Some More Uses of Duality

Last revised – October 8, 2007

- The **Max-Flow Min-Cut Theorem** is a just a special case of the main duality theorem

- Feasible solutions to *dual LPS* can provide lower bounds to associated *ILP*s.
  We will see how this can be used to design an H_n-approximation algorithm for the **Weighted Set-Cover** problem.

The Max-Flow Min-Cut Theorem

Let $N = (s, t, V, E, b)$ be a flow network with
s, t the source,sink
$n = |V|$, the # of vertices
$m = |E|$, the # of edges
$b(x, y)$, the capacity of edges (x, y).

We will use (x, y) to denote flow in (x, y).
Let A be the node-arc incidence matrix of (V, E).
An $s - t$ flow of value v can be written as

$$\begin{aligned} Af &= \begin{cases} +v & \text{Row } s \\ -v & \text{Row } t \\ 0 & \text{other rows} \end{cases} \\ f &\leq b \\ f &\geq 0 \end{aligned}$$

An $s - t$ flow of value v can be written as

$$
\begin{aligned}
Af &= \begin{cases} +v & \text{Row } s \\ -v & \text{Row } t \\ 0 & \text{other rows} \end{cases} \\
f &\leq b \\
f &\geq 0
\end{aligned}
$$

Define vector

$$
d_i = \begin{cases} -1 & i = s \\ +1 & i = t \\ 0 & \text{otherwise} \end{cases}
$$

Then maximizing v can be written as LP

$$
\begin{aligned}
& \max v \\
Af + dv &= 0 \\
f &\leq b \\
f &\geq 0
\end{aligned}
$$

Maximizing v can be written as LP1

$$\begin{array}{rcl} & \max v & \\ Af + dv & = & 0 \\ f & \leq & b \\ f & \geq & 0 \end{array}$$

We now take the dual of this LP, which will be the primal LP2 . The first n equations of LP1 will correspond to n variables in LP2; $\pi(x)$ for $x \in V$. Since the first n equations are equalities, these variable are free.

The last m equations of LP1 will correspond to m variables in LP2; $\gamma(x,y)$ for $(x,y) \in E$. Since the last m equations are inequalities, these variable are constrained.

LP2 is then

$$\begin{array}{rl} \min \sum_{(x,y)\in E} \gamma(x,y)b(x,y) & \\ \pi(x) - \pi(y) + \gamma(x,y) \geq 0 & \forall (x,y) \in E \\ -\pi(s) + \pi(t) \geq 1 & \\ \pi(x) \gtrless 0 & \\ \gamma(x,y) \geq 0 & \end{array}$$

$$\begin{aligned} \min \textstyle\sum_{(x,y)\in E} \gamma(x,y) b(x,y) \\ \pi(x) - \pi(y) + \gamma(x,y) &\geq 0 \quad \forall (x,y) \in E \\ -\pi(s) + \pi(t) &\geq 1 \\ \pi(x) &\gtrless 0 \\ \gamma(x,y) &\geq 0 \end{aligned}$$

A **cut** is a partition $(W, \bar{W})$ of the vertices V with $s \in W$ and $t \in \bar{W}$. The *capacity* of a cut is

$$C(W, \bar{W}) = \sum_{\substack{(i,j)\in E \\ s.t.\ i \in W, j \in \bar{W}}} b(i,j)$$

Theorem Every s-t cut determines a feasible solution with cost $C(W, \bar{W})$ to LP2 as follows:

$$\gamma(x,y) = \begin{cases} 1 & (x,y) \text{ such that } x \in W,\ y \in \bar{W} \\ 0 & \text{otherwise} \end{cases}$$

$$\pi(x) = \begin{cases} 0 & x \in W \\ 1 & x \in \bar{W} \end{cases}$$

We have just shown that every $(W, \bar{W})$ has an associated solution to primal LP2 with cost $C(W, \bar{W})$.

This proves a (weak) form of the
Max-Flow Min-Cut Theorem, i.e,

> **Theorem:**
> The value v of any s-t flow is no greater than the capacity $C(W, \bar{W})$ of any s-t cut.
>
> Furthermore, if $v = C(W, \bar{W})$,
> then v is a max-flow
> and $(W, \bar{W})$ is a min-cut.

Recall the **Set Covering Problem**. Let X be a set and $\mathcal{F}$ a family of subsets of X such that $X = \cup_{F \in \mathcal{F}} F$.

For example $X = \{1, 2, 3, 4, 5, 6\}$ and $\mathcal{F}$ contains the subsets

$$\begin{aligned} F_1 &= \{1, 3, 5\} \\ F_2 &= \{2, 3, 6\} \\ F_3 &= \{2, 5, 6\} \\ F_4 &= \{2, 3, 4, 6\} \\ F_5 &= \{1, 4\} \end{aligned}$$

A subset $F \in \mathcal{F}$ *covers* its elements.

The problem is to find a *minimum-size subset* $\mathcal{C} \subseteq \mathcal{F}$ that covers X, i.e., $X = \cup_{F \in \mathcal{C}} F$.

For example $\{F_1, F_2, F_4\}$ covers X but is not a minimal size solution.

$\mathcal{C} = \{F_1, F_4\}$ is a minimal size solution.

Finding a minimal-size set cover is NP-Hard.

We now generalize the problem to the *Weighted Set Cover problem* where each set F has a *weight* $Cost(F) = C(F)$, and the problem is to find a *Set Cover* of $\mathcal{C}$ of *Minimum Weight*, $Cost(\mathcal{C}) = \sum_{F \in \mathcal{C}} C(F)$.

For example $X = \{1, 2, 3, 4, 5, 6\}$ and $\mathcal{F}$ contains the subsets

$$
\begin{aligned}
F_1 &= \{1, 3, 5\}; & C(F_1) = 1 \\
F_2 &= \{2, 3, 6\}; & C(F_2) = 1 \\
F_3 &= \{2, 5, 6\}; & C(F_3) = 3 \\
F_4 &= \{2, 3, 4, 6\}; & C(F_4) = 5 \\
F_5 &= \{1, 4\}; & C(F_5) = 1
\end{aligned}
$$

For example $\mathcal{C} = \{F_1, F_4\}$ is a minimal cardinality solution but not a minimum weight one. $\mathcal{C} = \{F_1, F_2, F_5\}$ is a minimum weight solution.

The *Weighted Set Cover problem* is NP-Hard so being able to find an optimal solution is unlikely. We can find an H_n approximation algorithm, though where $n = |X|$ and $H_n = \sum_{i=1}^{n} \frac{1}{i} \sim \ln n$.

This means that, for every input, our algorithm will generate a cover $\mathcal{C}$ such that

$$Cost(\mathcal{C}) \leq H_n \cdot OPT$$

where OPT is the cost of the real optimal solution (which we do not know).

Question: If we do not know OPT how can we guarantee the approximation?

Answer: Using Duality

Question: If we do not know OPT how can we guarantee the approximation?

Answer: Using Duality.

1. Write *Weighted Set Cover* as mimimization ILP, P'.
Let OPT be cost of optimal solution to P'.

2. Relax the ILP to a LP, P.
Let z^* be cost of optimal solution to P.
Note that $z^* \leq OPT$.

3. Let D be the dual LP to P.
Construct some *feasible* solution π to D.
Let w be the cost of π.
Duality says that $w \leq z^* \leq OPT$.

4. Our algorithm will be to create a $\mathcal{C}$ satisfying

$$Cost(\mathcal{C}) \leq H_n \cdot w.$$

This will guarantee

$$Cost(\mathcal{C}) \leq H_n \cdot w \leq H_n \cdot z^* \leq H_n \cdot OPT$$

A Greedy Set-Cover algorithm

$U \leftarrow X$

$\mathcal{C} \leftarrow \emptyset$

  select a $F \in \mathcal{F}$ that minimizes $\frac{Cost(F)}{|F \cap U|}$

  $U \leftarrow U - F$

  $\mathcal{C} \leftarrow \mathcal{C} \cup \{F\}$

  For each $e \in F \cap U$ set $price(e) = \frac{Cost(F)}{|F \cap U|}$

return($\mathcal{C}$)

The value $\frac{Cost(F)}{|F \cap U|}$ is the *cost-effectiveness* of the set. It is the average cost of adding each item in $|F \cap U|$ to the cover.

$price(e)$ will contain this average value (used later in the analysis). Note that the set cover constructed has cost $\sum_{e \in U} price(e)$.

Note that if $Cost(F) = 1$ for all sets F then the algorithm always picks F that *maximizes* $|F \cap U|$. This is a *greedy* algorithm for set cover.

The integer LP will be

Minimize $\sum_{F \in \mathcal{F}} C(F) x_F$
subject to conditions
$\forall e \in U, \qquad \sum_{e \in F} x_F \geq 1$
$\forall F \in \mathcal{F} \qquad x_F \in \{0, 1\}$

The relaxation of the LP is

Minimize $\sum_{F \in \mathcal{F}} C(F) x_F$
subject to conditions
$\forall e \in U, \qquad \sum_{e \in F} x_F \geq 1$
$\forall F \in \mathcal{F}, \qquad 1 \geq x_F \geq 0$

Note that *this is the same as*

Minimize $\sum_{F \in \mathcal{F}} C(F) x_F$
subject to conditions
$\forall e \in U, \qquad \sum_{e \in F} x_F \geq 1$
$\forall F \in \mathcal{F}, \qquad x_F \geq 0$

We now introduce a variable y_e for all $e \in U$.
The *dual* of the relaxed LP is then

Maximize $\sum_{e \in U} y_e$
subject to conditions
$\forall F \in \mathcal{F}, \qquad \sum_{e \in F} y_e \leq C(F)$
$\forall e \in U, \qquad y_e \geq 0$

Maximize $\sum_{e \in U} y_e$
subject to conditions
$\forall F \in \mathcal{F}, \quad \sum_{e \in F} y_e \le C(F)$
$\forall e \in U, \quad y_e \ge 0$

Theorem: The setting $y_e = \frac{price(e)}{H_n}$ is a feasible solution to the dual problem D where $n = |U|$.

Proof: Consider some set $F \in \mathcal{F}$. Let $k = |F|$. Number the elements of F in the order in which they are covered by the algorithm as $e_1, e_2, \ldots, e_k$, breaking ties arbitrarily.

Let us examine the step of the algorithm at which item e_i is covered. Before this step starts F contains at least $k - i + 1$ uncovered elements.

Therefore, at this step F itself can cover e_i with cost-effectiveness at most $\frac{c(F)}{k-i+1}$. Since the algorithm chooses a set F' with minimal cost-effectiveness this implies $price(e_i) \le \frac{c(F)}{k-i+1}$. Thus

$$y_{e_i} \le \frac{1}{H_n} \cdot \frac{C(F)}{k - i + 1}$$

so

$$\sum_{i=1}^{k} y_{e_i} \le \frac{C(F)}{H_n} \left(\frac{1}{k} + \frac{1}{k-1} + \cdots \frac{1}{1} \right) = \frac{H_k}{H_n} \cdot C(F) \le C(F)$$

Theorem: The approximation algorithm is an H_n-approximation algorithm.

Proof:

- The theorem on the previous page states that the setting $\forall e \in U,\ y_e = \frac{price(e)}{H_n}$ is a feasible solution for the dual LP D. The objective function for the dual was $\sum_{e\in U} y_e$ which for this setting has value $w = \sum_{e\in U} \frac{price(e)}{H_n}$.

- From the *Duality Theorem* we have that $w \le z^*$ where z^* is optimal solution of the primal, P.

- $z^* \le OPT$ by definition of LP relaxation.

- Then $w \le z^* \le OPT$ or $\sum_{e\in U} \frac{price(e)}{H_n} \le OPT.$

- Recall that

$$Cost(\mathcal{C}) = \sum_{F\in\mathcal{C}} Cost(F) = \sum_{e\in U} price(e).$$

  This proves $Cost(\mathcal{C}) \le H_n \cdot OPT.$