Polynomial **OP**timization

Lasserre's moment-SOS hierarchy

Sparsity

Igor Klep, University of Ljubljana

Gaspard Monge Visiting Professor

21 February 2025

Outline

Optimization Problems (OPs)

A minimization problem is of the form:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in S \end{aligned}$$

where:

- S is the feasible region
- $f : S \to \mathbb{R}$ is the objective function

Usually S is also described using functions:

$$S = \{x \in \mathbb{R}^n \mid g_i(x) \geq 0,\ i = 1, \dots, m,\ h_j(x) = 0,\ j = 1, \dots, p\}$$

Optimization Problems (OPs)

A minimization problem is of the form:

$$\begin{aligned} \min \quad & f(x) \\ \text{s. t.} \quad & g_i(x) \geq 0, \; i = 1, \ldots, m \\ & h_j(x) = 0, \; j = 1, \ldots, p \end{aligned} \tag{OP}$$

Each $x \in \mathbb{R}^n$ satisfying the constraints is called feasible

Usually we are also interested in minimizers that solve (OP), i.e., all feasible $x^\star \in \mathbb{R}^n$ that minimize $f(x)$

Polynomial Optimization Problems (POPs)

A minimization problem is of the form:

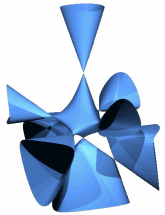
$$\begin{aligned} \min \quad & f(x) \\ \text{s. t.} \quad & g_i(x) \geq 0, \; i = 1, \dots, m \\ & h_j(x) = 0, \; j = 1, \dots, p \end{aligned} \qquad \text{(POP)}$$

where f, g_i, h_j are polynomials

The feasible region

$$S = \{x \in \mathbb{R}^n \mid g_i(x) \geq 0, \; i = 1, \dots, m, \; h_j(x) = 0, \; j = 1, \dots, p\}$$

is called a (basic closed) semialgebraic set



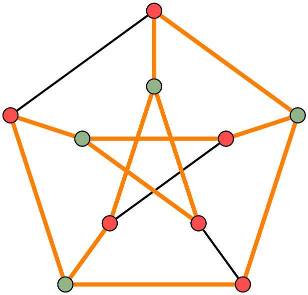
Source: Bruce Hunt
A Gallery of Algebraic Surfaces

Why do we care, I

Combinatorial optimization

Assign to each vertex i a value $x_i \in \{-1, 1\}$ in such a way as to maximize

$$\sum_{(i,j)\in E(G)} w_{ij} \frac{1 - x_i x_j}{2}.$$



People care about max cut?

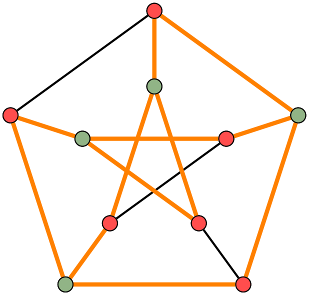
- easy to state
- it's NP-hard
- many hard optimization problems that arise in practice reduce to max cut

Why do we care, I

Combinatorial optimization

Assign to each vertex i a value $x_i \in \{-1, 1\}$ in such a way as to maximize

$$\sum_{(i,j)\in E(G)} w_{ij} \frac{1 - x_i x_j}{2}.$$



People care about max cut?

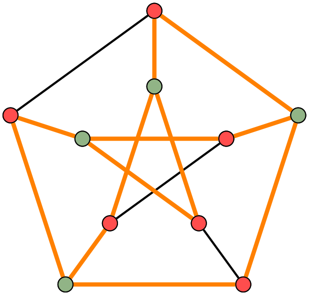
- easy to state
- it's NP-hard
- many hard optimization problems that arise in practice reduce to max cut
- correlation clustering ⇝ machine learning for unsupervised clustering, computer vision, bioinformatics (clustering genes based on expression data)

Why do we care, I and II

Combinatorial optimization and Physics

Assign to each vertex i a value $x_i \in \{-1, 1\}$ in such a way as to maximize

$$\sum_{(i,j)\in E(G)} w_{ij} \frac{1 - x_i x_j}{2}.$$



People care about max cut?

- easy to state
- it's NP-hard
- Finding the ground state of an Ising spin system is equivalent to solving a max cut problem.

↑	↑	↓	↑	↑	↑
↑	↑	↑	↓	↑	↑
↓	↓	↓	↑	↓	↓
↓	↑	↓	↓	↑	↓
↓	↑	↑	↑	↓	↓
↑	↓	↓	↓	↓	↓

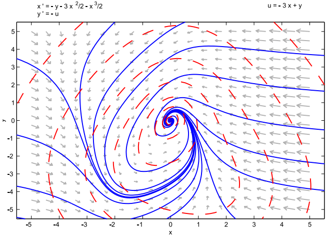
Why do we care, III

Control theory – stability of dynamical systems

- Given a system of ODEs
  $\dot{x}(t) = f\big(x(t)\big), \quad x(0) = x_0$
- Want to prove stability, i.e., that solutions converge to the origin for all initial conditions
- To prove this we need an energy-like Lyapunov function V satisfying

$$V(x) geq 0,$$

$$\dot{V}(x) = \left(\frac{\partial V}{\partial x}\right)^T f(x) \geq 0$$



Source: Pablo Parrilo

- For linear systems $\dot{x} = Ax$, can use quadratic Lyapunov functions:

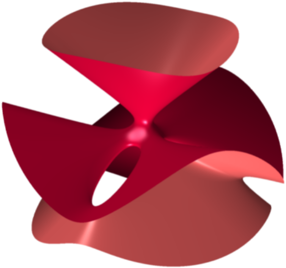
$$V(x) = x^T P x$$

where

$$P \succ 0, \quad A^T P + PA \prec 0$$

- In general, looking for a polynomial Lyapunov function V is a POP

Challenges in POP



Source: Wikipedia

- Non-convexity & too many local minima
- Complexity: NP-hard
- Solving large-scale problems
- Exploiting structure (symmetries, sparsity, ...)
- Numerical stability

Linear programming

Linear Optimization Problem is one of the form

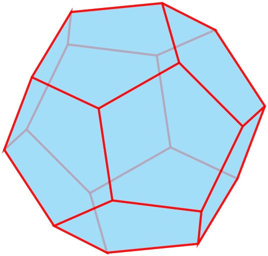
$$\begin{aligned} \min \quad & f(x) \\ \text{s. t.} \quad & g_i(x) \geq 0,\ i = 1, \dots, m \\ & h_j(x) = 0,\ j = 1, \dots, p \end{aligned} \tag{LP}$$

where the f, g_i, h_j are linear polynomials.

The feasible region S is a convex polyhedron.

Khachiyan (1979) showed that LPs can be solved quickly (in polynomial time).

In practice, one can today solve LPs with $n \approx m \approx ? \cdot 10^6$.



https://tikz.net/dodecahedron/

Linearizing POP

Naively

$$\begin{aligned} \min \quad & x_1^2x_2^2 + x_1x_2 + x_2^2 - 2x_2 + 2 \\ \text{s.t.} \quad & 1 - x_2^2 - x_1^2x_2^2 \geq 0 \end{aligned}$$

Linearizing POP

Naively

$$\begin{aligned} \min \quad & x_1^2 x_2^2 + x_1 x_2 + x_2^2 - 2x_2 + 2 \\ \text{s.t.} \quad & 1 - x_2^2 - x_1^2 x_2^2 \geq 0 \end{aligned}$$

Linearizing POP

Naively

$$
\begin{aligned}
\min \quad & x_1^2 x_2^2 + y_{12} + x_2^2 - 2x_2 + 2 \\
\text{s.t.} \quad & 1 - x_2^2 - x_1^2 x_2^2 \geq 0
\end{aligned}
$$

Linearizing POP

Naively

$$\begin{aligned} \min \quad & x_1^2 x_2^2 + y_{12} + x_2^2 - 2x_2 + 2 \\ \text{s.t.} \quad & 1 - x_2^2 - x_1^2 x_2^2 \geq 0 \end{aligned}$$

Linearizing POP

Naively

$$\begin{aligned} \min \quad & x_1^2 x_2^2 + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s. t.} \quad & 1 - y_{22} - x_1^2 x_2^2 \geq 0 \end{aligned}$$

Linearizing POP

Naively

$$\begin{aligned} \min \quad & x_1^2 x_2^2 + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s. t.} \quad & 1 - y_{22} - x_1^2 x_2^2 \geq 0 \end{aligned}$$

Linearizing POP

Naively

$$\begin{aligned} \min \quad & y_{1122} + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s.t.} \quad & 1 - y_{22} - y_{1122} \geq 0 \end{aligned}$$

Linearizing POP

Naively

$$\begin{aligned} \min \quad & y_{1122} + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s.t.} \quad & 1 - y_{22} - y_{1122} \geq 0 \end{aligned}$$

Linearizing POP

Slightly less naively

$$\begin{array}{ll} \min & y_{1122} + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s. t.} & 1 - y_{22} - y_{1122} \geq 0 \end{array}$$

Add redundant obvious constraints

$$(a_0 + a_1 x_1 + a_2 x_2 + a_{11} x_1^2 + a_{12} x_1 x_2 + a_{22} x_2^2)^2 \geq 0$$

$$\begin{pmatrix} a_0 & a_1 & a_2 & a_{11} & a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix} \begin{pmatrix} 1 & x_1 & x_2 & x_1^2 & x_1 x_2 & x_2^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_{11} \\ a_{12} \\ a_{22} \end{pmatrix} \geq 0$$

Linearizing POP

Slightly less naively

$$\begin{array}{ll} \min & y_{1122} + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s. t.} & 1 - y_{22} - y_{1122} \geq 0 \end{array}$$

Add redundant obvious constraints

$$(a_0 + a_1 x_1 + a_2 x_2 + a_{11} x_1^2 + a_{12} x_1 x_2 + a_{22} x_2^2)^2 \geq 0$$

$$\begin{pmatrix} a_0 & a_1 & a_2 & a_{11} & a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} 1 & x_1 & x_2 & x_1^2 & x_1 x_2 & x_2^2 \\ x_1 & x_1^2 & x_1 x_2 & x_1^3 & x_1^2 x_2 & x_1 x_2^2 \\ x_2 & x_1 x_2 & x_2^2 & x_1^2 x_2 & x_1 x_2^2 & x_2^3 \\ x_1^2 & x_1^3 & x_1^2 x_2 & x_1^4 & x_1^3 x_2 & x_1^2 x_2^2 \\ x_1 x_2 & x_1^2 x_2 & x_1 x_2^2 & x_1^3 x_2 & x_1^2 x_2^2 & x_1 x_2^3 \\ x_2^2 & x_1 x_2^2 & x_2^3 & x_1^2 x_2^2 & x_1 x_2^3 & x_2^4 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_{11} \\ a_{12} \\ a_{22} \end{pmatrix} \geq 0$$

Linearizing POP

Slightly less naively

$$\begin{aligned} \min \quad & y_{1122} + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s. t.} \quad & 1 - y_{22} - y_{1122} \geq 0 \end{aligned}$$

Add redundant obvious constraints

$$\begin{pmatrix} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 \\ x_1 & x_1^2 & x_1x_2 & x_1^3 & x_1^2x_2 & x_1x_2^2 \\ x_2 & x_1x_2 & x_2^2 & x_1^2x_2 & x_1x_2^2 & x_2^3 \\ x_1^2 & x_1^3 & x_1^2x_2 & x_1^4 & x_1^3x_2 & x_1^2x_2^2 \\ x_1x_2 & x_1^2x_2 & x_1x_2^2 & x_1^3x_2 & x_1^2x_2^2 & x_1x_2^3 \\ x_2^2 & x_1x_2^2 & x_2^3 & x_1^2x_2^2 & x_1x_2^3 & x_2^4 \end{pmatrix} \succeq 0$$

Linearizing POP

Slightly less naively

$$\begin{aligned} \min \quad & y_{1122} + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s. t.} \quad & 1 - y_{22} - y_{1122} \geq 0 \end{aligned}$$

Add redundant obvious constraints and linearize

$$\begin{pmatrix} 1 & x_1 & x_2 & y_{11} & y_{12} & y_{22} \\ x_1 & y_{11} & y_{12} & y_{111} & y_{112} & y_{122} \\ x_2 & y_{12} & y_{22} & y_{112} & y_{122} & y_{222} \\ y_{11} & y_{111} & y_{112} & y_{1111} & y_{1112} & y_{1122} \\ y_{12} & y_{112} & y_{122} & y_{1112} & y_{1122} & y_{1222} \\ y_{22} & y_{122} & y_{222} & y_{1122} & y_{1222} & y_{2222} \end{pmatrix} \succeq 0$$

Linearizing POP

Slightly less naively

$$\begin{array}{ll} \min & y_{1122} + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s. t.} & 1 - y_{22} - y_{1122} \geq 0 \end{array}$$

Add redundant obvious constraints

$$(a_0 + a_1 x_1 + a_2 x_2)^2 (1 - x_2^2 - x_1^2 x_2^2) \geq 0$$

Linearizing POP

Slightly less naively

$$\begin{aligned} \min \quad & y_{1122} + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s.t.} \quad & 1 - y_{22} - y_{1122} \geq 0 \end{aligned}$$

Add redundant obvious constraints

$$\begin{pmatrix} a_0 & a_1 & a_2 \end{pmatrix} (1 - x_2^2 - x_1^2 x_2^2) \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} \begin{pmatrix} 1 & x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \geq 0$$

Linearizing POP

Slightly less naively

$$\begin{aligned} \min \quad & y_{1122} + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s.t.} \quad & 1 - y_{22} - y_{1122} \geq 0 \end{aligned}$$

Add redundant obvious constraints

$$\begin{pmatrix} a_0 & a_1 & a_2 \end{pmatrix} (1 - x_2^2 - x_1^2 x_2^2) \begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & x_1^2 & x_1 x_2 \\ x_2 & x_1 x_2 & x_2^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \geq 0$$

Linearizing POP

Slightly less naively

$$\begin{array}{ll} \min & y_{1122} + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s. t.} & 1 - y_{22} - y_{1122} \geq 0 \end{array}$$

Add redundant obvious constraints

$$\begin{pmatrix} a_0 & a_1 & a_2 \end{pmatrix} \begin{pmatrix} 1 - x_1^2 x_2^2 - x_2^2 & -x_2^2 x_1^3 - x_2^2 x_1 + x_1 & -x_1^2 x_2^3 - x_2^3 + x_2 \\ -x_2^2 x_1^3 - x_2^2 x_1 + x_1 & -x_2^2 x_1^4 - x_2^2 x_1^2 + x_1^2 & -x_1^3 x_2^3 - x_1 x_2^3 + x_1 x_2 \\ -x_1^2 x_2^3 - x_2^3 + x_2 & -x_1^3 x_2^3 - x_1 x_2^3 + x_1 x_2 & -x_1^2 x_2^4 - x_2^4 + x_2^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \geq 0$$

Linearizing POP

Slightly less naively

$$\begin{aligned} \min \quad & y_{1122} + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s. t.} \quad & 1 - y_{22} - y_{1122} \geq 0 \end{aligned}$$

Add redundant obvious constraints

$$\begin{pmatrix} 1 - x_1^2 x_2^2 - x_2^2 & -x_2^2 x_1^3 - x_2^2 x_1 + x_1 & -x_1^2 x_2^3 - x_2^3 + x_2 \\ -x_2^2 x_1^3 - x_2^2 x_1 + x_1 & -x_2^2 x_1^4 - x_2^2 x_1^2 + x_1^2 & -x_1^3 x_2^3 - x_1 x_2^3 + x_1 x_2 \\ -x_1^2 x_2^3 - x_2^3 + x_2 & -x_1^3 x_2^3 - x_1 x_2^3 + x_1 x_2 & -x_1^2 x_2^4 - x_2^4 + x_2^2 \end{pmatrix} \succeq 0$$

Linearizing POP

Slightly less naively

$$\begin{aligned} \min \quad & y_{1122} + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s.t.} \quad & 1 - y_{22} - y_{1122} \geq 0 \end{aligned}$$

Add redundant obvious constraints and linearize

$$\begin{pmatrix} 1 - y_{22} - y_{1122} & x_1 - y_{122} - y_{11122} & x_2 - y_{222} - y_{11222} \\ x_1 - y_{122} - y_{11122} & y_{11} - y_{1122} - y_{111122} & y_{12} - y_{1222} - y_{111222} \\ x_2 - y_{222} - y_{11222} & y_{12} - y_{1222} - y_{111222} & y_{22} - y_{2222} - y_{112222} \end{pmatrix} \succeq 0$$

Linearizing POP

Lasserre's hierarchy (an example)

$$\begin{aligned} \min \quad & y_{1122} + y_{12} + y_{22} - 2x_2 + 2 \\ \text{s.t.} \quad & 1 - y_{22} - y_{1122} \geq 0 \end{aligned}$$

$$\begin{pmatrix} 1 & x_1 & x_2 & y_{11} & y_{12} & y_{22} \\ x_1 & y_{11} & y_{12} & y_{111} & y_{112} & y_{122} \\ x_2 & y_{12} & y_{22} & y_{112} & y_{122} & y_{222} \\ y_{11} & y_{111} & y_{112} & y_{1111} & y_{1112} & y_{1122} \\ y_{12} & y_{112} & y_{122} & y_{1112} & y_{1122} & y_{1222} \\ y_{22} & y_{122} & y_{222} & y_{1122} & y_{1222} & y_{2222} \end{pmatrix} \succeq 0$$

$$\begin{pmatrix} 1 - y_{22} - y_{1122} & x_1 - y_{122} - y_{11122} & x_2 - y_{222} - y_{11222} \\ x_1 - y_{122} - y_{11122} & y_{11} - y_{1122} - y_{111122} & y_{12} - y_{1222} - y_{111222} \\ x_2 - y_{222} - y_{11222} & y_{12} - y_{1222} - y_{111222} & y_{22} - y_{2222} - y_{112222} \end{pmatrix} \succeq 0$$

This is a semidefinite program (SDP), a far-reaching extension of LP

POP

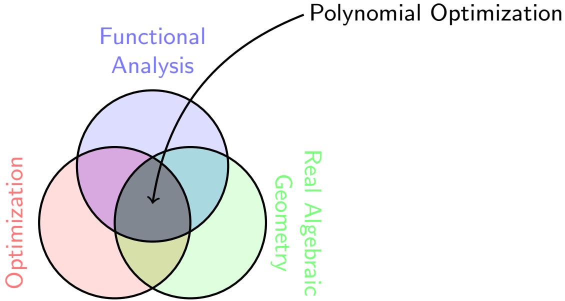
Some theory

- Positivstellensatz of Krivine (1964):

  to each *infeasible POP*, using the (less naive) linearization procedure, one can always add finitely many "redundant" inequalities such that the resulting *SDP* is *infeasible*.

- Schmüdgen's (1991) and Putinar's (1993) Positivstellensatz:

  for POP with *compact* feasible region, *optimal values* of the POP and of the resulting "infinite" SDP *coincide*.



Notation

- $x = (x_1, \ldots, x_n)$ commutative variables
- products of the x_j are monomials
- $[x]_k$ will denote (a vector of) monomials of degree $\leq k$
  If $n = k = 2$, then $[x]_2 = \begin{pmatrix} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 \end{pmatrix}^T$
- $\mathbb{R}[x]$ = all polynomials
- $\Sigma^2 = \left\{ \sum h_j^2 \mid h_j \in \mathbb{R}[x] \right\}$ convex cone of sums of squares (SOS)
- Given $g = (g_1, \ldots, g_m) \in \mathbb{R}[x]^m$ the feasible set

$$S(g) = \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$$

  is a (basic closed) semialgebraic set
- $\mathrm{QM}(g) = \Sigma^2 + \Sigma^2 g_1 + \cdots + \Sigma^2 g_m$
  is the quadratic module (weighted SOS) generated by $g = (g_1, \ldots, g_m)$

Observe: $f \in \mathrm{QM}(g) \;\Rightarrow\; f \geq 0$ on $S(g)$

Putinar's Positivstellensatz

Theorem (Putinar (1993))

Assume

- $S(g)$ *is bounded*
- g *contains a ball constraint* $R - \sum x_j^2 \geq 0$ *for some* $R \in \mathbb{R}$

If $f > 0$ *on* $S(g)$*, then* $f \in \mathrm{QM}(g)$

⚠ $f \geq 0$ on $S(g)$ does not imply $f \in \mathrm{QM}(g)$

Positivstellensätze

You are only as strong as your Positivstellensatz

- Artin's (1926) solution to Hilbert's 17th problem (1900)

$$f \geq 0 \text{ on } \mathbb{R}^n \iff f \in \Sigma^2\mathbb{R}(x) \iff \exists 0 \neq q \in \mathbb{R}[x] : \; q^2 p \in \Sigma^2$$

- Krivine Positivstellensatz (1964)

$$f > 0 \text{ on } S(g) \iff \exists q_1, q_2 \in \mathrm{QM}(\prod g) : \; q_1 p = 1 + q_2$$

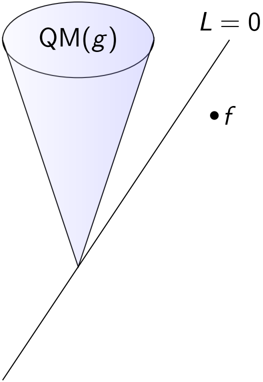
- Schmüdgen Positivstellensatz (1991)
  Assume $S(g)$ is compact. Then $f > 0$ on $S(g)$ implies $f \in \mathrm{QM}(\prod g)$

- Many further variants

Putinar's Positivstellensatz

Proof

$f > 0$ on $S(g)$; assume $f \notin \mathrm{QM}(g)$.



- $L : \mathbb{R}[x] \to \mathbb{R}$, $L(\mathrm{QM}(g)) \subseteq [0, \infty)$, $L(f) \leq 0$
- inner product $\langle a, b \rangle = L(ab)$ on $\mathbb{R}[x]$
- define $\hat{X}_j : \mathbb{R}[x] \to \mathbb{R}[x]$, $p \mapsto x_j p$
- by compactness, $\hat{X}_j$ are bounded, so extend to the Hilbert space completion $\mathcal{H}$ of $\mathbb{R}[x]$
- for $p \in \mathbb{R}[x]$ we have $L(p) = \langle p, 1 \rangle = \langle p(\hat{X})1, 1 \rangle$
- by the spectral theorem (for a tuple of commuting self-adjoint operators $\hat{X}$), there exists measure μ supported on $S(g)$ s.t.

$$L(p) = \int p \, \mathrm{d}\mu \quad \text{for all } p$$

- finally, $0 \geq L(f) = \int f \, \mathrm{d}\mu > 0$ $\lightning$

Sums of Squares (SOS)

Lemma

$f \in \mathbb{R}[x]_{2k}$ is a sum of squares iff there is $G \succeq 0$ s.t. $f = [x]_k^T G [x]_k$.

Proof.

- If $f = \sum_i g_i^2 \in \Sigma^2$, then $\deg g_i \leq k$ for all i
- write $g_i = G_i^T [x]_k$, where G_i^T is for row vector of the coefficients of g_i
- then $g_i^2 = [x]_k^T G_i G_i^T [x]_k$
- setting $G := \sum_i G_i G_i^T$, we have $f = [x]_k^T G [x]_k$

- For the converse, every PsD matrix G admits a Cholesky factorization $G = \sum_{i=1}^r G_i G_i^T$ for row vectors G_i
- letting $g_i := G_i^T [x]_k$, we get $f = \sum g_i^2$

□

SOS

An example

$$f = 4x_1^4 + 4x_2x_1^3 - 4x_1^3 + x_2^2x_1^2 - 2x_2x_1^2 + x_1^2 + x_2^2 - 2x_2 + 1$$

- Write $f = [x]_2^T G [x]_2$ for a symmetric matrix G
-

$$G = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \end{array} \begin{matrix} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 \end{matrix}$$

$$G = \begin{pmatrix} 1 & 0 & -1 & g_{14} & g_{15} & g_{16} \\ 0 & 1-2g_{14} & -g_{15} & -2 & g_{25} & g_{26} \\ -1 & -g_{15} & 1-2g_{16} & -g_{25}-1 & -g_{26} & 0 \\ g_{14} & -2 & -g_{25}-1 & 4 & 2 & g_{46} \\ g_{15} & g_{25} & -g_{26} & 2 & 1-2g_{46} & 0 \\ g_{16} & g_{26} & 0 & g_{46} & 0 & 0 \end{pmatrix} \begin{matrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_2^2 \end{matrix}$$

SOS

An example

$$f = 4x_1^4 + 4x_2x_1^3 - 4x_1^3 + x_2^2x_1^2 - 2x_2x_1^2 + x_1^2 + x_2^2 - 2x_2 + 1$$

- Write $f = [x]_2^T G[x]_2$ for a symmetric matrix G
- $$G = \begin{pmatrix} 1 & 0 & -1 & g_{14} & g_{15} & g_{16} \\ 0 & 1-2g_{14} & -g_{15} & -2 & g_{25} & g_{26} \\ -1 & -g_{15} & 1-2g_{16} & -g_{25}-1 & -g_{26} & 0 \\ g_{14} & -2 & -g_{25}-1 & 4 & 2 & g_{46} \\ g_{15} & g_{25} & -g_{26} & 2 & 1-2g_{46} & 0 \\ g_{16} & g_{26} & 0 & g_{46} & 0 & 0 \end{pmatrix}$$

SOS

An example

$$f = 4x_1^4 + 4x_2x_1^3 - 4x_1^3 + x_2^2x_1^2 - 2x_2x_1^2 + x_1^2 + x_2^2 - 2x_2 + 1$$

- Write $f = [x]_2^T G[x]_2$ for a symmetric matrix G
- $$G = \begin{pmatrix} 1 & 0 & -1 & g_{14} & g_{15} & g_{16} \\ 0 & 1-2g_{14} & -g_{15} & -2 & g_{25} & g_{26} \\ -1 & -g_{15} & 1-2g_{16} & -g_{25}-1 & -g_{26} & 0 \\ g_{14} & -2 & -g_{25}-1 & 4 & 2 & g_{46} \\ g_{15} & g_{25} & -g_{26} & 2 & 1-2g_{46} & 0 \\ g_{16} & g_{26} & 0 & g_{46} & 0 & 0 \end{pmatrix}$$

SOS

An example

$$f = 4x_1^4 + 4x_2x_1^3 - 4x_1^3 + x_2^2x_1^2 - 2x_2x_1^2 + x_1^2 + x_2^2 - 2x_2 + 1$$

- Write $f = [x]_2^T G [x]_2$ for a symmetric matrix G
- $$G = \begin{pmatrix} 1 & 0 & -1 & g_{14} & g_{15} & 0 \\ 0 & 1-2g_{14} & -g_{15} & -2 & g_{25} & 0 \\ -1 & -g_{15} & 1 & -g_{25}-1 & 0 & 0 \\ g_{14} & -2 & -g_{25}-1 & 4 & 2 & 0 \\ g_{15} & g_{25} & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

SOS

An example

$$f = 4x_1^4 + 4x_2x_1^3 - 4x_1^3 + x_2^2x_1^2 - 2x_2x_1^2 + x_1^2 + x_2^2 - 2x_2 + 1$$

- Write $f = [x]_2^T G [x]_2$ for a symmetric matrix G
- $$G = \begin{pmatrix} 1 & 0 & -1 & g_{14} & g_{15} & 0 \\ 0 & 1-2g_{14} & -g_{15} & -2 & g_{25} & 0 \\ -1 & -g_{15} & 1 & -g_{25}-1 & 0 & 0 \\ g_{14} & -2 & -g_{25}-1 & 4 & 2 & 0 \\ g_{15} & g_{25} & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

  $$\begin{pmatrix} 0 & 0 & 0 & 1 & -2 & 0 \end{pmatrix} \cdot G = 0$$

  leads to

  $$g_{15} = \frac{1}{2} g_{14}, \quad g_{25} = -1$$

SOS

An example

$$f = 4x_1^4 + 4x_2x_1^3 - 4x_1^3 + x_2^2x_1^2 - 2x_2x_1^2 + x_1^2 + x_2^2 - 2x_2 + 1$$

- Write $f = [x]_2^T G[x]_2$ for a symmetric matrix G

$$G = \begin{pmatrix} 1 & 0 & -1 & g_{14} & \frac{g_{14}}{2} & 0 \\ 0 & 1-2g_{14} & -\frac{g_{14}}{2} & -2 & -1 & 0 \\ -1 & -\frac{g_{14}}{2} & 1 & 0 & 0 & 0 \\ g_{14} & -2 & 0 & 4 & 2 & 0 \\ \frac{g_{14}}{2} & -1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

SOS

An example

$$f = 4x_1^4 + 4x_2x_1^3 - 4x_1^3 + x_2^2x_1^2 - 2x_2x_1^2 + x_1^2 + x_2^2 - 2x_2 + 1$$

- Write $f = [x]_2^T G[x]_2$ for a symmetric matrix G

$$G = \begin{pmatrix} 1 & 0 & -1 & g_{14} & \frac{g_{14}}{2} & 0 \\ 0 & 1-2g_{14} & -\frac{g_{14}}{2} & -2 & -1 & 0 \\ -1 & -\frac{g_{14}}{2} & 1 & 0 & 0 & 0 \\ g_{14} & -2 & 0 & 4 & 2 & 0 \\ \frac{g_{14}}{2} & -1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \cdot G = 0$$

leads to

$$g_{14} = 0$$

SOS

An example

$$f = 4x_1^4 + 4x_2x_1^3 - 4x_1^3 + x_2^2x_1^2 - 2x_2x_1^2 + x_1^2 + x_2^2 - 2x_2 + 1$$

- Write $f = [x]_2^T G [x]_2$ for a symmetric matrix G

$$G = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 & -1 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 4 & 2 & 0 \\ 0 & -1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

SOS

An example

$$f = 4x_1^4 + 4x_2x_1^3 - 4x_1^3 + x_2^2x_1^2 - 2x_2x_1^2 + x_1^2 + x_2^2 - 2x_2 + 1$$

- Write $f = [x]_2^T G [x]_2$ for a symmetric matrix G

$$G = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 & -1 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 4 & 2 & 0 \\ 0 & -1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \\ 0 & 1 \\ 2 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & -1 \\ -1 & 0 \\ 0 & 1 \\ 2 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}^T \succeq 0$$

Hence

$$f = (-x_1 + 2x_1^2 + x_1x_2)^2 + (-1 + x_2)^2 \in \Sigma^2$$

QM

. . . meet SDP?

Checking whether a polynomial $f \in \mathbb{R}[x]_{2k}$ is SOS is a feasibility SDP:

$$f \in \Sigma^2 \iff \exists G \succeq 0 : \; f = [x]_k^T G [x]_k$$

Checking whether a polynomial $f \in \mathbb{R}[x]_{2k}$ is in QM(g) is not an SDP:

$$f \in \mathrm{QM}(g) \iff \exists k_0, \dots, k_m \in \mathbb{N} \, \exists G_0, \dots, G_m \succeq 0 : \\ f = [x]_{k_0}^T G_0 [x]_{k_0} + [x]_{k_1}^T G_1 [x]_{k_1} \cdot g_1 + \cdots + [x]_{k_m}^T G_m [x]_{k_m} \cdot g_m$$

⚠ there is no control on the degrees k_j

QM

. . . meet SDP

Let $\delta_j = \deg(g_j)$.

We define the k-th truncation of $\mathrm{QM}(g)$ as follows

$$\begin{aligned}\mathrm{QM}(g)_k &= \Sigma_k^2 + \Sigma^2_{k-\lfloor \frac{1}{2}\delta_1 \rfloor} \cdot g_1 + \cdots + \Sigma^2_{k-\lfloor \frac{1}{2}\delta_m \rfloor} \cdot g_m \\ &= \Big\{ [x]_k^T G_0 [x]_k + [x]^T_{k-\lfloor \frac{1}{2}\delta_1 \rfloor} G_1 [x]_{k-\lfloor \frac{1}{2}\delta_1 \rfloor} g_1 + \cdots \\ &\qquad + [x]^T_{k-\lfloor \frac{1}{2}\delta_m \rfloor} G_m [x]_{k-\lfloor \frac{1}{2}\delta_m \rfloor} g_m \;\Big|\; G_1, \ldots, G_m \succeq 0 \Big\} \subseteq \mathbb{R}[x]_{2k}\end{aligned}$$

Then

$$\mathrm{QM}(g) = \bigcup_{k \in \mathbb{N}} \mathrm{QM}(g)_k$$

Testing membership in $\mathrm{QM}(g)_k$ is an SDP

⚠ Beware, $\mathrm{QM}(g) \cap \mathbb{R}[x]_{2k} \supsetneq \mathrm{QM}(g)_k$

Lasserre hierarchy

To each POP

$$\begin{aligned} \min \quad & f(x) \\ \text{s. t.} \quad & g_i(x) \geq 0,\ i = 1, \ldots, m \end{aligned}$$

we assign the sequence of SDP

$$\begin{aligned} \max \quad & \lambda \\ \text{s. t.} \quad & f - \lambda \in \mathrm{QM}(g)_k \end{aligned} \qquad (\mathrm{Lass}_k)$$

with optimal values λ_k

Theorem (Lasserre (2001))

Assume $S(g)$ is compact and g contains a ball constraint. Then

$$\lambda_k \nearrow \min_{S(g)} f$$

Lasserre hierarchy

Moment-SOS hierarchy

Using standard Lagrange duality from convex optimization,
we can obtain the dual SDP to (Lass_k)

$$\begin{array}{lll} \min & L(f) & \\ \text{s. t.} & L : \mathbb{R}[x]_{2k} \to \mathbb{R} \text{ linear} & \qquad (\mathrm{Lass}'_k) \\ & L(\mathrm{QM}(g)_k) \subseteq \mathbb{R}_{\geq 0}, \;\; L(1) = 1 & \end{array}$$

with optimal values λ_k

Values $y_\alpha = L(x^\alpha)$ are called pseudomoments,
and we build a Hankel matrix $H(L)$ indexed by monomials of degree $\leq k$,

$$H(L)_{\alpha,\beta} = L(x^{\alpha+\beta}) = y_{\alpha+\beta}$$

To each constraint g_j we also build the localizing Hankel matrix,

$$H(g_j L)_{\alpha,\beta} = L(x^{\alpha+\beta} g_j)$$

Lemma

L is feasible for (Lass'_k) iff $H(L) \succeq 0,\; H(g_1 L) \succeq 0, \ldots, H(g_m L) \succeq 0.$

Lasserre hierarchy

Moment-SOS hierarchy

Using standard Lagrange duality from convex optimization,
we can obtain the dual SDP to (Lass_k)

$$\begin{array}{ll} \min & L(f) \\ \text{s. t.} & L : \mathbb{R}[x]_{2k} \to \mathbb{R} \text{ linear} \\ & L(\mathrm{QM}(g)_k) \subseteq \mathbb{R}_{\geq 0}, \;\; L(1) = 1 \end{array} \qquad (\mathrm{Lass}'_k)$$

with optimal values λ_k

Lemma

L is feasible for (Lass'_k) iff $H(L) \succeq 0, \; H(g_1 L) \succeq 0, \; \ldots, \; H(g_m L) \succeq 0.$

We can now rewrite (Lass'_k) to make it look like an SDP:

$$\begin{array}{ll} \min & \mathrm{Tr}(H(L)\, G_f) \\ \text{s. t.} & H(L)_{0,0} = 1 \\ & H(L) \succeq 0, \; H(g_1 L) \succeq 0, \; \ldots, \; H(g_m L) \succeq 0 \end{array} \qquad (\mathrm{Lass}'_k)$$

Lasserre hierarchy

Moment-SOS hierarchy

Using standard Lagrange duality from convex optimization,
we can obtain the dual SDP to (Lass_k)

$$\begin{array}{ll} \min & L(f) \\ \text{s. t.} & L : \mathbb{R}[x]_{2k} \to \mathbb{R} \text{ linear} \\ & L(\mathrm{QM}(g)_k) \subseteq \mathbb{R}_{\geq 0}, \quad L(1) = 1 \end{array} \qquad (\mathrm{Lass}'_k)$$

with optimal values $\reflectbox{\ensuremath{\lambda}}_k$

Theorem

The primal-dual pair (Lass_k) *and* (Lass'_k) *satisfy strong duality:* $\reflectbox{\ensuremath{\lambda}}_k = \lambda_k$ *and*

$$\reflectbox{\ensuremath{\lambda}}_k \nearrow \min_{S(g)} f$$

Lasserre hierarchy

Extracting optimizers

$$\begin{aligned} \min \quad & \operatorname{Tr}(H(L)\, G_f) \\ \text{s. t.} \quad & H(L)_{0,0} = 1 \\ & H(L) \succeq 0,\ H(g_1 L) \succeq 0, \ldots, H(g_m L) \succeq 0 \end{aligned} \qquad (\mathrm{Lass}'_k)$$

Let $\delta = \max \delta_j$, where $\delta_j = \deg(g_j)$.

Theorem (Curto-Fialkow (1991), Henrion-Lasserre (2003))

Assume $H(L)$ is δ-flat (aka rank loop condition), i.e.,

$$\operatorname{rank} H(L)_k = \operatorname{rank} H(L)_{k - \lceil \frac{1}{2}\delta \rceil}$$

Then

- $\lambda_k = \min_{S(g)} f$
- *Gelfand-Naimark-Segal (GNS) construction + matrix diagonalization extracts a minimizer $x^\star \in S(g)$ for f*

Lasserre hierarchy

Extracting optimizers

$$\operatorname{rank} H(L)_k = \operatorname{rank} \begin{pmatrix} H(L)_{k-\lceil \frac{1}{2}\delta \rceil} & B \\ B^* & C \end{pmatrix} = \operatorname{rank} H(L)_{k-\lceil \frac{1}{2}\delta \rceil}$$

- Let E be the range = column space of $H(L)_{k-\lceil \frac{1}{2}\delta \rceil}$.

  Index columns of $H(L)_{k-\lceil \frac{1}{2}\delta \rceil}$ by monomials x^α of degree $|\alpha| \leq k - \lceil \frac{1}{2}\delta \rceil$.
- $H(L)$ induces a (semi-)inner product on E: $\langle \alpha, \beta \rangle = L(x^{\alpha+\beta}) = H(L)_{\alpha,\beta}$
- x_i act on E to produce a linear map $X_i : E \to E$.
- $X_i : E \to E$ are pairwise commuting symmetric matrices, so can be simultaneously diagonalized,

$$X_1 = \begin{pmatrix} d^1_{11} & & \\ & \ddots & \\ & & d^1_{ss} \end{pmatrix}, \qquad \ldots, \qquad X_n = \begin{pmatrix} d^n_{11} & & \\ & \ddots & \\ & & d^n_{ss} \end{pmatrix}$$

- Then $x^\star = (d^1_{ii}, \ldots, d^n_{ii})$ is a minimizer.

  Uses that L was a optimal solution of a step in the Lasserre hierarchy.

Lasserre hierarchy

Software

Plethora of available software options

- YALMIP (Löfberg)
  `https://yalmip.github.io/`
- GloptiPoly 3 (Henrion, Lasserre, Löfberg)
  `https://homepages.laas.fr/henrion/software/gloptipoly3/`
- SOSTOOLS (Papachristodoulou, Anderson, Valmorbida, Prajna, Seiler, Parrilo, Peet, Jagt)
  `https://github.com/oxfordcontrol/SOSTOOLS`

- Julia
  `https://julialang.org/`

⚠ All of these will require a separate SDP solver, such as
MOSEK, SeDuMi, COSMO, SDPA, SDPT3, CSDP, SDPNAL+, DSDP, ...

Lasserre hierarchy – Example in Julia

We solve the following POP using Lasserre moment-SOS hierarchy

$$\min\ x^2y^2 + xy + y^2 - 2y + 2 \quad \text{s. t.} \quad 1 - y^2 - x^2y^2 \geq 0$$

```
using SumOfSquares
using DynamicPolynomials #Enables symbolic variables
using MosekTools        #Mosek SDP solver

# Create an SOS optimization model
model = SOSModel(Mosek.Optimizer)

# Define polynomial variables x and y
@polyvar x y

# Define a decision variable t
@variable(model, t)

# Define the constraint set
S = @set 1 - y^2 - x^2 * y^2 >= 0

# Add the SOS relaxation constraint:
@constraint(model, x^2 * y^2 + x * y + y^2 - 2 * y + 2 >= t,
           domain = S, maxdegree = 8) #maxdegree controls relaxation

# Set the objective to maximize t (tightest lower bound)
@objective(model, Max, t)

# Solve the SDP relaxation and Print the optimal solution
optimize!(model)
println("Solution: $(value(t))")
```

Lasserre hierarchy

Some up-to-date results

- Tightened Lasserre relaxations (Nie, 2013) – Lagrangian or Jacobian form
- Finite convergence of Lasserre hierarchy holds generically (Nie, 2012)
- Unless P=NP there does not exist a poly-time algorithm to decide whether the Lasserre hierarchy has finite convergence (Vargas, 2024)
-

$S(g)$ **(compact)**	**error**	**certificate**	**reference**
w/ ball constraint	$O(1/\log(r)^c)$	$\mathrm{QM}(g)$	Nie, Schweighofer 2007
w/ ball constraint	$O(1/r^c)$	$\mathrm{QM}(g)$	Baldi, Mourrain, Parusinski 2022, 2023
General	$O(1/r^c)$	$\mathrm{QM}(\prod g)$	Schweighofer 2004
$[-1,1]^n$	$O(1/r)$	$\mathrm{QM}(g)$	Baldi, Slot 2024
$\mathbb{S}^{n-1}$	$O(1/r^2)$	$\mathrm{QM}(g)$	Fang, Fawzi 2021
B^n	$O(1/r^2)$	$\mathrm{QM}(g)$	Slot 2022
Δ^n	$O(1/r^2)$	$\mathrm{QM}(\prod g)$	Slot 2022
$[-1,1]^n$	$O(1/r^2)$	$\mathrm{QM}(\prod g)$	Laurent, Slot 2023

Table: Asymptotic error of Lasserre's hierarchies

Lasserre hierarchy

Some up-to-date results

- The opposite Lasserre hierarchy (Lasserre, 2011):
  a sequence of upper bounds λ^r converging to the minimum

$S(g)$ **(compact)**	**error**	**measure** μ	**reference**
Geometric assumption	$O(1/\sqrt{r})$	Lebesgue	de Klerk, Laurent, Sun 2017
Convex body	$O(1/r)$	Lebesgue	de Klerk, Laurent 2018
Semialgebraic with dense interior, convex body	$O(\log^2(r)/r^2)$	Lebesgue	Slot, Laurent 2021
$\mathbb{S}^{n-1}$	$O(1/r)$	uniform	Doherty, Wehner 2013
$\mathbb{S}^{n-1}$	$O(1/r^2)$	uniform	de Klerk, Laurent 2022
$[-1,1]^n$	$O(1/r^2)$	$\prod_i (1-x_i)^\lambda dx$	de Klerk, Laurent, Slot 2020, 2022
'Round' convex body	$O(1/r^2)$	Lebesgue	Slot, Laurent 2022
B^n			
Δ^n			

Table: Asymptotic error of Lasserre's hierarchy of upper bounds

Relaxing SOS

LP based

- We say $p \in \mathbb{R}[x]$ is diagonally-dominant-SOS (ddSOS) if

$$p(x) = \sum_i \alpha_i m_i^2(x) + \sum_{i,j} \beta_{ij}^+ (m_i(x) + m_j(x))^2 + \sum_{i,j} \beta_{ij}^- (m_i(x) - m_j(x))^2,$$

  for some monomials $m_i(x), m_j(x) \in [x]$ and some $\alpha_i, \beta_{ij}^+, \beta_{ij}^- \in \mathbb{R}_{\geq 0}$.

- ddSOS_{2d} = polynomials of degree $\leq 2d$ that are ddSOS

- A symmetric matrix $A = (a_{ij})$ is diagonally dominant (dd) if

$$a_{ii} \geq \sum_{j \neq i} |a_{ij}| \quad \text{for all } i.$$

- We denote the set of $n \times n$ dd matrices with DD_n.

(Gershgorin's circle theorem) dd matrices are PsD

Relaxing SOS

LP based

Theorem (Ahmadi–Majumdar (2017))

$p \in \mathbb{R}[x]_{2d}$ is ddSOS iff it admits a representation

$$p(x) = [x]_d^T Q [x]_d$$

for a dd matrix Q.

- Can test for ddSOS using LP

Indeed, that Q be dd can be imposed, e.g., by a set of linear inequalities

$$Q_{ii} \geq \sum_{j \neq i} z_{ij}, \forall i,$$
$$-z_{ij} \leq Q_{ij} \leq z_{ij}, \forall i, j, i \neq j$$

in variables Q_{ij} and z_{ij}.

Relaxing SOS

SOCP based

- We say $p \in \mathbb{R}[x]$ is scaled diagonally-dominant-SOS (sddSOS) if

$$p(x) = \sum_i \alpha_i m_i^2(x) + \sum_{i,j} (\hat{\beta}_{ij}^+ m_i(x) + \tilde{\beta}_{ij}^+ m_j(x))^2 + \sum_{i,j} (\hat{\beta}_{ij}^- m_i(x) - \tilde{\beta}_{ij}^- m_j(x))^2,$$

  for some monomials $m_i(x), m_j(x) \in [x]$ and some scalars $\alpha_i, \hat{\beta}_{ij}^+, \tilde{\beta}_{ij}^+, \hat{\beta}_{ij}^-, \tilde{\beta}_{ij}^-$ with $\alpha_i \geq 0$.

- sddSOS_{2d} = polynomials of degree $\leq 2d$ that are sddSOS

- A symmetric matrix A is scaled diagonally dominant (sdd) if there exists a diagonal matrix D, with positive diagonal entries, such that DAD is dd.
- We denote the set of $n \times n$ sdd matrices with SDD_n.

Gershgorin's circle theorem implies that sdd matrices are PsD

Relaxing SOS

SOCP based

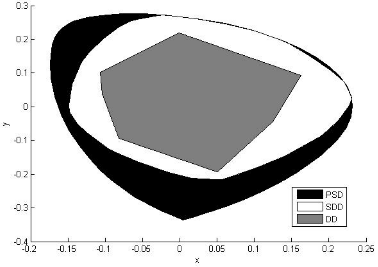
Theorem (Ahmadi–Majumdar (2017))

$p \in \mathbb{R}[x]_{2d}$ is sddSOS iff it admits a representation

$$p(x) = [x]_d^T Q [x]_d$$

for an sdd matrix Q.

- Can test for ddSOS using SOCP



A section of the cone of 5×5 dd, sdd, PsD matrices. Optimization over these sets can respectively be done by LP, SOCP, and SDP.

Source: Ahmadi–Majumdar

Sparse POP

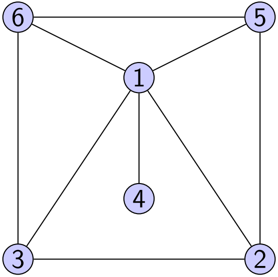
Correlative sparsity

Consider a sparse POP

$$\begin{aligned} \min \quad & f(x) \\ \text{s. t.} \quad & g_i(x) \geq 0, \; i = 1, \ldots, m \end{aligned} \qquad \text{(sparsePOP)}$$

Here sparse means few links between the variables.

- e.g. $f = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$
- Assign to f the correlative sparsity pattern (csp) graph



- vertices $= \{1, \ldots, n\}$ corresponding to the n variables
- $(i, j) \in$ edges iff x_ix_j appears in f

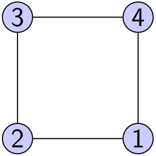
Sparse POP

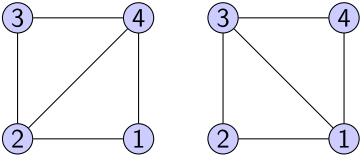
Intermezzo – chordal graphs

- chord = edge between two nonconsecutive vertices in a cycle
- chordal graph = all cycles of length ≥ 4 have at least one chord

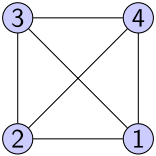
💡 any non-chordal graph can be extended to a chordal one by adding edges

- chordal extension is not unique





minimal



maximal

Sparse POP

Intermezzo – chordal graphs

- chord = edge between two nonconsecutive vertices in a cycle
- chordal graph = all cycles of length ≥ 4 have at least one chord

any non-chordal graph can be extended to a chordal one by adding edges

- chordal extension is not unique

(Gavril (1972), Vandenberghe–Andersen (2015))
The maximal cliques of a chordal graph can be enumerated
in linear time in the number of vertices and edges.

Sparse POP

Theorem (Running intersection Property (RiP) for chordal graphs (Blair–Peyton (1993))

For a chordal graph with maximal cliques $I_1, \ldots, I_p$:

$$\forall k < p: \quad I_{k+1} \cap (I_1 \cup \cdots \cup I_k) \subseteq I_\ell \quad \textit{for some } \ell \leq k$$

possibly after reordering

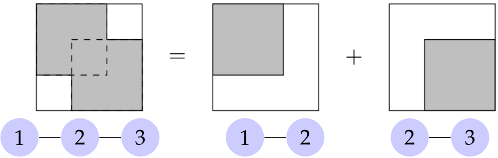
Sparse SDP matrices

Theorem [Griewank Toint '84, Agler et al. '88]

Chordal graph G with n vertices & maximal cliques I_1, I_2

$Q_G \succcurlyeq 0$ with nonzero entries corresponding to edges of G

$\Longrightarrow \; Q_G = {P_1}^T Q_1 P_1 + {P_2}^T Q_2 P_2$ with $Q_k \succcurlyeq 0$ indexed by I_k



What are P_1, P_2? $P_1 \in \mathbb{R}^{|I_1| \times n}$

$$P(i,j) = \begin{cases} 1 & \text{if } I(i) = j \\ 0 & \text{otherwise} \end{cases}$$

$$I_1 = (1,2) \Rightarrow P_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad I_2 = (2,3) \Rightarrow P_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

${P_1}^T Q_1 P_1$ inflates a $|I_1| \times |I_1|$ matrix Q_1 into a sparse $n \times n$ matrix

Sparse POP

Sparse Putinar

Consider (sparsePOP), where

- each g_j depends only on $x(I_k)$ for some k
- $f = \sum_k f_k$, where f_k depends only on $x(I_k)$
- RiP holds for I_ks
- ball constraint holds for each $x(I_k)$

Theorem (Sparse Putinar Positivstellnsatz (Lasserre, 2006))

If $f > 0$ on $S(g)$, then

$$f = \sum_k \sigma_{0k} + \sum_{j \in I_k} \sigma_{jk} g_j,$$

where σ_{jk} is SOS in $x(I_k)$.

Sparse POP

Sparse Putinar – the proof

Let $S(g) = \{x \mid g_j(x) \geq 0\}$ be compact and $f = \sum_k f_k$, with f_k depending on $x(I_k)$, and $f > 0$ on X.

$$\begin{aligned} S_k &= \{x(I_k) \mid g_j(x) \geq 0 \; \forall j \in I_k\} \\ &= \text{the subset of } S(g) \text{ which only ``sees'' variables indexed by } I_k \end{aligned}$$

Lemma (Grimm et al., 2007])

If RiP holds for (I_k), then:

$$f = \sum_k h_k$$

with h_k depending on $x(I_k)$, and $h_k > 0$ on S_k.

- Lemma is proved by induction on the number of subsets I_k
- Sparse Putinar is obtained by applying original Putinar to each h_k

Sparse POP

Beare: sparse SOS $\neq$ SOS sparse

$$
\begin{aligned}
f &= (x_1 + x_2 + x_3)^2 \\
&= \underbrace{\frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + 2x_1x_2}_{f_1 \in \mathbb{R}[x_1, x_2]} + \underbrace{\frac{1}{2}x_1^2 + \frac{1}{2}x_3^2 + 2x_1x_3}_{f_2 \in \mathbb{R}[x_1, x_3]} + \underbrace{\frac{1}{2}x_2^2 + \frac{1}{2}x_3^2 + 2x_2x_3}_{f_3 \in \mathbb{R}[x_2, x_3]}
\end{aligned}
$$

But

$$f \neq \sigma_1^2 + \sigma_2^2 + \sigma_3^2$$

for $\sigma_1 \in \mathbb{R}[x_1, x_2]$, $\sigma_2 \in \mathbb{R}[x_1, x_3]$, $\sigma_3 \in \mathbb{R}[x_2, x_3]$.

⚠ $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$ do not satisfy RiP

Sparse POP

Beare: sparse SOS $\neq$ SOS sparse

$$x_1^2 - 2x_1x_2 - 2x_1^2x_2 + 3x_2^2 + 2x_1^2x_2^2 - 2x_2x_3 + 18x_2^2x_3 + 6x_3^2 - 54x_2x_3^2 + 142x_2^2x_3^2$$

- is sparse w.r.t. $\{1,2\}$, $\{2,3\}$
- is not SOS(x_1, x_2)+ SOS(x_2, x_3)

Outro

Take away messages

- **Polynomial Optimization (POP)** is a powerful framework for solving non-convex problems
- **Challenges in POP:** Non-convexity, NP-hardness, scalability, and numerical stability
- **Lasserre's Hierarchy:** A systematic way to approximate POP using Semidefinite Programming (SDP)
- **Applications:** Used in combinatorial optimization, control theory, quantum information, machine learning, and statistics & finance
- **Software Tools:** Popular options include YALMIP, GloptiPoly, SOSTOOLS, and Julia-based SumOfSquares.jl
- **Key Takeaway:** Lasserre's SDP-based moment-SOS relaxations provide a tractable way to solve hard polynomial problems.