Linear Program

Set Cover

Set Cover

- *Given* a universe U of n elements, a collection of subsets of U, $S = \{S_1, \ldots, S_k\}$, and a cost function c: $S \rightarrow \mathbf{Q}^+$.
- *Find* a minimum cost subcollection of S that covers all elements of U.

Frequency

- Define the ***frequency*** f_i of an element e_i to be the number of sets it is in.
- Let $f = \max_{i=1,\ldots,n} f_i$.

IP (Set Cover)

$$\text{minimize} \quad \sum_{S \in \Omega} c(S) x_S$$

$$\text{s.t.} \quad \sum_{S : e \in S} x_S \geq 1, \qquad e \in U \quad (1)$$

$$x_S \in \{0, 1\} \qquad S \in \Omega$$

LP-relaxation (Set Cover)

$$\text{minimize} \quad \sum_{S\in\Omega} c(S)x_S$$

$$\text{s.t.} \quad \sum_{S:e\in S} x_S \geq 1, \qquad e\in U \qquad (1)$$

$$x_S \geq 0 \qquad S\in\Omega$$

Algorithm 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.

f-approximation

Theorem 10.1

Algorithm LP-rounding achieves an approximation factor of f for the set cover problem.

Proof.

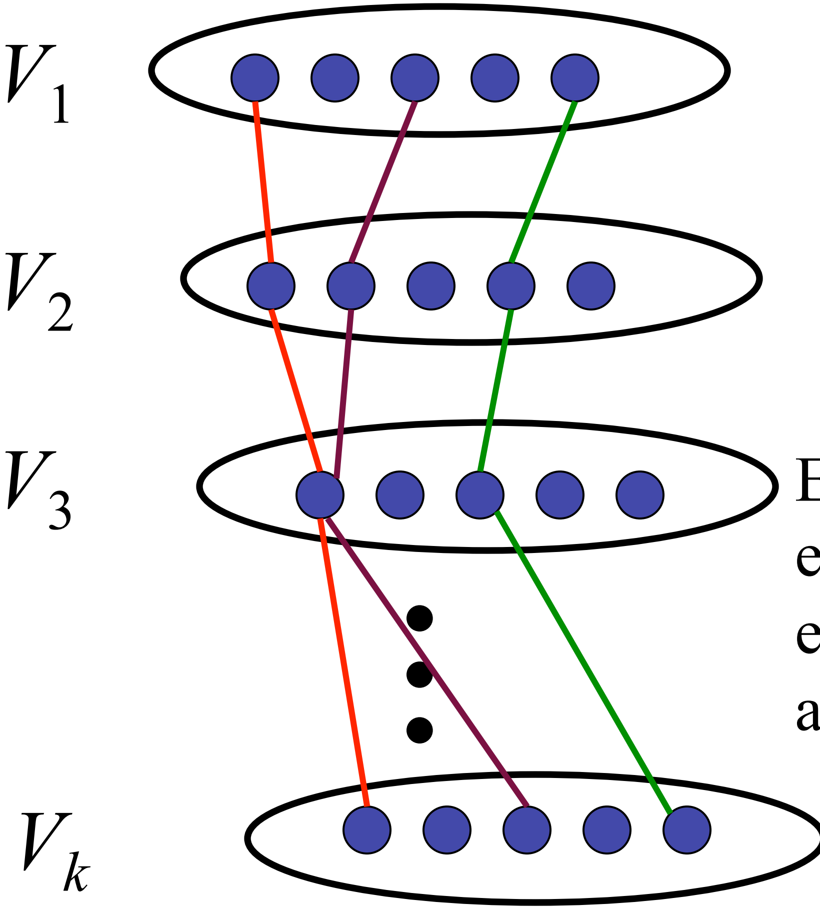
- Consider an arbitrary element e. Each element is in at most f sets.
- $e \in U$: (1) $\Rightarrow \exists\, x_S \geq 1/f\ (e \in S) \Rightarrow$ e is covered.
- We have $x_S \geq 1/f$ for every picked set S. Therefore the cost of the solution is at most f times the cost of fractional cover.

2-approximation

Corollary 10.2

Algorithm LP-rounding achieves an approximation factor of f for the vertex cover problem.

Tight example (hypergraph)



$V = V_1 \cup V_2 \cup \ldots \cup V_k$

n^k hyperedges

Each hyperedge picks one vertex from each V_i. In the set cover instance, elements correspond to hyperedges and sets correspond to vertices.

Primal and Dual programs

$$\sum_{j=1}^{n} c_j x_j \rightarrow \min$$

$$s.t. \quad \sum_{j=1}^{n} a_{ij} x_j \geq b_i, \qquad i = 1, \ldots, m$$

$$x_j \geq 0, \qquad j = 1, \ldots, n$$

$$\sum_{i=1}^{m} b_i y_i \rightarrow \max$$

$$s.t. \quad \sum_{i=1}^{m} a_{ij} y_i \leq c_i, \qquad j = 1, \ldots, n$$

$$y_i \geq 0, \qquad i = 1, \ldots, m$$

The 1-st LP-Duality Theorem

The primal program has finite optimum iff its dual has finite optimum. Moreover, if $x^*=(x_1^*,\ldots,x_n^*)$ and $y^*=(y_1^*,\ldots,y_m^*)$ are optimal solutions for the primal and dual programs, respectively, then

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

Weak Duality Theorem

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

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

- **Proof**. Since $\boldsymbol{y}$ is dual feasible and x_j are nonnegative,

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

- Similarly, since $\boldsymbol{x}$ is primal feasible and y_i are nonnegative,

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

Weak Duality Theorem(2)

- We obtain

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

- By the 1-st LP-Duality theorem, $\boldsymbol{x}$ and $\boldsymbol{y}$ are both optimal solutions iff both inequalities hold with equality. Hence we get the following result about the structure of optimal solutions.

The 2-nd LP-Duality Theorem

- Let ***x*** and ***y*** be primal and dual feasible solutions , respectively. Then, ***x*** and ***y*** are both optimal iff all of the following conditions are satisfied:

$$\forall j:\ x_j\left(c_j - \sum_{i=1}^{m} a_{ij}y_i\right) = 0.$$

$$\forall i:\ y_i\left(\sum_{i=1}^{m} a_{ij}x_j - b_i\right) = 0.$$

Primal-Dual Schema

- The primal-dual schema is the method of choice for designing approximation algorithms since it yields combinatorial algorithms with good approximation factors and good running times.
- We will first present the central ideas behind this schema and then use it to design a simple f factor algorithm for set cover, where f is the frequency of the most frequent element.

Central idea

- Most known approximation algorithms using the primal–dual schema run by ensuring one set of conditions and suitably relaxing the other. In the following description we capture both situations by relaxing both conditions. Eventually, if primal conditions are ensured, we set $\alpha = 1$, and if dual conditions are ensured, we set $\beta = 1$.

Complementary slackness conditions

$$\sum_{j=1}^{n} c_j x_j \to \min$$

$$s.t. \quad \sum_{j=1}^{n} a_{ij} x_j \geq b_i, \qquad i = 1, \ldots, m$$

$$x_j \geq 0, \qquad j = 1, \ldots, n$$

$$\sum_{j=1}^{n} b_i y_i \to \max$$

$$s.t. \quad \sum_{i=1}^{m} a_{ij} y_i \leq c_i, \qquad j = 1, \ldots, n$$

$$y_i \geq 0, \qquad i = 1, \ldots, m$$

Primal complementary slackness conditions

$$(\alpha \geq 1), \forall j, 1 \leq j \leq n : x_j = 0 \vee \frac{c_j}{\alpha} \leq \sum_{i=1}^{m} a_{ij} y_i \leq c_j.$$

Dual complementary slackness conditions

$$(\beta \geq 1), \forall i, 1 \leq i \leq m : y_i = 0 \vee b_i \leq \sum_{i=1}^{m} a_{ij} x_j \leq \beta b_i.$$

- **Proposition 10.3**

  If $\boldsymbol{x}$ and $\boldsymbol{y}$ are primal and dual feasible solutions satisfying the conditions stated above then

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

Proof

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

$$= \alpha \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right) y_i \overset{dual}{\leq} \alpha\beta \sum_{i=1}^{m} b_i y_i .$$

Primal-Dual Scheme

- The algorithm starts with a primal infeasible solution and dual feasible solution; these are usually trivial solutions $\boldsymbol{x} = 0$ and $\boldsymbol{y} = 0$.
- It iteratively improves the feasibility of the primal solution, and the optimality of the dual solution, ensuring that in the end a primal feasible solution is obtained and all conditions stated above, with a suitable choice of α and β are satisfied.
- The primal solution is always extended integrally, thus ensuring that the final solution is integral.
- The improvements to the primal and the dual go hand-in-hand : the current primal solution is used to determine the improvement to the dual, and vice versa.
- Finally, the cost of the dual solution is used as a lower bound on OPT.

LP-relaxation (Set Cover)

$$\text{minimize} \quad \sum_{S \in \Omega} c(S) x_S$$

$$\text{s.t.} \quad \sum_{S : e \in S} x_S \geq 1, \qquad e \in U$$

$$x_S \geq 0 \qquad S \in \Omega$$

Dual Program (Set Cover)

$$\text{maximize} \quad \sum_{e \in U} y_e$$

$$\text{s.t.} \quad \sum_{e : e \in S} y_e \leq c(S), \qquad S \in \Omega$$

$$y_e \geq 0, \qquad e \in U$$

$$\alpha = 1, \beta = f$$

Primal complementary slackness conditions

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

Dual complementary slackness conditions

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

Set will be said to be **tight**, if $\sum_{e:e \in S} y_e = c(S)$.

Pick only tight sets in the cover.

Each element having a nonzero dual value can be covered at most f times.

Primal-Dual Algorithm

0) **Input** $(U, \Omega, c\colon \Omega \rightarrow \mathbf{Q}^{+})$
1) $\boldsymbol{x} \leftarrow 0$, $\boldsymbol{y} \leftarrow 0$.
2) **Until** all elements are covered, **do:**
   - Pick an uncovered element, say e, and raise y_e until some set goes tight.
   - Pick all tight sets in the cover and update $\boldsymbol{x}$.
   - Declare all the elements occurring in these sets as "covered".
3) **Output** $(\boldsymbol{x})$

f-factor approximation

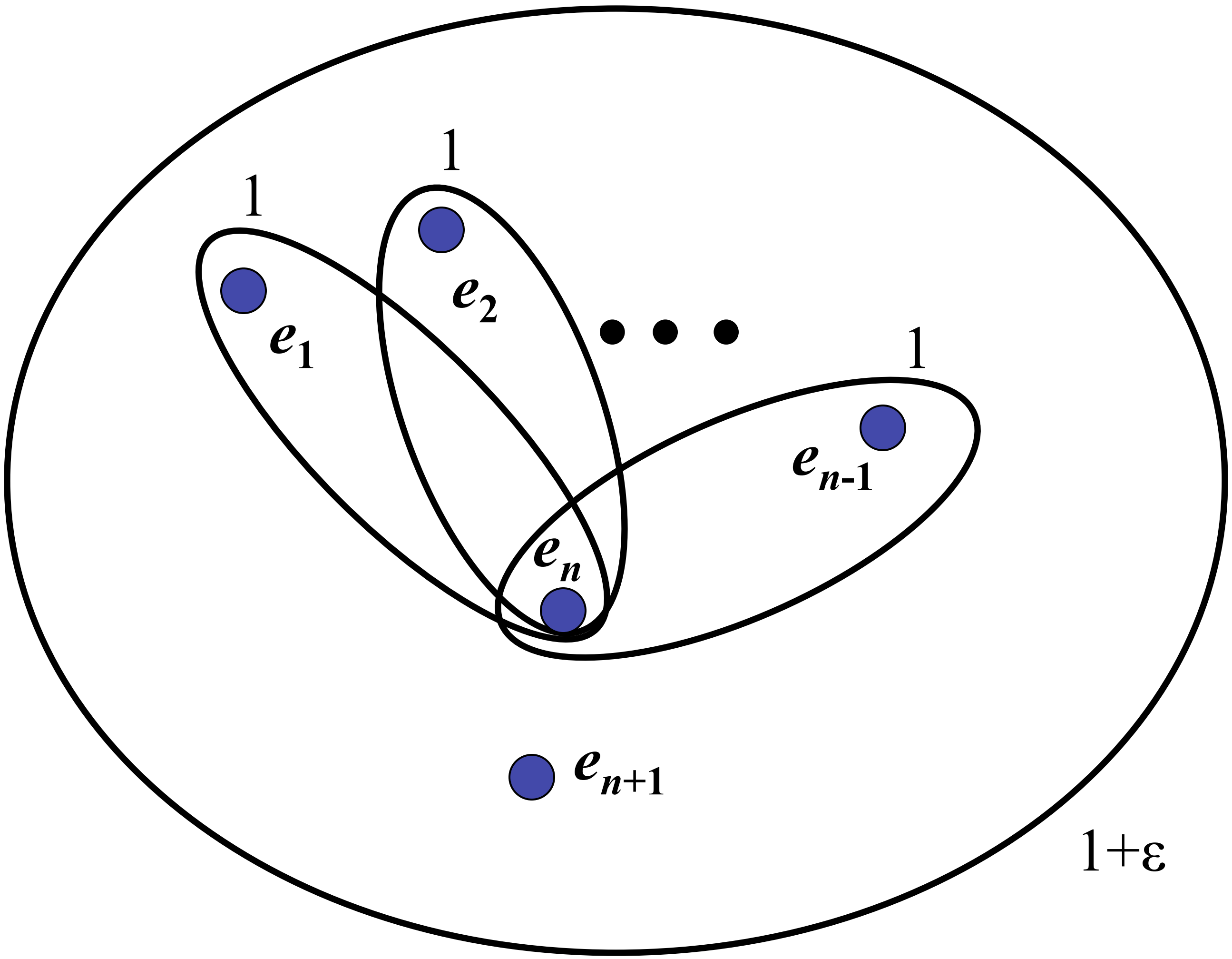
- **Theorem 10.4**

  Primal-Dual Algorithm achieves an approximation factor of f.

Proof

- The algorithm terminates when all elements are covered (feasibility).
- Only tight sets are picked in the cover by the algorithm. Values of y_e in tight sets no longer change (feasibility and primal condition).
- Each element having a nonzero dual value can be covered at most f times (dual condition).
- By proposition 10.3 the approximation factor is f.

Tight example



$f = n$

Vertex cover

- *Given* an undirected graph $G = (V, E)$, and a cost function on vertices c: $V \rightarrow \mathbf{Q}^+$.
- *Find* a minimum cost vertex cover.

IP (Vertex Cover)

$$\text{minimize} \quad \sum_{v \in V} c(S) x_v$$

$$\text{s.t.} \quad x_u + x_v \geq 1 \qquad (u,v) \in E$$

$$x_v \in \{0,\ 1\} \qquad v \in V$$

LP-relaxation (Vertex Cover)

$$\begin{array}{llll} \text{minimize} & & \displaystyle\sum_{v\in V} c(S) x_v & \\ \text{s.t.} & x_u + x_v \geq 1 & & (u,v) \in E \\ & x_v \geq 0 & & v \in V \end{array}$$

Half-integral solution

- Recall that *an extreme point solution* of a set of linear inequalities is a feasible solution that cannot be expressed as convex combination of two other feasible solutions.
- A *half-integral solution* to LP is a feasible solution in which each variable is 0, 1, or 1/2.

Property of extreme points

Lemma 10.5

- Let x be a feasible solution to the LP-relaxation that is not half- integral.
- Then, x is the convex combination of two feasible solutions and is therefore not an extreme point solution for the set of inequalities in LP.

Proof

- Consider the set of vertices for which solution x does not assign half-integral values. Partition this set as follows.

$$V_+ = \left\{ v : \frac{1}{2} < x_v < 1 \right\}, \quad V_- = \left\{ v : 0 < x_v < \frac{1}{2} \right\}.$$

- For ε > 0, define the following two solutions.

$$y_v = \begin{cases} x_v + \varepsilon, & x_v \in V_+ \\ x_v - \varepsilon, & x_v \in V_- \\ x_{v,} & otherwise \end{cases} \qquad z_v = \begin{cases} x_v - \varepsilon, & x_v \in V_+ \\ x_v + \varepsilon, & x_v \in V_- \\ x_{v,} & otherwise \end{cases}$$

Proof (2)

- By assumption, $V_+ \cup V_- \neq \varnothing$, and so x is distinct from y and z.
- Furthermore, x is a convex combination of y and z, since $x = 1/2(y + z)$.
- We will show, by choosing $\varepsilon > 0$ small enough, that y and z are both feasible solutions for LP, thereby establishing the lemma.

Proof (3)

- Ensuring that all coordinates of y and z are nonnegative is easy.
- Suppose $x_u + x_v > 1$. Choose ε small enough, such that y and z do not violate the constraint for such an edge.
- $x_u + x_v = 1$. There are essentially three possibilities for x_u and x_v.
  - $x_u = x_v = 1$;
  - $x_u = 0, x_v = 1$;
  - $u \in V_+, v \in V_-$.
- In all three cases, for any choice of ε,

$$x_u + x_v = y_u + y_v = z_u + z_v = 1.$$

Property of extreme points

Theorem 10.6

- Any extreme point solution for the set of inequalities in LP is half-integral.

- Theorem 10.6 directly leads to a factor 2 approximation algorithm for vertex cover: find an extreme point solution, and pick all vertices that are set to half or one in this solution.

Exercise

- Modify Algorithm LP-rounding so that it picks all sets that are nonzero in the fractional solution. Show that the algorithm also achieves a factor of f.
- Hint: Use the primal complementary slackness conditions to prove this.