LP-based Approximation Algorithms

Stefano Leonardi

Sapienza Universitá di Roma

Theoretical Computer Science, Academic Year 2010/2011

Lecture Outline

Part I	LP-based Relaxation and Rounding
Part II	The Primal-dual Method
Part III	Primal-dual Method for Vertex Cover
Part IV	Set Cover via Dual Lifting
Part V	Randomized Rounding

LP-based Relaxation and Rounding

Optimization Problems

Approximation algorithm for problem P:

- I : Instance of problem P
- $A(I)$: value of A's solution on I
- $OPT(I)$: value of the optimal solution on I
- A is r-approximate if
  - Minimization: $\forall I, A(I) \le r \times OPT(I)$
  - Maximization: $\forall I, A(I) \ge \frac{1}{r} \times OPT(I)$

Relate to the Optimum

- The optimum may be hard to compute or to characterize
- However in some cases we can relate to the optimum, e.g., Christofides

$$ALG(I) \leq MST(I) + \texttt{Mathing}(I) \leq \left(1 + \frac{1}{2}\right) OPT(I)$$

Relate to the optimum of a relaxed problem:

$$OPT(I) = \texttt{min}_{x \in S(I)} f(x)$$

Relaxation: $S(I) \subseteq R(I)$

$$LB(I) = \texttt{min}_{x \in R(I)} g(x)$$

$\forall I, x \in S(I), g(x) \leq f(x)$

Use the solution of the relaxed problem

$$\texttt{If } \forall I,\ ALG(I) \leq r \times LB(I)$$
$$\texttt{then } \forall I,\ ALG(I) \leq r \times OPT(I)$$

Optimum of $LB(I)$:

$$x^* : g(x^*) = \texttt{min}_{x \in R(I)} g(x)$$

The optimum to the relaxed problem must be easy to compute.

Rounding:

Round x^* to a $\bar{x} \in S(I)$: $f(\bar{x}) \leq r \times g(x^*)$

LP formulation for Vertex Cover

Vertex Cover

- $G = (V, E)$, $w(u) \in \Re^+$, $u \in V$
- Find a set $U \subseteq V$ of min total cost $\sum_{u \in U} w(u)$ such that:
- $\forall e = (u, v) \in E$, either $u \in U$ or $v \in U$

atleast one covered.

$$
\begin{array}{llcll}
\textbf{min} & \sum_{v \in V} x(v) w(v) & & & \\
\text{s.t.} & x(u) + x(v) & \geq & 1 & \forall (u, v) \in E \\
 & x(u) & \in & \{0, 1\} & u \in V
\end{array}
$$

LP-relaxation for Vertex Cover

$$
\begin{array}{rrcll}
\text{min} & \sum_{v \in V} x(v) w(v) & & & \\
\text{s.t.} & x(u) + x(v) & \geq & 1 & \forall (u,v) \in E \\
 & x(u) & \in & [0,1] & u \in V
\end{array}
$$

- The fractional LP program can be computed in polynomial time
- All vertex covers are still feasible solution to the LP relaxation
- The optimum to the LP relaxation is a lower bound to the optimum vertex cover

Rounding

- Find an optimal solution $x^* \in R$ to the relaxed problem in polynomial time
- Round $x^* \in R$ to a $\bar{x} \in S$
  - $\bar{x}(u) = 1$ if $x^*(u) \geq \frac{1}{2}$
  - $\bar{x}(u) = 0$ if $x^*(u) < \frac{1}{2}$
- The solution $\bar{x}$ is feasible:

$$\forall e = (u, v), \bar{x}(u) + \bar{x}(v) \geq 1$$

since either $x^*(u) \geq \frac{1}{2}$ or $x^*(v) \geq \frac{1}{2}$

Approximation

- The solution is 2-approximated:

$$\sum_{u \in U} w(u)\bar{x}(u) \leq 2 \times \sum_{u \in V} w(u)x^*(u) \leq 2 \times OPT,$$

since $\bar{x}(u) \leq 2 \times x^*(u)$.

Integrality Gap

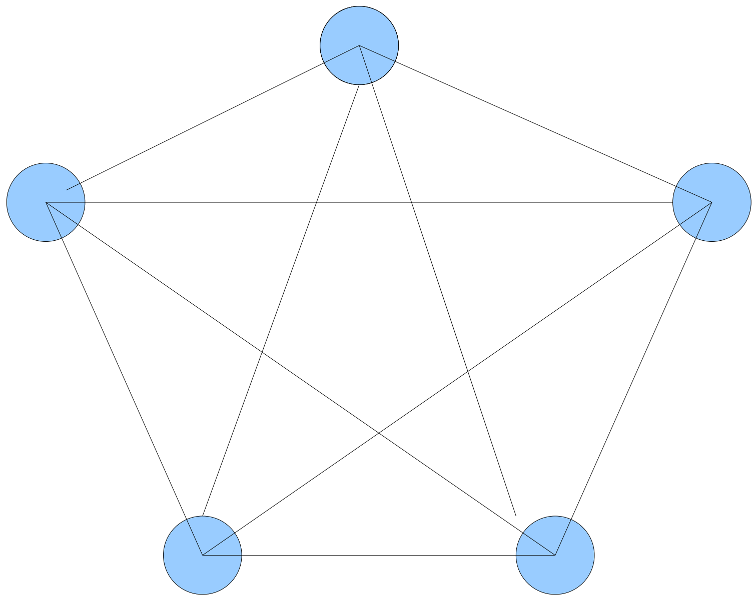
Definition: largest ratio on all instances between the optimum integral solution and the optimum relaxed solution.

One cannot hope to achieve an approximation ratio better than the integrality gap of the relaxation

The rounding step should pay a factor at least equal to the integrality gap of the relaxation

Integrality Gap for Vertex Cover

- On a clique graph the optimal vertex cover is of size $n-1$
- $x^*(u) = \frac{1}{2}$ is a feasible fractional solution of value $n/2$
- The integrality gap is equal to $2\left(1-\frac{1}{n}\right)$
- It is not possible to prove better than 2 approximation for Vertex Cover with this LP



LP rounding for Set Cover

Given:

- U: universe of n elements $\{e_1, \ldots, e_n\}$
- $\mathcal{S} = \{S_1, \ldots, S_m\}$: collection of m subsets of U
- $c : S_i \to \Re^+$: cost function for sets

Goal:

Find a subcollection of minimum cost that covers U

LP formulation for Set Cover

$$\begin{array}{llcll} \text{min} & \sum_{S \in \mathcal{S}} c(S)x(S) & & & \\ \text{s.t.} & \sum_{S: e \in S} x(S) & \geq & 1 & \forall e \in U \\ & x(S) & \in & \{0,1\} & S \in \mathcal{S} \end{array}$$

LP relaxation:

$$\begin{array}{llcll} \text{min} & \sum_{S \in \mathcal{S}} c(S)x(S) & & & \\ \text{s.t.} & \sum_{S: e \in S} x(S) & \geq & 1 & \forall e \in U \\ & x(S) & \in & [0,1] & S \in \mathcal{S} \end{array}$$

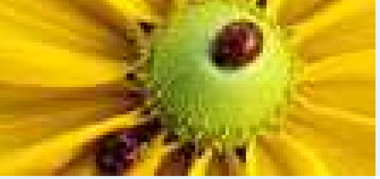
f-approximation for Set Cover

Let $f = \max_{e \in U} |\{S \in \mathcal{S} | e \in S\}|$

1. Round to 1 all variables $x(S) \geq \frac{1}{f}$
2. The solution is feasible since every elements appears in at least one set with $x(S) \geq \frac{1}{f}$
3. $ALG \leq f \times OPT^{LP}$

For Vertex cover we have $f = 2$.

An $O(\log n)$ approximation algorithm for Weighted Set Cover will follow in this lecture



More basic techniques

- Primal-dual scheme: consider the dual of a relaxation of the problem:

$$\max\{h(y) : y \in D\} \leq \min\{g(x) : x \in R)$$

Construct a fesible solution $x \in S$ for the primal from $y \in D$ such that

$$f(x) \leq r \times h(y) \leq r \times h(y^*) \leq r \times f(x^*) \leq r \times OPT$$

- Use Randomization in the rounding step: Interpret x^* as a set of probabilistic values to guide the rounding step. E.g., $\bar{x}(u) = 1$ with pb $x^*(u)$.

- Use different relaxations: e.g., semi-definite programming relaxations

The Primal-Dual Method

LP Duality

$$\min \quad 7x_1 + x_2 + 5x_3$$

$$\begin{aligned} \text{s.t.} \quad x_1 - x_2 + 3x_3 &\geq 10 \\ 5x_1 + 2x_2 - x_3 &\geq 6 \\ x_1, x_2, x_3 &\geq 0 \end{aligned}$$

Lower bounds on OPT:

$$\begin{aligned} 7x_1 + x_2 + 5x_3 &\geq x_1 - x_2 + 3x_3 \geq 10 \\ 7x_1 + x_2 + 5x_3 &\geq x_1 - x_2 + 3x_3 \\ &+ \; 5x_1 + 2x_2 - x_3 \geq 16 \end{aligned}$$

Best lower bound on OPT

$$\min \quad 7x_1 + x_2 + 5x_3$$

$$\text{s.t.} \quad \begin{aligned} x_1 - x_2 + 3x_3 &\geq 10 \\ 5x_1 + 2x_2 - x_3 &\geq 6 \\ x_1, x_2, x_3 &\geq 0 \end{aligned}$$

$$\max \quad 10y_1 + 6y_2$$

$$\text{s.t.} \quad \begin{aligned} y_1 + 5y_2 &\leq 7 \\ -y_1 + 2y_2 &\leq 1 \\ 3y_1 - y_2 &\leq 5 \\ y_1, y_2 &\geq 0 \end{aligned}$$

Primal-Dual Formulation

$$
\begin{array}{llll}
\text{min} & \displaystyle\sum_{j=1}^{n} c_j x_j & & \\
\text{s.t.} & \displaystyle\sum_{j=1}^{n} a_{ij} x_j \geq b_i & & i = 1, \ldots, m \\
& x_j \geq 0 & & j = 1, \ldots, n
\end{array}
$$

$$
\begin{array}{llll}
\text{max} & \displaystyle\sum_{i=1}^{m} b_i y_i & & \\
\text{s.t.} & \displaystyle\sum_{i=1}^{m} a_{ij} y_i \leq c_j & & j = 1, \ldots, n \\
& y_i \geq 0 & & i = 1, \ldots, m
\end{array}
$$

Strong Duality Theorem

Theorem 1 *If the Primal has finite optimum then the Dual has finite optimum. Let x^* and y^* be the primal and the dual optimum solutions. Then*

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

.

LP Duality

Weak Duality Theorem

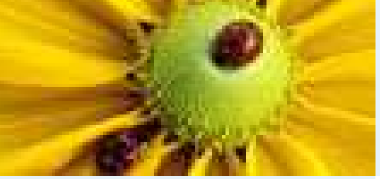
Theorem 2 *If x is feasible for the Primal and y is feasible for the Dual, then*

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

$$\sum_{j=1}^{n} c_j x_j \geq \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} y_i \right) x_j = \tag{1}$$

$$= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right) y_i \geq \sum_{i=1}^{m} b_i y_i \tag{2}$$

Corollary 3 *x and y are optimal for the Primal and the Dual if and only if (1) and (2) hold with equality.*



Complementary Slackness Conditions

x and y are optimal solutions if and only if:

(1) Primal Complementary Slackness Condition

$$\forall 1 \leq j \leq n \quad \texttt{either} \quad x_j = 0$$
$$\texttt{or} \quad \sum_{i=1}^{m} a_{ij} y_i = c_j$$

(2) Dual Complementary Slackness Condition

$$\forall 1 \leq i \leq m \quad \texttt{either} \quad y_i = 0$$
$$\texttt{or} \quad \sum_{j=1}^{n} a_{ij} x_j = b_i$$

The Primal-Dual Schema

Ensure Primal Complementary Slackness Condition:

$$\forall 1 \le j \le n \quad \texttt{either} \quad x_j = 0$$

$$\texttt{or} \quad \sum_{i=1}^{m} a_{ij} y_i = c_j$$

Relax Dual Complementary Slackness Condition

$$\forall 1 \le i \le m \quad \texttt{either} \quad y_i = 0$$

$$\texttt{or} \quad \sum_{j=1}^{n} a_{ij} x_j \le r \times b_i$$

The Primal-Dual Schema

Theorem 4 *If x and y satisfy the conditions of the Primal-Dual schema then*

$$\sum_{j=1}^{n} c_j x_j \leq r \times \sum_{i=1}^{m} b_i y_i$$

Proof:

$$\sum_{j=1}^{n} c_j x_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} y_i \right) x_j =$$

$$= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right) y_i \leq r \times \sum_{i=1}^{m} b_i y_i$$

□

Application of the Primal-Dual schema

- Primal is a relaxation of a problem P.
- x is integral feasible for P
- It follows:

$$\sum_{j=1}^{n} c_j x_j \leq r \times \sum_{i=1}^{m} b_i y_i \leq r \times \sum_{i=1}^{m} b_i y_i^*$$

$$= r \times \sum_{j=1}^{n} c_j x_j^* \leq r \times OPT$$

Primal-dual schema gives r-approximation algorithm

Primal-dual Method for Vertex Cover

LP formulation for Vertex Cover

Given a graph $G = (V, E)$, $\delta(v) = \{e = (v, u) \in E\}$

Primal: min $\sum_{v \in V} x(v)w(v)$ *

at least one

$$\text{s.t.} \quad \begin{array}{rcll} x(u) + x(v) & \geq & 1 & \forall e = (u, v) \in E \\ x(u) & \geq & 0 & u \in V \end{array}$$

$x(u) \leq 1$ in the fractional relaxation can be omitted

Dual: max $\sum_{e \in E} y(e)$ *

$$\text{s.t.} \quad \begin{array}{rcll} \sum_{e \in \delta(v)} y(e) & \leq & w(v) & \forall v \in V \\ y(e) & \geq & 0 & \forall e \in E \end{array}$$

The Primal-Dual Algorithm for Vertex Cover

1. Increase variable $y(e)$ for an edge $e = (u, v)$ until

$$\sum_{e \in \delta(u)} y(e) = w(u) \left(\texttt{or} \sum_{e \in \delta(v)} y(e) = w(v) \right)$$

2. Set $x(u) = 1$ (or $x(v) = 1$)
3. Remove all edges adjacent to u (or v)

Repeat until all edges are removed

Proof of 2-approximation

Primal complementary slackness condition holds

$$\forall u \in V \quad \texttt{either} \quad x(u) = 0$$
$$\texttt{or} \quad \sum_{e \in \delta(u)} y(e) = w(u)$$

Dual complementary slackness condition is 2-relaxed

$$\forall e \in E \quad \texttt{either} \quad y(e) = 0$$
$$\texttt{or} \quad x(u) + x(v) \leq 2$$

Set Cover via Dual Lifting

Set Cover

Given:

- U: universe of n elements $\{e_1, \ldots, e_n\}$
- $\mathcal{S} = \{S_1, \ldots, S_m\}$: collection of m subsets of U
- $c : S_i \rightarrow \Re^+$: cost function for sets

Goal:

Find a subcollection of minimum cost that covers U

The greedy algorithm achieves an $O(\log n)$ approximation.

Greedy algorithm for Set Cover

- Pick at any iteration the most cost-effective set
- C_i: set of elements yet not covered before set S_i is selected by Greedy
- $c(S)/(C_i \cap S)$: cost-effectiveness of set S

1. $C_0 = U$
2. `While` $C_i \neq \emptyset$ `do`

   Find the set S with min $\alpha = c(S)/(C_i \cap S)$

   Pick set S and $\forall e \in S \cap C_i$, $price(e) = \alpha$

   $C_{i+1} = C_i / S$
3. Output the picked sets

Analysis of Greedy

Assume U covered by Greedy in order $\{e_1, \ldots, e_n\}$

Lemma 5 $price(e_j) \leq \frac{OPT}{n-j+1}$

Proof:

- At any iteration the optimal solution covers C_i at cost at most OPT
- There exists a set of OPT with $\alpha \leq \frac{OPT}{C_i}$
- When e_j is covered at iteration i, $C_i \geq n - j + 1$
- Since e_j is covered by the most cost-effective set:

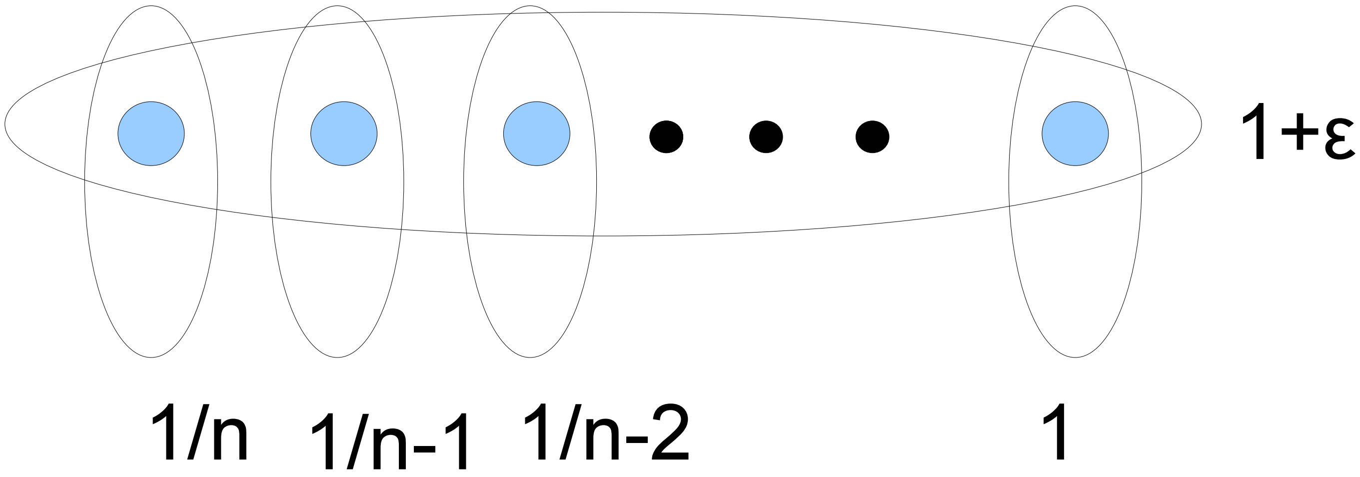
$$price(e_j) \leq \frac{OPT}{C_i} \leq \frac{OPT}{n-j+1}$$

□

Theorem 6

$$ALG = \sum_{j=1}^{n} price(e_j) \leq OPT \times \sum_{j=1}^{n} \frac{1}{n-j+1} = OPT \times H_n$$

A tight example for Greedy



Greedy outputs all singleton sets with cost

$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}$

Optimum cost is $1 + \epsilon$.

Dual-fitting analysis of Greedy Set Cover

Dual-fitting method:

- Show that the integral solution is fully paid by an unfeasible dual solution
- The dual solution can be made feasible by scaling down each variable by a factor f
- $$ALG = DUAL^{unf} = f \times DUAL^{feas} \leq f \times OPT^{LP} \leq f \times OPT$$
- Alternative to argue about complementary slackness conditions

LP formulation for Set Cover

$$\begin{aligned}
\text{Primal:}\quad \min\ & \sum_{S\in\mathcal{S}} c(S)x(S) \\
\text{s.t.}\ & \sum_{S:e\in S} x(S) \geq 1 \quad \forall e \in U \\
& x(S) \geq 0 \quad S \in \mathcal{S}
\end{aligned}$$

Constraint $x(S) \leq 1$ can be omitted

$$\begin{aligned}
\text{Dual:}\quad \max\ & \sum_{e\in U} y(e) \\
\text{s.t.}\ & \sum_{e\in S} y(e) \leq c(S) \quad \forall e \in U \\
& y(e) \geq 0 \quad \forall e \in U
\end{aligned}$$

Analysis of Greedy

Interpret the Greedy solution as an unfeasible dual:

$$y(e) = price(e), ALG = \sum_{e \in U} y(e)$$

Lemma 7 $y'(e) = \frac{price(e)}{H_n}$ *is dual feasible.*

- Consider any set $S = \{e_1, \ldots, e_k\}$ with elements numbered by the order they are covered by Greedy.
- S can cover e_i at price $\leq \frac{c(S)}{k-i+1}$ when i is covered.
- Since Greedy picks the most cost-effective set: $price(e_i) \leq \frac{c(S)}{k-i+1}$
- Dual variables $y'(e) \leq \frac{1}{H_n} \frac{c(S)}{k-i+1}$
- Dual constraint for set S:

$$\sum_{i=1}^{k} y'(e_i) \leq \frac{c(S)}{H_n} \left(\frac{1}{k} + \frac{1}{k-1} + \ldots + 1 \right) = c(S) \frac{H_k}{H_n} \leq c(S)$$

Randomized Rounding for Set Cover

Randomized Runding

- Interpret primal variables in the fractional relaxation as probabilities
- Obtain a primal solution by setting variables to 1 independently with probability equal to the fractional value
- The expected cost of the solution is equal to the fractional optimum ...... but
- Solution may not be feasible
- Repeat as many times as needed to enforce feasibility

Randomized Rounding for Set Cover

$$\begin{array}{llll} \text{min} & \sum_{S \in \mathcal{S}} c(S)x(S) & & \\ \text{s.t.} & \sum_{S: e \in S} x(S) \geq 1 & & \forall e \in U \\ & x(S) \in [0,1] & & S \in \mathcal{S} \end{array}$$

- Pick set S with probability $p(S) = x^*(S)$
- $E[ALG] = \sum_{S \in \mathcal{S}} c(S)p(S) = \sum_{S \in \mathcal{S}} c(S)x^*(S) = OPT^{LP} \leq OPT$
- Is the solution feasible?

Make the solution feasible

- For an element a: $\{S_1, \ldots, S_k\} = \{S \in \mathcal{S} : a \in S\}$

$$\begin{aligned} Pr[a \texttt{ is covered}] &= 1 - (1 - p(S_1)) \times \ldots \times (1 - p(S_k)) \\ &\geq 1 - \left(1 - \frac{1}{k}\right)^k \\ &\geq 1 - \frac{1}{e} \end{aligned}$$

since $p(S_1) + \ldots + p(S_k) \geq 1$

- Each element $a \in U$ is covered with $Pb \geq 1 - \frac{1}{e}$

Make the solution feasible

- Pick $d \log n$ subcollections $C' = C_1 \cup \ldots \cup C_{d \log n}$ with d such that:

$$Pr[a \texttt{ not covered}] \leq \left(\frac{1}{e}\right)^{d \log n} \leq \frac{1}{4n}$$

- $E[COST(C')] \leq d \times \log n \, OPT^{LP}$
- $Pr[COST(C') \geq 4d \times \log n OPT^{LP}] \leq \frac{1}{4}$
- $Pr[C' \texttt{ not easible}] \leq n \times \frac{1}{4n} \leq \frac{1}{4}$
- $Pr[COST(C') \leq 4d \times \log n \, OPT^{LP} \texttt{AND } C' \texttt{feasible}] \geq \frac{1}{2}$
- Expected[number of repetitions] = 2