Lecture 4 Goemans & Williamson

Tool #1 Semidefinite Programs

Lemma 1: The set of PSD matrices forms a convex cone.

Proof: Plug in definition for $\lambda A + \lambda' A' \succeq 0$

Functional form:

"Primal" $\min_X \operatorname{tr} CX$

s.t. $\operatorname{tr} A_i X \leq b_i \quad \forall i$

$X \succeq 0$

"Dual" $\min c^T x$

s.t. $\sum_i x_i M_i \succeq M_0$

Convert via Lagrange function

Detour #2 Second order Cone progr.

→ constraint set of the form

$\|Ax + b\|_2 \leq c^T x + d$

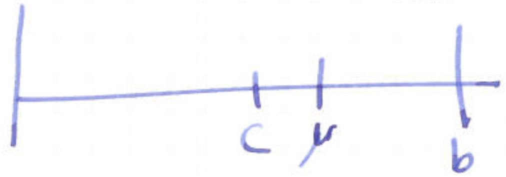
Can define thus via ellipsoid

Randomized Algorithm:

$x \sim p(x)$; $\mathbb{E}[x] = \mu$
then $\Pr\{x > c\} \leq \frac{\mu}{c}$ } Markov ineq.

So if we want algorithm to achieve $\geq \mu$ start with $x \leq b$ for some $b > \mu$

~~then $\Pr\{x \ldots\} \geq \frac{b-\mu}{b-\mu c}$~~

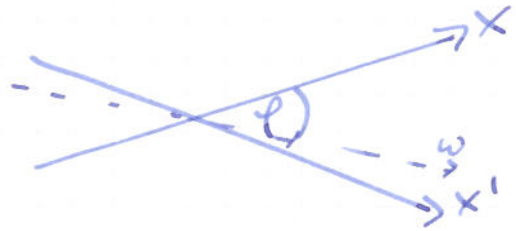


$\Pr\{x < c\} \leq \frac{b-\mu}{b-c}$

$\Rightarrow \Pr\{x \geq c\} \geq \frac{\mu - c}{b - c} > 0$ ✓

how we only need to try often enough to get a sufficiently good result

Grothendieck identity:



$\theta = \arccos(\langle x, x' \rangle)$

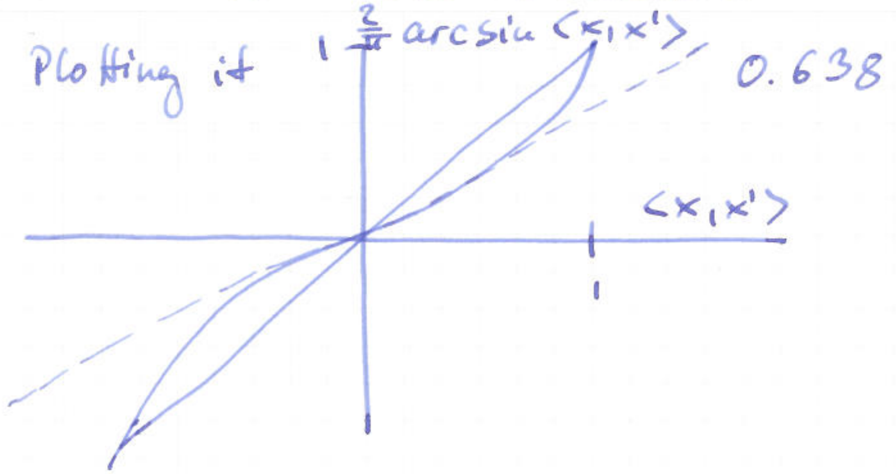
$\Pr_w\{\operatorname{sgn}\langle x, w\rangle \operatorname{sgn}\langle x', w\rangle > 0\} = 1 - \frac{\theta}{\pi}$

$\Rightarrow \mathbb{E}_w[\operatorname{sgn}\langle x, w\rangle \operatorname{sgn}\langle x', w\rangle] = 1 - \frac{2}{\pi}\theta = \frac{2}{\pi}\arcsin\langle x, x'\rangle$

$\Rightarrow \mathbb{E}_w[1 - \operatorname{sgn}\langle\rangle \operatorname{sgn}\langle\rangle] = \frac{2}{\pi}\arccos\langle x, x'\rangle$

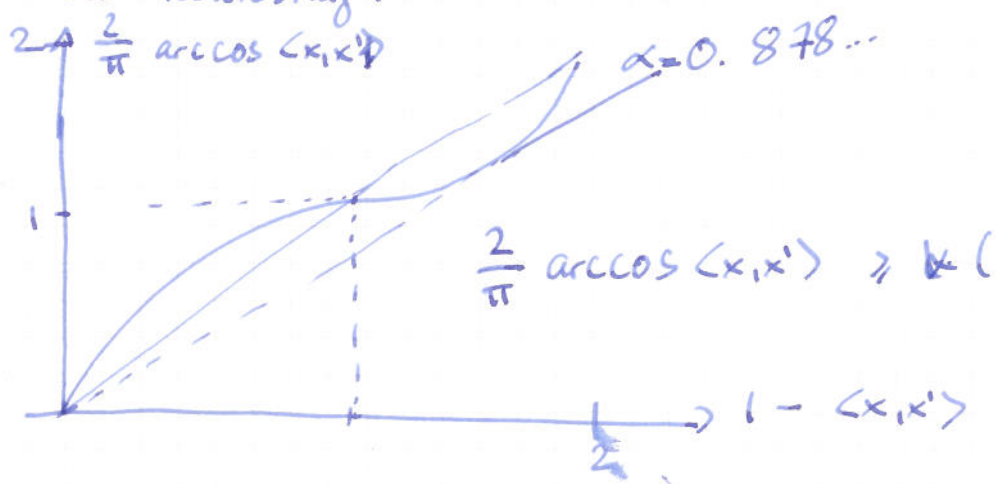
Lecture 4 Goemans & Williamson

Plotting it



→ this means that arcsin is good approx of id

more interesting:



$\frac{2}{\pi}\arccos\langle x,x'\rangle \geq \alpha(1-\langle x,x'\rangle)$

more general:

$\operatorname{sgn}\langle x,w\rangle \longrightarrow \frac{Wx}{\|Wx\|_2}$ for $W\in\mathbb{R}^{p\times d}$

[this gives lower-dimensional approximation for the original problem

MAXCUT problem

graph $G(V,E)$ with edge-weights $A_{ij}\geq 0$
find a cut Y such that

$$\sum_{ij}\underbrace{\tfrac{1}{2}(1-y_iy_j)}_{\substack{0 \text{ if } y_i=y_j\\ 1 \text{ otherwise}}} A_{ij} \quad \text{is maximized}$$

$$\max_Y \sum_{ij}\tfrac{1}{2}(1-y_iy_j)A_{ij} \quad \text{s.t. } y_i\in\{\pm 1\}$$

$$\Leftrightarrow \tfrac{1}{2}\sum_{ij}A_{ij} - \tfrac{1}{2}\min_Y \operatorname{tr}(YY^t)A \quad \text{s.t. } y_i\in\{\pm 1\}$$

$$\Leftrightarrow \tfrac{1}{2}\sum_{ij}A_{ij} - \tfrac{1}{2}\min_K \operatorname{tr} K\cdot A \quad \text{s.t. } K_{ii}=1,\ \operatorname{rank} K = 1,\ K\succeq 0$$

relaxation: drop the rank-1 constraint

Algorithm

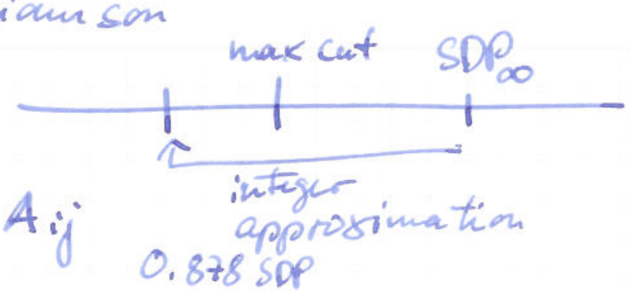
a) $\min_K$ imize $\operatorname{tr} K\cdot A$ s.t. $K_{ii}=1$; $K\succeq 0$

b) cholesky decomposition $VV^t = K$

c) compute $Y = \operatorname{sgn} V\cdot w$ for w isotropic

(if needed, draw several w and pick the best) lecture 4 Goemans & Williamson

Analysis



$$\max_{y \in \{\pm 1\}^n} \sum_{ij} \frac{1}{2}(1 - y_i y_j) A_{ij}$$

$$\leq \max_{\substack{K \succeq 0 \\ k_{ii} = 1}} \sum_{ij} \frac{1}{2}(1 - k_{ij}) A_{ij}$$

$$\leq \frac{1}{0.878} \mathbb{E}_{\tilde{y}} \sum_{ij} \frac{1}{2}(1 - \tilde{y}_i \tilde{y}_j) A_{ij}$$

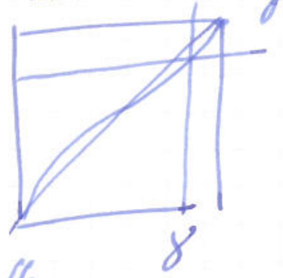
$$\tilde{y}_i = [V \cdot w]_i$$

Modifications:

a) y_i, y_j in the same class $\Rightarrow k_{ij} = 1$

b) y_i, y_j in opposite classes $\Rightarrow k_{ij} = -1$

Tighter bound:



if maximum cut is large then the guarantee is also better

define $$A := \frac{\frac{1}{2}\sum_{ij}(1 - k_{ij}) W_{ij}}{\sum_{i \neq j} W_{ij}}$$

if $A > \gamma$ the bound $\geq \frac{\arccos(1 - 2A)}{A\pi}$

Proof sketch: use the fact that [this gets better] ③

(details in G&W)

Negative weights $A_{ij} < 0$

Lemma:

$$\mathbb{E}\left[\frac{1}{2}\sum_{ij} A_{ij}[1 - y_i y_j] - A_-\right] \geq \alpha \frac{1}{2}\left[\sum_{ij} A_{ij}(1 - k_{ij}) - A_-\right]$$

here $$A_- = \frac{1}{2}\sum_{\substack{ij : i \neq j \\ A_{ij} < 0}} A_{ij}$$

Proof:

LHS $= \frac{1}{2}\mathbb{E}\left[\sum_{ij : A_{ij} > 0} A_{ij}[1 - y_i y_j]\right]$

$+ \frac{1}{2}\mathbb{E} \sum_{ij : A_{ij} < 0} |A_{ij}| [y_i y_j]$

RHS: ditto $\frac{1}{2}\sum_{ij : A_{ij} > 0} A_{ij}[1 - k_{ij}]$

$+ \frac{1}{2}\sum_{ij\, A_{ij} < 0} |A_{ij}| [1 + k_{ij}]$

next bound $\mathbb{E}[y_i y_j] \geq \alpha [1 + k_{ij}]$

by change of variables $\frac{2}{\pi} \arcsin t \geq \alpha [1 + t]$

Theorem: $$\mathbb{E}\left[\frac{1}{2}\sum_{ij} A_{ij}[1 - y_i y_j]\right] \geq \alpha \frac{1}{2}\sum_{ij} A_{ij}(1 - k_{ij}) + (1 - \alpha) A_-$$

(small contribution)

lecture 4

Research question 1: learning to cut

→ assume that we have vertex attributes $\ell(i)$

→ want to learn $y(i)$ such that

$$\mathbb{E}_{i,j^*} \left[1 - y(i) y(j) \right] A_{ij}$$

is large, even for unseen A_{ij} and (i,j)

→ Ansatz $y(i) = \mathrm{sgn}(\langle \ell(i), u \rangle)$

where $u = Vw$

estimate $K = \ell(X)^+ \tilde{K} \ell(X)$ for $\mathrm{tr}\, A \left[\ell(x)^{+} \tilde{K} \ell(x)\right]$

factorize $\tilde{K} = VV^+$ and then use $\ell(X) V$ in lieu of V

Lemma: the same guarantees hold relative to the SDP solution but no guarantee of optimality w.r.t. full K (unless same subspace?)

Research question 2: kernel target alignment (HSIC / KTA)

given matrices K, L with $\mathrm{tr} KL > C$

find binary rank matrix YY^+ for $\mathrm{sgn}(V \cdot w) = Y$

to get large $\mathrm{tr}\, YY^+ L$ => all rank classifiers are good?

→ deal with neg. weights

→ complexity for rank hyperplanes / boosting / HSIC => good perf.

Max 2SAT: $\max_y \sum_{ij} w_{ij}\, \mathrm{bool}(y_i, y_j) + \sum_i w_i\, \mathrm{bool}(y_i)$

add y_0 ~~as~~ to break symmetry

y_i TRUE => $\frac{1 + y_0 y_i}{2}$

y_i FALSE => $\frac{1 - y_0 y_i}{2}$

$y_i = y_j$ => $\frac{(1 + y_i y_j)}{2}$

y_i AND y_j = $\frac{(1 + y_i y_0)(1 + y_j y_0)}{4}$

$= \frac{1}{4}(1 + y_i y_0 + y_j y_0 + y_i y_j)$

y_i OR $y_j = 1 - \frac{(1 - y_i y_0)(1 - y_j y_0)}{2}$

y_i XOR y_j = y_i AND NOT y_j + y_j AND NOT y_i ...

$$\max \sum_{ij} a_{ij} \underbrace{(1 - y_i y_j)} + b_{ij}(1 + y_i y_j)$$

$a_{ij} \frac{2}{\pi} \arccos \langle x_i, x_j \rangle + b_{ij} \frac{2}{\pi} \arcsin \langle x_i, x_j \rangle$

α-approx of MAX 2-sat

MAX SAT see G&W