Optimality of sum-of-squares

In this lecture, we show that sum-of-squares achieves the best possible approximation guarantees for every constraint satisfaction problems (CSPs) among all polynomial-size semidefinite programming (SDP) relaxations (Lee et al. [2015]).

The first step is to formalize the notion of SDP relaxations for constraint satisfaction problem. For simplicity, we restrict ourselves to Boolean CSPs, that is, every variable takes values from the alphabet $\{0,1\}$. There are many equivalent for definitions for the kind of relaxations we want to study. It turns out that the most convenient definition for the purpose of our proof is in terms of certain certificates of nonnegativity, which generalize the notion of sum-of-squares certificates.

1. Definition (subspace certificates). Let U be a linear subspace of the set of real-valued functions on $\{0,1\}^n$. We say that $f\colon \{0,1\}^n \to \mathbb{R}$ has a U-certificate (of nonnegativity) if there exists functions $g_1, \dots, g_r \in U$ such that

$$f = g_1^2 + \cdots + g_r^2 . \tag{1}$$

We can view the task of approximating a boolean constraint satisfaction problem in terms of certifying the nonnegativity of a collection of functions. For example, we can consider nonnegative functions of the form

$$x \mapsto \frac{\max f_G}{0.878} - f_G(x), \tag{2}$$

where G ranges over all graphs and $f_G(x)$ is the function that counts the number of edges cut by the bipartition x. The fact that degree-2 sos achieves approximation ratio 0.878 for Max Cut is equivalent to the fact that this set of functions has degree-2 sos certificates of nonnegativity.

The following definition formalizes what it means for a set of nonnegative functions to have small subspace certifcates of nonnegativity.

2. Definition (SDP size). Let $\mathcal{F}$ be a family of real-valued functions on hypercubes. For $n \in \mathbb{N}$, let $\mathcal{F}_n$ be the restriction of $\mathcal{F}$ to functions on $\{0,1\}^n$. For a function $s\colon \mathbb{N} \to \mathbb{N}$, we write $\mathcal{F} \in \mathrm{SIZE}_{\mathrm{SDP}}(s)$ if there exists a constant C such that for every $n \in \mathbb{N}$, there exists a subspace $U \subseteq \mathbb{R}^{\{0,1\}^n}$ of dimension at most $C \cdot s(n)$ such that every function $f \in \mathcal{F}_n$ has a U-certificate of nonnegativity.

Since subspace certificates generalize sos certificates, any family $\mathcal{F}$ of real-valued functions on hypercubes with degree-d sos certificates satisfies $\mathcal{F} \in \mathrm{SIZE}_{\mathrm{SDP}}(n^d)$.

The following theorem gives a partial converse to that statement. For a function $f\colon \{0,1\}^m \to \mathbb{R}$, we define extensions$(f)$ to be the set of functions $g\colon \{0,1\}^n \to \mathbb{R}$ with $n \geq m$ such that $g(x) = f(x_S)$ for some $S \subseteq [n]$ with $|S| = m$. In other words, extensions(f) consists of all functions on hypercubes that compute f on a subset of input bits.

3. Theorem (Optimality of sum-of-squares). *Let $f\colon \{0,1\}^m \to \mathbb{R}$. Suppose* extensions$(f) \in \mathrm{SIZE}_{\mathrm{SDP}}(n^d)$ *for some $d \in \mathbb{N}$. Then, f has a degree-$10d$ sos certificate of nonnegativity.*

For the theorem, it is important to pass to extensions because every nonnegative function f on the hypercube satisfies $\{f\} \in \mathrm{SIZE}_{\mathrm{SDP}}(1)$.[1].

[1] To see that every nonnegative function satisfies $\{f\} \in \mathrm{SIZE}_{\mathrm{SDP}}(1)$ consider the 1-dimensional subspace spanned by the function $\sqrt{f}$.

Theorem 3 shows the optimality of sos for CSPs because the families $\mathcal{F}$ of functions that correspond to CSPs are closed under extensions, so that extensions$(\mathcal{F}) = \mathcal{F}$. For families $\mathcal{F}$ with this closure property, the above theorem says that $\mathcal{F} \in \mathrm{SIZE}_{\mathrm{SDP}}(n^d)$ implies that $\mathcal{F}$ has degree-$10d$ sos certficiates.

We remark that the proof of Theorem 3 also gives some bounds for super constant d (i.e., d is allowed to depend on n). However, the resulting bounds appear to be far from tight.

Positive matrix functions

The following lemma gives a characterization of SDP size that is algebraically more concise. This alternative characterization will be convenient for the purposes of the proof of Theorem 3. We say that a matrix function $Q\colon \{0,1\}^n \to \mathbb{R}^{N\times N}$ is *positive* if $Q(x) \succeq 0$ for all $x \in \{0,1\}^n$.

4. Lemma (positive matrix function). *Let U be a linear subspace of functions on $\{0,1\}^n$ of dimension N. Let $\mathcal{F}_U$ be the set of functions with a U-certificate of nonnegativity. Then, there exists a positive matrix function $Q\colon \{0,1\}^n \to \mathbb{R}^{N\times N}$ such that*

$$\mathcal{F}_U = \{x \mapsto \operatorname{Tr} P \cdot Q(x) \mid P \succeq 0\}\,. \tag{3}$$

To get some intuition about this lemma, observe that functions of the form $x \mapsto \operatorname{Tr} P \cdot Q(x)$ are indeed nonnegative because the product

of two positive semidefinite matrices has nonnegative trace. To prove this inequality note that $\operatorname{Tr} AB = \|\sqrt{A}\sqrt{B}\|_F^2 \geq 0$ for all $A, B \succeq 0$.

Proof. Let $h_1, \dots, h_N$ be a basis for U and let $h\colon \{0,1\}^n \to \mathbb{R}^N$ be the vector-valued function $h = (h_1, \dots, h_N)$. Note that every function $g \in U$ has a representation $g(x) = \langle h(x), v\rangle$ for $v \in \mathbb{R}^N$. Choose Q such that $Q(x) = h(x)h(x)^\intercal$. Let f be any function with a U-certificate of nonnegativity so that $f = g_1^2 + \cdots + g_r^2$ for $g_1, \dots, g_r \in U$. Let $v_1, \dots, v_r$ be the coordinates of these functions so that $g_i(x) = \langle h(x), v_i\rangle$ for all $i \in [r]$. Then,

$$f(x) = g_1(x)^2 + \cdots + g_r(x)^2 = \sum_{i=1}^r \langle h(x), v_i\rangle^2 = \operatorname{Tr}\left(\sum_{i=1}^r v_i v_i^\intercal\right) Q(x)\,. \tag{4}$$

We conclude $f \in \mathcal{F}_U$ as desired. □

Proof strategy

Let $f\colon \{0,1\}^m \to \mathbb{R}$ be such that $\text{extensions}(f) \in \text{SIZE}_{\text{SDP}}(n^d)$. We are to show that f has a degree-$10d$ sos certificate. To this end, we will show that every $10d$ pseudo-distribution $\mu\colon \{0,1\}^m \to \mathbb{R}$ satisfies $\tilde{\mathbb{E}}_\mu f \geq 0$.

Fix $n \in \mathbb{N}$ large enough. Let $Q\colon \{0,1\}^n \to \mathbb{R}^{N\times N}$ for $N \leq C \cdot n^d$ be a positive matrix function such that every extensions $x \mapsto f(x_S)$ of f to $\{0,1\}^n$ has a representation $f(x_S) = \operatorname{Tr} P_S \cdot Q(x)$ as in Lemma 4. Note that $\tilde{\mathbb{E}}_\mu f = \tilde{\mathbb{E}}_{\mu(x_S)} f(x_S) = \sum_x \mu(x_S) \operatorname{Tr} P_S Q(x)$ for every $S \subseteq [n]$ with $|S| = m$. In the following, let S be a random subsets of $[n]$ with cardinality m. We will show that

$$\mathop{\mathbb{E}}_S \sum_x \mu(x_S) \operatorname{Tr} P_S Q(x) \geq -\varepsilon\,, \tag{5}$$

where $\varepsilon > 0$ tends to 0 as n grows. Note that the inequality Eq. (5) would be easy to show if μ is a probability distribution.[2] Hence, the task of proving inequality Eq. (5) boils down to proving that pseudo-distributions behave like actual probability distributions in certain ways. (We have encountered many tasks of this kind in earlier lectures.) It is instructive to consider the special case that the function $x \mapsto \sqrt{Q(x)}$ has degree at most d.[3] In this case, $\operatorname{Tr} P_S Q(x) = \|\sqrt{P_S} \cdot \sqrt{Q(x)}\|^2$ is a sum of squares of polynomials of degree at most d.

[2] If μ is a probability distribution, then the left-hand side of Eq. (5) is a sum of nonnegative terms.

[3] We define the degree of a matrix valued function $M(x)$ to the maximum degree of an entry $M_{i,j}(x)$.

5. Theorem (Learning low-degree positive matrix function). *Let $f\colon \{0,1\}^m \to \mathbb{R}$ be such that $\text{extensions}(f) \in \text{SIZE}_{\text{SDP}}(n^d)$ and let*

$\mu\colon \{0,1\}^m \to \mathbb{R}$ be a degree-10d pseudo-distribution. Then, there exists $C \geq 1$ (depending on f and μ) such that

$$\mathop{\mathbb{E}}_{\mu} f \geq \mathop{\mathbb{E}}_{S} \mathop{\mathbb{E}}_{\mu(x_S)} \operatorname{Tr} P_S \tilde{Q}(x) \tag{6}$$

OLD

Lee et al. [2015] have shown that sos is *optimal* among all similar sized SDP's of certain form for constraint satisfaction problems. In this lecture we will sketch the proof of a representative theorem. We start with a definition. Let $\mathcal{U}$ be a subspace of the functions from $\{0,1\}^n$ to $\mathbb{R}$. We say that a function $g\colon \{0,1\}^n \to \mathbb{R}$ has a $\mathcal{U}$ *certificate of non-negativity* if there are some $u_1, \ldots, u_k \in \mathcal{U}$ such that $g(x) = \sum_{i=1}^k u_i(x)^2$ for every $x \in \{0,1\}^n$. This is a generalization of sos certificates, which correspond to the subspace of all polynomials of degree at most d. We say that g is an *n extension* of $f\colon \{0,1\}^m \to \mathbb{R}$ if $g(x) = f(x_S)$ for some m-sized subset $S \subseteq [n]$. Clearly, if f is non negative then every extension of it is also non-negative.

6. Theorem. *Suppose that $f\colon \{0,1\}^m \to \mathbb{R}$ is a bounded[4] function such that there is a degree d pseudo-distribution μ over $\{0,1\}^m$ satisfying $\tilde{\mathbb{E}}_\mu f = -0.1$. Then there is some $n = m^{O(1)}$ such that for every subspace $\mathcal{U}$ of $\mathbb{R}^{\{0,1\}^n}$ of dimension $n^{o(d)}$, there is an n-extension g of f such that g does not have a $\mathcal{U}$ certificate of non-negativity.*

[4] In this lecture we say that a function $f\colon \{0,1\}^m \to \mathbb{R}$ is *bounded* if there are constant degree sos polynomials s, s' such that $-1 + s(x) = f(x) = 1 - s'(x)$ for every $x \in \{0,1\}^m$.

For example, the lower bound of Grigoriev [2001] yields a 3XOR instance on m variables where no assignment satisfies more than 0.6 fraction of the constraints but there is a degree $\Omega(m)$ pseudo-distribution that pretends to satisfy all of them. Thus if we let $f(x)$ equal 0.6 minus the number of satisfied constraints then we see that we can turn this into a lower bound on 3XOR for general $\mathcal{U}$ proofs. Note that passing to an extension is inherent, as if f is non-negative we can always simply add the function $\sqrt{f}$ to our subspace $\mathcal{U}$.

The heart of the proof turns out to lower bound a notion known as the *PSD rank* of a matrix related to all possible extensions of f.

7. Definition (PSD rank). Let M be an $N \times N'$ non-negative matrix. We say that M has *PSD rank* at most r if there are $r \times r$ PSD matrices $P_1, \ldots, P_N$, $Q_1, \ldots, Q_{N'}$ such that $M_{i,j} = \operatorname{Tr}(P_i Q_j)$ for all $i \in [N], j \in [N']$.

Theorem 6 turns out to be a corollary of the following theorem:

8. Theorem. *Suppose that $f\colon \{0,1\}^m \to \mathbb{R}$ is a bounded function such that there is a degree d pseudo-distribution μ over $\{0,1\}^m$ satisfying $\tilde{\mathbb{E}}_\mu f = -0.1$. Then there is some $n = m^{O(1)}$ such that the $\binom{n}{m} \times 2^n$ matrix M where $M_{S,x} = f(x_S)$ has PSD rank $m^{\Omega(d)}$.*

Theorem 6 follows from Theorem 8 by noting that if *every* n-extension g of f has a $\mathcal{U}$ proof then we can express each entry of this matrix as a product of PSD matrices whose image is in $\mathcal{U}$. Thus our focus from now on would be to prove Theorem 8.

Proof of the PSD rank lower bound

We now prove Theorem 8. Under our assumptions, the pseudo-distribution μ satisfies

$$\mathop{\mathbb{E}}_{m} uf = \sum_{w\in\{0,1\}^m} \mu(w)f(w) = 2^m \mathop{\mathbb{E}}_{w\in\{0,1\}^m} \mu(w)f(w) \leq -0.1 \qquad (7)$$

which in particular means that

$$\mathop{\mathbb{E}}_{S\sim\binom{[m]}{n}} \mathop{\mathbb{E}}_{x\in\{0,1\}^n} \mu(x_S)f(x_S) \leq -0.12^{-m}\,. \qquad (8)$$

Under our assumptions there are $r \times r$ PSD matrices $P(S)$,$Q(x)$ such that $f(x_S) = \mathrm{Tr}(P(S)Q(x))$ and so we can write Eq. (8) as $\mathrm{Tr}(MQ) \leq -0.1$ where M, Q are two $r2^n \times r2^n$ block matrices whose x^{th} blocks equal $2^m\mu(x_S)P(S)$ and $2^{-n}Q(x)$ respectively.

The main technical claim is that we can find $\tilde{Q}$ with the same trace as Q of the form $\tilde{Q} = q(M)^2$ with q being an $\tilde{O}(1)$ degree polynomial with $\mathrm{Tr}(M\tilde{Q}) \leq -0.05 \cdot 2^{-m}$ Since every entry of the x^{th} block of M is a polynomial in x of degree at most m, this means that there is some degree $\tilde{O}(m)$ matrix valued polynomial R such that

$$\mathop{\mathbb{E}}_{S\sim\binom{[m]}{n}} \mathop{\mathbb{E}}_{x\in\{0,1\}^n} \mu(x_S)\,\mathrm{Tr}(P(S)R^2(x)) \leq -0.05 \cdot 2^{0-m} \qquad (9)$$

we then change the order of expectations and write this as

$$\mathop{\mathbb{E}}_{w\in\{0,1\}^m} \mu(w) \mathop{\mathbb{E}}_{x'\in\{0,1\}^{n-m}, S\sim\binom{[m]}{n}} \mathrm{Tr}(P(S)R^2(w,x')) \qquad (10)$$

and this *random restriction* will turn out to imply that R is approximately a polynomial of degree $o(d)$ in the variables w. Now for every matix valued polynomial R of degree d' and PSD matrix P, the value $\mathrm{Tr}(PR^2(w))$ is a degree $2d'$ sos polynomial (as can be evidenced by writing it as $\sum_{i,j}(\sqrt{P}R(w))^2_{i,j}$) and hence we get a contradiction to the fact that μ is a pseudo-distribution.

References

Dima Grigoriev. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoret. Comput. Sci.*, 259(1-2):613–622, 2001. ISSN 0304-3975. doi: 10.1016/S0304-3975(00)00157-2. URL http://dx.doi.org/10.1016/S0304-3975(00)00157-2.

James R. Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *STOC*, pages 567–576. ACM, 2015.