The Lasserre hierarchy for binary polynomial optimization

Lucas Slot Monique Laurent

CWI

MINOA
MIXED-INTEGER NON-LINEAR OPTIMISATION: ALGORITHMS AND APPLICATIONS

We consider the problem of computing:

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

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:

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

- BPOP is NP-hard in general
- Many techniques exist for approximation
- Today: two semidefinite hierarchies due to Lasserre

Observation

We can rewrite:

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

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 SDP

Question

What can be said of the quality of $f_{(r)}$, i.e., can we bound $f_{\min} - f_{(r)}$?

- Finite convergence: $f_{(r)} = f_{\min}$ when $r \geq \frac{n+d-1}{2}$
  [Fawzi, Saunderson, Parrilo 2016 $(d = 2)$] [Sakaue et al. 2017 $(d > 2)$]
- But, nothing is known for $r < \frac{n+d-1}{2}$, when the bound is not exact.

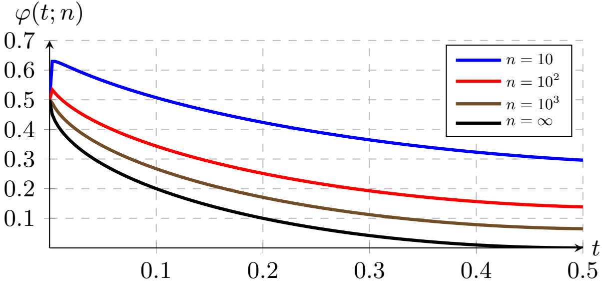
Theorem (Main result on the outer hierachy)

Let $f \in \mathbb{R}[x]_d$ *and choose* $r \in \mathbb{N}$ *such that* $t := r/n \in [0, 1/2]$*. Then:*

$$\frac{f_{\min} - f_{(r)}}{\|f\|_\infty} \leq C_d \underbrace{\left(1/2 - \sqrt{t(1-t)} + \frac{t^{1/6}\sqrt{1-t}}{n^{1/3}}\right)}_{\varphi(t;n)}$$

when $d(d+1) \cdot \varphi(t;n) \leq 1/2$.

- This analysis applies in the regime $r \approx t \cdot n$, and becomes sharper as $n \to \infty$



The inner (measure-based) Lasserre hierachy

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

Theorem (Main result on the inner hierachy)

Let $f \in \mathbb{R}[x]_d$ and choose $r \in \mathbb{N}$ such that $t := r/n \in [0, 1/2]$. Then:

$$\frac{f^{(r)} - f_{\min}}{\|f\|_\infty} \leq \frac{1}{2} C_d \left(1/2 - \sqrt{t(1-t)} + \frac{t^{1/6}\sqrt{1-t}}{n^{1/3}} \right)$$

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:

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

- We wish to bound:

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

- We focus on the outer hierarchy, but the inner hierarchy will play an important role in the proof

Key steps for analyzing the outer hierarchy

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)

Observation

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

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

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

- If we assume $\mathbf{K}(1) = 1$, we know that $\mathbf{K}^{-1}(f+\lambda) \geq 0$ on $\mathbb{B}^n$ if:

$$\|\mathbf{K}^{-1}p - p\|_\infty \leq \lambda \text{ for all } p \in \mathbb{R}[x]_d \text{ with } \|p\|_\infty = 1.$$

- We will bound this quantity by considering the eigenvalues of $\mathbf{K}$

Funk-Hecke formula

The eigenvalues of $K(x,y) = q^2(d(x,y))$ are given by the coefficients λ_i in the expansion of q^2 into the Krawtchouk polynomials $\mathcal{K}_i$:

$$q^2(t) = \sum_{i=0}^{2r} \lambda_i \mathcal{K}_i(t)$$

Step 2. Fourier analysis on $\mathbb{B}^n$ and symmetry reduction

Characters and Krawtchouk polynomials

- For $a \in \mathbb{B}^n$ define the character $\chi_a(x) := (-1)^{a \cdot x}$
- The characters form an ONB for the space $\mathcal{R}[x] := \mathbb{R}[x]/(x_i^2 = x_i)$ of polynomials on $\mathbb{B}^n$.
- Then, $\mathcal{R}[x]$ decomposes as:

$$\mathcal{R}[x] = H_0 \perp H_1 \perp \cdots \perp H_n, \quad H_i = \{\chi_a : |a| = i\}$$

  The components H_i are invariant and irreducible under the symmetries of $\mathbb{B}^n$ (permutations and bit-flips)
- We can write $p \in \mathcal{R}[x]_d$ as (harmonic decomposition):

$$p = p_0 + p_1 + \cdots + p_d \quad (p_i \in H_i)$$

- The Krawtchouk polynomials $\mathcal{K}_i$ are the orthogonal polynomials w.r.t. the measure $\omega = \sum_{t=0}^{n} \binom{n}{t} \delta_t$, with $\langle f, g \rangle_\omega = \int_0^n f \cdot g d\omega = \sum_{t=0}^{n} \binom{n}{t} f(t) g(t)$.
- Key fact: For $x, y \in \mathbb{B}^n$ with $d(x, y) = k$, we have:

$$\sum_{|a|=i} \chi_a(x)\chi_a(y) = \mathcal{K}_i(k)$$

Theorem (Funk-Hecke)

Let $q \in \mathbb{R}[t]_r$, and decompose q^2 as $q^2(t) = \sum_{i=0}^{2r} \lambda_i \mathcal{K}_i(t)$. Then the kernel $K(x,y) = q^2(d(x,y))$ satisfies:

$$\mathbf{K}p = \lambda_0 p_0 + \lambda_1 p_1 + \cdots + \lambda_d p_d \qquad \textit{for } p \in \mathcal{R}[x]_d$$

Proof.

Apply the key fact to show that $\mathbf{K}\chi_a = \lambda_{|a|}\chi_a$ for all $a \in \mathbb{B}^n$ □

Recall

We want to choose q such that $\mathbf{K}(1) = 1$ and

$$\|\mathbf{K}^{-1}p - p\|_\infty \text{ is small for all } p \in \mathbb{R}[x]_d \text{ with } \|p\|_\infty = 1$$

Upshot

Using Funk-Hecke, we find that if $\lambda_0 = 1$, then $\mathbf{K}(1) = 1$ and:

$$\|\mathbf{K}^{-1}p - p\|_\infty \leq \max_{k=0}^{d} \|p_k\|_\infty \cdot \sum_{i=1}^{d} |1 - \lambda_i^{-1}| \leq 2C_d \sum_{i=1}^{d} (1 - \lambda_i)$$

So we want a q with $\lambda_0 = 1$, and λ_i as close as possible to 1.

Step 3. Connection to the inner hierarchy

Goal

Find a univariate $q \in \mathbb{R}[t]_r$ for which the coefficients in $q^2(t) = \sum_{i=0}^{2r} \lambda_i \mathcal{K}_i(t)$ satisfy:

$$\lambda_0 = 1 \text{ and } \sum_{i=1}^{d} (1 - \lambda_i) \text{ is small}$$

- Recall that the $\mathcal{K}_i$ are orthogonal w.r.t. $\omega = \sum_{t=0}^{n} \binom{n}{t} \delta_t$ and so we have

$$\lambda_i = \langle \widehat{\mathcal{K}}_i, q^2 \rangle_\omega := \int_0^n \widehat{\mathcal{K}}_i \cdot q^2 d\omega,$$

  where $\widehat{\mathcal{K}}_i := \mathcal{K}_i / \|\mathcal{K}_i\|_\omega^2 = \mathcal{K}_i / \mathcal{K}_i(0)$.
- We thus achieve our goal by solving:

$$\inf_{q \in \mathbb{R}[t]_r} \left\{ \int_0^n g \cdot q^2 d\omega : \int_0^n q^2 d\omega = 1 \right\}, \text{ with } g(t) := d - \sum_{i=1}^{d} \widehat{\mathcal{K}}_i(t).$$

- This is just the inner Lasserre hierarchy for minimizing g on $[0, n]$ w.r.t. the measure ω!
- To summarize, we have: $f_{\min} - f_{(r)} \leq 2C_d \big(g_\omega^{(r)} - g_{\min} \big)$

Step 4. Analyzing the inner hierachy

Theorem (special case of de Klerk, Laurent 2020)

Let $\hat{g}(t) = ct$, $c > 0$ be a linear polynomial. Then:

$$\hat{g}_\omega^{(r)} - \hat{g}_{\min} = c \cdot \xi_{r+1},$$

where ξ_{r+1} is the least root of $\mathcal{K}_{r+1}$.

- Problem: $g(t) = d - \sum_{i=1}^{d} \widehat{\mathcal{K}}_i(t)$ is not linear!
- But, it is upper estimated by its linear approximation at $t = 0$:

$$g(t) \leq \hat{g}(t) := d(d+1) \cdot (t/n) \quad (t = 0, 1, \ldots, n)$$

- We may conclude:

$$g_\omega^{(r)} - g_{\min} \leq \hat{g}_\omega^{(r)} - \hat{g}_{\min} = d(d+1) \cdot (\xi_{r+1}/n)$$

Theorem (Levenshtein 1995)

The least root ξ_r of $\mathcal{K}_r$ satisfies:

$$\xi_r^n / n \leq \frac{1}{2} - \sqrt{(1-t)t} + \frac{t^{1/6}\sqrt{1-t}}{n^{1/3}}$$

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 $r \in \mathbb{N}$
- The entire analysis carries over the q-ary cube $Q^n = \{0, 1, \ldots, q-1\}^n$
- 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.