The sos algorithm over general domains

So far our focus has been on the task of optimizing some n-variate polynomial f over the Boolean cube. But the sos algorithm is applicable in a much more general setting. In particular, we can replace the Boolean cube with any set $\Omega \subseteq \mathbb{R}^n$ that is defined by polynomial inequalities. This is important to capture computational problems that go beyond those that we have considered so far. We now make the appropriate general definitions of sos proofs, pseudo-distributions, and state the more general theorem regarding the sos algorithm.

It turns out that in the most general setting some issues arise that do not come up for the hypercube. So far these issues seem to be pathological in the sense that one can construct bad examples but they don't play a role when designing algorithms based on sum-of-squares. We will highlight an important special case that avoids those issues altogether and still significantly generalizes the hypercube.

General sum-of-squares proofs

Let $\mathbb{R}[x]$ be the ring of polynomials with real coefficients in variables $x = (x_1, \ldots, x_n)$. Let $\mathcal{A} = \{f_1 \geq 0, \ldots, f_m \geq 0\}$ be a system of *polynomial constraints* with $f_1, \ldots, f_m \in \mathbb{R}[x]$. We will be interested in two kinds of problems related to $\mathcal{A}$:

- decide if $\mathcal{A}$ has a solution or if it is infeasible,
- given a polynomial $g \in \mathbb{R}[x]$, decide if g is nonnegative over the set of solutions to $\mathcal{A}$.

We say that a polynomial $p \in \mathbb{R}[x]$ is *sum-of-squares (sos)* if there are polynomials $q_1, \ldots, q_r \in \mathbb{R}[x]$ such that $p = q_1^2 + \cdots + q_r^2$.s

1. Definition (Sum-of-squares proof). A *sum-of-squares (sos) proof* that the system of polynomial constraints $\mathcal{A}$ implies the constraint $\{g \geq 0\}$ consists of sum-of-squares polynomials $(p_S)_{S \subseteq [m]}$ in $\mathbb{R}[x]$ such that[1]

$$g = \sum_{S \subseteq [m]} p_S \cdot \prod_{i \in S} f_i \,. \tag{1}$$

[1] In Eq. (1) we adopt the convention that empty products are 1, so that $\prod_{i \in S} f_i = 1$ for $S = \emptyset$.

We say that this proof has degree at most ℓ if each summand has degree at most ℓ, i.e., every set $S \subseteq [m]$ satisfies $\deg(p_S \cdot \Pi_{i \in f_i}) \leq \ell$. If there exists such a proof of degree at most ℓ, we write

$$\mathcal{A} \vdash_\ell \{g \geq 0\} \,. \tag{2}$$

The identity Eq. (1) implies that g is nonnegative for every point x that satisfies the system $\mathcal{A}$ because each summand on the right-hand side is nonnegative. Given $\mathcal{A}$, g, and sos polynomials $(p_S)_{S\subseteq[m]}$ (together with their sos representations), we can efficiently verify that Eq. (1) holds by comparing coefficients on the left and on the right.

In order to emphasize the variables for the proofs, we sometimes write Eq. (2) as $\{f_1(x) \geq 0, \ldots, f_m(x) \geq 0\} \vdash_{x,\ell} \{g(x) \geq 0\}$.[2]

[2] This convention is useful when the functions $f_1, \ldots, f_m$ and g involve variables besides the "proof variables". The distinction is important sometimes because non-proof variables do not count toward the degree of the proof and are treated as constants.

The following theorem (Krivine'64, Stengle'74) shows that sum-of-squares proofs are enough to decide the infeasibility of systems of polynomial constraints. However, the proof is highly non-constructive and certainly does not give any bounds on the degree (which we would need in order to use sum-of-squares proofs for the design of algorithms).

2. Theorem (Positivstellensatz). *For every system of polynomial constraints $\mathcal{A} = \{f_1 \geq 0, \ldots, f_m \geq 0\}$, either there exists a solution or there exists a sum-of-squares proof $\mathcal{A} \vdash_\ell \{-1 \geq 0\}$ for some $\ell \in \mathbb{N}$.*

We refer to a sos proof of the form $\mathcal{A} \vdash_\ell \{-1 \geq 0\}$ as a *degree-d sum-of-squares refutation for $\mathcal{A}$*.

It turns out that sum-of-squares proofs capture many kinds of "real-world mathematical proofs". In particular, they obey the following intuitive inference rules, for all system of polynomial constraints $\mathcal{A}, \mathcal{B}, \mathcal{C}$, polynomials $f, g\colon \mathbb{R}^n \to \mathbb{R}$ and $F\colon \mathbb{R}^n \to \mathbb{R}^m, G\colon \mathbb{R}^n \to \mathbb{R}^k, H\colon \mathbb{R}^p \to \mathbb{R}^n$,

$$\begin{aligned}
\text{addition: } & \frac{\mathcal{A} \vdash_\ell \{f \geq 0, g \geq 0\}}{\mathcal{A} \vdash_\ell \{f + g \geq 0\}}, \\
\text{multiplication: } & \frac{\mathcal{A} \vdash_\ell \{f \geq 0\},\ \mathcal{A} \vdash_{\ell'} \{g \geq 0\}}{\mathcal{A} \vdash_{\ell+\ell'} \{f \cdot g \geq 0\}}, \\
\text{transitivity: } & \frac{\mathcal{A} \vdash_\ell \mathcal{B},\ \mathcal{B} \vdash_{\ell'} \mathcal{C}}{\mathcal{A} \vdash_{\ell\cdot\ell'} \mathcal{C}}, \\
\text{substitution: } & \frac{\{F \geq 0\} \vdash_\ell \{G \geq 0\}}{\{F(H)) \geq 0\} \vdash_{\ell\cdot\deg(H)} \{G(H) \geq 0\}}.
\end{aligned} \tag{3}$$

Exercises

3. Exercise (Univariate polynomials). Let $p \in \mathbb{R}[t]$ be a univariate degree-d polynomial such that $p(t) \geq 0$ for all $t \in \mathbb{R}$. Show that

$$\vdash_d \{p \geq 0\}. \tag{4}$$

4. Exercise (Some bound). Let $d \in \mathbb{N}$ and $f \in \mathbb{R}[x]$ be a polynomial of degree at most d. Show that there exists a scalar $M > 0$ such that

$$\{\|x\|^2 \leq 1\} \vdash_d \{f \leq M\}. \tag{5}$$

Pseudo-distributions

We can represent a finitely supported probability distribution over $\mathbb{R}^n$ by its probability mass function $\mu \colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ so that $\mu(x)$ is equal to the probability of the point $x \in \mathbb{R}^n$ under the distribution. We define pseudo-distributions as generalizations of such probability mass functions. First, we define the formal expecation of a function $f \colon \mathbb{R}^n \to \mathbb{R}$ under a finitely supported function μ,

$$\tilde{\mathbb{E}}_{\mu} f = \sum_{x \in \text{support}(\mu)} f(x) \cdot \mu(x). \tag{6}$$

Note that since μ is finitely supported there are no issues of measurability or convergence here.

5. Definition (pseudo-distribution). A finitely supported function $\mu \colon \mathbb{R}^n \to \mathbb{R}$ is a *degree-d pseudo-distribution* if $\tilde{\mathbb{E}}_{\mu} 1 = 1$ and $\tilde{\mathbb{E}}_{\mu} f^2 \geq 0$ for all polynomials f on $\mathbb{R}^n$ with $\deg f \leq d/2$.

Furthermore, if $\text{support}(\mu) \subseteq \Omega$, we say that μ is a *degree-d pseudo-distribution over Ω*.

Note that these definitions are consistent with our previous definition for pseudo-distributions over the hypercube.

Unlike for actual probability distributions, many computational problems for pseudo-distributions admit efficient algorithms. The following characterization in terms of positive semidefinite matrices is the main structure that those algorithms exploit.

6. Lemma (pseudo-moments). *Let $\mu \colon \mathbb{R}^n \to \mathbb{R}$ be finitely supported and $\tilde{\mathbb{E}}_{\mu} 1 = 1$. Then, μ is a degree-d pseudo-distribution if and only if formal degree-d moment matrix $\tilde{\mathbb{E}}_{\mu(x)} \left((1, x)^{\otimes d/2}\right) \left((1, x)^{\otimes d/2}\right)^{\intercal}$ is positive semidefinite.*

Proof. Suppose that the degree-d moment matrix M is positive semidefinite. Let p be any polynomial of degree at most $d/2$. Let v be a vector such that $p(x) = \langle p, (1, x)^{\otimes d/2} \rangle$. Then, $\tilde{\mathbb{E}}_{\mu} p^2 = \langle v, Mv \rangle \geq 0$. It follows that μ is a degree-d pseudo-distribution.

On the other hand, suppose that this moment matrix M is not positive semidefinite. Then, there exists a vector v such that $\langle v, Mv \rangle < 0$.

Let p be the polynomial $p(x) = \langle v, (1,x)^{\otimes d/2}\rangle$. Then, $\tilde{\mathbb{E}}_\mu p^2 = \langle v, M, v\rangle < 0$. It follows that μ is not a degree-d pseudo-distribution. □

The following definition formalizes what it means for a pseudo-distribution to satisfy a system of polynomial constraints.

7. Definition (Model for polynomial constraints). Let $\mathcal{A} = \{f_1 \geq 0, \ldots, f_m \geq 0\}$ be a set of polynomial constraints over $\mathbb{R}^n$. Let $\mu\colon \mathbb{R}^n \to \mathbb{R}$ be a pseudo-distribution. We say that *μ satisfies $\mathcal{A}$ at degree ℓ*, denoted $\mu \models_\ell \mathcal{A}$, if every set $S \subseteq [m]$ and every sos polynomial p on $\mathbb{R}^n$ with $\deg h + \sum_{i\in S} \max\{\deg f_i, \ell\} \leq d$ satisfies[3]

$$\tilde{\mathbb{E}}_\mu h \cdot \prod_{i\in S} f_i \geq 0\,. \tag{7}$$

[3] The parameter ℓ here allows us some additional freedom which will be important for stating the interaction between sum-of-squares proof and pseudo-distributions. The idea behind the condition $\sum_{i\in S} \max\{\deg f_i, \ell\}$ is that it bounds the degree of the polynomial $h \cdot \prod_{i\in S} f_i$ by d, even when we treat each f_i to have degree $\max\{\deg f_i, \ell\}$.

We write $\mu \models \mathcal{A}$ (without further specifying the degree) if $\mu \models_0 \mathcal{A}$.

Note that if a degree-d pseudo-distribution μ satisfies $\mu \models \{f \geq 0\}$ for a polynomial f with $\deg f \leq d$, then $\tilde{\mathbb{E}}_\mu f \geq 0$. However, if μ only satisfies $\mu \models_\ell \{f \geq 0\}$, then we can only conclude $\tilde{\mathbb{E}}_\mu f \geq 0$ if $\ell \leq d$.

Duality between pseudo-distributions and sum-of-squares proofs

The following theorem shows a duality between pseudo-distributions and sos proofs for systems of polynomial constraints that are explicitly bounded in the sense that they contain a constraint that implies that every variable is restricted to a finite interval.[4]

[4] The technical term for this property of systems of polynomial constraints is *Archimedian*.

8. Theorem (Duality of pseudo-distributions and sos proofs). *Let $\mathcal{A}$ be a system of polynomial constraints over $\mathbb{R}[x]$ that contains a constraint of the form $\|x\|^2 \leq M$ for some scalar $M \geq 0$. Then for every even $d \in \mathbb{N}$ and every polynomial $f \in \mathbb{R}[x]_{\leq d}$, exactly one of the following conditions is satisfied:*[5]

- *for every $\varepsilon > 0$, there exists a degree-d sos proof $\mathcal{A} \vdash_d \{f \geq -\varepsilon\}$,*
- *there exists a degree-d pseudo-distribution $\mu\colon \mathbb{R}^n \to \mathbb{R}$ such that $\mu \models \mathcal{A}$ and $\mathbb{E}_\mu f \leq 0$.*

[5] Note that the duality between pseudo-distributions and sos proofs for the Boolean hypercube is stronger because it is possible to choose $\varepsilon = 0$ below.

A direct consequence of this theorem is that for all $\mathcal{A}$ and f as above, the supremum of the set $\{c \in \mathbb{R} \mid \mathcal{A} \vdash_d f \geq c\}$ is equal to the minimum value of $\mathbb{E}_\mu f$ over all degree-d pseudo-distributions μ such that $\mu \models \mathcal{A}$.

An important special case of the above theorem is that $f = -1$. In this case the theorem says that either $\mathcal{A}$ has a degree-d sos refutation so that $\mathcal{A} \vdash_d \{-1 \geq 0\}$ or there exists a degree-d pseudo-distribution μ that satisfies the constraints $D \models A$.

Proof. Let $C \subseteq \mathbb{R}[x]_{\leq d}$ be the cone of polynomials g such that $\mathcal{A} \vdash_d \{g \geq 0\}$. We will show that if f is in the closure of C, then $\mathcal{A} \vdash_d \{f \geq -\varepsilon\}$ for all $\varepsilon > 0$ and thatif f is not in the closure of C, then there exists a degree-d pseudo-distribution μ such that $\mu \models \mathcal{A}$ and $\tilde{\mathbb{E}}_\mu f < 0$.

If f is in the closure of C, then by convexity there exists a polynomial $g \in C$ such that $(1-\varepsilon)f + \varepsilon g \in C$ for all $\varepsilon > 0$, which means that $\mathcal{A} \vdash_d \{f \geq -\varepsilon(g-f)\}$ for all $\varepsilon > 0$. Since $\mathcal{A}$ contains a constraint of the form $\|x\|^2 \leq M$, it follows that $\mathcal{A} \vdash_d \{g - f \leq M'\}$ for some scalar $M' > 0$.[6] Putting these sos proofs together, we get $\mathcal{A} \vdash_d \{f \geq -\varepsilon\}$ for all $\varepsilon > 0$.

[6] *Exercise:* Show that for every polynomial $f \in \mathbb{R}[x]_{\leq d}$, there exists a scalar $M > 0$ such that $\{\|x\|^2 \leq 1\} \vdash_d \{f \leq M\}$.

If f is not in the closure of C, then there exists a separating linear functional ϕ such that $\phi[f] < 0$ and $\phi[g] \geq 0$ for every $g \in C$. We claim that by rescaling we may assume $\phi[1] = 1$. It is enough to show that $\phi[1] > 0$. Indeed, since $\mathcal{A}$ contains a constraint of the form $\|x\|^2 \leq M$, we have $\mathcal{A} \vdash_d \{f \leq M'\}$ for some scalar $M' > 0$. Therefore, $M' - f \in C$ and $0 \leq \phi[M' - f] = M' \cdot \phi[1] - \phi[f] < M' \cdot \phi[1]$, which means that $\phi[1] > 0$.

Using multivariate interpolation, we can represent the linear functional ϕ as a linear combination of point evaluations, i.e., there exist points $y^{(1)}, \ldots, y^{(m)} \in \mathbb{R}^n$ and scalars $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ such that $\phi[g] = \sum_{i=1}^m \alpha_i \cdot g(y^{(i)})$ for all $g \in \mathbb{R}[x]_{\leq d}$. Let $\mu \colon \mathbb{R}^n \to \mathbb{R}$ be the finitely-supported function such that $\mu(y^{(i)}) = \alpha_i$ for all $i \in [m]$ and $\mu(y) = 0$ for $y \in \mathbb{R}^n \setminus \{y^{(1)}, \ldots, y^{(m)}\}$. Then, μ is a degree-d pseudo-distribution with $\mu \models \mathcal{A}$ and $\tilde{\mathbb{E}}_\mu f < 0$. □

Soundness and completeness

It turns out that the duality between pseudo-distributions and sum-of-squares proofs is related to the idea that sum-of-squares proofs are a sound and complete proof system when allowing pseudo-distributions as models.

We remark that the following lemmas about soundness and completeness do not contain significant new ideas beyond the duality of pseudo-distributions and sos proofs. However, they are useful

in order to reason about pseudo-distributions and sos proofs in a composable way.

9. Lemma (Soundness). *Let μ be a pseudo-distribution and let $\mathcal{A}, \mathcal{B}$ be systems of polynomial constraints. Suppose μ satisfies $\mu \models_\ell \mathcal{A}$ and there exists a sum-of-squares proof $\mathcal{A} \vdash_{\ell'} \mathcal{B}$. Then, μ satisfies $\mu \vdash_{\ell \cdot \ell'} \mathcal{B}$.*

This soundness lemma shows that sum-of-squares proofs allow us to reason about properties of pseudo-distributions.

The following completeness theorem shows that sum-of-squares proof allow us to reason about all properties of pseudo-distributions (under the mild technical assumption that the system of polynomial constraints is explicitly bounded).

10. Lemma (completeness). *Let $d \geq \ell' \geq \ell$, and let $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}[x]$ be systems of polynomial constraints such that $\mathcal{A}$ contains a constraint of the form $M - \sum_{i=1}^n x_i^2 \geq 0$ for some $M \geq 0$. Suppose every degree-d pseudo-distribution μ that satisfies $\mu \models_\ell \mathcal{A}$ also satisfies $\mu \models_{\ell'} \mathcal{B}$, then for every $\varepsilon > 0$ there exists a sum-of-squares proof $\mathcal{A} \vdash_d \mathcal{B}_\varepsilon$, where $\mathcal{B}_\varepsilon$ is the system obtained from $\mathcal{B}$ by weakening each constraint by ε.*

General sum-of-squares algorithm

The following theorem shows that we can efficiently search through pseudo-distributions satisfying a system of polynomial constraints.

11. Theorem (general sum-of-squares algorithm). *There exists an algorithm that given d and a satisfiable, explicitly bounded system of polynomial constraints $\mathcal{A}$ over $\mathbb{R}^n$, outputs in time $n^{O(d)}$ a degree-d pseudo-distribution that approximately satisfies $\mu \models \mathcal{A}$ up to error 2^{-n}.*[7]

[7] We have not defined what it means for a pseudo-distribution to approximately satisfy a system of polynomial constraints. The idea is that we allow a small slack for all of the equations Eq. (7).

About approximation errors and bit complexity: We can still use the same kind of sum-of-squares proofs in order to argue about properties of pseudo-distributions that approximately satisfy a system of polynomial constraints. However, one caveat is that in order for approximation errors not to amplify we need to ensure that the bit complexity of the coefficients of the sum-of-squares proofs we apply are polynomially bounded. While there are examples that require sum-of-squares proofs with exponential bit complexity (O'Donnell [2016]), the proofs that arise in the settings we consider for designing algorithms have small bit complexity.

Sum-of-squares certificates over instance-independent variety

In this section, we discuss properties of pseudo-distributions and sos proofs in a setting that generalizes the Boolean hypercube but avoids some of issues that arise in the general case.

Let $\Omega \subseteq \mathbb{R}^n$ be an algebraic set (defined by a system of polynomial equations) such that $f_1, \ldots, f_m$ are a linear basis for the set of polynomials of degree at most d that vanish over Ω,

$$\text{Span}\{f_1, \ldots, f_m\} = \mathbb{R}[x]_{\leq d} \cap I(\Omega)\,. \tag{8}$$

(Here, $I(\Omega)$ denotes the set of polynomials that vanish over Ω.) For many interesting choices of Ω (e.g., the Boolean hypercube and the Euclidean sphere), we can construct such a basis $f_1, \ldots, f_m$ in time $n^{O(d)}$.

We will show that if we are given such a basis for the degree-d part of the ideal of Ω, then much about sum-of-squares proofs and pseudo-distributions works like for the hypercube.

12. Theorem (Sum-of-squares duality over simple varieties). *Let Ω and $f_1, \ldots, f_m \in \mathbb{R}[x]_{\leq d}$ be as in Eq. (8). Let $\mathcal{A} = \{f_1 = 0, \ldots, f_m = 0\}$. Then, every $f \in \mathbb{R}[x]_{\leq d}$ satisfies exactly one of the following conditions:*

- *there exists a degree-d sos proof $\mathcal{A} \vdash_d \{f \geq 0\}$,*
- *there exists a degree-d pseudo-distribution $\mu\colon \mathbb{R}^n \to \mathbb{R}$ such that $\mu \models \mathcal{A}$ and $\tilde{\mathbb{E}}_\mu f < 0$.*

As for the hypercube, we can make this theorem algorithmic.

13. Theorem (Sum-of-squares algorithm over simple varieties). *For every even $d \in \mathbb{N}$, there exists an $n^{O(d)}$-time algorithm that given a basis $f_1, \ldots, f_m$ for $\mathbb{R}[x]_{\leq d} \cap I(\Omega)$ as in Eq. (8) and a polynomial $f \in \mathbb{R}[x]_{\leq d}$ outputs,*

- *either a degree-d sos proof $\mathcal{A} \vdash_d \{f \geq 0\}$,*
- *or a degree-d pseudo-distribution $\mu : \mathbb{R}^n \to \mathbb{R}$ such that $\mu \models \mathcal{A}$ and $\tilde{\mathbb{E}}_\mu f \leq 2^{-n}$.*

We remark that under mild assumptions on Ω, for every degree-d pseudo-distribution $\mu\colon \mathbb{R}^n \to \mathbb{R}$ such that $\mu \models \mathcal{A}$, there exists a degree-d pseudo-distribution $\mu'\colon \Omega \to \mathbb{R}$ with the same first d moments, so that $\tilde{\mathbb{E}}_{\mu(x)}(1, x)^{\otimes d} = \tilde{\mathbb{E}}_{\mu'(x)}(1, x)^{\otimes d}$.

References

Ryan O'Donnell. SOS is not obviously automatizable, even approximately. *Electronic Colloquium on Computational Complexity (ECCC)*, 23: 141, 2016.