Saarland
Informatics Campus

Script of Lecture 4, Approximation Algorithms
Summer term 2017

Tobias Mömke, Hang Zhou
http://www-cc.cs.uni-saarland.de/course/61/

Written by Hang Zhou

1 Set Cover (revisited)

Input: a ground set of elements $E = \{e_1, \dots, e_n\}$, subsets of those elements $S_1, \dots, S_m$ where each $S_j \subseteq E$, and a nonnegative weight $w_j \geq 0$ for each subset S_j.

Output: minimum weight collection of subsets that covers all elements of E.

2 Linear Program and Integer Program

Linear program (LP):

- *decision variables*: representing decisions that need to be made (e. g., whether a subset S_j is chosen in the solution; we create a variable x_j where $x_j = 1$ indicates that S_j is chosen and $x_j = 0$ indicates that S_j is not chosen).
- *constraints*: linear inequalities and equalities containing the decision variables (e. g., $0 \leq x_j \leq 1$ for each S_j; $\sum_{j: e_i \in S_j} x_j \geq 1$ for each e_i).
- *feasible solution*: an assignment of real numbers to the decision variables such that all the constraints are satisfied.
- *objective function*: a linear function of the decision variables (e. g., $\sum_{j=1}^{m} w_j x_j$).
- *optimal solution*: a feasible solution that minimizes or maximizes the objective function.

Integer program (IP): In addition, every decision variable has an integer value (e. g., $x_i \in \{0, 1\}$).

Finding a minimum-weight set cover is equivalent to solving the following integer program:

$$\begin{aligned} \text{minimize} & \sum_{j=1}^{m} w_j x_j & \\ \text{subject to} & \sum_{j: e_i \in S_j} x_j \geq 1, & i = 1, \dots, n, \\ & x_j \in \{0, 1\}, & j = 1, \dots, m. \end{aligned}$$

In general, integer programs cannot be solved in polynomial time. (Note that the set cover problem is NP-hard.) However, linear programs are polynomial time solvable, but we are not allowed to require that the decision variables are integers. So we replace the constraints $x_j \in \{0, 1\}$ by the constraints $x_j \geq 0$ and we obtain the following linear program, which can be solved in polynomial time:

$$
\begin{aligned}
\text{minimize} & \sum_{j=1}^{m} w_j x_j \\
\text{subject to} & \sum_{j:e_i \in S_j} x_j \geq 1, & i = 1, \ldots, n, \\
& x_j \geq 0, & j = 1, \ldots, m.
\end{aligned}
$$

The constraints $x_j \leq 1$ are not necessary, because for any solution to the linear program, we can reduce any $x_j > 1$ to $x_j = 1$ without affecting the feasibility of the solution and without increasing the cost.

We say that the linear program is a *relaxation* of the integer program, which means that every feasible solution to the integer program is feasible for the linear program, and that its value remains the same for the linear program.

3 A Deterministic Rounding Algorithm

Suppose we solve the linear program for the set cover problem and obtain an optimal solution x^*. How do we construct a solution to the set cover problem?

Let f be the maximum number of sets in which any element appears, i. e., $f := \max_{i \in \{1,\ldots,n\}} f_i$, where f_i is the number of sets containing the element e_i. For every set S_j, we include S_j in our solution if and only if $x_j \geq 1/f$. Let I be the collection of the sets in our solution.

Lemma 1. *I is a solution to the set cover problem.*

Proof. We need to show that every element e_i is contained in some set from I. By definition, the number of sets S_j containing e_i is at most f. From the constraint $\sum_{j:e_i \in S_j} x_j^* \geq 1$ in the linear program, we know there must be some j such that $e_i \in S_j$ and that $x_j^* \geq 1/f$. Thus S_j is a set in I that contains e_i. □

Lemma 2. *The cost of I is at most f times the minimum cost for the set cover problem.*

The proof is left as an exercise.

From the above two lemmas, we know that the rounding algorithm leads to an f-approximate solution to the set-cover problem.

4 Dual of the Linear Program

For each element e_i, associate a price $y_i \geq 0$ for its coverage, such that:

$$
\sum_{i:e_i \in S_j} y_i \leq w_j.
$$

The following linear program aims to find the highest total price that the elements can be charged:

$$
\begin{aligned}
\text{maximize} & \sum_{i=1}^{n} y_i \\
\text{subject to} & \sum_{i:e_i \in S_j} y_i \leq w_j, & j = 1, \ldots, m, \\
& y_i \geq 0, & i = 1, \ldots, n.
\end{aligned}
$$

This is called the *dual* linear program of the linear program for set cover (*primal* linear program).

Observation: a variable in the dual LP corresponds to a constraint in the primal LP; and a constraint in the dual LP corresponds to a variable in the primal LP.

Weak duality: any feasible solution to the dual LP has a value no greater than the optimal solution to the primal LP.

To see this, consider any feasible dual solution y and any feasible primal solution x. We have:

$$\begin{aligned}\sum_{i=1}^{n} y_i &\le \sum_{i=1}^{n} y_i \sum_{j:e_i\in S_j} x_j && \text{(constraints in primal LP)}\\ &= \sum_{j=1}^{m} x_j \sum_{i:e_i\in S_j} y_i \\ &\le \sum_{j=1}^{m} x_j w_j. && \text{(constraints in dual LP)}\end{aligned}$$

Strong duality: As long as there exist feasible solutions to both the primal and dual linear programs, their optimal values are equal. In other words, if x^* is an optimal solution to the primal linear program, and y^* is an optimal solution to the dual linear program, then we have

$$\sum_{j=1}^{m} w_j x_j^* = \sum_{i=1}^{n} y_i^*.$$

5 Rounding a Dual Solution

Let y^* be an optimal solution to the dual linear program. We construct a set cover solution I' as follows: We include in I' a set S_j if and only if the constraint corresponding to S_j in the dual LP is *tight*, i. e., $\sum_{i:e_i\in S_j} y_i^* = w_j$. We will show that I' is an f-approximate solution to the set cover problem.

Lemma 3. *I' is a solution to the set cover problem.*

Proof. We prove by contradiction. Suppose there exists some uncovered element e_k. Then for every subset S_j containing e_k we have

$$\sum_{i:e_i\in S_j} y_i^* < w_j.$$

Define $\varepsilon := \min_{j:e_k\in S_j}(w_j - \sum_{i:e_i\in S_j} y_i^*)$. We have $\varepsilon > 0$. Construct a new dual solution y' where $y'_k = y_k^* + \varepsilon$ and $y'_i = y_i^*$ for $i \ne k$. Then y' is a feasible solution for the dual linear program and $\sum_{i=1}^{n} y'_i > \sum_{i=1}^{n} y_i^*$, which contradicts with the optimality of y^*. □

Lemma 4. *The cost of I' is at most f times the minimum cost for the set cover problem.*

The proof is left as an exercise.

6 Primal-Dual Method

To obtain an f-approximate solution based on the dual LP, we could solve the dual LP directly (in polynomial time). But there is a much faster algorithm (Algorithm 1), called a *primal-dual* algorithm. Recall that in the previous section, we need three properties: (1) y is a feasible dual solution; (2) a set S_j is included in I' if and only if the corresponding constraint is tight; (3) I' covers every element e_i. Algorithm 1 is inspired by the proof of Lemma 3.

Algorithm 1 Primal-dual algorithm for the set cover problem

```
1: y ← 0
2: I′ ← ∅
3: while there exists e_i ∉ ∪_{j∈I′} S_j do
4:     Increase the dual variable y_i until there is some ℓ with e_i ∈ S_ℓ such that ∑_{j:e_j∈S_ℓ} y_j = w_ℓ
5:     I′ ← I′ ∪ {ℓ}
   return I′
```

Theorem 1. *Algorithm 1 is an f-approximation algorithm for the set cover problem.*