Sum-of-squares hierarchies for polynomial optimization

Lucas Slot Monique Laurent

November 26, 2021

Polynomial optimization

We consider the problem of computing:

$$f_{\min} := \min_{x \in \mathbf{X}} f(x), \tag{POP}$$

where:

- $f \in \mathbb{R}[x]$ is a polynomial of degree d.
- $S = \{x \in \mathbb{R}^n : g_i(x) \geq 0 \text{ for } i \in [m]\}$ is a compact semialgebraic set.

In this talk:

- Binary hypercube: $\{0,1\}^n = \{x : 1 - x_i^2 = 0 \text{ for } i \in [n]\}$
- Hypersphere: $S^{n-1} = \{x : 1 - \|x\|^2 = 0\}$
- Ball: $B^n = \{x : 1 - \|x\|^2 \geq 0\}$
- Simplex: $\Delta^n = \{x : x \geq 0,\ 1 - \sum_i x_i \geq 0\}$

Sum-of-squares hierarchies

The problem (POP) can be reformulated as:

$$f_{\min} = \max\{\lambda \in \mathbb{R} : f - \lambda \in \mathcal{P}(S)\}, \quad \mathcal{P}(S) := \{p : p(x) \geq 0 \text{ for } x \in S\}$$

- Relax by choosing a smaller and simpler set $Q \subseteq \mathcal{P}(S)$:

$$f_{\min} \geq \max\{\lambda \in \mathbb{R} : f - \lambda \in Q\}$$

- We get the Putinar-type hierachy $f_{\text{sos},r} \leq f_{\min}$ if we choose the quadratic module:

$$Q = \mathcal{Q}_r(S) := \{\sum_{i=0}^{m} g_i\sigma_i : \sigma_i \in \Sigma[x],\ \deg(g_i\sigma_i) \leq 2r\}$$

- We get the Schmüdgen-type hierarchy $\overline{f}_{\text{sos},r} \leq f_{\min}$ if we choose the preordering:

$$Q = \mathcal{T}_r(S) := \{\sum_{I \subseteq [m]} g_I\sigma_I : \sigma_I \in \Sigma[x],\ \deg(g_I\sigma_I) \leq 2r\} \quad (g_I := \prod_{i \in I} g_i)$$

Convergence of the hierachies

We know that:

- $f_{\mathrm{sos},r} \to f_{\min}$ as $r \to \infty$ for Archimedean S. (Putinar's Positivstellensatz)
- $\overline{f}_{\mathrm{sos},r} \to f_{\min}$ as $r \to \infty$ for compact S. (Schmüdgen's Positivstellensatz)

Question: Can we quantify this convergence? That is, can we analyze the errors:

$$f_{\min} - f_{\mathrm{sos},r} \quad \text{and} \quad f_{\min} - \overline{f}_{\mathrm{sos},r} \quad ?$$

Convergence of the hierarchies

S	**relaxation**	**order of convergence**	**reference**
'compact'	$\mathcal{Q}_r$	$1/(\log r/c)^c \quad (c>0)$	[Schweighofer, 2004]
compact	$\mathcal{T}_r$	$1/r^c \quad (c>0)$	[Nie, Schweighofer, 2007]
$[-1,1]^n$	$\mathcal{T}_r$	$1/r$	[de Klerk, Laurent, 2010]
$[-1,1]^n$	$\mathcal{T}_r$	$1/r^2$	[Laurent, **S**., 2021]
$\{0,1\}^n$	$\mathcal{Q}_r \ (=\mathcal{T}_r)$	'Krawtchouk'	[Laurent, **S**., 2021]
S^{n-1}	$\mathcal{Q}_r \ (=\mathcal{T}_r)$	$1/r^2$	[Fang, Fawzi, 2020]
B^n	$\mathcal{Q}_r \ (=\mathcal{T}_r)$	$1/r^2$	[**S**., 2021]
Δ^n	$\mathcal{T}_r$	$1/r$	[Kirschner, de Klerk, 2021]
Δ^n	$\mathcal{T}_r$	$1/r^2$	[**S**., 2021]

The convergence rates in $O(1/r^2)$ for $\{0,1\}^n$, S^{n-1}, B^n and Δ^n are all established using the same proof technique:

1. Make use of Fourier analysis and reproducing kernels to define a parameter:
$$f_{\mathrm{harm},q} \leq f_{\min},$$
which depends on a given univariate sum of squares q.
2. Show that if q has degree $2r$, we then have:
$$f_{\mathrm{harm},q} \leq f_{\mathrm{sos},r} \leq f_{\min}$$
3. Use results on the roots of orthogonal polynomials to find q with:
$$f_{\mathrm{harm},q} - f_{\min} = O(1/r^2).$$

$$\implies f_{\mathrm{sos},r} - f_{\min} = O(1/r^2).$$

Orthogonal decompositions

- Let $\mathbf{X} \subseteq \mathbb{R}^n$ be a compact set and μ a finite Borel measure supported on $\mathbf{X}$.
- We have an inner product on the space $\mathcal{R}$ of polynomials on $\mathbf{X}$:

$$\langle p, q \rangle := \int_{\mathbf{X}} p(\mathbf{x}) q(\mathbf{x}) d\mu(\mathbf{x}).$$

- We write $\{P_\alpha : \alpha \in \mathbb{N}^n\}$ for an orthonormal basis of $\mathcal{R}$ with $\deg(P_\alpha) = |\alpha|$.
- The space $\mathcal{R}$ decomposes as:

$$\mathcal{R} = \bigoplus_{k=0}^{\infty} H_k, \quad H_k := \operatorname{span}\{P_\alpha : |\alpha| = k\}$$

- A polynomial $p \in \mathcal{R}$ can be written as: $p = p_0 + p_1 + \ldots + p_d, \quad p_k \in H_k.$

Example

For the hypersphere S^{n-1} with the uniform probability measure, we have:

$$\mathcal{R} = \bigoplus_{k=0}^{\infty} \mathrm{Harm}_k, \quad \mathrm{Harm}_k := \operatorname{span}\{p : p \text{ is homogeneous, harmonic of degree } k\}$$

Kernel operators

- A kernel $\mathrm{K} : \mathbf{X} \times \mathbf{X} \to \mathbb{R}$ on $\mathbf{X}$ induces a linear operator $\mathbf{K} : \mathcal{R} \to \mathcal{R}$ via:

$$\mathbf{K}p(\mathbf{x}) = \langle \mathrm{K}(\mathbf{x}, \cdot), p(\cdot) \rangle = \int_{\mathbf{X}} \mathrm{K}(\mathbf{x}, \mathbf{y}) p(\mathbf{y}) d\mu(\mathbf{y}).$$

- The kernel $\mathrm{CD}_d(\mathbf{x}, \mathbf{y}) := \sum_{|\alpha| \le d} P_\alpha(\mathbf{x}) P_\alpha(\mathbf{y})$ has the property that:

$$\mathbf{CD}_d p(\mathbf{x}) = p(\mathbf{x}) \quad (p \text{ has degree } d).$$

- This kernel CD_d is called the reproducing kernel or Christoffel-Darboux kernel of degree d.
- For highly symmetrical $\mathbf{X}$, the kernel $\mathrm{CD}_d(\mathbf{x}, \mathbf{y})$ has a simpler expression.

Example (Funk-Hecke formula)

For the hypersphere S^{n-1} with the uniform probability measure, we have:

$$\mathrm{CD}_d(\mathbf{x}, \mathbf{y}) = \sum_{k=0}^{d} \mathcal{G}_k^n (\mathbf{x} \cdot \mathbf{y}),$$

where $\mathcal{G}_k^n$ is the Gegenbauer polynomial of degree k (and parameter n).

The harmonic bound on S^{n-1}

- Let $f \in \mathcal{R}$ be a polynomial of degree d, to be minimized on S^{n-1}.
- Consider a univariate polynomial:

$$q(x) = \sum_{k=0}^{2r} \widehat{q}_k \mathcal{G}_k^n(x) \quad (\widehat{q}_k \neq 0 \text{ for } k \leq d).$$

- Define the kernel $\mathrm{K}_q(\mathbf{x}, \mathbf{y}) := q(\mathbf{x} \cdot \mathbf{y})$ on S^{n-1}.
- By the Funk-Hecke formula, we have:

$$\mathbf{K}_q f(\mathbf{x}) = \sum_{k=0}^{d} \widehat{q}_k f_k(\mathbf{x}),$$

and

$$\mathbf{K}_q^{-1} f(\mathbf{x}) = \sum_{k=0}^{d} (1/\widehat{q}_k) f_k(\mathbf{x}).$$

- If the $\widehat{q}_k \approx 1$, then K_q approximates the Christoffel-Darboux kernel (in some sense).

The harmonic bound on S^{n-1} (cont.)

Proposition (Fang and Fawzi, 2020)

Suppose that q is a sum of squares of degree $2r$, and that $\widehat{q}_0 = 1$. Then we have:

$$f_{\mathrm{harm},q} := \min_{\mathbf{x} \in S^{n-1}} \mathbf{K}_q^{-1} f(\mathbf{x}) \leq f_{\mathrm{sos},r}.$$

Proof.

The kernel $\mathrm{K}_q(\mathbf{x}, \mathbf{y}) = q(\mathbf{x} \cdot \mathbf{y})$ lies in the quadratic module $\mathcal{Q}_r(S^{n-1})$ for fixed $\mathbf{y}$. This gives a representation of $f - f_{\mathrm{harm},q}$ in $\mathcal{Q}_r(S^{n-1})$ via:

$$\begin{aligned} f(\mathbf{x}) - f_{\mathrm{harm},q} &= \mathbf{K}_q \mathbf{K}_q^{-1} f(\mathbf{x}) - f_{\mathrm{harm},q} \\ &= \mathbf{K}_q \left(\mathbf{K}_q^{-1} f(\mathbf{x}) - f_{\mathrm{harm},q} \right) \\ &= \int_{S^{n-1}} \underbrace{\mathrm{K}_q(\mathbf{x}, \mathbf{y})}_{\in \mathcal{Q}_r(S^{n-1})} \cdot \underbrace{(\mathbf{K}_q^{-1} f(\mathbf{y}) - f_{\mathrm{harm},q})}_{\geq 0} d\mu(\mathbf{y}). \end{aligned}$$

□

This does not give an explicit representation of $f(\mathbf{x}) - f_{\mathrm{harm},q}$.

Analyzing the bound on S^{n-1}

Theorem (Fang and Fawzi, 2020)

For $r \geq d$, there exist a sum of squares q of degree $2r$ with $\widehat{q}_0 = 1$ and:

$$f_{\min} - f_{\text{harm},q} = f_{\min} - \min_{\mathbf{x} \in S^{n-1}} \mathbf{K}_q^{-1} f(\mathbf{x}) = O(1/r^2).$$

As a result: $f_{\min} - f_{\text{sos},r} = O(1/r^2)$.

Proof.

Recall that: $\mathbf{K}_q^{-1} f(\mathbf{x}) = \sum_{k=0}^{d} (1/\widehat{q}_k) f_k(\mathbf{x})$. Therefore, for any $\mathbf{x} \in S^{n-1}$, we have:

$$\begin{aligned} |f(\mathbf{x}) - \mathbf{K}_q^{-1} f(\mathbf{x})| &= |\sum_{k=0}^{d} (1 - 1/\widehat{q}_k) f_k(\mathbf{x})| \\ &\leq \sum_{k=0}^{d} |1 - 1/\widehat{q}_k| \cdot \max_{\mathbf{x} \in S^{n-1}} |f_k(\mathbf{x})|. \end{aligned}$$

So: $f_{\min} - f_{\text{harm},q} \leq \sum_{k=0}^{d} |1 - 1/\widehat{q}_k| \cdot \max_{\mathbf{x} \in S^{n-1}} |f_k(\mathbf{x})|$.

Proof (cont.)

It remains to analyze $\sum_{k=0}^{d} |1 - 1/\widehat{q}_k|$ and $\max_{\mathbf{x} \in S^{n-1}} |f_k(\mathbf{x})|$.

Claim: There exists a sum of squares q of degree $2r$ with $\widehat{q}_0 = 1$ and:

$$\tau(q) := \sum_{k=0}^{d} |1 - 1/\widehat{q}_k| = O(1/r^2).$$

- Best choice: $\tau(q) \approx$ least root of $\mathcal{G}_{r+1}^n$ (the degree $r+1$ Gegenbauer polynomial)
- Is a special case of Lasserre's measure-based hierarchy of upper bounds for the minimization of $h(x) := d - \sum_{k=0}^{d} \mathcal{G}_k^n(x)$ on $[-1, 1]$.

Claim: We may bound:

$$\max_{\mathbf{x} \in S^{n-1}} |f_k(\mathbf{x})| \leq \gamma_d \max_{\mathbf{x} \in S^{n-1}} |f(\mathbf{x})| \quad (k \leq d)$$

where, $\gamma_d > 0$ is a constant depending only on the degree d of f (but not on n).

- Proven by Fang and Fawzi for S^{n-1} using an inductive proof.
- Also true on $\{0,1\}^n$ (our proof applies to S^{n-1} with a better constant).
- Weaker statement holds for the unit ball and simplex.

□

Intermezzo: connection to last week's talk

Proposition (last week's talk by Mauricio Velasco)

Let $\mathcal{W} = \{(\mathbf{y}_j, w_j) : j \in [N]\}$ be a positive cubature rule for S^{n-1} of strength $2r + d$. Suppose that q is a sum of squares of degree $2r$, and that $\widehat{q}_0 = 1$. Then we have:

$$f_{\mathrm{cub},q} := \min_{j=1,2,\dots,N} \mathbf{K}_q^{-1} f(\mathbf{y}_j) \leq f_{\mathrm{sos},r}.$$

Proof.

The kernel $\mathrm{K}_q(\mathbf{x}, \mathbf{y}) = q(\mathbf{x} \cdot \mathbf{y})$ lies in the quadratic module $\mathcal{Q}_r(S^{n-1})$ for fixed $\mathbf{y}$. This gives a representation of $f - f_{\mathrm{cub},q}$ in $\mathcal{Q}_r(S^{n-1})$ via:

$$\begin{aligned} f(\mathbf{x}) - f_{\mathrm{cub},q} &= \int_{S^{n-1}} \mathrm{K}_q(\mathbf{x}, \mathbf{y}) \cdot (\mathbf{K}_q^{-1} f(\mathbf{y}) - f_{\mathrm{cub},q}) d\mu(\mathbf{y}) \\ &= \sum_{j=1}^{N} \underbrace{\mathrm{K}_q(\mathbf{x}, \mathbf{y}_j)}_{\in \mathcal{Q}_r(S^{n-1})} \cdot \underbrace{w_j (\mathbf{K}_q^{-1} f(\mathbf{y}_j) - f_{\mathrm{cub},q})}_{\geq 0} \end{aligned}$$

□

So we have:

$$f_{\mathrm{harm},q} = \min_{\mathbf{x} \in S^{n-1}} \mathbf{K}_q^{-1} f(\mathbf{x}) \leq f_{\mathrm{cub},q} \leq f_{\mathrm{sos},r}$$

What about the binary cube, the unit ball and the simplex?

- The binary cube $\{0,1\}^n$ is very similar to S^{n-1}. The reproducing kernel is given by:

$$\mathrm{CD}_d(\mathbf{x},\mathbf{y}) = \sum_{k=0}^{d} \mathcal{K}_k\big(d_{\mathrm{ham}}(\mathbf{x},\mathbf{y})\big),$$

  where $\mathcal{K}_k$ is the Krawtchouk polynomial of degree k.
- Gives an analysis of $f_{\min} - f_{\mathrm{sos},r}$ in terms of roots of Krawtchouk polynomials.
- Proof for the unit ball and the simplex follows the same structure, but the expressions of the reproducing kernel are more complicated.
- Convergence rate in $O(1/r^2)$ for $f_{\min} - f_{\mathrm{sos},r}$ on B^n (Putinar = Schmüdgen).
- Convergence rate in $O(1/r^2)$ for $f_{\min} - f_{\mathrm{sos},r}$ on Δ^n (Schmüdgen).

The unit ball

Theorem (Xu, 1999)

The reproducing kernel $\mathrm{CD}_d(\mathbf{x},\mathbf{y})$ *on the unit ball w.r.t. the measure* $(1-\|\mathbf{x}\|^2)^{-1/2}d\mathbf{x}$ *is given by:*

$$\mathrm{CD}_d(\mathbf{x},\mathbf{y}) = \frac{1}{2}\sum_{k=0}^{d}\Big(\mathcal{G}_k^{n+1}(\mathbf{x}\cdot\mathbf{y}+\sqrt{1-\|\mathbf{x}\|^2}\sqrt{1-\|\mathbf{y}\|^2}) \quad + \\ \mathcal{G}_k^{n+1}(\mathbf{x}\cdot\mathbf{y}-\sqrt{1-\|\mathbf{x}\|^2}\sqrt{1-\|\mathbf{y}\|^2})\Big)$$

- For a univariate polynomial $q(x)=\sum_{k=0}^{2r}\widehat{q}_k\mathcal{G}_k^{n+1}(x)$, we set:

$$\mathrm{K}_q(\mathbf{x},\mathbf{y}) := \frac{1}{2}\Big(q(\mathbf{x}\cdot\mathbf{y}+\sqrt{1-\|\mathbf{x}\|^2}\sqrt{1-\|\mathbf{y}\|^2}) \quad + \\ q(\mathbf{x}\cdot\mathbf{y}-\sqrt{1-\|\mathbf{x}\|^2}\sqrt{1-\|\mathbf{y}\|^2})\Big)$$

- Like before, we then have:

$$\mathbf{K}_q f = \sum_{k=0}^{d}\widehat{q}_k f_k \quad \text{and} \quad \mathbf{K}_q^{-1} f = \sum_{k=0}^{d}(1/\widehat{q}_k) f_k$$

Proposition (**S.**, 2021)

Suppose that q is a sum of squares of degree $2r$, and that $\widehat{q}_0 = 1$. Then we have:

$$f_{\mathrm{harm},q} := \min_{\mathbf{x} \in B^n} \mathbf{K}_q^{-1} f(\mathbf{x}) \leq f_{\mathrm{sos},r}.$$

Proof.

We only need to show that $\mathrm{K}_q(\mathbf{x}, \mathbf{y}) \in \mathcal{Q}_r(B^n)$ for fixed $\mathbf{y}$. The terms in Xu's formula cancel out precisely as we need them to. □

- The rest of the analysis is essentially the same:

$$f_{\min} - f_{\mathrm{harm},q} \leq \sum_{k=0}^{d} |1 - 1/\widehat{q}_k| \cdot \max_{\mathbf{x} \in B^n} |f_k(\mathbf{x})|$$

- For the first factor we use Fang and Fawzi's result again.
- For the second factor we need a new argument.
- Proof for the simplex is analogous, but with a different expression for the reproducing kernel.

Conclusion

Summary

- For semialgebraic sets $\mathbf{X}$ with nice symmetric structure, we can analyze the behaviour of the sum-of-squares hierarchy using Fourier analysis and reproducing kernels.
- For S^{n-1} and $\{0, 1\}^n$, one uses the classical Funk-Hecke formula.
- For the unit ball and simplex, we uses expressions derived by Xu.
- The convergence rate links back to the roots of orthogonal polynomials (or: to the measure-based hierarchy).

Open questions

- Can we extend to more sets? Perhaps by adding simple (e.g. linear) contraints?
- Can we improve our choice of univariate polynomial q? Perhaps by using the function f somehow?
- Can we apply this technique in the non-commutative setting?