Lecture 11. Applications of duality; Descent method

Zero-sum games: rock-scissors-paper game.

$$\begin{array}{c|ccc} & R & S & P \\ \hline R & 0 & 1 & -1 \\ S & -1 & 0 & 1 \\ P & 1 & -1 & 0 \end{array}$$

$G \in \mathbb{R}^{n\times n}$: payoff matrix.

x, y : strategy distribution over $\{R, S, P\}$.

expected payoff : $\mathbb{E}[\text{payoff}] = \sum_{i,j} G_{ij} x_i y_j$.

$x = y = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. $\mathbb{E}[\text{payoff}] = 0$.

$x = (0, 0, 1)$. $y = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$. $\mathbb{E}[\text{payoff}] = \frac{1}{4}$.

$x = (0, 0, 1)$. $y = (0, 1, 0)$. $\mathbb{E}[\text{payoff}] = -1$.

Player X : for fixed x, player Y's best strategy is to minimize $\sum_{i,j} G_{ij} x_i y_j$. So X's strategy is to max min.

Player Y : for fixed y, player X's best strategy is to maximize $\sum_{i,j} G_{ij} x_i y_j$. So Y's strategy is to min max.

Min max Theorem: $\max_x \min_y x^T G y = \min_y \max_x x^T G y$.

Example. $G = \begin{pmatrix} 3 & -1 \\ -2 & 1 \end{pmatrix}$.

if X first, with strategy (x_1, x_2). payoff of Y is $\begin{pmatrix} 3x_1 - 2x_2 \\ -x_1 + x_2 \end{pmatrix}$

goal of X : $\max_{x_1 + x_2 = 1} \min \{3x_1 - 2x_2, -x_1 + 2x_2\}$

$\max z.$ s.t. $3x_1 - 2x_2 \geq z$. $-x_1 + x_2 \geq z$. $x_1 + x_2 = 1$. $x_1, x_2 \geq 0$

if Y first with strategy (y_1, y_2). payoff of X is $\begin{pmatrix} 3y_1 - y_2 \\ -2y_1 + y_2 \end{pmatrix}$

goal of Y: $\min\limits_{y_1+y_2=1} \max \{3y_1 - y_2, \ -2y_1 + y_2\}$.

$\min w$. s.t. $3y_1 - y_2 \leq w$. $-2y_1 + y_2 \leq w$. $y_1 + y_2 = 1$. $y_1 y_2 \geq 0$.

minmax is the dual of maxmin. so equality by strong duality.

Matching. vertex cover and fractional matching / vertex cover.

matching: a subset of edges. $S \subseteq E$. s.t. $\forall v \in V$.

at most one of incident edges is in S.

vertex cover: a subset of vertices. $T \subseteq V$. s.t. $\forall e \in E$.

at least one of incident vertices is in T.

max matching: $\max \sum x_e$. s.t. $\sum\limits_{e \sim v} x_e \leq 1$. $\forall v \in V$

min vertex cover: $\min \sum y_v$ s.t. $\sum\limits_{v \sim e} y_v \geq 1$. $\forall e \in E$.

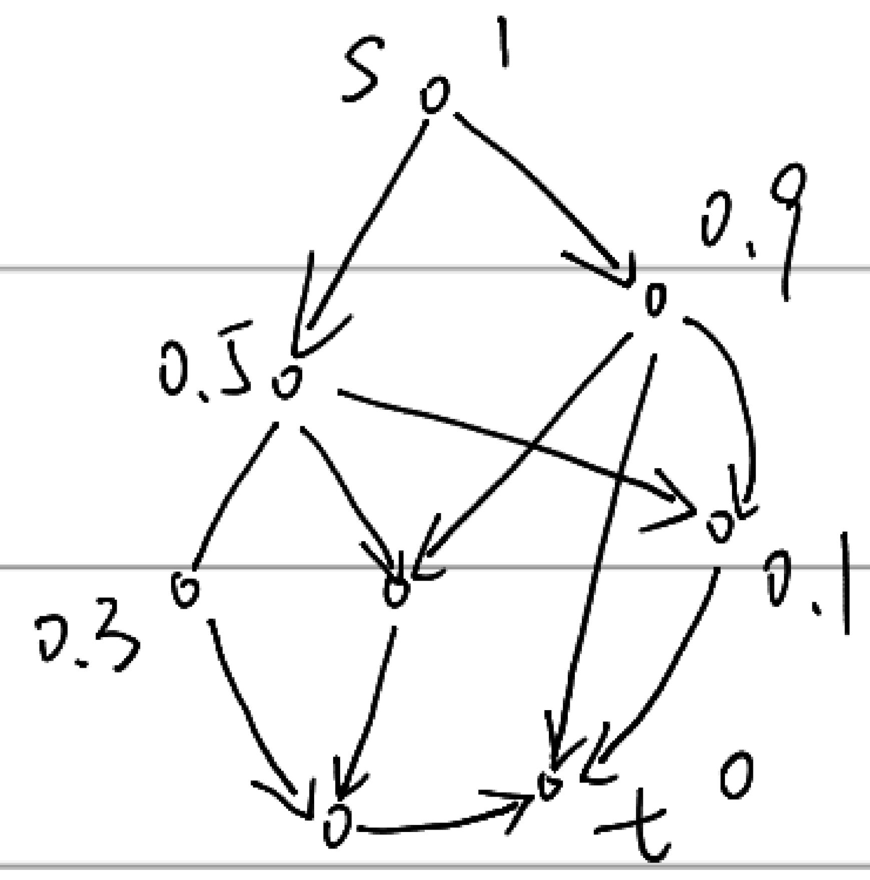
if $x_e, y_v \in \{0, 1\}$. done. otherwise define fractional.

optimal solution at vertex of LP $\Rightarrow$ no need to consider irrational.

Thm: max fractional matching = min fractional vertex cover.

in particular, in bipartite graph. max matching = min vertex cover.

Max flow and min cut: consider DAG for convenience.



DAG having a source and a terminal.

each edge has a capacity. c_e.

for each vertex except s.t. flow-in = flow-out.

max flow: $\max \sum_u x_{su}$. s.t. $0 \le x_{uv} \le c_{uv}$. $\sum_u x_{uv} - \sum_w x_{vw} = 0$. $\forall v$.

cut: select a subset S of vertices. s.t. $s \in S$. $t \notin S$.

$\text{cut}(S, \bar{S}) = \{(u,v) : u \in S, v \in \bar{S}\}$. $\iff$ a subset T of edges E. s.t. $\forall$ path P from s to t. $T \cap P \neq \phi$.

min cut: $\text{minimize}_S \sum_{e \in \text{cut}(S)} c_e$.

dual of max flow: $\min \sum_{u,v} c_{uv} w_{uv}$ s.t. $\begin{cases} w_{su} + y_u \ge 1 \\ w_{vt} - y_v \ge 0. \\ w_{uv} + y_v - y_u \ge 0. \end{cases}$

let $y_s = 1$. $y_t = 0$. then $w_{uv} \ge y_u - y_v$.

if $y_v \in \{0,1\}$. $w_{uv} \in \{0,1\}$ for any u, v. it is min cut.

Select a cut randomly (the probabilistic method).

pick $\rho \in (0,1)$ uniformly at random. let $S = \{v : y_v > \rho\}$.

$\mathbb{E}\left[\sum_{(u,v) \in \text{cut}(S)} c_{uv}\right] = \mathbb{E}\left[\sum_{(u,v)} c_{uv} \cdot \mathbb{1}[(u,v) \in \text{cut}(S)]\right]$

① $\mathbb{1}[P]$ is an indicator random variable that is 1 if P is true and 0 otherwise.

$= \sum_{(u,v)} c_{uv} \cdot \mathbb{E}\left[\mathbb{1}[(u,v) \in \text{cut}(S)]\right]$

② $\mathbb{E}[X] \le y \Rightarrow \exists \omega \in \Omega.\ X(\omega) \le Y$.

$= \sum_{(u,v)} c_{uv} \cdot \Pr((u,v) \in \text{cut}(S)) = \sum_{(u,v)} c_{uv} \cdot (y_u - y_v) \le \sum c_{uv} w_{uv}$. $\square$

Unconstrained convex optimization: gradient descent.

Consider an unconstrained smooth covex optimization: $\min f(x)$.

Recall the first-order condition for optimality: $\nabla f(x^*) = 0$.

Example: $\min_w \|y - Xw\|^2$ $\nabla f(w) = 2(w^T X^T X - y^T X)$.

$\nabla f(w^*) = 0 \Rightarrow w^* = (X^T X)^{-1} X^T y$. how about algorithms?

Consider the unary function $f(x) = x^2$ if at x with $f'(x) = -1$. ?

derivative indicates the reverse direction to the optimal value.

$x \to x - t \cdot f'(x)$ for some t. in general. $x \to x - t \cdot d_x$

Hope: $-d_x$ is a descent direction for some small enough $t > 0$

let $g(t) = f(x - t \cdot d_x) < f(x) = g(0)$. Suppose f differentiable

$\Rightarrow g'(0) = -d_x^T \nabla f(x) \le 0$. is $-d_x^T \nabla f(x) < 0$ sufficient?

For convex f. by the first-order condition. for convexity.

$$f(x) > f(x - t \cdot d_x) \ge f(x) - t \cdot d_x^T \nabla f(x) \Rightarrow d_x^T \nabla f(x) > 0.$$

For convex differentiable f. $-d_x$ descent direction $\Leftrightarrow d_x^T \nabla f(x) > 0$.

Gradient descent: choose $d_x = \nabla f(x)$ $d_x^T \nabla f(x) > 0$ unless $\nabla f(x) = 0$.

In fact, $-\nabla f(x)$ is the max rate descending direction.

w.l.o.g. assume $\|d_x\| = 1$. consider the derivative at direction d_x.

$$\nabla_{d_x} f(x) = \lim_{t \downarrow 0} \frac{1}{t}\left(f(x+td_x) - f(x)\right) = d_x^T \nabla f(x).$$

By Cauchy-Schwarz. $d_x^T \cdot \nabla f(x) = \langle d_x, \nabla f(x) \rangle \le \|d_x\| \cdot \|\nabla f(x)\|$

with equality iff $d_x = (\pm) \nabla f(x) / \|\nabla f(x)\|$. $\ge -\|d_x\| \cdot \|\nabla f(x)\|$.

Gradient descent algorithm: initialize from a starting point $x \in \mathbb{R}^n$

repeat. $x \leftarrow x - t \nabla f(x)$ until some stopping criterion.

Two questions: 1. how to choose t? 2. how to determine stopping time?

Ideal stopping criterion: $\nabla f(x) = 0$. stopping at optimal.

Practical: $\|\nabla f(x)\| < \delta$. $|f(new) - f(old)| < \delta$ or $\delta f(old)$.

can also stop if maximum # of iterations is reached.

more difficult: choose t, or in general. choose step size.

two constraints: whether converge / rate of convergence.