LP techniques for set cover

Chs. 13, 14, 15

Risto Hakala

risto.m.hakala@tkk.fi

March 10, 2008

Outline

- Recap of linear programming and LP-duality
- Set cover via dual fitting
- Rounding applied to set cover
- Set cover via the primal-dual schema

Linear programming and LP-duality

Linear programs

- Minimization linear program:

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{n} c_j x_j & \\ \text{subject to} \quad & \sum_{j=1}^{n} a_{ij} x_j \geq b_i, & i = 1, \ldots, m \\ & x_j \geq 0, & j = 1, \ldots, n, \end{aligned}$$

  where a_{ij}, b_i, and c_j are given rational numbers.
- Feasible solutions $\vec{x} = (x_1, \ldots, x_n)$ to this program provide Yes certificates for the question "Is the optimum value less than or equal to α?"

Linear programs

- Maximization linear program:

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{m} b_i y_i \\ \text{subject to} \quad & \sum_{i=1}^{m} a_{ij} y_i \leq c_j, \qquad j = 1, \ldots, n \\ & y_i \geq 0, \qquad i = 1, \ldots, m, \end{aligned}$$

  where a_{ij}, b_i, and c_j are given rational numbers.
- Feasible solutions $\vec{y} = (y_1, \ldots, y_m)$ to this program provide No certificates for the question "Is the optimum value less than or equal to α?"

LP-duality

Let a minimization linear program be the primal program.

Theorem 12.2 (Weak duality theorem)

If $\vec{x} = (x_1, \ldots, x_n)$ and $\vec{y} = (y_1, \ldots, y_m)$ are feasible solutions for the primal and dual program, respectively, then

$$\sum_{j=1}^{n} c_j x_j \geq \sum_{i=1}^{m} b_i y_i. \tag{1}$$

By the LP-duality theorem, (1) holds with equality iff both $\vec{x}$ and $\vec{y}$ are optimal solutions.

Set cover via dual fitting

Dual fitting

- In order to establish the approximation guarantee, the cost of the solution produced by the algorithm needs to be compared with the cost of an optimal solution.
- Since it is **NP**-hard to find the cost of an optimal solution of a minimization (resp. maximization) problem, we try to get around this by coming up with a polynomial time computable lower (resp. upper) bound on OPT.
- Dual fitting is a powerful method which helps finding a good bound on OPT using LP-duality theory.
- In this presentation, dual fitting is used to *analyze* the natural greedy algorithm for the set cover problem.

Idea behind dual fitting

- Dual fitting uses the linear programming relaxation of the problem and its dual to find the approximation guarantee of the algorithm.
- It is shown that the objective function value of the primal solution found by the algorithm is at most the objective function value of the dual computed; however, the dual is infeasible.
- The approximation guarantee is obtained by scaling the dual solution by a suitable factor F such that the solution becomes feasible.
- The shrunk dual is a lower bound on OPT by the weak duality theorem (Theorem 12.2), and the factor F is the approximation guarantee.

Set cover via dual fitting

Problem 2.1 (Set cover)

Given a universe U of n elements, a collection of subsets of U, $\mathcal{S} = \{S_1, \ldots, S_k\}$, and a cost function $c\colon \mathcal{S} \to \mathbb{Q}^+$, find a minimum cost subcollection of $\mathcal{S}$ that covers all elements of U.

Theorem 2.4

The greedy set cover algorithm (Algorithm 2.2) is an H_n factor approximation algorithm for the minimum set cover problem, where $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$.

It is shown how the approximation factor H_n is derived via dual fitting.

Greedy set cover algorithm

Algorithm 2.2 (Greedy set cover algorithm)

1. $C \leftarrow \emptyset$
2. *While* $C \neq U$ *do*
   Find the set S *whose cost-effectivness* $c(S)/|S - C|$ *is smallest.*
   Let $\alpha = c(S)/|S - C|$.
   Pick S, *and for each* $e \in S - C$, *set price*$(e) = \alpha$.
   $C \leftarrow C \cup S$.
3. *Output the picked sets.*

Set cover problem as an integer program

- Let $x_S \in \{0, 1\}$ be a variable which is set to 1 iff set $S \in \mathcal{S}$ is picked in the set cover.
- The set cover problem can be stated then as an integer linear program:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{S \in \mathcal{S}} c(S) x_S & \\
\text{subject to} \quad & \sum_{S\,:\,e \in S} x_S \geq 1, & e \in U \\
& x_S \in \{0, 1\}, & S \in \mathcal{S}
\end{aligned}
$$

LP-relaxation of the set cover problem

- The LP-relaxation of this integer program is obtained by letting the domain of variables x_S be $[0, \infty[$:

$$\text{minimize} \quad \sum_{S \in \mathcal{S}} c(S) x_S \quad \text{subject to} \quad \sum_{S\,:\,e \in S} x_S \geq 1, \;\; x_S \geq 0.$$

- Introducing the variable y_e for each $e \in U$, we obtain the dual program:

$$\text{maximize} \quad \sum_{e \in U} y_e \quad \text{subject to} \quad \sum_{e\,:\,e \in S} y_e \leq c(S), \;\; y_e \geq 0.$$

Analysis of the greedy set cover algorithm

- The original algorithm defines dual variables price(e) for each element e.
- This leads to (generally) infeasible dual solutions such that

$$\sum_{S \in \mathcal{S}} c(S) x_S = \sum_{e \in U} \text{price}(e),$$

i.e., the cost of the primal solution is at most the cost of the dual computed.

- We get a feasible solution by defining dual variables y_e as

$$y_e = \frac{\text{price}(e)}{H_n}, \quad e \in U.$$

Analysis of the greedy set cover algorithm

Lemma 13.2

The vector $\vec{y}$ defined as $y_e = price(e)/H_n$, $e \in U$, is a feasible solution for the dual program of the LP-relaxed set cover problem.

Proof.

Consider a set $S \in \mathcal{S}$ consisting of k elements. Number the elements in the order in which they are covered by the algorithm, say $e_1, \ldots, e_k$. Consider the iteration in which the algorithm covers element e_i. In this case, at most $i - 1$ elements have been covered by the cover C. Hence, S covers e_i at an average cost of at most $c(S)/|S - C| = c(S)/(k - (i - 1))$.

Analysis of the greedy set cover algorithm

Proof, cont'd.

Since the algorithm chooses the most cost-effective set in this iteration, $\text{price}(e_i) \leq c(S)/(k-i+1)$. Thus,

$$y_{e_i} = \frac{\text{price}(e_i)}{H_n} \leq \frac{1}{H_n} \cdot \frac{c(S)}{k-i+1}.$$

Summing over all elements in S,

$$\sum_{i=1}^{k} y_{e_i} \leq \frac{c(S)}{H_n} \cdot \left(\frac{1}{k} + \frac{1}{k-1} + \cdots + \frac{1}{1}\right) = \frac{H_k}{H_n} \cdot c(S) \leq c(S).$$

Therefore, S is not overpacked. □

Analysis of the greedy set cover algorithm

Theorem 13.3

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

Proof.

The cost of the set cover picked is

$$\sum_{e \in U} \mathrm{price}(e) = H_n \left(\sum_{e \in U} y_e \right) \leq H_n \cdot \mathrm{OPT}_f \leq H_n \cdot \mathrm{OPT},$$

where the first inequality follows from the weak LP-duality theorem and the fact that $\vec{y}$ is feasible. □

Analysis of the greedy set cover algorithm

- As a corollary, we get an upper bound of H_n on the integrality gap of the LP-relaxation.
- This bound is essentially tight, so H_n is indeed the best approximation factor one can achieve using this relaxation.
- The greedy algorithm and its analysis using dual fitting extend naturally to several generalizations of the set cover problem.

Constrained set multicover via dual fitting

Constrained set multicover problem

Each element e in the universe U needs to be covered a specific number r_e of times. Each set $S \in \mathcal{S}$ is allowed to be picked at most once.

The corresponding integer program is derived as before.

$$
\begin{aligned}
\text{minimize} \quad & \sum_{S \in \mathcal{S}} c(S) x_S & \\
\text{subject to} \quad & \sum_{S\,:\,e \in S} x_S \geq r_e, & e \in U \\
& x_S \in \{0, 1\}, & S \in \mathcal{S}
\end{aligned}
$$

LP-relaxation of constrained set multicover

- The constraint $x_S \leq 1$ in the LP-relaxation is no longer redundant because each set should be picked at most once:

$$
\begin{array}{lll}
\text{minimize} & \displaystyle\sum_{S \in \mathcal{S}} c(S) x_S & \\
\text{subject to} & \displaystyle\sum_{S\,:\,e \in S} x_S \geq r_e, & e \in U \\
& -x_S \geq -1, & S \in \mathcal{S} \\
& x_S \geq 0, & S \in \mathcal{S}
\end{array}
$$

LP-relaxation of constrained set multicover

- Introducing y_e for each $e \in U$ and z_S for each $S \in \mathcal{S}$, we obtain the dual program:

$$
\begin{array}{lll}
\text{maximize} & \displaystyle\sum_{e \in U} r_e y_e - \sum_{S \in \mathcal{S}} z_S & \\
\text{subject to} & \displaystyle\sum_{e\,:\,e \in S} y_e - z_S \leq c(S), & S \in \mathcal{S} \\
& y_e \geq 0, & e \in U \\
& z_S \geq 0, & S \in \mathcal{S}
\end{array}
$$

A greedy algorithm for constrained set multicover

- Let us say that an element e is alive if it occurs in fewer than r_e times of the picked sets.
- In each iteration, the algorithm picks the most cost-effective unpicked set, where the cost-effectiveness is defined as the average cost at which it covers alive elements.
- The algorithm halts when there are no more alive elements.
- The approximation guarantee of H_n is achieved again.

- The analysis of this algorithm is similar as with set cover, but more technical.

Constrained set multicover via dual fitting

- Set price(e, j_e) to be the cost-effectiveness of the set S which covers e for the j_eth time.
- The algorithm gives an infeasible dual solution $(\vec{\alpha}, \vec{\beta})$, where

$$\alpha_e = \text{price}(e, r_e) \quad \text{and} \quad \beta_S = \sum_{e\,:\,e \in S} (\text{price}(e, r_e) - \text{price}(e, j_e)).$$

- A feasible solution $(\vec{y}, \vec{z})$ is obtained by scaling

$$y_e = \frac{\alpha_e}{H_n} \quad \text{and} \quad z_S = \frac{\beta_S}{H_n}.$$

Rounding applied to set cover

Rounding applied to set cover

- LP-rounding technique is used to *design* approximation algorithms for the set cover problem.
- The first rounding algorithm achieves an approximation guarantee of f, where f is the frequency of the most frequent element.
- The second algorithm, achieving a guarantee of $O(\log n)$, illustrates the use of randomization in rounding.

A simple rounding algorithm

Algorithm 14.1 (Set cover via LP-rounding)

1. *Find an optimal solution to the LP-relaxation.*
2. *Pick all sets S for which $x_S \geq 1/f$ in this solution.*

Analysis of the simple rounding algorithm

Theorem 14.2

Algorithm 14.1 achieves an approximation factor of f for the set cover problem.

Proof.

Let $\mathcal{C}$ be the collection of picked sets. An element e is in at most f sets. It is covered by $\mathcal{C}$ because one set must be picked to the extend of at least $1/f$ in the fractional cover. Hence, $\mathcal{C}$ is a valid set cover. Rounding increases x_S by a factor of at most f. Therefore, the cost of $\mathcal{C}$ is at most f times the cost of the fractional cover. □

Randomized rounding applied to set cover

- Fractions in an optimal fractional solution are viewed as probabilities.
- Rounding is done by flipping coins with these biases and rounding accordingly.
- Repeating this process $O(\log n)$ times, and picking a set if it is chosen in any of the iterations, we get a set cover with high probability, by a standard coupon collector argument.
- The expected cost of the cover is

$$O(\log n) \cdot \mathrm{OPT}_f \leq O(\log n) \cdot \mathrm{OPT}.$$

Set cover via the primal-dual schema

Primal-dual schema

- Primal-dual schema is another method for *designing* approximation algorithms using linear programming.
- Optimal solutions to linear programs are characterized by the fact that they satisfy all the complementary slackness conditions (Theorem 12.3).
- Primal-dual schema is driven by a relaxed version of these conditions: a solution is constructed iteratively such that it satisfies the relaxed versions of complementary slackness conditions at all times.
- Another factor f algorithm for the set cover problem is presented.

Relaxed complementary slackness conditions

Primal complementary slackness conditions

Let $\alpha \geq 1$.
For each $1 \leq j \leq n$: either $x_j = 0$ or $c_j/\alpha \leq \sum_{i=1}^{m} a_{ij} y_i \leq c_j$.

Dual complementary slackness conditions

Let $\beta \geq 1$.
For each $1 \leq i \leq m$: either $y_i = 0$ or $b_i/ \leq \sum_{j=1}^{n} a_{ij} x_j \leq \beta \cdot b_i$.

By Theorem 12.3, solutions $\vec{x}$ and $\vec{y}$ are both optimal iff $\alpha = 1$ and $\beta = 1$.

Overview of the schema

Proposition 15.1

If $\vec{x}$ and $\vec{y}$ are primal and dual feasible solutions satisfying the slackness conditions, then

$$\sum_{j=1}^{n} c_j x_j \leq \alpha\beta \sum_{i=1}^{m} b_i y_i.$$

Proof.

From slackness conditions, we get $c_j x_j \leq \alpha x_j \sum_{i=1}^{m} a_{ij} y_i$ and $\alpha y_i \sum_{j=1}^{n} a_{ij} x_j \leq \alpha\beta b_i y_i$. It follows that

$$\sum_{j=1}^{n} c_j x_j \leq \sum_{i=1}^{m} \alpha y_i \sum_{j=1}^{n} a_{ij} x_j \leq \alpha\beta \sum_{i=1}^{m} b_i y_i.$$

Overview of the schema

- Pick a primal infeasible solution $\vec{x}$, and a dual feasible solution $\vec{y}$, such that the slackness conditions are satisfied for chosen α and β.
- Iteratively improve the feasibility of $\vec{x}$ (integrally) and the optimality of $\vec{y}$, such that the conditions remain satisfied, until $\vec{x}$ becomes feasible.
- An approximation guarantee of $\alpha\beta$ is achieved using this schema, since

$$\sum_{j=1}^{n} c_j x_j \leq \alpha\beta \sum_{i=1}^{m} b_i y_i \leq \alpha\beta \cdot \mathrm{OPT}_f \leq \alpha\beta \cdot \mathrm{OPT}$$

by Proposition 15.1 and the LP-duality theorem.

Primal-dual schema applied to set cover

Set $\alpha = 1$ and $\beta = f$. Set S is called *tight* if $\sum_{e\,:\,e \in S} y_e = c(S)$.

Primal conditions, "Pick only tight sets in the cover"

$$\forall S \in \mathcal{S} : x_S \neq 0 \Rightarrow \sum_{e\,:\,e \in S} y_e = c(S)$$

Dual conditions, "Each e, $y_e \neq 0$, can be covered at most f times"

$$\forall e : y_e \neq 0 \Rightarrow \sum_{S\,:\,e \in S} x_S \leq f$$

Primal-dual schema applied to set cover

Algorithm 15.2 (Set cover – factor f)

1. *Initialization:* $\vec{x} \leftarrow \vec{0}$; $\vec{y} \leftarrow \vec{0}$.
2. *Until all elements are covered, do*
   Pick an uncovered element e, *and raise* y_e *until some set goes tight.*
   Pick all tight sets in the cover and update $\vec{x}$.
   Declare all the elements occuring in these sets as "covered".
3. *Output the set cover* $\vec{x}$.

Primal-dual schema applied to set cover

Theorem 15.3

Algorithm 15.2 achieves an approximation factor of f.

Proof.

Clearly, there will be no uncovered and no overpacked sets in the end. Thus, primal and dual solutions will be feasible. Since they satisfy the relaxed complementary slackness conditions with $\alpha = 1$ and $\beta = f$, the approximation factor is f by Proposition 15.1. □

Summary

- Dual fitting provides a way for analyzing approximation algorithms.
- Rounding and the primal-dual schema can be used to design approximation algorithms.
- These methods were applied in analysis of the set cover problem.
- LP-duality theory proved to be extremely useful.