Proofs vs Algorithms: The SoS Method.

Lecture 1

Administrivia 3 hour lecture

Around 5:30 : 10-15 mins

Evaluation: 4-6 HW

Off. hours: 10-11 a.m. of Wednesdays

2-3 pm. Monday (Peter)

What is this course about? The Sum-of-Squares Method.

1) Checking feasibility of poly inequality system over $\mathbb{R}$.

$$\left.\begin{array}{l} P_1(x) \geq 0 \\ \vdots \\ P_m(x) \geq 0 \end{array}\right| \quad (1) \qquad x = (x_1, \ldots, x_n)$$

Is there a x s.t (1) is satisfied.

2) Checking Positivity

Is $q(x) \geq 0$ $\forall$ x : satisfying (1).

Very expressive

Max-Cut :: your input: $G=(V,E)$

$|V|=n$

find $S \subseteq V$ s.t. $|E(S,\bar{S})|$ is maximized.

for some β $\rightarrow$

$$\frac{1}{4}\cdot\sum_{\{i,j\}\in E}(x_i-x_j)^2=\beta\cdot|E|$$

$$\underbrace{x_i^2=1 \quad \forall\, 1\leq i\leq n.}_{x\in\{-1,1\}^n}$$

Ex. Max-clique, Max-3-SAT, Knapsack, ..

Fact: (1) is NP-Hard.

Goal: Analyze a relaxation for the feasibility problem — Sum-of-Squares Method — try to find interesting problems/instances where one can get poly time algorithms.

Sum-of-Squares Method

Comp. Complexity: use of semi-definite programming programming. Lovász '79, Goemans-Williamson '95,

... approx algos for CSPs.

Khot's UGC, optimal approx-algos for CSPs

↳ Special cases of SOS method.

Proof Complexity: Positivstellensatz proof system

Grigoriev-Vorobjov '98.

Parillo '00: Sequence of ~~of~~ SDP relaxations

Lasserre '00: Sequence of SDP relaxations
"Generalized Moment Problem"

→ Nesterov, Shor

Modern View: (2010s): Combinatorial optimization → CSPs

$\boxed{2014}$: average-case problems
→ Machine learning, stats, ...

Our Plan

1) Worst-Case Algo Design: Max-Cut, Coloring

Raghavendra's Thm

Grothendieck problem, ...

2) Average-Case Algo Design: finding planted communities, clustering mixtures of Gaussians

3) Lower bounds: Geometry, Pseudocalibration

4) Connections: Extended Formulations
Statistical Query Model
Low-deg poly Method.
Spectral Algorithms

Problem 1: Given $f: \{-1,1\}^n \to \mathbb{R}$ with rational coeffs. Goal is to decide if $f \geq 0$ on all $x \in \{-1,1\}^n$ or find an $x \in \{-1,1\}^n$ s.t. $f(x) < 0$.

E.g. Max-Cut.

$G: G(V,E)$.

$$f_G(x) : \frac{1}{4} \sum_{\{i,j\} \in E} (x_i - x_j)^2$$

decide if max-cut $(G) \leq C$.

$\Updownarrow$

$C - f_G(x) \geq 0$ on all $x \in \{-1,1\}^n$

representation: f_G is given as a vector of coefficients.

2. <u>Certifying Non-negativity</u>:

given $f : \{-1,1\}^n \to \mathbb{R}$
find an "efficiently verifiable" certificate
of non-negativity.

Proof
$\equiv$
Sum-of-Squares Certificates (of non-negativity)

<u>Def</u>: A degree d SoS cert. of non-neg of
$f : \{-1,1\}^n \to \mathbb{R}$ is a list of polynomials
$g_1, \dots, g_r : \{-1,1\}^n \to \mathbb{R}.$, $\deg(g_i) \leq \frac{d}{2}$

<u>and</u>: $f(x) = \sum_{i \leq r} g_i^2(x) \quad \forall \ x \in \{-1,1\}^n!$

<u>Efficiently verifiable?</u>

(1) how large is r? $\quad (\leq n^d$ WLOG$)$

(2) how large are coeffs of g_i?

Prop: Suppose $r \leq n^d$, all coeffs of g_i are bounded in magnitude by $2^{poly(n)}$. Then the identity $f = \sum_i g_i^2$ over all $x \in \{-1,1\}^n$ can be checked in poly(n^d) time.

Pf: Given $g_i \longrightarrow$ compute g_i^2

$\longrightarrow$ $\sum g_i^2$

$$\underline{f} - \underline{\sum g_i^2} = 0 \quad \forall\ x \in \{-1,1\}^n?$$

Fact: $\forall\ f : \{-1,1\}^n \to \mathbb{R}$ — there exists a unique representation of $f = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \prod_{i \in S} x_i$

(≡ coeff vector rep^n is unique)

Fourier Transform

Prop: Let $f : \{-1,1\}^n \to \mathbb{R}$ be non-neg. over $\{-1,1\}^n$. Then, $\exists$ a deg $2n$ SoS certificate of non-neg:

Proof: Consider $g, \{-1,1\}^n \to \mathbb{R}$

Proof: Consider

$$g(x) = \sqrt{f(x)}$$

Every function on $\{-1,1\}^n$ is a polynomial of deg $\leq n$

$$\underline{f = g^2} \rightarrow \text{deg } 2n \text{ SoS certificate!}$$

Tensor Notation: $v \in \mathbb{R}^n$

$$v^{\otimes 2} \in \mathbb{R}^{n^2}$$

$\rightarrow$ coordenates (i,j)

$$v^{\otimes 2}(i,j) = v_i \cdot v_j$$

$$v^{\otimes k} \in \mathbb{R}^{n^k}$$

v: $\quad (1, v) \in \mathbb{R}^{n+1}$

Fact: $\forall A \succeq 0$: $\exists B \in \mathbb{R}^{n \times n}$

$A \in \mathbb{R}^{n \times n}$ s.t. $A = B^T \cdot B$

Theorem (PSD Matrices vs SoS Certificates)

f, has deg d SoS cert of n.n. iff $\exists A$,

$$A \succeq 0 \quad \& \quad f(x) = \left\langle (1,x)^{\otimes \frac{d}{2}}, A \cdot (1,x)^{\otimes \frac{d}{2}} \right\rangle$$

Proof: "if" $f(x) = \langle (1,x)^{\otimes d/2}, (B^T B)(1,x)^{\otimes d/2} \rangle$

$$= \langle \underbrace{B(1,x)^{\otimes d/2}}, \underbrace{B(1,x)^{\otimes d/2}} \rangle$$

$$= \| B(1,x)^{\otimes d/2} \|_2^2$$

$$g_i(x) = \langle e_i, B(1,x)^{\otimes d/2} \rangle$$

poly of deg $\leq \frac{d}{2}$. $\therefore f(x) = \sum_{i=1}^{(n+1)^{d/2}} g_i^2(x)$

"only if": Suppose f has a deg d SoS cert

$$f = \sum_{i \leq r} g_i^2$$

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

deg $\leq \frac{d}{2}$

$$f(x) = \sum_i \langle v_i, (1,x)^{\otimes d/2} \rangle^2$$

$$= \langle (1,x)^{\otimes \frac{d}{2}}, \underbrace{\left(\sum_i v_i v_i^T \right)}_{\text{always psd}} (1,x)^{\otimes \frac{d}{2}} \rangle$$

Corollary: f has a deg d cert

then in fact it has a cert with $r \leq (n+1)^{d/2}$

Lemma (Bit complexity of SOS proofs)

Suppose f has a deg d SoS cert over $\{-1,1\}^n$.
Then, f has a deg d SoS cert. with bit complexity poly(n^d).

Proof: $f(x) = \langle (1,x)^{\otimes d/2}, A \cdot (1,x)^{\otimes d/2} \rangle$

$f(x) = \sum_{S,T} A_{S,T} \cdot x_S \cdot x_T$

for some A.

$$f = \sum_U \hat{f}_U \cdot x_U \cdot$$

$$x_U = \prod_{i \in U} x_i$$

$$\hat{f}_U = \sum_{\substack{S,T \\ \boxed{S \Delta T = U} \\ |S|,|T| \leq \frac{d}{2}}} A_{S,T} \cdot$$

$$\hat{f}_\phi = \sum_S A_{S,S} = tr(A).$$

$$\hat{f}_\phi = tr(A) = \sum_i \lambda_i(A)$$

$$\sum_{S,T} A_{S,T}^2 = \|A\|_F^2 = \sum_i \lambda_i^2(A) \leq n^d \cdot \sum_i \lambda_i$$

$\rightarrow \quad A = LDL^T$ decomposition.

L: triangular, diagonal $= 1$

D: +ve diagonal

$$f(x) = \langle \underbrace{L^T(1,x)^{\otimes d/2}}, \underbrace{DL^T(1,x)^{\otimes d/2}} \rangle$$

$$= \dots$$

What happens when f doesn't have a deg d SoS cert?

$\rightarrow \exists x : f(x) < 0$

$\rightarrow$ when $d < 2n$, f may be non-neg & yet there is no deg d SoS cert.

Pseudo-distribution

<u>Fact</u>: $SoS_d \{ f \mid f$ has a deg d cert. of n.neg $\}_{2^n} \subseteq \mathbb{R}^{2^n}$

is closed, $\underbrace{\text{convex}}$, cone.

K: convex, closed, <u>compact</u>

$v \notin K$.

$\exists$ H. $= \{ x \mid \underbrace{\langle u, x \rangle} \geq \tau \}$ — normal vector.

$K \subseteq H$.

$v \notin H$.

"normal vector"

$p \notin SOS_d$: $\mu : \{-1,1\}^n \to \mathbb{R}$

$\forall f : \langle \mu, f \rangle = \sum_{x \in \{-1,1\}^n} f(x) \cdot \mu(x)$.

$\exists \mu$; $\sum_x \mu(x) = 1$, $\sum_x \mu(x) \cdot p(x) < 0$ $\to \mathbb{E}_\mu p(x)$

& $\sum_x \mu(x) \cdot f(x) \geq 0$ $\to \mathbb{E}_\mu f$

$\forall f \in SOS_d$.

If $\mu \geq 0$, then it describes a prob distⁿ over $\{-1,1\}^n$.

$\exists$ a distⁿ $\mathbb{E}_\mu p < 0$, $\forall f \in SOS_d$: $\mathbb{E}_\mu f \geq 0$

Notation (Pseudo-distribution / Pseudo-expectation)

$$\tilde{\mathbb{E}}_\mu f = \sum_{x \in \{-1,1\}^n} \mu(x) \cdot f(x)$$

Pseudodistⁿ. (Def.)

A deg d p.d. over $\{-1,1\}^n$ is a function $\mu : \{-1,1\}^n \to \mathbb{R}$ s.t. formal expectation $\tilde{\mathbb{E}}_\mu$ satisfies: (1) $\tilde{\mathbb{E}}_\mu \cdot 1 = 1$

(2) $\forall f : \deg \leq \frac{d}{2}$, $\tilde{\mathbb{E}}_\mu \cdot f^2 \geq 0$.

Fact: Every deg $\geq 2n$ pseudo-distⁿ is an actual probability distribution.

Sketch: $f_y : \{-1,1\}^n \to \mathbb{R}$, $f_y(y) = 1$
$f_y(x) = 0 \;\forall x \neq y$

μ: deg $\geq 2n$, $\tilde{\mathbb{E}}_\mu f_y^2 \geq 0$

$$\boxed{\text{prob}(y) = \tilde{\mathbb{E}}_\mu \cdot f_y^2}$$

Claim: $= \mu(y)$.

Specifying pseudo-dist^{n}s

Lemma: $\forall$ deg d p.d. μ, $\exists$ a deg d multilinear polynomial $\mu' : \{-1,1\}^n \to \mathbb{R}$

s.t. $\tilde{\mathbb{E}}_{\mu}\, p(x) = \tilde{\mathbb{E}}_{\mu'}\, p \quad \forall p$ of deg $\le d$.

<u>Proof:</u> $\tilde{\mathbb{E}}_{\mu(x)}\, p(x) = \langle \mu, p \rangle$

$$\mu(x) = \sum_{S \subseteq [n]} \hat{\mu}_S \cdot X_S$$

$$p(x) = \sum_{\substack{S \subseteq [n] \\ |S| \le d}} \hat{p}_S \cdot X_S .$$

$$\mu = \mu' + \mu''$$

$\mu' \to$ <u>deg $\le d$ part of μ</u>

<u>Claim:</u> $\forall\, S \neq T$

$$\sum_{x \in \{-1,1\}^n} X_S \cdot X_T = 0 .$$

$$\langle \mu, p \rangle = \langle \mu', p \rangle$$

$\mu' \to$ our multilinear polynomial.

<u>Notation:</u> $\tilde{\mathbb{E}}_{\mu} (1, x)^{\otimes d} \to$ pseudo-moments

expectations of deg $\leq d$ monomials.

<u>Prop:</u> μ is a deg d p.d. iff 1) $\tilde{\mathbb{E}}_\mu \cdot 1 = 1$. ✓

and 2) $\tilde{\mathbb{E}}_\mu (1,x)^{\otimes d/2} \cdot \left((1,x)^{\otimes d/2}\right)^T \succcurlyeq 0$.

$\longrightarrow M_{d/2}$

deg d moment matrix

<u>S,T:</u> $\tilde{\mathbb{E}}_\mu \cdot x_S \cdot x_T = \tilde{\mathbb{E}}_\mu x_{S \Delta T}$

<u>Proof:</u> $f: \{-1,1\}^n \to \mathbb{R}$, deg $d/2$ polynomial.

<u>Claim:</u> $\underbrace{\tilde{\mathbb{E}}_\mu \cdot f^2}$ $= (\hat{f})^T \cdot M_{d/2} \cdot (\hat{f})$.

$$\tilde{\mathbb{E}}_\mu f^2 = \tilde{\mathbb{E}}_\mu \left(\sum_S \hat{f}_S \cdot x_S\right)^2 = \underline{\tilde{\mathbb{E}}_\mu} \sum_{S,T} \hat{f}_S \cdot \hat{f}_T \cdot x_S \cdot x_T$$

$$= \sum_{S,T} \hat{f}_S \cdot \hat{f}_T \cdot \tilde{\mathbb{E}}_\mu x_S \cdot x_T$$

<u>Duality</u>

<u>Thm:</u> for every f, every even $d \in \mathbb{N}$,

$\exists$ a deg d SoS cert. of n-neg of f

if & only if

$\forall$ deg d p.d. μ over $\{-1,1\}^n$, $\tilde{\mathbb{E}}_\mu f \geq 0$.

Proof: If f has a deg d cert

$$f = \sum_i g_i^2$$

$$\tilde{\mathbb{E}}_\mu f = \sum_i \tilde{\mathbb{E}}_\mu g_i^2 \geq 0. \checkmark$$

Now Suppose f is not contained in deg d SOS cone

$\exists$ halfspace $\rightarrow \mu$ "normal vector"

$$\begin{cases} \tilde{\mathbb{E}}_\mu f < 0 \\ \tilde{\mathbb{E}}_\mu g^2 \geq 0 \quad \forall g \text{ of deg } d/2 \end{cases}$$

Here I didn't explicity note that by scaling $g \to$ the threshold must be 0.

Need: $\tilde{\mathbb{E}}_\mu 1 > 0$

Prop: $\forall f : \{-1,1\}^n \to \mathbb{R}$, deg $\leq d$ ← even

$\exists$ a $L > 0$ s.t. $L + f$ has a deg d

SoS certificate.

Proof assuming prop

$$\tilde{\mathbb{E}}_\mu f < 0$$

$$\tilde{\mathbb{E}}_\mu (L+f) \geq 0$$

$$\underbrace{\tilde{\mathbb{E}}_\mu L}\geq -\tilde{\mathbb{E}}_\mu f$$

$$L\cdot\tilde{\mathbb{E}}_\mu 1\geq -\tilde{\mathbb{E}}_\mu f$$

$$\tilde{\mathbb{E}}_\mu 1\geq \frac{-\hat{\mathbb{E}}_\mu f}{L}>0. \quad \square$$

Proof of Prop:

$f=\sum_{|S|\leq d}\hat{f}_S\cdot X_S$. First let's handle monomials.

Claim* $(1-X_S)$ has a deg d certificate.
$(1+X_S)$ " " " " "

Pf assuming: $\sum_S|\hat{f}_S|(1+(\mathrm{sign}(\hat{f}_S)\cdot X_S))$ has a deg cert

$$=\underbrace{\sum_S|\hat{f}_S|}+\underbrace{\sum_S|\hat{f}_S|\cdot\mathrm{sgn}(\hat{f}_S)\cdot X_S}$$

$$\approx\underbrace{\sum_S|\hat{f}_S|}_{\hookrightarrow L}+f$$

Proof of claim: $x_S=X_{T_1}\cdot X_{T_2}$
for $|T_1|,|T_2|\leq d/2$

Then,

$$\underbrace{(x_{T_1} - x_{T_2})^2} = x_{T_1}^2 + x_{T_2}^2 - 2 \cdot x_{T_1} \cdot x_{T_2}$$
$$= 2 - 2\, x_{T_1} \cdot x_{T_2}$$

$$1 - x_{T_1} x_{T_2} = \frac{1}{2} (x_{T_1} - x_{T_2})^2$$

$$|||$$

$$1 - x_S = \frac{1}{2} (x_{T_1} - x_{T_2})^2$$

$$(x_{T_1} + x_{T_2})^2 = 2 + 2\, x_{T_1} x_{T_2}$$

$$|||$$

$$1 + x_S = \frac{1}{2} (x_{T_1} + x_{T_2})^2$$

$\square$