Lec 1. SoS basics: from proof system to optimization.

(I) Setting:

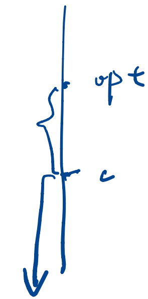
- $f: \{0,1\}^n \mapsto \mathbb{R}$
  - Fact: f can be written as a polynomial with $\deg \le n$

    eg $\quad f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \prod_{i \in S} x_i$

- $\forall$ polyn $p(x)$, $x \in \{0,1\}^n$, $\quad p(x) = \langle v_p, (1,x)^{\otimes d} \rangle$, $v_p \in \mathbb{R}^{n^{O(d)}}$
  - $(1,x)^{\otimes d}$: e.g. $x = (x_1, x_2, x_3)$. $(1,x)^{\otimes 2} = (1,x) \otimes (1,x)$
  - $\dim = n^{O(d)}$. contain all $\deg \le d$ monomial in x_i.

    $$p(x) = 1 + x_1 x_2 + 2x_1^2 + 2x_2^2 = \langle \cdots \rangle \begin{pmatrix} x_1 \\ x_1 x_2 \\ \vdots \\ \vdots \end{pmatrix}$$

(II) Optimization $\Longrightarrow$ Certification



$$\min_{x\in\{0,1\}^n} \quad f(x)$$

- Cert version: decide if $opt > c$, for given $c \in \mathbb{R}$.
- SoS cert: $\{g_i\}_{i=1}^m$ $\quad f - c = g_1^2 + g_2^2 + \dots + g_m^2$

$\Rightarrow \quad f - c \geq 0 \equiv opt \geq c$

deg-d SoS

$\Rightarrow \deg(g_i^2) \leq d$

- size of SoS cert: $\sim \deg(g_i)$

$$g = \sum \hat{g} \prod x_i = \langle v_p, \underbrace{(1,x)^{\otimes d}}_{n^{O(d)}} \rangle$$

- $d = O(1)$, $O(\log)$, $O(n^{\varepsilon})$, $o(n)$

Thm 1: If $f-c \geq 0$ has deg-d SoS cert, then it can be found in $n^{O(d)}$ time.

Pf sketch: $f-c = g_1^2 + \cdots + g_m^2$. s.t. $\deg(g_i) \leq d \ \forall i$.

$$g_i^2 = \langle v_i, (1,x)^{\otimes d/2} \rangle^2 = \left((1,x)^{\otimes \frac{d}{2}}\right)^T v_i v_i^T (1,x)^{\otimes d/2}$$

$$\Rightarrow \quad f-c = \left((1,x)^{\otimes d/2}\right)^T \underline{\left(\sum_{i=1}^m v_i v_i^T\right)} (1,x)^{\otimes d/2}$$

$\exists$ psd K s.t. $f-c = \langle x, Kx \rangle$

SDP: solve for K
$$\begin{cases} K \succeq 0 \\ \text{match coeff of } f-c \text{ : linear constraint} \end{cases}$$

$\square$

Exmp: $f = x_1^2 + x_2^2 - 2x_1x_2$

$$\begin{cases} x_1^2 + x_2^2 - 2x_1x_2 = \langle (1, x_1, x_2),\ K(1, x_1, x_2) \rangle \\ K \succeq 0 \end{cases}$$

RHS

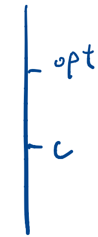
$$K = \begin{pmatrix} K_{00} & K_{01} & K_{02} \\ \ddots & K_{11} & K_{12} \\ & \ddots & K_{22} \end{pmatrix}$$

$LHS = x_1^2 + x_2^2 - 2x_1x_2$

$RHS = K_{00} + 2K_{01}x_1 + 2K_{02}x_2 + K_{11}x_1x_2$

$$\begin{cases} K_{00} = 0 \\ K_{11} = -2 \\ K_{.} = \cdots \end{cases}$$

(II) Back to optimization

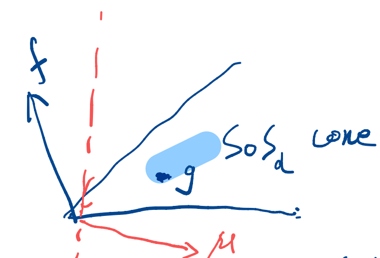


$\min_{x} f(x) \Rightarrow \max_{c \in \mathbb{R}} c$ s.t. $f(x) - c$ has SoS_d cert   opt1

$\Leftarrow\Rightarrow \min_{c \in \mathbb{R}} c$ s.t. $f(x) - c$ has no SoS_d cert   opt2

- What is "no SoS_d cert"?
  - View f as a vector
    - $\langle f, g \rangle = \frac{1}{2^n} \sum_x f(x) g(x)$
  - $f \notin \text{cone}(SoS_d) \Rightarrow \exists \mu$ separate f

$$\begin{cases} \langle \mu, f \rangle < 0 \\ \langle \mu, g \rangle \geq 0 \quad \forall g \in \text{cone}(SoS_d) \end{cases}$$



$\forall g \in \text{cone}(SoS_d)$

$g = \sum_i \lambda_i P_i(x)$ $i \in n^{odd}$

- $\lambda_i \geq 0$
- P_i : SoS_d polyn

$\cdot\ \langle \mu, 1 \rangle = 1$

Def: μ is called deg-d pseudo-distribution.

$\tilde{\mathbb{E}}_\mu[f] := \langle \mu, f \rangle$ $\quad \tilde{\mathbb{E}}$: pseudo-expectation

Fact: f has no SoS_d cert $\iff \exists$ deg-d μ. $\tilde{\mathbb{E}}_\mu[f] \le 0$

$$\min_{c \in \mathbb{R}} \ c \quad \text{s.t.} \quad \underline{f(x) - c \text{ has no } SoS_d \text{ cert}}$$

$$\iff \exists \mu \text{ s.t. } \langle \mu, f - c \rangle < 0$$

$$\Rightarrow \quad \min_{\mu, c} \ c$$

$$\text{s.t.} \quad \underbrace{\langle \mu, f - c \rangle \le 0}_{= \tilde{\mathbb{E}}_\mu[f] \le c}, \quad \mu \text{ is pseudo-distr}$$

$\Rightarrow \quad c^* = \min_{\mu} \tilde{\mathbb{E}}_\mu[f]$

$$\min_{x \in \{0,1\}^n} f(x) \xrightarrow{\text{SoS}_d \text{ relaxation}} \min_{\mu} \tilde{\mathbb{E}}_\mu[f]$$

s.t. μ is deg-d pseudo-distr

- $n^{O(d)}$ time solvable. (in most interesting cases)
- Rounding : $\mu \xrightarrow{\text{extract}} x \in \{0,1\}^n$

  $\mu : \{0,1\}^n \mapsto \mathbb{R}$

- What about constrained optimization ?

$$\min f(x) \quad \text{s.t.} \quad P_1(x) \geq 0, \ \ldots, \ P_m(x) \geq 0$$
$$g_1(x) = 0, \ \ldots, \ g_k(x) = 0$$

- SoS cert for opt $\geq c$: a set of deg-$\frac{d}{2}$ polyns $\{g_i\}$ s.t.

$$f - c = g_0^2 + \sum_{i=1}^{m} p_i g_i^2 + \sum_i q_i g_i^2$$

- $$\min_{x \in \{0,1\}} f(x) \quad \text{s.t.} \quad p(x) \geq 0, \quad q(x) = 0 \quad \xRightarrow{SoS_d\text{-relax}} \quad \min_{\mu} \tilde{\mathbb{E}}_\mu[f] \quad \text{s.t.} \quad \tilde{\mathbb{E}}_\mu[p(x) g^2(x)] \geq 0. \ \ \forall \deg(g^2) + \deg(p) \leq d. \quad \tilde{\mathbb{E}}_\mu[q(x) g^2(x)] = 0.$$

- "$\{f \geq c\}$ is SoS-deduced from axiom $\{P_i \geq 0 : i \in [m]\}$"

Q: what if the feasible region is $\emptyset$, e.g. $P_1(x) = x_1^2 - 4$?

Ans: Sometimes SoS is unable to tell: The 3XOR SoS LB.

Def (3XOR) : $\left\{ \begin{array}{l} x_i \oplus x_j \oplus x_k = a_{ijk} \\ \quad\quad \vdots \end{array} \right.$ $x \in \{0,1\}^n$, $a_{ijk} \in \{0,1\}$

$$\max_{x \in \{\pm 1\}^n} \sum_{ijk} x_i x_j x_k \cdot a_{ijk}$$

Known: [Håstad]

$(\frac{1}{2}+\delta, 1-\delta)$-apx is NP-hard

Thm (SoS LB) $\exists$ 3XOR inst φ. $\mathrm{opt}(\varphi) \leq \frac{1}{2}+\varepsilon$, $\mathrm{SoS}_{O(n)}(\varphi) \geq 1-\varepsilon$, i.e.

$$\max_{\mu} \tilde{\mathbb{E}}_{\mu}\left[\sum x_i x_j x_k a_{ijk}\right] \geq 1-\varepsilon$$

(III) SoS as proof system

- Certification: proposition: $f(x) \geq 0$

  $n^{O(d)}$-time solvable
  $$\begin{cases} \text{proof}: & f = g_1^2 + \cdots + g_2^2 \\ \text{refutation}: & cf + g_1^2 + \cdots + g_2^2 = -1, \quad c \geq 0 \end{cases}$$

- $\{f \geq 0\} \vdash_d \{g \geq h\}$: $g - h = f \sum_{i=1}^{m} g_i^2 \quad (\deg(g_i^2) + \deg(f) \leq d)$

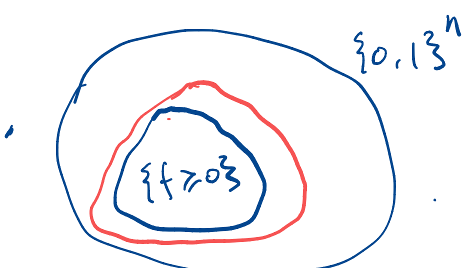
  f, g, h : polyn

- "Imply ($\vdash_d$)" : $\forall \mu \in SoS_d$ relax of $\{f \geq 0\}$

  there's $\tilde{\mathbb{E}}_{\mu}[g(x) - h(x)] \geq 0$

  $f : \{0,1\}^n \to \mathbb{R}$

  $x \in \{0,1\}^n$

- 

Rounding paradigm

- μ : "pretend" μ is a real distr over $\{x : f(x) \geq 0\}$
- Try to round/sample μ to a integral $x \in \{0,1\}^n$
  - randomized rounding $\Rightarrow$ feasible x w.p. $p(x)$ & $p(x) \geq \frac{1}{2}$.
  - If further : $\{f(x) \geq 0\} \vdash_d \{p(x) \geq \frac{1}{2}\}$. Then pseudo-distr suffice.

Exmp: $\max\limits_{x \in \{0,1\}^n} \underbrace{\sum\limits_{(i,j) \in E} (x_i - x_j)^2}_{f(x)}$

$\Rightarrow \max \ 1.$

s.t. $\sum\limits_{i,j} (x_i - x_j)^2 > c$ $\overset{SoS_2}{\Longrightarrow}$ $\mu : \{0,1\}^n \mapsto \mathbb{R}$

pretend μ is a distr over x s.t. $f(x) > c$

Rounding:

- $\tilde{\mathbb{E}}_\mu[x_i] = s_i$, $\tilde{\mathbb{E}}_\mu[x_i x_j] = \sigma_{ij}$

- $\xi_{ij} \sim \mathcal{N}\left(\begin{pmatrix} s_i \\ s_j \end{pmatrix}, \begin{pmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ij} & \sigma_{jj} \end{pmatrix}\right)$, $\xi_{ij} \in \mathbb{R}^2$, $(i,j) \in E$

- random $\alpha \in \mathbb{R}^n$,

Lec II: SoS_2 for Max Cut.

Prob: Given $G=(V=[n],E)$. $\max_{x\in\{0,1\}^n} \frac{1}{|E|}\sum_{(i,j)\in E}(x_i-x_j)^2 = f(x)$

Thm 1: $\forall$ Max Cut inst and SoS_2 sol μ, one can find a real distr μ' (in poly-time) s.t. $\mathbb{E}_{\mu'} f(x) \geq 0.878\, \tilde{\mathbb{E}}_{\mu}[f(x)]$

Conjecture (UGC): achieving $(0.878+\varepsilon)$-apx is NP-hard.

$$= \min_{\rho\in[-1,1]} \frac{2\arccos\rho}{1-\rho}$$

$\equiv$ achieving $(0.878+\varepsilon)$-apx need $\Omega(n)$-deg SoS

Known: $\Omega(\log n)$-deg SoS.

Lmm 1: $\forall$ deg ≥ 2 pseudo-distr μ on $\{0,1\}^n$, $\exists$ Gaussian μ' on $\mathbb{R}^n$ matching the 1st and 2nd moment of μ.

$$\tilde{\mathbb{E}}_{\mu}[x_i] = \mathbb{E}_{\mu'}[x_i] \quad . \quad \forall\, i, j.$$

$$\tilde{\mathbb{E}}_{\mu}[x_i x_j] = \mathbb{E}_{\mu'}[x_i x_j]$$

W.l.o.g. assume $\tilde{\mathbb{E}}_\mu[x_i] = \frac{1}{2}, \ \forall i \in [n]$.

- If otherwise, let $\mu_0 = \frac{1}{2}\mu(x) + \frac{1}{2}\mu(\mathbb{1} - x)$, then $\tilde{\mathbb{E}}_{\mu_0}[x_i] = \frac{1}{2}$ and $\tilde{\mathbb{E}}_{\mu_0}[f] = \tilde{\mathbb{E}}_\mu[f]$: since $f(x) = f(\mathbb{1} - x)$

Alg:

- By thm 1, we have $g \in \mathbb{R}^n$, $g \sim \mathcal{N}(\frac{1}{2}\cdot\mathbb{1}_n, \tilde{\Sigma})$
  where $\tilde{\Sigma} = \tilde{\mathbb{E}}_\mu (x - \frac{1}{2}\mathbb{1}_n)(x - \frac{1}{2}\mathbb{1}_n)^T$
- Out $\hat{x} \in \{0,1\}^n$, $\hat{x}_i = \mathbb{1}[g_i > \frac{1}{2}]$ $\rightarrow \mu'$

Thm 1: $\mathop{\mathbb{E}}_{(i,j)\sim E} \mathop{\mathbb{E}}_{g} (\hat{x}_i - \hat{x}_j)^2 \geq 0.878 \mathop{\mathbb{E}}_{(i,j)\sim E} \tilde{\mathbb{E}}_\mu [(x_i - x_j)^2]$

Pf: fix $(i,j) \in E$.

$$\mathbb{E}_g (\hat{x}_i - \hat{x}_j)^2 = \Pr[(g_i > \tfrac{1}{2} \wedge g_j \le \tfrac{1}{2}) \text{ or } (g_i \le \tfrac{1}{2} \wedge g_j > \tfrac{1}{2})]$$

$$= \Pr[(g_i - \tfrac{1}{2})(g_j - \tfrac{1}{2}) \le 0]$$

(Let $\xi_i = 2g_i - 1$) $= \Pr[\xi_i \xi_j < 0]$

$\xi \sim \mathcal{N}(0, 4\tilde{\Sigma})$

- $\Pr[\xi_i \xi_j < 0]$. $\Rightarrow$ distr of (ξ_i, ξ_j)

$$(\xi_i, \xi_j) \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, 4\begin{pmatrix}\tilde{\Sigma}_{ii} & \tilde{\Sigma}_{ij} \\ \tilde{\Sigma}_{ij} & \tilde{\Sigma}_{jj}\end{pmatrix}\right)$$

- $\tilde{\Sigma}_{ii} = \tilde{\mathbb{E}}_\mu[(x_i - \frac{1}{2})^2] = \tilde{\mathbb{E}}_\mu x_i^2 - (\tilde{\mathbb{E}}_\mu x_i)^2 = \frac{1}{4}$

$\tilde{\Sigma}_{ij} = \tilde{\mathbb{E}}_\mu[(x_i - \frac{1}{2})(x_j - \frac{1}{2})] = \tilde{\mathbb{E}} x_i x_j - \frac{1}{4}$

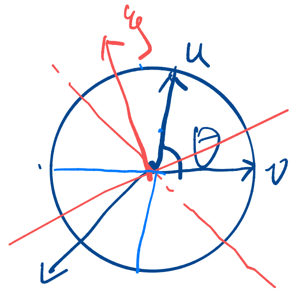
$\Rightarrow \quad (\xi_i, \xi_j) \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{pmatrix}\right)$ (Let $\rho_{ij} = 4\tilde{\Sigma}_{ij}$ $4\tilde{\mathbb{E}} x_i x_j - 1$)

(ρ-corr Gaussian)

- How to sample from

$\Rightarrow$ ① fix $u, v \in \mathbb{R}^2$, $\|u\| = \|v\| = 1$, $\langle u, v \rangle = \rho_{ij}$

② pick $\xi \sim \mathcal{N}(0, I_2)$. output $\xi_i = \langle \xi, u \rangle$. $\xi_j = \langle \xi, v \rangle$



$\theta = \arccos \rho$

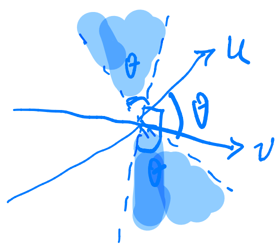
$\xi \sim N(0, I_2)$

claim: $(\xi_i, \xi_j) \sim N\left(0, \begin{pmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{pmatrix}\right)$

- Hyperplane rounding:

$$\Pr[\xi_i \xi_j < 0]$$

$$= \Pr[\langle \xi, u\rangle \langle \xi, v\rangle < 0] = \frac{\theta}{\pi}$$



$$\mathbb{E}_g (\hat{x}_i - \hat{x}_j)^2 = \frac{\theta}{\pi} = \frac{\arccos \rho_{ij}}{\pi}$$

Recall $\tilde{\mathbb{E}}_\mu[(x_i - x_j)^2] = \frac{1}{2}(1 - \rho_{ij})$ $\quad (\rho_{ij} = 4\tilde{\mathbb{E}} x_i x_j - 1)$

$$\Rightarrow \text{apx ratio} \geq \min \frac{\mathbb{E}_g (\hat{x}_i - \hat{x}_j)^2}{\tilde{\mathbb{E}} (x_i - x_j)^2} = \min_{\rho_{ij}} \frac{2 \arccos \rho_{ij}}{\pi (1 - \rho_{ij})} = 0.878$$

$\square$

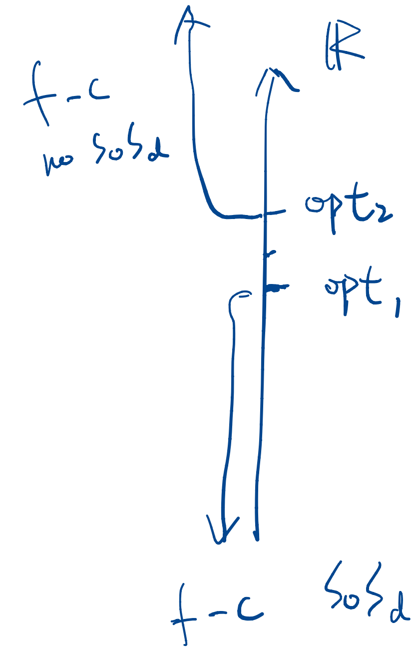
Goemans-Williamson rounding, $\alpha_{GW} = 0.878$

$\max_{c \in \mathbb{R}} \; c$ s.t. $f(x) - c$ has SoS_d cert

opt1

$\min_{c \in \mathbb{R}} \; c$ s.t. $f(x) - c$ has no SoS_d cert

opt2



$opt2 - \varepsilon$

$\Rightarrow f - c$ SoS_d

if $opt2 - \varepsilon > opt_1$: contradiction

$\Rightarrow opt2 - \varepsilon \le opt_1$