CS-GY 6763/CS-UY 3943: Lecture 8
Linear programming and relaxations

NYU Tandon School of Engineering, R. Teal Witter

Given:

- n ground elements $[n] = \{1, 2, \ldots, n\}$ ← water
- m sets $S_1, S_2, \ldots, S_m$ where $S_j \subseteq [n]$ ← sensor
- non-negative weights $w_j \geq 0$ ← unit

for $j \in [m]$.

$$\sum_{j \in C} w_j \leq k \qquad \max \left| \bigcup_{j \in C} S_j \right|$$

Find:

$$\min_{C \subseteq [m]} \sum_{j \in C} w_j \qquad \text{subject to} \qquad \cup_{j \in C} S_j = [n].$$

APPLICATION: COMPUTER VIRUS DETECTION

Given:

- ground elements are known viruses
- sets are three-byte code sequences that occur in viruses so

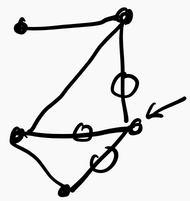
$$S_j = \{\text{viruses that contain } j\text{th three-byte sequence}\}$$

- non-negative weights $w_j \geq 0$

for $j \in [m]$.

Which sequences should we use to identify viruses?

Given:



- ground elements are edges
- sets are nodes so

$$S_j = \{\text{edges adjacent to } j\text{th node}\}$$

- $w_j = 1$

for $j \in [m]$.

What vertices should we choose so that all edges are connected to at least one chosen vertex?

LINEAR PROGRAMMING

Let $\mathbf{x} \in \mathbb{R}^m$ be a vector of decision variables and $\mathbf{c} \in \mathbb{R}^n$ be a vector of constraints. Then the primal linear programs is:

$$\text{Minimize} \quad \underbrace{\mathbf{b}^{\mathsf{T}}}_{1\times m}\underbrace{\mathbf{x}}_{m\times 1} \quad \text{subject to} \quad \underbrace{\mathbf{A}}_{n\times m}\underbrace{\mathbf{x}}_{m\times 1} \geq \underbrace{\mathbf{c}}_{n\times 1}, \quad \mathbf{x} \geq 0 \quad (\textit{Primal})$$

Question: What are the dimensions of $\mathbf{A}$ and $\mathbf{b}$?

LP FOR SET COVER: OBJECTIVE

$x \in \mathbb{R}^m$

Let $x_j = 1$ iff $j \in C$.

The objective is to minimize the sum of weights in C:

$$\min_{C \subseteq [m]} \sum_{j \in C} w_j \quad \Leftrightarrow \quad \min_{\mathbf{x}} \sum_{j=1}^{m} w_j x_j \quad \Leftrightarrow \quad \min_{\mathbf{x}} \mathbf{b}^T \mathbf{x}$$

Question: What is **b**?

$b = \begin{bmatrix} w_1 \\ w_j \\ w_m \end{bmatrix}$

LP FOR SET COVER: CONSTRAINT

$c =$ $\quad\quad$ $\underset{n\times m}{A}x =$

$A_{ij} = 1$ iff $i \in S_j$

The constraint is that C covers the ground elements:

$$[n] = \cup_{j \in C} S_j \quad \Leftrightarrow \quad \sum_{j : i \in S_j}^{m} x_j \geq 1 \;\; \forall i \in [n] \quad \Leftrightarrow \quad \mathbf{A}\mathbf{x} \geq \mathbf{c}$$

Question: What are **A** and **c**?

LP FOR SET COVER: STATEMENT

assume
$x \geq 0$

$$\text{Minimize} \quad \mathbf{b}^{\mathsf{T}}\mathbf{x} \quad \text{subject to} \quad \mathbf{A}\mathbf{x} \geq \mathbf{c}$$

$$\text{Minimize} \quad \sum_{j=1}^{m} w_j x_j \quad \text{subject to} \quad \sum_{j:i \in S_j} x_j \geq 1 \qquad x \in \{0,1\}^m$$

Integer program

We can solve linear program in polynomial time with e.g. Interior Point, Ellipsoid Method.

But there's a problem ...what is it?

Definition (Relaxation)

A linear program (where $\mathbf{x} \in \mathbb{R}^m$) is a *relaxation* of an integer program (where $\mathbf{x} \in \{0,1\}^m$) if

- a feasible solution to the IP is a feasible solution to the LP and
- the value of the feasible solution in the IP has the same value in the LP.

Therefore $OPT_{LP} \leq OPT_{IP} = OPT_{SC}$.

Theorem

Let $\mathbf{x}^*$ *be optimal solution to LP. Define*

sets that contain i

$$f = \max_{i \in [n]} |\{j : i \in S_j\}|.$$

Choose C *so that* $j \in C$ *iff* $x_j^* \geq 1/f$.

Then C *is feasible and gives an* f*-approximation to set cover.*

Question: What approximation do we get for vertex cover?

LP TO SET COVER: PROOF

Claim: C is feasible.

Fix i. We know

$j \in C$ iff $x_j^* \geq 1/f$

$$\sum_{j : i \in S_j} x_j^* \geq 1.$$

Then

$\exists\ j$ so that $x_j^* \geq 1/f_i \geq 1/f$

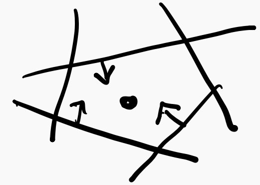
Claim: C gives an f-approximation to set cover.

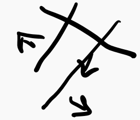
$j \in C$ iff $x_j^* \geq 1/f$

$f x_j^* \geq 1$

our solution to set cover $= \sum_{j \in C} w_j \leq \sum_{j \in C} f w_j x_j^*$

$\leq f \sum_{j=1}^{m} w_j x_j^* = f \cdot OPT_{LP} \leq f \cdot OPT_{SC}$





m $\qquad$ n

Minimize b^Tx subject to $Ax \overset{1}{\geq} c$, $x \geq 0$ (*Primal*)

n $\qquad$ m

Maximize c^Ty subject to $A^Ty \leq b$, $y \geq 0$ (*Dual*)

Minimize b^Tz " " $Az \geq c$ $z = x$

Weak duality: $c^Ty \leq b^Tx$

Proof:

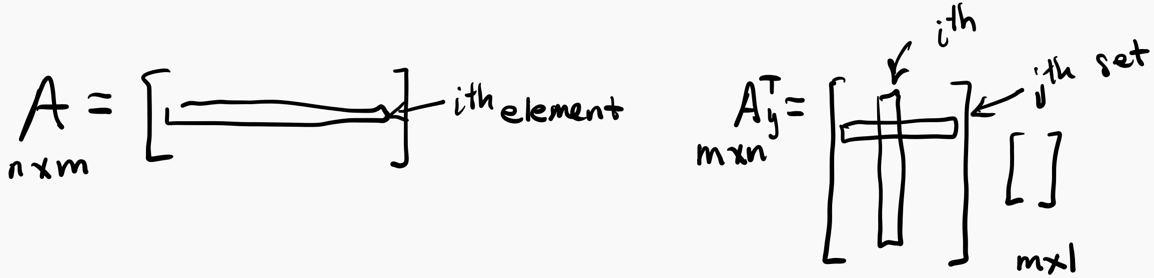
$$c^Ty = y^Tc \overset{1}{\leq} y^TAx = (A^Ty)^Tx = x^TA^Ty \leq x^Tb$$

Note: strong duality holds if primal and dual are bounded and feasible.

DUAL FOR SET COVER

$c = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ $\qquad$ $b = \begin{bmatrix} w_1 \\ w_j \\ w_m \end{bmatrix}$

$$\text{Maximize} \quad \mathbf{c}^T\mathbf{y} \quad \text{subject to} \quad \mathbf{A}^T\mathbf{y} \leq \mathbf{b} \quad (\mathit{Dual})$$



$$\rightarrow \text{Maximize} \quad \sum_{i=1}^{n} y_i \quad \text{subject to} \quad \sum_{i:i\in S_j} y_i \leq w_j \quad \forall j$$

Intuition: y_i represents how much we pay for ground element i.

Theorem

Let $\mathbf{y}^*$ be optimal solution to dual. Choose C' so that $j \in C'$ iff

$$\sum_{i:i \in S_j} y_i^* = w_j.$$

Then C' is feasible and gives an f-approximation to set cover.

$j \in C'$ iff $\sum_{i:i\in S_j} y^*_i = w_j$

Claim: C' is feasible.

Suppose for contradiction there's an uncovered ground element k.

Then for all subsets S_j containing k, we have $\sum_{i:i\in S_j} y_i^* < w_j$.

Define $\epsilon = \min_{j:k\in S_j}(w_j - \sum_{i:i\in S_j y_i^*})$.

Now let $\mathbf{y}'$ be $\mathbf{y}^*$ except with $y'_k = y^*_k + \epsilon$.

Then $\quad \forall j \quad \sum_{i:i\in S_j} y'_i \le w_j$

$$\sum_{i=1}^{n} y'_i = \epsilon + \sum_{i=1}^{n} y^*_i$$

DUAL TO SET COVER: PROOF

$j \in C' \iff \sum_{i : i \in S_j} y_i^* = w_j$

Claim: C' gives an f-approximation to set cover.

our soln to set cover $= \sum_{j \in C'} w_j = \sum_{j \in C'} \sum_{i : i \in S_j} y_i^*$

$\leq \sum_{j=1}^{n} \sum_{i : i \in S_j} y_i^* = \sum_{i=1}^{n} y_i^* \, |\{j : i \in S_j\}|$

$\leq f \sum_{i=1}^{n} y_i^* = f\,OPT_{dual} = f\,OPT_{primal}$

$\leq f \cdot OPT_{SC}$

Question: Is C' better than C? $\iff$ $c' \supseteq c$

constraint

$$\sum_{i=1}^{n} y_i \leq \sum_{i=1}^{n} y_i \sum_{j:i\in S_j} x_j = \sum_{j=1}^{m} x_j \sum_{i:i\in S_j} y_i \leq \sum_{j=1}^{m} x_j w_j$$

OPT_{dual} $\qquad$ OPT_{LP}

Strong duality: the above inequalities are tight for $\mathbf{x}^*$ and $\mathbf{y}^*$.

- If $y_i^* > 0$ then $\sum_{j:i\in S_j} x_j^* = 1$

$j \in C' \iff \sum_{i:i\in S_j} y^*_i = w_j \Rightarrow \exists i'\ y^*_{i'} > 0 \Rightarrow \sum_{j:i'\in S_j} x^*_{j} = 1$

$\exists j'$ s.t. $x^*_{j'} \geq 1/f$

- If $x_j^* > 0$ then $\sum_{i:i\in S_j} y_i^* = w_j$

$j \in C \iff x^*_j \geq 1/f > 0 \Rightarrow \sum_{i:i\in S_j} y^*_i = w_j \iff j \in C'$

PRIMAL VS. DUAL ALGORITHM

We know we can use an LP solver for primal.

Can we do better for dual?