Polynomial Optimization

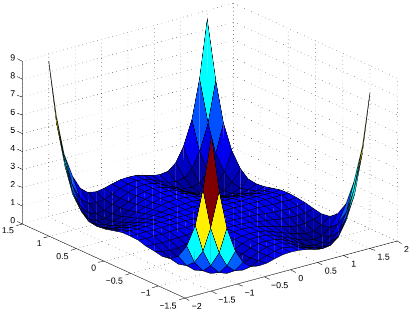
Sums of squares and moments

Part 1

Monique Laurent

POEMA 1st Workshop – Florence, January 2020

What is polynomial optimization?



(P)

Minimize a polynomial function f over a region

$$K = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$$

defined by polynomial inequalities

$\rightsquigarrow$ K is a **basic closed semialgebraic set**

Some hard instances

Testing nonnegative polynomials: $f \geq 0$ on K?

The unconstrained quadratic case is easy:

- A symmetric matrix M is **positive semidefinite** ($M \succeq 0$) if and only if $f_M = x^\mathsf{T} M x \geq 0$ on $\mathbb{R}^n$

Can test whether $M \succeq 0$ in **polynomial time**, using Gaussian elimination

The quartic case is hard:

- A symmetric matrix M is **copositive** if $f_M = x^\mathsf{T} M x \geq 0$ on $\mathbb{R}^n_+$

Equivalently, the polynomial $F_M = \sum_{i,j=1}^n M_{ij} x_i^2 x_j^2$ is nonnegative over $\mathbb{R}^n$

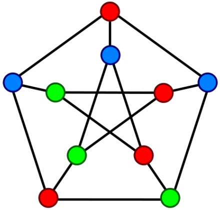
Testing matrix copositivity is co-NP complete [Kabadi-Murty 1987]

- A polynomial f is **convex** if and only if its Hessian matrix $H_f(x) = (\partial^2 f(x)/\partial x_i \partial x_j)_{i,j=1}^n$ is positive semidefinite

Equivalently, the polynomial $y^\mathsf{T} H_f(x) y$ is nonnegative over $\mathbb{R}^n \times \mathbb{R}^n$

Testing convexity is NP-hard [Ahmadi et al. 2013]

Some hard combinatorial problems over graphs



$\alpha = 4, \ \omega = 2, \ \chi = 3$

- **stability number** $\alpha(G)$:

maximum cardinality of a set of pairwise **non-adjacent** vertices (**stable set**)

- **clique number** $\omega(G)$:

maximum cardinality of a set of pairwise **adjacent** nodes (**clique**)

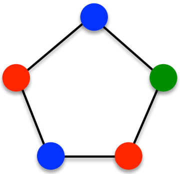
- **coloring number** $\chi(G)$:

minimum number of colors needed to properly color the vertices of G

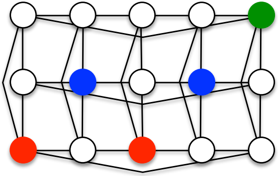
Computing $\alpha(G)$, $\omega(G)$, $\chi(G)$ is hard NP-complete [Karp 1972]

Easy relations: $\omega(G) \leq \chi(G)$ and $\omega(G) = \alpha(\overline{G})$

Reducing coloring to the stability number



G is 3-colorable



$G \square K_3$ has a stable set of cardinality $|V(G)|$

$\chi(G)$ is the smallest integer c such that $\alpha(G \square K_c) = |V(G)|$

Polynomial optimization formulations for $\alpha(G)$

- Basic formulation:

$$\alpha(G) = \max \sum_{v \in V} x_v \;\; \text{s.t.} \;\; x_u x_v = 0 \; (uv \in E), \; x_v^2 = x_v \; (v \in V)$$

- Motzkin-Straus formulation:

$$\frac{1}{\alpha(G)} = \min \; x^T(I + A_G)x \;\; \text{s.t.} \;\; \sum_{v \in V} x_v = 1, \; x_v \geq 0 \; (i \in V)$$

- Copositive formulation:

$$\alpha(G) = \min \; \lambda \;\; \text{s.t.} \;\; \lambda(I + A_G) - J \text{ is copositive}$$

A first bound for $\alpha(G)$ and $\chi(G)$

The **theta number** [Lovász 1979] of a graph $G = (V = [n], E)$:

$$\vartheta(G) = \max_{X \in \mathcal{S}^n} \langle J, X \rangle \;\text{ s.t. } X_{uv} = 0 \;\forall uv \in E,\; Tr(X) = 1,\; X \succeq 0$$

$\rightsquigarrow$ expressed via a **semidefinite program**

$\rightsquigarrow$ can be computed in polynomial time (to arbitrary precision)

(Lovász sandwich) Theorem: $\alpha(G) \;\leq\; \vartheta(G) \;\leq\; \chi(\overline{G})$

How to get **stronger bounds**?

What is semidefinite programming?

Semidefinite programming (SDP) is linear optimization over the cone of positive semidefinite matrices.

LP		SDP
vector variable $x \in \mathbb{R}^n$	$\rightsquigarrow$	symmetric matrix variable $X \in \mathcal{S}^n$
$x \geq 0$		$X \succeq 0$ [positive semidefinite]

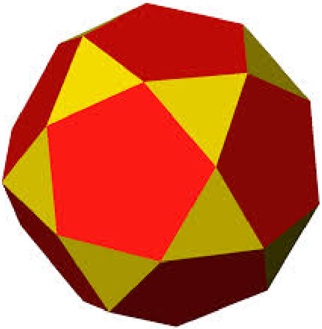
LP
$$\begin{array}{ll} \max_x & \langle c, x \rangle \\ \text{s.t.} & \langle a_j, x \rangle = b_j \quad (j = 1, \ldots, m) \\ & x \geq 0 \end{array}$$

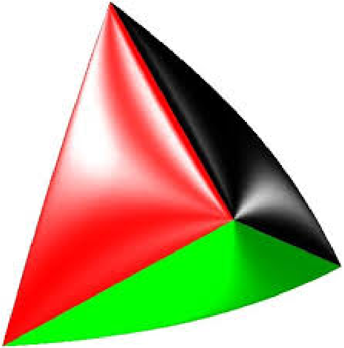
SDP
$$\begin{array}{ll} \sup_X & \langle C, X \rangle \\ \text{s.t.} & \langle A_j, X \rangle = b_j \quad (j = 1, \ldots, m) \\ & X \succeq 0 \end{array}$$

Input data: $b_j \in \mathbb{R}$, $c, a_j \in \mathbb{R}^n$, inner product $\langle c, x \rangle = c^{\mathsf{T}} x = \sum_{i=1}^n c_i x_i$

$C, A_j \in \mathcal{S}^n$, with trace inner product: $\langle C, X \rangle = Tr(C^{\mathsf{T}} X) = \sum_{i,j=1}^n C_{ij} X_{ij}$

Geometrically





LP

SDP

Optimization over a polyhedron

a spectrahedron

SDP duality

SDP (primal form):

$$\text{(P)} \quad \boxed{\begin{array}{lll} p^* := & \sup_{X \in \mathcal{S}^n} & \langle C, X \rangle \\ & \text{s.t.} & \langle A_j, X \rangle = b_j \quad (j = 1, \ldots, m), \ X \succeq 0 \end{array}}$$

SDP (dual form):

$$\text{(D)} \quad \boxed{\begin{array}{lll} d^* := & \inf_{y \in \mathbb{R}^m} & b^\mathsf{T} y \\ & \text{s.t.} & \sum_{j=1}^m y_j A_j - C \succeq 0 \end{array}}$$

Theorem:

- **Weak duality:** $p^* \leq d^*$
- **Strong duality:** $p^* = d^*$ holds in any of the two cases:
  1. (D) is **bounded** ($d^* > -\infty$) and **strictly feasible** ($\exists y$ with $\sum_{j=1}^m y_j A_j - C \succ 0$); then (P) has an optimal solution (sup is max)
  2. (P) is **bounded** ($p^* < \infty$) and **strictly feasible** ($\exists X \succeq 0$ primal feasible); then (D) has an optimal solution (inf is min).

Example

Recall the SDP defining the theta number $\vartheta(G)$:

Primal SDP:

$$\text{(P)} \quad \max \ \langle J, X\rangle \ \text{ s.t. } \ Tr(X) = 1, \ X_{ij} = 0 \ (ij \in E(G)), \ X \succeq 0$$

Dual SDP:

$$\text{(D)} \quad \min \ y \ \text{ s.t. } \ yI + \sum_{ij \in E} z_{ij} E_{ij} - J \succeq 0$$

Observations:

- both (P) and (D) are strictly feasible and bounded
- One can reformulate $\vartheta(G)$ as

$$\vartheta(G) = \min \ y \ \text{ s.t. } \ yI + Z - J \succeq 0, \ Z_{ij} = 0 \text{ if } i = j \text{ or } ij \in E(\overline{G})$$

$$\vartheta(G) = \min \ y \ \text{ s.t. } \ yI - B \succeq 0, \ B_{ij} = 1 \text{ if } i = j \text{ or } ij \in E(\overline{G})$$

$$\vartheta(G) = \min \ \lambda_{\max}(B) \ \text{ s.t. } \ B_{ij} = 1 \text{ if } i = j \text{ or } ij \in E(\overline{G})$$

Algorithms for LP vs. SDP

1940's: Dantzig simplex algorithm for LP.

Works well in practice, but is it **efficient (= poly-time)**?

1980's: **efficient algorithms** for LP and SDP:

Khachiyan: ellipsoid method (not practical)

Karmarkar, Nemirovski-Nesterov: interior-point algorithms (practical)

LP is widely used, also in industrial applications.

SDP has a greater modeling power:

- combinatorial optimization [approximation algorithms]
- **sums of squares of polynomials** [real algebraic geometry]
- quantum information
- many more ...

Testing sums of squares of polynomials with SDP

$$f(x) = \sum_{|\alpha| \leq 2d} f_\alpha x^\alpha$$ is a **sum of squares of polynomials**

$$\Updownarrow$$

$f(x) = \sum_i p_i(x)^2$ [write $p_i(x) = \overline{p_i}^T [x]_d$]

$$\Updownarrow$$

$$f(x) = \sum_i [x]_d^T \overline{p_i}\ \overline{p_i}^T [x]_d = [x]_d^T \Big(\underbrace{\sum_i \overline{p_i}\ \overline{p_i}^T}_{X \succeq 0} \Big) [x]_d$$

$$\Updownarrow$$

The SDP: $\left\{ \begin{array}{l} \displaystyle\sum_{\beta,\gamma | \beta+\gamma=\alpha} X_{\beta,\gamma} = f_\alpha \quad (|\alpha| \leq 2d) \\ X \succeq 0 \end{array} \right.$ **is feasible**

Gram-matrix method [Powers-Wörmann 1998]

Example

$$f(x,y) = x^4 + 2x^3y + 3x^2y^2 + 2xy^3 + 2y^4 \text{ is SOS?}$$

$$\boxed{f(x,y) = (x^2\ xy\ y^2)\underbrace{\begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}}_{X \succeq 0?}\begin{pmatrix} x^2 \\ xy \\ y^2 \end{pmatrix}}$$

Equate coefficients on both sides:

x^4: $a = 1$ $\quad x^3y$: $2b = 2$ $\quad x^2y^2$: $2c + d = 3$ $\quad xy^3$: $2e = 2$ $\quad y^4$: $f = 2$

$$X = \begin{pmatrix} 1 & 1 & c \\ 1 & 3-2c & 1 \\ c & 1 & 2 \end{pmatrix} \succeq 0 \iff -1 \leq c \leq 1$$

$$c = -1 \rightsquigarrow f = (x^2 + xy - y^2)^2 + (y^2 + 2xy)^2$$

$$c = 0 \rightsquigarrow f = (x^2 + xy)^2 + \tfrac{3}{2}(xy + y^2)^2 + \tfrac{1}{2}(xy - y^2)^2$$

General approach to polynomial optimization

Strategy

(P) $$\boxed{f_{\min} = \min_{x \in K} f(x)}$$

Approximate **(P)** by a hierarchy of convex (semidefinite) relaxations

Shor (1987), Nesterov (2000), **Lasserre, Parrilo** (2000–)

Such relaxations can be constructed using

representations of nonnegative polynomials as sums of squares

and

the dual theory of moments

Sums of squares approach

Strategy (use sums of squares)

(P) $$\boxed{f_{\min} = \min_{x \in K} f(x) = \sup_{\lambda \in \mathbb{R}} \ \lambda \ \text{ s.t. } f(x) - \lambda \geq 0 \ \forall x \in K}$$

Testing whether a polynomial f is nonnegative is **hard**

but one can test the *sufficient condition*:

f is a sum of squares of polynomials (SOS)

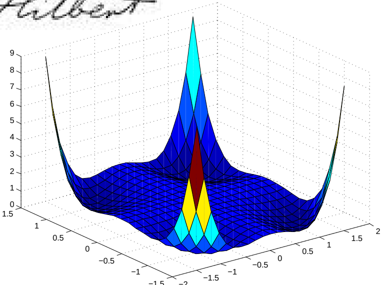
using **semidefinite programming**

Are all nonnegative polynomials SOS?

Hilbert [1888]: *Every nonnegative polynomial in n variables and even degree d is a sum of squares of polynomials* *if and only if* $n = 1$, *or* $d = 2$, *or* ($n = 2$ *and* $d = 4$).

Hilbert's 17th problem [1900]: *Is every nonnegative polynomial a sum of squares of* **rational** *functions?*

Artin [1927]: Yes



Motzkin [1967]:

$p = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$

is nonnegative on $\mathbb{R}^2$,

not a sum of squares, but

$(x^2 + y^2)^2 p$ is SOS!

Another example

Horn matrix:

$$M = \begin{pmatrix} 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 \end{pmatrix} \quad \rightsquigarrow \quad F_M = \sum_{i,j=1}^{5} M_{ij} x_i^2 x_j^2$$

- M is copositive, i.e., F_M is nonnegative
- F_M is **not** a sum of squares (M is **not** psd)
- $(\sum_{i=1}^5 x_i^2) F_M$ is a sum of squares

$$\begin{aligned}(\textstyle\sum_{i=1}^5 x_i^2) f_M = \quad & x_1^2(x_1^2 - x_2^2 + x_3^2 + x_4^2 - x_5^2)^2 \\ & + x_2^2(x_2^2 - x_3^2 + x_4^2 + x_5^2 - x_1^2)^2 \\ & + x_3^2(x_3^2 - x_4^2 + x_5^2 + x_1^2 - x_2^2)^2 \\ & + x_4^2(x_4^2 - x_5^2 + x_1^2 + x_2^2 - x_3^2)^2 \\ & + x_5^2(x_5^2 - x_1^2 + x_2^2 + x_3^2 - x_4^2)^2 \\ & + 4(x_1^2 x_2^2 x_4^2 + x_2^2 x_3^2 x_5^2 + x_3^2 x_4^2 x_1^2 + x_4^2 x_5^2 x_2^2 + x_5^2 x_1^2 x_3^2)\end{aligned}$$

[Parrilo 2000]

Many sums of squares ?

Theorem: [Blekherman 2003]

Few SOS polynomials

when fixing the degree and letting the number of variables grow:

$$\boxed{\frac{\mathrm{vol}(\mathrm{POS}_{n,2d})}{\mathrm{vol}(\mathrm{SOS}_{n,2d})} = \Theta(n^{\frac{d-1}{2}D})} \qquad [D = \binom{n+2d-1}{2d} - 1]$$

Theorem: [Lasserre 2006] [Lasserre-Netzer 2006]

SOS polynomials are **dense** within nonnegative polynomials,

when fixing the number of variables and letting the degree grow:

If $f \geq 0$ on $[-1,1]^n$, then

$\forall \epsilon > 0 \;\; \exists k \in \mathbb{N}$ such that $f + \epsilon \left(1 + \sum_{i=1}^n x_i^{2k}\right)$ is SOS

Positivity certificates over K

$$\boxed{K = \{x \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}} \qquad \mathbf{g} = \{g_1, \ldots, g_m\}$$

Quadratic module: $\mathcal{M}(\mathbf{g}) = \{s_0 + s_1 g_1 + \ldots + s_m g_m \mid s_j \text{ SOS}\}$

Preordering: $\mathcal{P}(\mathbf{g}) = \{\sum_{e \in \{0,1\}^m} s_e g_1^{e_1} \cdots g_m^{e_m} \mid s_e \text{ SOS}\} \supseteq \mathcal{M}(g)$

Theorem: Assume K is compact.

1. [Schmüdgen'91] $f > 0$ on $K \Longrightarrow f \in \mathcal{P}(\mathbf{g}) \Longrightarrow f \geq 0$ on K
2. [Putinar'93] **Archimedean cond.**: $R^2 - \sum_{i=1}^n x_i^2 \in \mathcal{M}(\mathbf{g})$ for some $R > 0$.
   $f > 0$ on $K \Longrightarrow f \in \mathcal{M}(\mathbf{g}) \Longrightarrow f \geq 0$ on K

Positivstellensatz for general K [Krivine 1964, Stengle 1974]

$$f > 0 \text{ on } K \iff \exists p, q \in \mathcal{P}(\mathbf{g}) \quad pf = q + 1$$

$$f \geq 0 \text{ on } K \iff \exists p, q \in \mathcal{P}(\mathbf{g}) \ \exists k \in \mathbb{N} \quad pf = f^{2k} + q$$

SOS relaxations for (P)

Truncated quadratic module:

$$\mathcal{M}(g)_{2t} := \{ \underbrace{s_0}_{\deg \leq 2t} + \underbrace{s_1 g_1}_{\deg \leq 2t} + \ldots + \underbrace{s_m g_m}_{\deg \leq 2t} \mid s_j \ \text{ SOS} \}$$

Replace

(P) $$\boxed{f_{\min} = \inf_{x \in K} f(x) = \sup \ \lambda \ \text{ s.t. } f - \lambda \geq 0 \ \text{ on } K}$$

by

(SOSt) $$\boxed{f_{\mathrm{sos},t} = \sup \ \lambda \ \text{ s.t. } f - \lambda \in \mathcal{M}(g)_{2t}}$$

- $f_{\mathrm{sos},t} \leq f_{\mathrm{sos},t+1} \leq f_{\min}$, $f_{\mathrm{sos},t}$ can be computed with SDP
- If K compact (+ Archimedean), then **asymptotic convergence:** $\lim_{t \to \infty} f_{\mathrm{sos},t} = f_{\min}$ [Lasserre 2001]

Moment approach

$$f_{\min} = \inf_{x \in K} f(x) \quad = \inf_{\mu} \int_K f(x) d\mu \ \text{ s.t. } \mu \text{ is a probability measure on } K$$

$$= \inf_{L \in \mathbb{R}[x]^*} L(f) \ \text{ s.t. } L \text{ has a representing measure } \mu \text{ on } K$$

μ is a **representing measure** of L on K if

$L(p) = \int_K p(x) d\mu(x)$ for all $p \in \mathbb{R}[x]$

Deciding if a linear functional $L \in \mathbb{R}[x]^*$ has a representing measure μ on K

is the (difficult, classical) **moment problem**

But one can use the (easier) **necessary condition**:

L is nonnegative on the quadratic module $\mathcal{M}(g) = \{s_0 + \sum_j s_j g_j : s_j \text{ SOS}\}$:

$L(p^2) \geq 0, \ \ L(p^2 g_j) \geq 0$ for all $p \in \mathbb{R}[x]$ and $j \in [m]$

Moment matrices

$L \in \mathbb{R}[x]^*$ is determined by its values on the monomial base:

$$
\begin{array}{lrcl}
L: & \mathbb{R}[x] & \to & \mathbb{R} \\
 & x^\alpha & \mapsto & L(x^\alpha) =: y_\alpha \\
 & f = \sum_\alpha f_\alpha x^\alpha & \mapsto & L(f) = \sum_\alpha f_\alpha y_\alpha = \overline{f}^{\mathsf{T}} y
\end{array}
$$

Moment matrix: $M(y) := (L(x^\alpha x^\beta))_{\alpha,\beta\in\mathbb{N}^n} = (y_{\alpha+\beta})_{\alpha,\beta\in\mathbb{N}^n}$

Localizing moment matrix: $M(gy) = (L(g(x)x^\alpha x^\beta))_{\alpha,\beta} = (\sum_\gamma g_\gamma y_{\alpha+\beta+\gamma})_{\alpha,\beta}$

L is nonnegative on the quadratic module $\mathcal{M}(g) = \{s_0 + \sum_j s_j g_j : s_j \text{ SOS}\}$:

$$L(p^2) \geq 0 \quad \forall p \quad \Longleftrightarrow \quad M(y) \succeq 0$$

and, for all j, $\quad L(g_j p^2) \geq 0 \quad \forall p \quad \Longleftrightarrow \quad M(g_j y) \succeq 0$

Key facts: $\quad L(p^2) = \overline{p}^{\mathsf{T}} M(y) \overline{p}, \quad L(gp^2) = \overline{p}^{\mathsf{T}} M(gy) \overline{p}$

Examples

- For $n = 1$, $M_t(y)$ is a **Hankel matrix:**

$$\begin{pmatrix} 1 & x & x^2 & x^3 \\ x & x^2 & x^3 & x^4 \\ x^2 & x^3 & x^4 & x^5 \\ x^3 & x^4 & x^5 & x^6 \end{pmatrix} \rightsquigarrow M_3(y) = \begin{pmatrix} y_0 & y_1 & y_2 & y_3 \\ y_1 & y_2 & y_3 & y_4 \\ y_2 & y_3 & y_4 & y_5 \\ y_3 & y_4 & y_5 & y_6 \end{pmatrix}$$

- For $n = 2$, $M_t(y)$ is of **Hankel type**:

$$M_2(y) = \begin{array}{c} \\ 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_2^2 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 \end{array} \\ \begin{pmatrix} y_{00} & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{31} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{30} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{pmatrix} \end{array}$$

Moment relaxations for (P)

(P) $\boxed{f_{\min} = \inf_{L \in \mathbb{R}[x]^*} L(f) \ \text{ s.t. } \ L \text{ has a representing measure } \mu \text{ on } K}$

Truncate at degree $2t$: **(MOMt)**

$$\begin{aligned} f_{\mathrm{mom},t} &= \inf_{L \in \mathbb{R}[x]^*_{2t}} L(f) \ \text{ s.t. } \ L \geq 0 \text{ on } \mathcal{M}(g)_{2t} \\ &= \inf_{y=(y_\alpha)_{|\alpha|\leq 2t}} \overline{f}^{\mathsf{T}} y \ \text{ s.t. } \ M_t(y) \succeq 0, \ M_{t-d_j}(g_j y) \succeq 0 \ \ \forall j \in [m] \end{aligned}$$

(SOSt) $\boxed{f_{\mathrm{sos},t} = \sup \ \lambda \ \text{ s.t. } \ f - \lambda \in \mathcal{M}(g)_{2t}}$ $\rightsquigarrow$ dual SDP

$$\boxed{f_{\mathrm{sos},t} \leq f_{\mathrm{mom},t} \leq f_{\min}}$$

Asymptotic convergence if K is compact [+ Archimedean condition]

Optimality criterion for moment relaxation (MOMt)

$$\boxed{K = \{x \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}}$$

$d_K = \max_j \lceil \deg(g_j)/2 \rceil$

Theorem [CF 1996,2000 + Henrion-Lasserre 2005 + Lasserre-L-Rostalski 2008]

Assume y is an optimal solution of (MOMt) such that

$$\mathrm{rank}\ M_s(y) = \mathrm{rank}\ M_{s-d_K}(y) \text{ for some } d_K \leq s \leq t$$

- Then the relaxation is **exact:** $f_{\mathrm{mom},t} = f_{\min}$
- Moreover, one can compute the **global minimizers:**

$$V_{\mathbb{C}}(\mathrm{Ker} M_s(y)) \subseteq \{ \text{ global minimizers of } f \text{ on } K \},$$

with **equality** if $\mathrm{rank}\ M_t(y)$ is **maximum** (rank = # minimizers)

Remarks

- Many interior point algs for SDP give a **max rank** optimal solution
- Algorithm for computing the (finitely many) **real roots** of polynomial equations (and real radical ideals)
  [Lasserre-L-Rostalski 2008,2009]
- **Finite convergence** holds **generically** [Nie 2013]
- **Finite convergence** in the **convex case**
  [Lasserre 2009, de Klerk-L 2011]
- Several implementations: GloptiPoly [Henrion-Lasserre], SOSTOOLS [Prajna et al.], SparsePOP [Waki et al.], YALMIP [Löfberg]

Exploiting sparsity structure

For $I \subseteq [n]$ $x_I := (x_i)_{i \in I}$ denotes the group of variables indexed by I.

Consider sets $I_1, \ldots, I_L \subseteq [n]$ for which the polynomials p, g_j satisfy:

1. $p = p_1 + \ldots + p_L$, where each $p_\ell \in \mathbb{R}[x_{I_\ell}]$
2. Each g_j belongs to some $\mathbb{R}[x_{I_\ell}]$

Define the (weaker) bound: $\widehat{p_t^{\text{sos}}} = \sup \ \lambda$ s.t. $p - \lambda = s + \sum_{j=1}^m s_j g_j$,
where $s = \sum_{\ell=1}^L s_\ell$ with $s_\ell \in \Sigma[x_{I_\ell}]_{2t}$, and $s_j \in \Sigma[x_{I_\ell}]$ if $g_j \in \mathbb{R}[x_{I_\ell}]$

Then, $\widehat{p_t^{\text{sos}}} \leq p_t^{\text{sos}} \leq p_{\min}$ but $\widehat{p_t^{\text{sos}}}$ involves **smaller** psd matrices!

Theorem: [Lasserre 2006, Grimm-Netzer-Schweighofer 2007]
Assume the sets $I_1, \ldots, I_L$ satisfy the **running intersection property** (up to reordering):

$$\forall k \geq 2 \ \ \exists k_0 \leq k-1 \quad \text{s.t. } I_k \cap (I_1 \cup \ldots \cup I_{k-1}) \subseteq I_{k_0}$$

and, for each $l \leq L$, the polynomials g_j using variables in I_l generate an Archimedean quadratic module.

Then $\lim_{t \to \infty} \widehat{p_t^{\text{sos}}} = p_{\min}$.

Example

- The polynomial

$p(x) =$

$\sum_{i=1}^{n-3}(x_i + 10x_{i+1})^2 + 5(x_{i+2} - x_{i+3})^2 + (x_{i+1} - 2x_{i+2})^4 + 10(x_i - 10x_{i+3})^4$

has sparsity structure satisfying (RIP) for the sets

$I_\ell = \{\ell, \ell+1, \ell+2, \ell+3\}$ for $\ell = 1, \ldots, n-3$

- The polynomial

$$p(x) = \sum_{i=2}^{n}(a_i x_1 + b_i x_i)^4$$

has sparsity structure satisfying (RIP) for the sets

$I_\ell = \{x_1, x_\ell\}$ for $\ell = 2, \ldots, n$

Exploiting equations:
binary polynomial optimization

Back to the stability number $\alpha(G)$

$$\alpha(G) = \max\ p(x) := \sum_{i=1}^n x_i \text{ s.t. } x_i x_j = 0\ (ij \in E),\ x_i^2 = x_i\ (i \in [n])$$

Define the ideal $\mathcal{I}^G$ (and its truncations $\mathcal{I}_t^G$):

$$\mathcal{I}^G = (\{x_i x_j : ij \in E\} \cup \{x_i^2 - x_i : i \in [n)\})$$

Moment bound of order t:

$$\text{las}_t(G) := \max_{L \in (\mathbb{R}[x]_{2t})^*} L(p) \text{ s.t. } L(1) = 1,\ L \geq 0 \text{ on } \Sigma_{2t},\ L = 0 \text{ on } \mathcal{I}_{2t}^G$$

$$= \max \sum_{i=1}^n y_i \text{ s.t. } y_0 = 1,\ M_t(y) \succeq 0,\ L_y = 0 \text{ on } \mathcal{I}_{2t}^G$$

Lemma: Assume y is feasible for the moment bound of order t.

1. y_α depends only on the support $I = \{i \in [n] : \alpha_i \geq 1\}$ of α
   $\rightsquigarrow$ rename y_α as z_I.
2. $M_t(y)$ is a flat extension of $M_n(y)$ if $t \geq n+1$
3. $M_t(y)$ is a flat extension of $M_t^{01}(z) := (z_{I \cup J})_{|I|,|J| \leq t}$.

More economical reformulation of the moment bound $\mathrm{las}_t(G)$:

$$\max \sum_{i=1}^{n} z_i \text{ s.t. } z_\emptyset = 1,\ M_t^{01}(z) \succeq 0,\ z_I = 0 \text{ if } I \text{ contains an edge}$$

Facts:

1. The bound of order 1 is the **theta number**: $\mathrm{las}_1(G) = \vartheta(G)$
2. The bound of order $t = \alpha(G)$ is **exact**:
   $\mathrm{las}_t(G) = \alpha(G)$ if $t \geq \alpha(G)$.
3. Such more economical reformulation extends to any polynomial optimization problem over $\{0,1\}^n$ (or $\{\pm 1\}^n$)

Observation: This extends to $K \subseteq V_{\mathbb{C}}(I)$, where $V_{\mathbb{C}}(I)$ is finite. Then, finite convergence holds and one gets a SDP reformulation for $p_{\min}$ using matrices of size $N = \dim \mathbb{R}[x]/I$

Key idea: Work in the quotient space $\mathbb{R}[x]/I$ instead of $\mathbb{R}[x]$.

Extension to the general moment problem (GMP)

The general moment problem (GMP)

$$\mathrm{val} := \inf_{\mu} \int_K f_0 d\mu(x) \;\; \text{s.t.} \;\; \int_K f_k(x) d\mu(x) = b_k \;\, (k \leq m_0), \;\; \mu \text{ measure}$$

Some instances:

- Minimizing a **rational function** f_0/f_1 (assuming $f_1 > 0$ on K)

$$\min_{x \in K} \frac{f_0(x)}{f_1(x)} = \min_{\mu} \int_K f_0(x) d\mu(x) \text{ s.t. } \int_K f_1(x) d\mu(x) = 1$$

- **Polynomial cubature rules**: The (GMP) problem

$$\int_K x^\alpha d\mu(x) = \int_K x^\alpha dx \qquad \text{for all } |\alpha| \leq d$$

has a solution μ which is finite atomic with $N \leq \binom{n+d}{d}$ atoms

$\rightsquigarrow$ cubature rule with N points, exact for integrating degree $\leq d$ polynomials on K [Tchakaloff'57]

Duality for (GMP)

Primal (GMP):

$$\mathrm{val} := \inf_{\mu} \int_K f_0 d\mu(x) \;\; \text{s.t.} \;\; \int_K f_k(x) d\mu(x) = b_k \;\; (k \leq m_0)$$

Dual program:

$$\mathrm{val}^* = \sup_{y \in \mathbb{R}^{m_0}} \sum_{k=1}^{m_0} b_k y_k \text{ s.t. } f_0 - \sum_{k=1}^{m_0} y_k f_k \geq 0 \text{ on } K$$

Theorem: Assume (GMP) is feasible. Then, $\mathrm{val}^* \leq \mathrm{val}$,
with **equality** if there exists $z \in \mathbb{R}^{m_0+1}$ s.t. $\sum_{k=0}^{m_0} z_k f_k > 0$ on K.
Then, (GMP) has an optimal solution μ, which is finite atomic with at most m atoms.

Moment relaxations for (GMP)

Moment relaxation: for an integer $t \geq \max_k \lceil \deg(f_k)/2 \rceil$

$$\mathrm{val}_t = \inf L(f_0) \text{ s.t. } L(f_k) = b_k \ (k \leq m_0),\ L \geq 0 \text{ on } \mathcal{M}(\mathbf{g})_{2t}$$

Theorem: Assume the Archimedean condition holds for $\mathcal{M}(\mathbf{g})$ and there exists $z \in \mathbb{R}^{m_0+1}$ such that $\sum_{k=0}^{m_0} z_k f_k > 0$ on K. Then

$$\mathrm{val}^* = \sup_t \mathrm{val}_t = \mathrm{val}.$$

Proof: As $\mathrm{val}^* = \mathrm{val}$ and $\sup_t \mathrm{val}_t \leq \mathrm{val}$
it suffices to show $\mathrm{val}^* \leq \sup_t \mathrm{val}_t$.

Let $\epsilon > 0$ and y feasible for val^*, i.e., $f_0 - \sum_{k=1}^{m_0} y_k f_k \geq 0$ on K.

Then, $(\epsilon z_0 + 1) f_0 + \sum_{k=1}^{m_0} (\epsilon z_k - y_k) f_k > 0$ on K, and thus belongs to $\mathcal{M}(\mathbf{g})_{2t}$ for some t.

Then, $(\epsilon z_0 + 1) L(f_0) + \sum_{k=1}^{m_0} (\epsilon z_k - y_k) b_k \geq 0$ $\forall$ L feasible for val_t

Hence, $b^\mathsf{T} y \leq \epsilon b^\mathsf{T} z + (\epsilon z_0 + 1)\mathrm{val}_t \quad \rightsquigarrow \mathrm{val}^* \leq \sup_t \mathrm{val}_t$

Flatness and finite convergence

Theorem:
Let $t \geq d_K, \max_k \lceil \deg(f_k)/2 \rceil$.

Let y be an optimal solution to the relaxation val_t.

Assume that the flatness condition holds:

$$\mathrm{rank}\ M_s(y) = \mathrm{rank}\ M_{s-d_K}(y) =: r \quad \text{for some } s \ \text{ s.t. } d_K \leq s \leq t.$$

Then $\mathrm{val}_t = \mathrm{val}$ and (GMP) has an optimal solution μ which is finite atomic with r atoms.

For more on (GMP) see the monograph *'Moments, Positive Polynomials and Their Applications'* of Lasserre (2009).

Some references

P. Parrilo: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization, PhD thesis, 2000

J.-B. Lasserre: Global optimization with polynomials and the problem of moments, *SIAM J. Optimization*, 2001

J.-B. Lasserre: Moments, Positive Polynomials and their Applications, Imperial College Press, 2009

M. Laurent: Sums of squares, moment matrices and optimization over polynomials, in IMA volume 149, 2009

M. Anjos and J.-B. Lasserre (eds): Handbook on Semidefinite, Conic and Polynomial Optimization, Springer 2012

G. Blekherman, P. Parrilo, R. Thomas (eds): Semidefinite Optimization and Convex Algebraic Geometry, MOS-SIAM Series on Optim., 2012.

J.B. Lasserre: Introduction to Polynomial and Semi-Algebraic Optimization, Cambridge University Press, 2015