COSC2421 Aproximation Algorithms
Lecture # 10
lecturere: A. Borodin
Scribe: N. Mirmohammadi

May 4, 2005

Lecture 10 Scribe Notes

Preview:

In this lecture we will use LP duality to produce approximation algorithms for a number of problems

First, we will introduce the hitting set problem which is equivalant to the set cover problem, but lends itself more readily to model other problems. Then the primal dual approach to linear programming is introduced, and two algorithms are given for set cover and Steiner Forest problems using the primal dual schema.

Hitting Set Problem:

Input:

$< E, \mathcal{T}, c >$ where

- $E = \{e_1, e_2, \ldots, e_r\}$: a set of elements
- $\mathcal{T} = \{T_i \subseteq E\}$: set of subsets of E
- $c : E \to \mathbb{R}$: a cost function

Goal:

Find A, such that $A \subseteq E : A \cap T_i \neq \emptyset$, $\forall i$ so that A has minimum cost $C(A) = \sum_{e \in A} c_e$.

Hitting Set IP:

Minimize: $\sum_{e \in A} c_e x_e$

Subject to:

$\sum_{e:e\in T_i} x_e \geq 1$, for all i

$x_e \in \{0,1\}$

Compare this to the set cover problem.

Set Cover Problem:

Input:

$<\mathcal{C}, U, c>$ where

- $\mathcal{C} = \{S_1, S_2, \ldots, S_n\}$: collection of sets
- $U = \{u_1, u_2, \ldots, u_m\}$: universe of elements
- $c : \mathcal{C} \rightarrow \mathbb{R}$: a weight function

Goal:

Find set $\mathcal{A}$, such that $\mathcal{A} \subseteq \mathcal{C}$ and $\cup_{S_i \in \mathcal{A}} = U$ and the cost of $\mathcal{A}$ is minimal.

Set Cover IP:

Minimize: $\sum_s c_S x_S$

Subject to:

$\sum_{S:u\in S} x_S \geq 1$, for all $u \in U$

$x_S \in \{0,1\}$

To see that these are in fact equivalant, we give a 1-1 conversion between the two problems. Any instance of the set cover problem is an instance of the a hitting set problem:

$$\begin{aligned} E &= \mathcal{C} \\ T_i &= \{S_j | u_i \in S_j\} \subseteq E \end{aligned}$$

Any instance of the hitting set problem is an instance of the set cover problem:

$$\begin{aligned} U &= \{T_i | T_i \subseteq E\} \\ S_i &= \{T_j | e_i \in T_j\} \subseteq U \end{aligned}$$

Definition 1 *(f-frequency set cover probelem)*

An instance of the set cover problem is called f-frequency set cover problem if each u occurs in at most f sets.

Definition 2 *(f-cardinality hitting set)*

An instance of the hitting set problem is called a f-cardinality hitting set problem if each T_i has at most f elements.

The f-cardinality hitting set problem is equivalant to the f-frequency set cover problem. For example, the vertex cover problem is an instance of a 2-frequency set cover problem, or a 2-cardinality hitting set problem.

Generalizing what we know about vertex cover yields the following theorem.

Theorem 1 *The f-frequency set cover(f-cardinality hitting set) problem has a poly-time f-approximation using IP/LP relaxation and natural rounding.*

Proof

rounding $\bar{x}_S = \begin{cases} 1 & x_S^* \geqslant \frac{1}{f} \\ 0 & \text{otherwise} \end{cases}$

because each element can occur in at most f sets , this gives an f-approximation

□

But we don't have to solve this LP relaxation, which takes a bit of work, to get the answer. Instead, we could use LP-duality to get the f-approximation.

LP Duality:

The following is an example of an LP and its dual.

Example 1 Here is an example of an LP mimization in canonical form:

Minimize: $5x + 6y + 10z$

Subject to:

$6x - 7y \geq 8$

$3x + 2y - 4z \geq 3$

$x, y, z \geq 0$

Now, in order to find a lower bound on the objective function, note that the coefficient of each term in the objective function is greater than the coedfficeient of the same term in the second constraint. So we have:

$$5x + 6y + 10z > 3x + 2y - 4z \geq 3$$

But we want to ger as high a lower bound as possible, so we'd like to find some positive multipliers y_1, y_2 such that

$$\begin{aligned} x : 5 &\geq 6y_1 + 3y_2 \\ y : 6 &\geq -7y_1 + 2y_2 \\ z : 10 &\geq -4y_2 \end{aligned}$$

so that they maximize the lower bound : $8y_1 + 3y_2$

The multipliers have to be positive so that they don' t change the direction of the inequality in the constraints. So for any minimization (resp. maximization) LP in canonical form, there is a maximization (resp. minimization) LP which gives a lower (resp. upper) bound on the value of the original LP. The original LP is called the Primal LP, and the second LP is called it's Dual. The Dual of the Dual of any LP is the Primal LP.

LP-Duality:

Here is a LP minimization in it's canonical form:

Minimize: $\sum_{j=1}^{n} c_j x_j$

Subject to:

$\sum_{j=1}^{n} a_{\mathrm{ij}} x_j \geq b_i 1 \leq i \leq m$

$x_j \geq 0$

The dual to this problem is the following:

Maximize: $\sum_{i=1}^{m} b_i y_i$

Subject to:

$\sum_{j=1}^{n} a_{\mathrm{ij}} y_j \leq c_j 1 \leq j \leq n$

$y_i \geq 0$

So the value of any feasible dual solution is less than or equal to the value of any feasible primal solution, therefore $OPT_{dual} \leq OPT_{primal}$. This yields a lower bound on the optimal value.

Theorem 2 *(Strong Duality) The primal LP has a finite optimum iff its dual has a finite optimum. Furthermore, if $\vec{x}^*$ (respectively $\vec{y}^*$) is a primal (respectively dual) optimum then*

$$OPT_P = \sum c_j x_j^* = \sum b_i y_i^* = OPT_D$$

Theorem 3 *(Weak Duality) If $\vec{x}$ and $\vec{y}$ are primal/dual feasible then*

$$\sum c_j x_j \geq \sum b_i y_i$$

Equality holds if both $\vec{x}$ and $\vec{y}$ are optimal.

Corollary 1 *(Complementary Slackness Condition) If $\vec{x}$ and $\vec{y}$ are primal/dual optimal then*

$\forall j (1 \leq j \leq n) x_j \neq 0 \Rightarrow \sum_{j=1}^{n} a_{ij} y_j = c_j$ *Primal Slackness*

$\forall j (1 \leq i \leq m) y_i \neq 0 \Rightarrow \sum_{j=1}^{m} a_{ij} x_i = b_i$ *Dual Slackness*

Definition 3 *(Relaxed Definition of Slackness) $\vec{x}$ and $\vec{y}$ are primal/dual feasible satisfying α/β relaxed slackness if*
$\forall j (1 \leq j \leq n) x_j \neq 0 \Rightarrow \frac{C_j}{\alpha} \leq \sum_{j=1}^{m} a_{\mathrm{ij}} y_j \leq c_j$
$\forall j (1 \leq i \leq m) y_i \neq 0 \Rightarrow b_i \leq \sum_{j=1}^{n} a_{\mathrm{ij}} x_i \leq \beta . b_i$
α and β show how far we are from the optimal value.

Theorem 4 *(Relaxed Complementary Slackness Condition) If $\vec{x}$ and $\vec{y}$ are primal/dual feasible and satisfy α/β relaxed slackness, then $\vec{x}$ is a $\alpha \cdot \beta$ approximation; that is,*

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

Proof

$$\begin{aligned}
\sum_{j=1}^{n} c_j x_j &\leq \sum_{j=1}^{n} (\alpha \sum_{j=1}^{m} a_{ij} y_j) x_j \\
&\leq \alpha \sum_{j=1}^{n} (\sum_{j=1}^{m} a_{ij} y_j) x_j \\
&\leq \alpha \sum_{j=1}^{m} (\sum_{j=1}^{n} a_{ij} x_j) y_j \\
&\leq \alpha \sum_{j=1}^{m} \beta \cdot b_i y_j \\
&\leq \alpha \cdot \beta \sum_{i=1}^{m} b_i y_i
\end{aligned}$$

□

Definition 4 *(Tight Primal) We use $\alpha = 1$ for most applications. This is called tight primal.*

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

The Primal Dual Schema:

The basic Primal, Dual approach for constructing a feasible integral solution to an LP is as follows:

Start off with a (non-feasible) integral solution of the Primal LP. Then go back and forth between Primal and Dual, at each step improving the solution; the Dual always remains feasible, but gets better and better. The primal starts as non-feasible and ends up feasible.

In what follows, we present algorithms for the Set Cover problem and the Steiner Forest problems using the primal dual schema.

SET COVER:

We have already introduced the LP for set cover, here is the dual LP for that problem:

Maximize: $\sum_{e \in U} y_e$

Subject To: $\sum_{u: u \in S} y_u \leq c_S \quad \forall S \in \mathcal{C}$

To find the solution we use the folowing algorithm which uses the primal dual schema:

Algorithm:

1. Initialization: $x \leftarrow 0; y \leftarrow 0$
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 x.

   Declare all the elements occurring in these sets as "covered".

3. Output the set cover x.

Note: for every uncovered element e in the set, e's equation in the dual is not tight.

This algorithm used on the cover set problem yields $(1, f)$-relaxed slackness, and hence a f-approximation.

Claim: If the input to the problem contains enough information, that the algorithm can figure out which variable to raise, then this algorithm is in fact a priority algorithm.

THE STEINER FOREST PROBLEM:

Input:
$< G = (V, E), \mathcal{S}, c >$ where

- $G = (V, E)$ is an undirected graph.
- $c : E \rightarrow \mathbb{R}^{\geq 0}$
- $\mathcal{S} = \{S_1, S_2, \ldots, S_k\}$ where each $S_i \subseteq V$ and $S_i \cap S_j = \emptyset$ when $i \neq j$.

Goal:

Generate a forest $A \subseteq E$, such that all $u, v \in S_i$ are connected in (V, A) and A has minimum weight.

The LP for this problem and it's dual are given below, with the necessary definitions following it:

Primal:

Minimize: $\sum_{e \in E} c_e x_e$

Subject to:

$\sum_{e: e \in \delta(S)} x_e \geq f(S), S \subseteq V$

$x_e \geq 0$

Dual:

Maximize: $\sum_{S \subseteq V} f(S) \cdot y_S$

Subject to:

$\sum_{S: e \in \delta(S)} y_S \leq c_e, e \in E$

$y_S \geq 0, S \subseteq V$

Definition 5 *(Cut) A cut is defined as follows:*

$$\delta(S) = \{e | e = (u, v), u \in S, v \in \bar{S}\}$$

Definition 6 *The following function $f : 2^v \rightarrow \{0, 1\}$, specifies the minimum number of edges that must cross each cut in every feasible solution.*

$$f(S) = \begin{cases} 1 & \text{if } \exists u \in S, v \in \bar{S}, u, v \in S_i \text{ for some } i \\ 0 & \text{otherwise} \end{cases}$$

The following is an algorithm using the primal dual schema with synchronization to solve this problem. But first here are some more definitions:

Definition 7 *S is Active relative to F if S is a connected component in (V,F) and f(S)=1*

Definition 8 *if F is a set of edges,*
$\deg_F(S)$ = *# of edges in* $F \cap \delta(S)$

Algorithm:

1. (Initialization) $F \leftarrow \emptyset$; for each $S \subseteq V$, $y_S \leftarrow 0$

2. (Edge augmentation) while there exists an unsatisfied set do:

   simultaniously raise y_S for each active set, until some edge e becomes tight;

   $F \leftarrow F \cup \{e\}$.

3. (Pruning) return $F' = \{e \in F | F - \{e\}$ is primal infeasible $\}$

Theorem 5 *this algorithm gives a 2-approximation for the Steiner Forest Problem.*

Proof First note that F' (F after pruning) is still a feaible solution to the Primal problem.This is because in each iteration Dual variables of connected components only get raised, so no edge running within the same component can become tight,so F is acyclic(a forest) so each edge on any path $u - v$ in F is non redundant and is not deleted in step 3.

$\sum_{S \subseteq V} \deg_F(S) \cdot y_s \leq 2 \cdot \sum_{S \subseteq V} y_s$

$\Rightarrow$

$\sum_{e \in F'} c_e = \sum_{S \subseteq V} (\sum_{e \in \delta(S) \cap F'} y_S) = \sum_{S \subseteq V} \deg_F(S) \cdot y_s \leq 2 \cdot \sum_{S \subseteq V} y_s$ □

If the input items are edges represtented by their costs, their end points and whether these end points are required or Steiner nodes, then this algorithm can be seen to be a stack algorithm.

References

[1] V.J. Vazirani Approximation Algorithms. Springer-Verlag Berlin 2001