Lecture 21 Sum of Squares Algs

Nov 12

Q2a degree $|C|+1$, size ~~$|C|+1$~~ $O(|C|)$

$P = \{ p_1 \geq 0, \dots, p_m \geq 0 \}$ polynomial inequalities, $1 \geq 0 \in P$

For today, assume we are working modulo the Boolean axioms $\langle x_i^2 - x_i = 0 \rangle$.

→ So, all polys are multilinear, and algebra preserves multilinearity

e.g.

$$p = x_1 x_2 \qquad q = 3 x_2 x_3$$

$$pq = 3 x_1 x_2 x_3$$

An SOS proof of $q(x) \geq c$ from P looks like

$$\sum_{i=1}^{m} f_i p_i = q - c$$

where each f_i is a sum-of-squares, i.e.

$$f_i = \sum g_i^2 \qquad \text{for some polynomials } g_i.$$

Just like with Sherali-Adams there is an object (a "pseudo-expectation") that rules out the existence of low-degree SOS proofs.

Defn A function $\tilde{\mathbb{E}} : \{\text{multilinear polys}\} \to \mathbb{R}$ is a **degree-d SOS pseudo-expectation** for P if

(1) $\tilde{\mathbb{E}}[1] = 1$

(2) For all $p \in P \cup \{1\}$ and any poly q with

$\deg(pq^2) \leq d$,

$\tilde{\mathbb{E}}[pq^2] \geq 0$

(3) $\tilde{\mathbb{E}}$ is linear: so $\tilde{\mathbb{E}}[ap + bq] = a\tilde{\mathbb{E}}[p] + b\tilde{\mathbb{E}}[q]$ for p, q polys, $a, b \in \mathbb{R}$.

As with SA, if P has a degree-d SOS pseudo expectation then it does not have a degree-d SOS refutation.

Why? Suppose $\downarrow$ is a degree-d refutation

$$\sum_{i=1}^{m} f_i p_i = -1$$

$$\underbrace{\sum \tilde{\mathbb{E}}[f_i p_i]}_{\geq 0} = \tilde{\mathbb{E}}[-1] = -1$$

Contradiction!

Unlike SA, there isn't a linear program in general that searches for SOS pseudo-expectations.

The $\tilde{\mathbb{E}}[pq^2] \geq 0$ are not linear, but they can be captured by a semi-definite program.

Let's consider the constraint $\tilde{\mathbb{E}}[1 \cdot q^2] \geq 0$

Write $q(x) = \sum_{S \subseteq [n]} \vec{q}_S \prod_{i \in S} x_i = \sum_{S \in \binom{[n]}{\leq d}} \vec{q}_S \prod_{i \in S} x_i$

$$\tilde{\mathbb{E}}[q^2] = \tilde{\mathbb{E}}\Big[\Big(\sum_S \overset{\in \mathbb{R}}{\vec{q}_S} \prod_{i\in S} x_i\Big)\Big(\sum_T \vec{q}_T \prod_{i\in T} x_i\Big)\Big]$$

$$= \sum_S \sum_T \vec{q}_S \vec{q}_T \, \tilde{\mathbb{E}}\Big[\prod_{i\in S\cup T} x_i\Big] \geq 0$$

If $\vec{q} \in \mathbb{R}^{\binom{n}{\leq d}}$ and let $M_{\tilde{\mathbb{E}}}$ be the $\binom{n}{\leq d} \times \binom{n}{\leq d}$ matrix:

$$\begin{array}{cc} & \prod_{i\in T} x_i \\ \prod_{i\in S} x_i & \left(\begin{array}{c} \vdots \\ \cdots\ \tilde{\mathbb{E}}\Big[\prod_{i\in S\cup T} x_i\Big] \\ \end{array}\right) \end{array} = M_{\tilde{\mathbb{E}}}$$

then the $\tilde{\mathbb{E}}[q^2] \geq 0$ translates to

$$\vec{q}^{\,T} M_{\tilde{\mathbb{E}}} \vec{q} \geq 0$$

for all $\vec{q} \in \mathbb{R}^{\binom{n}{\leq d}}$.

i.e. the matrix $M_{\tilde{\mathbb{E}}}$ is positive semidefinite.

<u>Defn</u> A real symmetric matrix M is Positive semidefinite if

$$\vec{q}^{\,T} M q \geq 0$$

for all real vectors $\vec{q}$.

(Equivalently, all eigenvalues of M are non-negative.)

ex] Let $\mu \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n \times n}$, Σ is symmetric and PSD

then there is a Gaussian distribution over $\mathbb{R}^n$ with mean μ and covariance Σ.

Taken together, we can write this as a set of pseudo-expectation constraints PSD constraints of matrices over $\tilde{\mathbb{E}}[\prod_{i \in S} x_i)$ as variables:

$$\tilde{\mathbb{E}}[1] = 1$$

$$\forall p \in P \cup \{1\} \quad M_{\tilde{\mathbb{E}},p} \succeq 0 \quad (\text{i.e. } M_{\tilde{\mathbb{E}},p} \text{ is PSD}).$$

Called a semi-definite program.

Like linear programming, semi-definite programs can be* optimized over in polynomial time.

Thm Let P be a set of polynomial inequalities. For any d, ε there is an algorithm running in

$$n^{O(d)} \operatorname{poly}(\log(1/\varepsilon))$$

time solving

$$\min \quad \tilde{\mathbb{E}}[q]$$

$$\text{s.t.} \quad \tilde{\mathbb{E}} \in \operatorname{sos}_d(P)$$

The semidefinite constraints

for degree-d written above.

where each constraint is satisfied up to additive error ε.

SDPs also have a nice(-ish) duality theory — for us it means we can prove the following completeness theorems for SoS.

<u>Lemma</u> $\text{SOS}_n(P) = \text{conv}(\{0,1\}^n \cap P)$

<u>Thm</u> Let $P = \{p_1 \geq 0, \cdots, p_m \geq 0\}$ be a set of poly inequalities. Then, if $q(x) \geq c$ that is valid for all $\{0,1\}$-solutions of P then there is an SOS proof of $q - c$ from P.

Furthermore,

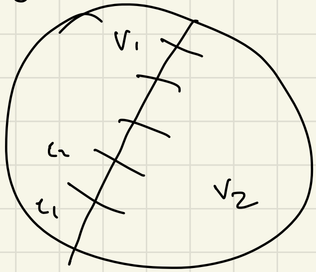
$$\max_c \left\{ q(x) \geq c \text{ is derivable in degree-}d \text{ SoS} \right\}$$

$$= \min_c \left\{ \tilde{\mathbb{E}}[q] = c : \tilde{\mathbb{E}} \in \text{SOS}_d(P) \right\}$$

Approx. Algorithms for SoS

ex) Max-Cut: Given graph $G=(V,E)$ with edge weights w_{ij}, goal is to partition $V = V_1 \cup V_2$ s.t. the total weight of edges crossing the partition is maximized.

G



$\sum_{e \text{ cross cut}} w_{ij} \leftarrow$ maximize

- LP gives $\leq$ 2-approximation
- SA (even up to degree $\Omega(n)$) also only gives a 2-approximation.
- We show degree-2 SOS gives a $1/0.878$-approximation — this is optimal assuming UGC.

For each vertex i introduce a variable $x_i \in \{\pm 1\}$.

Then Max-Cut is exactly

$$\max \sum_{i \neq j} w_{ij} \left(\frac{1 - x_i x_j}{2} \right)$$

$x_i = x_j \to x_i x_j = 1$

$x_i \neq x_j \to x_i x_j = -1$

$$\text{s.t.} \quad x_i^2 = 1 \quad (\text{i.e. } x_i \in \{\pm 1\})$$

Consider

$$\max \tilde{\mathbb{E}} \left[\sum_{i \neq j} w_{ij} \left(\frac{1 - x_i x_j}{2} \right) \right]$$

$$\text{s.t.} \quad \tilde{\mathbb{E}} \in SOS_2(\{1 \geq 0\})$$

$\tilde{\mathbb{E}}[x_i^2] = 1$ for all x_i

$$\max \sum_{i \neq j} w_{ij} \left(\frac{1 - \tilde{\mathbb{E}}[x_i x_j]}{2} \right)$$

$$\text{s.t.} \quad \tilde{\mathbb{E}} \in \mathrm{sos}_2(\{1 \geq 0\})$$

Let $\mathrm{opt}_{\mathrm{sos}_2}(G, w)$ be the optimal value given by the above SDP, and let $\mathrm{opt}(G, w)$ be the weight of the optimal cut.

since relaxation

$$\frac{1}{0.878} \mathrm{opt}(G,w) \geq \mathrm{opt}_{\mathrm{sos}_2}(G, w) \geq \mathrm{opt}(G, w)$$

[Goemans-Williamson 84]

Thm Given $\tilde{\mathbb{E}}$, the optimizer for the previous SDP, there is a polynomial-time algorithm that outputs

$$\alpha \in \{\pm 1\}^n$$

s.t. the value of this solution is at least

$$0.878 \, \mathrm{opt}_{\mathrm{sos}_2}(G, w).$$

Q. How do we use $\tilde{\mathbb{E}}$ to get a decent integral value?

Answer: We pretend $\tilde{\mathbb{E}}$ is an expectation over a real distribution of solutions to Max-Cut, and we sample from this "fake" distribution using $\tilde{\mathbb{E}}$.

Recall $\tilde{\mathbb{E}}$ gives values to all x_is and all products $x_i x_j$.

Let $\mu = (\tilde{\mathbb{E}}[x_1], \dots, \tilde{\mathbb{E}}[x_n]) \in \mathbb{R}^n$

$$\Sigma = \left(\tilde{\mathbb{E}}[x_i x_j]\right)_{i,j \in [n]} \in \mathbb{R}^{n \times n}$$

By SOS constraints Σ is PSD so $N(\mu, \Sigma)$ is a real Gaussian distribution.

<u>Lemma</u> Let $g, h \in \mathbb{R}$ be jointly distributed Gaussians and

$$\rho = \mathbb{E}[gh]$$

then

$$\mathbb{E}\left[\frac{1 - \operatorname{sign}(g)\operatorname{sign}(h)}{2}\right] \geq 0.878 \left(\frac{1-\rho}{2}\right)$$

Assume wlog that $\tilde{\mathbb{E}}[x_i] = 0 \ \forall i$, otherwise

$$\tilde{\mathbb{E}}'[p(x)] = \frac{1}{2}\left(\tilde{\mathbb{E}}[p(x)] + \tilde{\mathbb{E}}[p(-x)]\right)$$

satisfies this and gives the same value to the Max Cut polynomial.

Sample from Gaussian $\vec{g} \sim N(\vec{0}, (\tilde{\mathbb{E}}[x_i x_j]))$.

Then let $\alpha \in \{\pm 1\}^n$ be defined by

$$\alpha_i = \operatorname{sign}(\vec{g}_i).$$

$$\mathbb{E}\left[\sum w_{ij}\left(\frac{1-\alpha_i\alpha_j}{2}\right)\right] \geq 0.878 \sum w_{ij}\left(\frac{1-\tilde{\mathbb{E}}[x_i x_j]}{2}\right)$$

$$= 0.878 \; opt_{sos_2}. \qquad \square$$

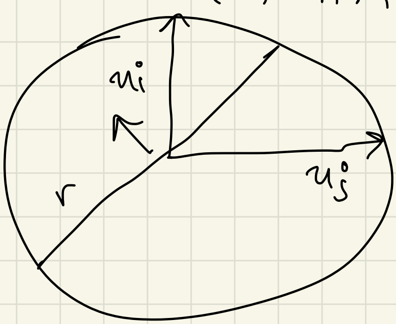
PSD — by properties of PSD matrix

$$\left(\tilde{\mathbb{E}}[x_i x_j]\right)_{i,j\in[n]} = U^T U$$

$U \in \mathbb{R}^{n\times n}$ $\quad u_i \in \mathbb{R}^n$ for each vertex $i \in [n]$

s.t. $\tilde{\mathbb{E}}[x_i x_j] = u_i \cdot u_j$

Now sample $r \sim N(0, I_n)$, output $\text{sign}(\langle r, u_i\rangle)$.



Recent ton of work in average-case statistical algs

- Clustering Gaussian Mixture Models
- Compressed Sensing
- Dictionary Learning
- ....