Sum-of-squares hierarchies for binary polynomial optimization

Lucas Slot Monique Laurent

Polynomial optimization on the binary cube

We consider the problem of computing:

$$f_{\min} := \min_{x \in \mathbb{B}^n} f(x),$$

where

- $f \in \mathbb{R}[x]$ is a polynomial of degree d.
- $\mathbb{B}^n := \{0, 1\}^n \subseteq \mathbb{R}^n$ is the boolean hypercube.

Example (MaxCut)

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

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

Example (StableSet)

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

$$\textsc{Stab}(G) = \max_{x \in \mathbb{B}^{|V|}} \sum_{i \in V} x_i - \sum_{\{i,j\} \in E} x_i x_j.$$

Any binary polynomial optimization problem with linear constraints can be reformulated as an unconstrained binary optimization problem.

Semidefinite hierarchies for polynomial optimization

- (Binary) polynomial optimization is generally intractable
- This motivates the search for efficient bounds on the optimum
- Many such bounds have been proposed based on lift-and-project methods
- Today: two semidefinite hierarchies due to Lasserre

The outer Lasserre hierachy (sum-of-squares hierarchy)

Observation
We can rewrite:

$$f_{\min} = \max\{\lambda \in \mathbb{R} : f - \lambda \geq 0 \text{ on } \mathbb{B}^n\}$$

- The nonnegativity condition can be relaxed to a sum-of-squares condition:

$$\max\{\lambda \in \mathbb{R} : f - \lambda \text{ is a sum-of-squares on } \mathbb{B}^n\},$$

  meaning $f(x) - \lambda = \sum_i p_i^2(x)$ for all $x \in \mathbb{B}^n$ for certain $p_i \in \mathbb{R}[x]$

- We can then put a bound on the degree of the p_i:

$$f_{(r)} = \max\{\lambda \in \mathbb{R} : f - \lambda \text{ is a sum-of-squares of degree } \leq 2r \text{ on } \mathbb{B}^n\},$$

  meaning $f(x) - \lambda = \sum_i p_i^2(x)$ for all $x \in \mathbb{B}^n$ for certain $p_i \in \mathbb{R}[x]$ of degree at most r

The outer Lasserre hierachy (sum-of-squares hierarchy)

Definition (Lasserre, 2001)

For $r \in \mathbb{N}$, define:

$$f_{(r)} = \max\{\lambda \in \mathbb{R} : f - \lambda \text{ is a sum-of-squares of degree } \leq 2r \text{ on } \mathbb{B}^n\}$$

- $f_{(r)} \leq f_{(r+1)} \leq f_{\min}$
- For fixed r, $f_{(r)}$ can be computed efficiently using semidefinite programming
- There is a one-to-one correspondence between semidefinite matrices and sum-of-squares polynomials:

$$p(x) \text{ is a sum-of-squares} \iff \exists M \succeq 0 \text{ with } p(x) = [x]^\top M[x],$$

where $[x] = (x^\alpha)_\alpha$ is the vector of monomials up to degree $\deg(p)/2$

Observation

We can rewrite:

$$f_{\min} = \min_{\nu} \left\{ \int_{\mathbb{B}^n} f d\nu : \int_{\mathbb{B}^n} d\nu = 1 \right\}$$

Definition (Lasserre, 2010)

Let μ be the uniform measure on $\mathbb{B}^n$. For $r \in \mathbb{N}$, define:

$$f^{(r)} = \min_{s \in \Sigma_r[x]} \left\{ \int_{\mathbb{B}^n} f \cdot s d\mu : \int_{\mathbb{B}^n} s d\mu = 1 \right\}$$

- $f^{(r)} \geq f^{(r+1)} \geq f_{\min}$
- For fixed r, $f^{(r)}$ can be computed efficiently using SDP
- In principle we could choose a different reference measure μ.

Summary

We have the hierarchies:

$$\text{(outer)} \quad f_{(r)} = \max\{\lambda \in \mathbb{R} : f - \lambda \text{ is sos of degree } \leq 2r \text{ on } \mathbb{B}^n\}$$

$$\text{(inner)} \quad f^{(r)} = \min_{s \in \Sigma_r[x]} \left\{ \int_{\mathbb{B}^n} f \cdot s d\mu : \int_{\mathbb{B}^n} s d\mu = 1 \right\}$$

satisfying:

$$f_{(r)} \leq f_{\min} \leq f^{(r)} \leq f_{\max}$$

Question

What can be said about the quality of these hierarchies? That is, can we bound:

$$\frac{f_{\min} - f_{(r)}}{\|f\|_\infty} \quad \text{and} \quad \frac{f^{(r)} - f_{\min}}{\|f\|_\infty} \quad ?$$

(here, $\|f\|_\infty := \max_{x \in \mathbb{B}^n} |f(x)|$)

Both hierarchies can be defined more generally for polynomial optimization over sets $K \subseteq \mathbb{R}^n$ other than the cube. Depending on K, their quality has been investigated:

(inner) $$\frac{f^{(r)} - f_{\min}}{\|f\|_\infty} = \begin{cases} O(1/r^2) & \text{if } K = [-1,1]^n \quad \text{(de Klerk, Laurent)} \\ O(1/r^2) & \text{if } K = S^{n-1} \quad \text{(de Klerk, Laurent)} \\ O(1/r^2) & \text{if } K = B^n, \Delta^n, \text{`nice'} \quad (\textbf{S.}, \text{Laurent}) \\ O(\log^2 r/r^2) & \text{for `general' } K \quad (\textbf{S.}, \text{Laurent}) \end{cases}$$

(outer) $$\frac{f_{\min} - f_{(r)}}{\|f\|_\infty} = \begin{cases} O(1/\log(r/c)^{1/c}) & \text{if } K \text{ `compact'} \quad \text{(Nie, Schweighofer)} \\ O(1/r) & \text{if } K = S^{n-1} \quad \text{(Doherty, Wehner)} \\ O(1/r^2) & \text{if } K = S^{n-1} \quad \text{(Fang, Fawzi)} \end{cases}$$

Note: The inner hierarchy is much better understood than the outer hierarchy

Back to the binary cube

$$\text{(outer)} \quad f_{(r)} = \max\{\lambda \in \mathbb{R} : f - \lambda \text{ is sos of degree } \leq 2r \text{ on } \mathbb{B}^n\}$$

$$\text{(inner)} \quad f^{(r)} = \min_{s \in \Sigma_r[x]} \left\{ \int_{\mathbb{B}^n} f \cdot s d\mu : \int_{\mathbb{B}^n} s d\mu = 1 \right\}$$

Known results

- Finite convergence for the outer hierachy:

$$f_{(r)} = f_{\min} \text{ when } r \geq \frac{n+d-1}{2} \approx \frac{1}{2} n$$

  [Fawzi, Saunderson, Parrilo 2016 $(d = 2)$] [Sakaue et al. 2017 $(d > 2)$]
- Finite convergence for the inner hierachy:

$$f^{(r)} = f_{\min} \text{ when } r \geq n$$

- But, nothing is known when the bounds are not exact

Theorem (Main result on the outer hierachy)

Let $f \in \mathbb{R}[x]_d$. Let ξ_{r+1}^n be the least root of the degree $r+1$ Krawtchouk polynomial (with parameter n). Then for any n and $r \leq \frac{1}{2}n$ large enough, we have:

$$\frac{f_{\min} - f_{(r)}}{\|f\|_\infty} \leq C_d \cdot (\xi_{r+1}^n / n)$$

Theorem (Main result on the inner hierachy)

For any n and $r \leq \frac{1}{2}n$, we have:

$$\frac{f^{(r)} - f_{\min}}{\|f\|_\infty} \leq \frac{1}{2} C_d \cdot (\xi_{r+1}^n / n)$$

Main new results

Theorem (Levenshtein)

For $t \in [0, 1/2]$, write:

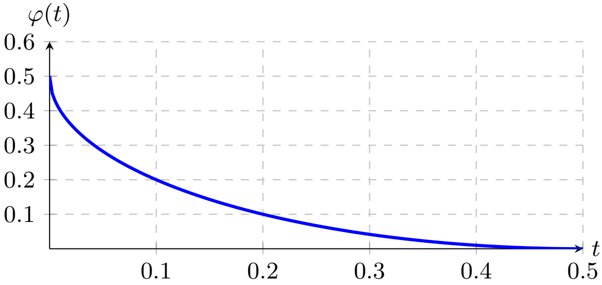
$$\varphi(t) = 1/2 - \sqrt{t(1-t)}.$$

Then the least root ξ_r^n of the degree $r+1$ Krawtchouk polynomial satisfies:

$$\xi_r^n / n \leq \varphi(r/n) + c \cdot (r/n)^{-1/6} \cdot n^{-2/3}$$

- So, roughly:

$$\frac{f_{\min} - f_{(r)}}{\|f\|_\infty} \lessapprox \varphi(r/n)$$



Key steps of the proof for the outer hierarchy

Observation

We may assume for the proof that $f_{\min} = f(0) = 0$ and $\|f\|_\infty = 1$.

Goal: Show that there exists a small $\lambda > 0$ such that $f + \lambda$ is a sum-of-squares on $\mathbb{B}^n$ of degree at most $2r$.

1. Use the polynomial kernel technique to produce sum-of-squares representations [Fang, Fawzi 2020]
2. Perform a symmetry reduction using classical Fourier analysis on $\mathbb{B}^n$
3. Link the reduced problem to an analysis of the inner hierarchy in a univariate setting
4. Exploit a known connection between the inner hierarchy and extremal roots of orthogonal polynomials (Krawtchouk)

Our result for the inner hierarchy can be extrapolated from steps 3 and 4.

Step 1: The polynomial kernel technique (Fang, Fawzi 2020)

Goal: Find a sum-of-squares representation of $f + \lambda$ for some small $\lambda > 0$.

- Consider a polynomial kernel of the form:

$$K(x, y) := q^2(d_{ham}(x, y)) \quad (x, y \in \mathbb{B}^n),$$

  with $q \in \mathbb{R}[t]_r$ a univariate polynomial to be chosen later
- The kernel K induces a linear operator on $\mathbb{R}[x]$ by:

$$\mathbf{K}p(x) := \int_{\mathbb{B}^n} p(y)K(x, y)d\mu(y) = \frac{1}{2^n} \sum_{y \in \mathbb{B}^n} p(y)K(x, y)$$

- When $p \geq 0$ on $\mathbb{B}^n$, then $\mathbf{K}p$ is sos of degree $\leq 2r$ on $\mathbb{B}^n$ (!)
- If we choose λ big enough s.t. $\mathbf{K}^{-1}(f + \lambda) \geq 0$ on $\mathbb{B}^n$, we find that

$$f + \lambda = \mathbf{K} \underbrace{\mathbf{K}^{-1}(f + \lambda)}_{\geq 0} \text{ is sos of degree } \leq 2r \text{ on } \mathbb{B}^n$$

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

Step 2, 3 and 4:

Problem: How do we ensure that $\mathbf{K}^{-1}(f+\lambda) \geq 0$ on $\mathbb{B}^n$?

- We want the eigenvalues of $\mathbf{K}$ to be as close as possible to 1 (so that $\mathbf{K} \approx \mathrm{Id}$).

2. If $K(x,y) = q^2(d(x,y))$, then the eigenvalues λ_i of $\mathbf{K}$ are given by the decomposition:

$$q^2(t) = \sum_{i=0}^{2r} \lambda_i \mathcal{K}_i(t) \quad \text{where } \mathcal{K}_i \text{ is the Krawtchouk polynomial}$$

3. Selecting the univariate polynomial q so that the λ_i are as close as possible to 1 corresponds to a univariate instance of the inner hierarchy
4. In the univariate case, the inner hierarchy is well understood, and the behaviour can be described using orthogonal polynomials [de Klerk, Laurent 2020]

Concluding remarks

- We have shown a guarantee on the outer hierarchy $f_{\min} - f_{(r)}$ using a connection to (a special case of) the inner hierachy
- The treatment of this special case can be extended to obtain our result on the inner hierarchy
- As far as we know, this is the first analysis in the setting $r < \frac{n+d-1}{2}$
- But, our results apply only in the setting $r \approx t \cdot n$. In particular they give no information for fixed $r \in \mathbb{N}$
- Open question: is it possible to add (linear) constraints?

Some references

Hamza Fawzi, James Saunderson, and Pablo A. Parrilo.
Sparse sums of squares on finite abelian groups and improved semidefinite lifts.
Math. Prog., 160(1-2):149–191, 2016.

Shinsaku Sakaue, Akiko Takeda, Sunyoung Kim, and Naoki Ito.
Exact semidefinite programming relaxations with truncated moment matrix for binary polynomial optimization problems.
SIAM Journal on Optimization, 27(1):565–582, 2017.

Kun Fang and Hamza Fawzi.
The sum-of-squares hierarchy on the sphere, and applications in quantum information theory.
Math. Prog., 2020.

Etienne de Klerk and Monique Laurent.
Worst-case examples for Lasserre's measure–based hierarchy for polynomial optimization on the hypercube.
Mathematics of Operations Research, 45(1):86–98, 2020

V. I. Levenshtein.
Krawtchouk polynomials and universal bounds for codes and designs in hamming spaces.
IEEE Transactions on Information Theory, 41(5):1303–1321, 1995.