The polynomial kernel method and sum-of-squares hierarchies for polynomial optimization

Lucas Slot Monique Laurent

November 1, 2021

Polynomial optimization

We consider the problem of computing:

$$f_{\min} := \min_{x \in S} f(x),$$

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

Some examples of semialgebraic sets:

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

Example (MaxCut)

For the complete graph K_n with edge-weights $w_{ij} \geq 0$, we have:

$$\textsc{MaxCut}(K_n, w) = \max_{x \in [-1,1]^n} \sum_{1 \leq i < j \leq n} w_{ij}(x_i - x_j)^2.$$

Example (StableSet)

The stability number of a graph $G = ([n], E)$ can be computed as:

$$\alpha(G) = \max_{x \in \{-1,1\}^n} \sum_{i \in [n]} x_i - \sum_{\{i,j\} \in E} x_i x_j.$$

Example (Motzkin, Straus (1965))

Alternatively, we have:

$$\frac{1}{\alpha(G)} = \min_{x \in \Delta^n} \sum_{i \in [n]} x_i^2 + 2 \sum_{\{i,j\} \in E} x_i x_j.$$

Polynomial optimization (cont.)

- Polynomial optimization is generally intractable
- This motivates the search for efficient bounds on the optimum
- In this talk: two hierarchies of lower bounds due to Lasserre
- Both are based on relaxing nonnegativity to a sum-of-squares condition.

Definition

A polynomial $p \in \mathbb{R}[x]$ is called a sum of squares if it can be written as:

$$p(x) = p_1(x)^2 + p_2(x)^2 + \ldots + p_k(x)^2.$$

Note that such a polynomial is globally nonnegative. We write $\Sigma[x]$ for the set of all sum-of-squares polynomials.

Two sum-of-squares hierarchies

We can rewrite:

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

- Connection between minimization and verifying nonnegativity for polynomials.
- Checking membership of $\mathcal{P}(S)$ is still hard.
- 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 a bound $f_{(r)} \leq f_{\min}$ by choosing the quadratic module:

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

- We get a (stronger) bound $\overline{f}_{(r)} \leq f_{\min}$ by choosing the preordering:

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

- Membership of $\mathcal{Q}_r(S)$ and $\mathcal{T}_r(S)$ can be checked with semidefinite programming

Recap

- We have two hierarchies of lower bounds on $f_{\min}$:

$$f_{(r)} \leq f_{(r+1)} \leq \overline{f}_{(r+1)} \leq \overline{f}_{(r+2)} \leq f_{\min}$$

- $f_{(r)}$ and $\overline{f}_{(r)}$ can be computed by solving a semidefinite program

Convergence of the hierachies

- $f_{(r)} \to f_{\min}$ as $r \to \infty$ for 'compact' S (Putinar's Positivstellensatz)
- $\overline{f}_{(r)} \to f_{\min}$ as $r \to \infty$ for compact S (Schmüdgen's Positivstellensatz)
- Question: Can we quantify this convergence? That is, can we analyze as a function of r the errors:

$$f_{\min} - f_{(r)} \quad \text{and} \quad f_{\min} - \overline{f}_{(r)} \quad ?$$

Convergence of the hierarchies

S	**relaxation**	**order of convergence**	**citation**
'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]
Δ^n	$\mathcal{T}_r$	$1/r$	[Kirschner, de Klerk, 2021]
S^{n-1}	$\mathcal{Q}_r \ (= \mathcal{T}_r)$	$1/r^2$	[Fang, Fawzi, 2020]
$\{-1, 1\}^n$	$\mathcal{Q}_r \ (= \mathcal{T}_r)$	'Krawtchouk'	[Laurent, **S**., 2021]
$[-1, 1]^n$	$\mathcal{T}_r$	$1/r^2$	[Laurent, **S**., 2021]
B^n, Δ^n	$\mathcal{T}_r$	$1/r^2$	[**S**., (work in progress)]

The last four results all use the polynomial kernel method

The polynomial kernel method [Fang, Fawzi 2020]

Goal: For a given polynomial $f \geq 0$ on S, show that there exists a small $\lambda > 0$ such that $f + \lambda$ lies in $\mathcal{Q}_r(S)$. This is equivalent to showing $f_{\min} - f_{(r)} \leq \lambda$.

- Consider a polynomial kernel $K(x, y)$ on S with:

$$x \mapsto K(x, y) \in \mathcal{Q}_r(S) \text{ for fixed } y \in S$$

- After choosing a measure μ on S, the kernel K induces a linear operator $\mathbf{K}$ on $\mathbb{R}[x]$ by:

$$\mathbf{K}p(x) := \int_S K(x, y)p(y)d\mu(y) \quad (p \in \mathbb{R}[x])$$

- If $p \geq 0$ on S, then $\mathbf{K}p$ lies in $\mathcal{Q}_r(S)$ (!)
- If we choose λ big enough s.t. $\mathbf{K}^{-1}(f + \lambda) \geq 0$ on S, we find that:

$$f + \lambda = \mathbf{K} \underbrace{\mathbf{K}^{-1}(f + \lambda)}_{\geq 0} \text{ lies in } \mathcal{Q}_r(S)$$

- This immediately implies: $f_{\min} - f_{(r)} \leq \lambda$

The polynomial kernel method (cont.)

Problems

- How do we ensure that $x \mapsto K(x, y) \in \mathcal{Q}_r(S)$?
  $\rightarrow$ case-by-case argument
- How do we ensure that $\mathbf{K}^{-1}(f + \lambda) \geq 0$ on S?
  $\rightarrow$ make sure that $\mathbf{K} \approx \mathrm{Id}$, meaning its eigenvalues are close to 1

Constructing kernels

- On the hypersphere and $\{-1, 1\}^n$, one can use Fourier analysis/symmetry to reduce to a univariate setting.
- On the unit ball and simplex, one can use closed forms of the Christoffel-Darboux kernel, again reducing to a univariate setting.
- On $[-1, 1]^n$, one can use the Jackson kernel, which is a well-known kernel from functional approximation.

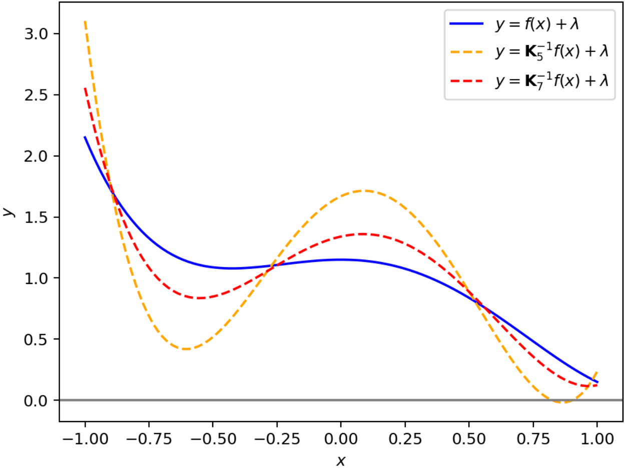
An example on $[-1, 1]$

- Consider the polynomial $f(x) = 1 - x^2 - x^3 + x^4$, which is nonnegative on $[-1, 1]$
- The Jackson kernel of degree r satisfies $x \mapsto K_r(x, y) \in \mathcal{T}_r([-1, 1])$
- If we set:

$$\mathbf{K}_r p(x) := \int_{-1}^{1} K_r(x, y) p(y) d\mu(y),$$

  it is known that $\mathbf{K}_r \to \mathrm{Id}$ as $r \to \infty$. To be more precise, its eigenvalues tend to 1 at a rate in $O(1/r^2)$.
- What happens if we apply the inverse operator $\mathbf{K}_r^{-1}$ to $f + \lambda$, setting $\lambda = 0.15$?

An example on $[-1, 1]$



Conclusion

Summary

- Polynomial optimization captures hard combinatorial problems.
- Sum-of-squares hierarchies provide tractable lower bounds on the optimum using semidefinite programming
- The polynomial kernel method allows one to show guarantees on the quality of these bounds in certain special cases
- Examples include the hypersphere, the binary cube the unit box, unit ball and simplex

Open questions

- So far, we mostly have results for the (expensive) bounds based on the preordering $\mathcal{T}_r$. Can we also get results for the bounds based on the quadratic module $\mathcal{Q}_r$?
- Can we add simple constraints (e.g. linear)?
- Can we extend to the noncommutative setting?