\}$ do not span \mathbb{U} . So this collection has nonzero nullspace inside \mathbb{U} .

$\exists e$ s.t. $e \perp e_i \forall i \in 1 \dots q \implies e \notin \bigcup_{i=1}^q S_{e_i}$ □

References.

For the section on dual-fitting, I relied heavily on Williamson and Shmoys [2011] and Vazirani [2001]. For the integrality gap, I used Williamson and Shmoys [2011] and Sinop and O'Donnell [2008].

References

Ali Kemal Sinop and Ryan O'Donnell. Lecture 2: Lp relaxations, randomized rounding. University Lecture, Jan 2008. URL <http://www.cs.cmu.edu/~anupamg/adv-approx/lecture2.pdf>.

Vijay V Vazirani. Approximation algorithms. *Approximation algorithms*, August 2001. URL <http://portal.acm.org/citation.cfm?id=500776>.

David P Williamson and David B Shmoys. The Design of Approximation Algorithms, 1st edition. *The Design of Approximation Algorithms, 1st edition*, May 2011. URL <http://portal.acm.org/citation.cfm?id=1971947>.