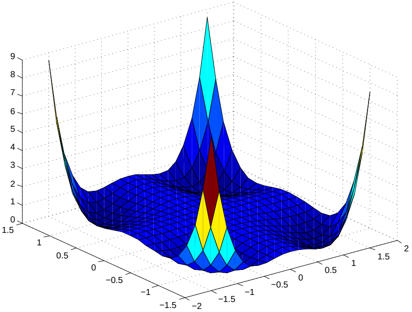
Optimization over Polynomials: Selected Topics

Monique Laurent

ICM 2014

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

Some instances

Testing nonnegativity of polynomials

The quadratic case is easy

The quadratic form $x^\mathsf{T} Mx$ is nonnegative over $\mathbb{R}^n$ if and only if the matrix M is positive semidefinite

This can be tested in **polynomial time**, using Gaussian elimination

The quartic case is hard

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

A symmetric matrix M is **copositive** if $x^\mathsf{T} Mx \geq 0$ for all $x \geq 0$

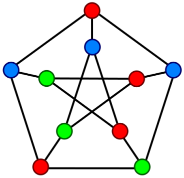
Equivalently, the polynomial $f_M = \sum_{i,j=1}^{n} M_{ij} x_i^2 x_j^2$ is nonnegative over $\mathbb{R}^n$

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

A polynomial $f(x)$ is **convex** if and only if its Hessian matrix $H(f)(x)$ is positive semidefinite

Equivalently, the polynomial $F = y^\mathsf{T} H(f)(x) y$ is nonnegative on $\mathbb{R}^n \times \mathbb{R}^n$

Two hard combinatorial problems over graphs



$\alpha = 4 \quad \chi = 3$

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

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

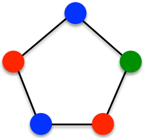
- **coloring number** $\chi(G)$:

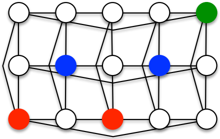
minimum number of colors needed to properly color the vertices of G.

$\alpha(G)$, $\chi(G)$ are NP-complete [Karp 1972]

Chvátal's reduction:

$\chi(G)$ 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:

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

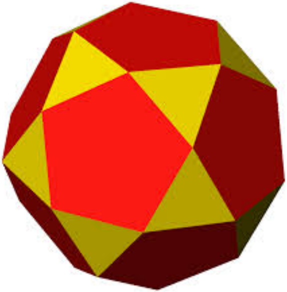
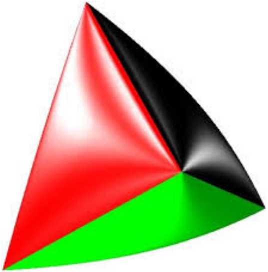
- Copositive formulation:

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

Bounds for $\alpha(G)$ and $\chi(G)$?

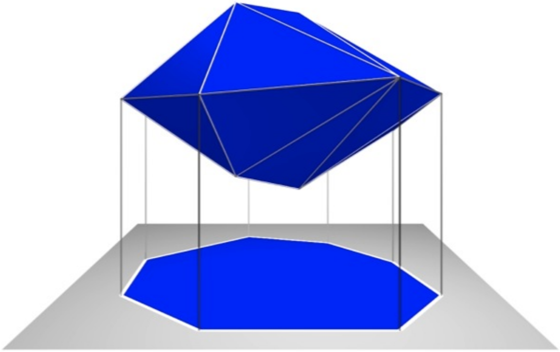
Linear Programming & Semidefinite Programming

Optimize a linear function over

a polyhedron	a convex set (spectrahedron)
$a_j^T x = b_j, \ x \geq 0$	$\langle A_j, X \rangle = b_j, \ X \succeq 0$
	
LP	SDP

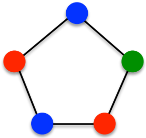
There are efficient algorithms to solve LP and SDP (up to any precision).

One more idea: Lift to higher dimensional space



Add new variables modeling all pairwise products $x_i x_j$ of original variables,
or higher order products $x_i x_j x_k$, etc.

Semidefinite bounds for $\alpha(G)$ and $\chi(G)$



S stable $\rightsquigarrow x = (1,0,0,1,0)^{\mathsf{T}} \rightsquigarrow X = \frac{1}{|S|} x x^{\mathsf{T}}$

$X \succeq 0$ (X is positive semidefinite)

Lovász' theta number: [Lovász 1979]

$$\vartheta(G) = \max \sum_{i,j \in V} X_{ij} \ \text{ s.t. } \ Tr(X) = 1, \ X_{uv} = 0 \ (uv \in E), \ X \succeq 0$$

Sandwich inequalities: $\alpha(G) \leq \vartheta(G) \leq \chi(\overline{G})$

Stronger bounds?

Stronger semidefinite bounds for $\alpha(G)$ and $\chi(G)$

$$S \text{ stable} \rightsquigarrow x = (1,0,0,1,0)^{\mathsf{T}} \rightsquigarrow X = \frac{1}{|S|}xx^{\mathsf{T}} \succeq 0, \ \geq 0$$

Stronger bound for $\alpha(G)$: [McEliece et al. 1978, Schrijver 1979]

$$\vartheta'(G) = \max \sum_{i,j \in V} X_{ij} \ \text{ s.t. } \ Tr(X) = 1, \ X_{uv} = 0 \ (uv \in E), \ X \succeq 0, \ X \geq 0$$

Stronger bound for $\chi(\overline{G})$: [Szegedy 1994]

$$\vartheta^{+}(G) = \max \sum_{i,j \in V} X_{ij} \ \text{ s.t. } \ Tr(X) = 1, \ X_{uv} \leq 0 \ (uv \in E), \ X \succeq 0$$

Sandwich inequalities: $\alpha(G) \leq \vartheta'(G) \leq \vartheta(G) \leq \vartheta^{+}(G) \leq \chi(\overline{G})$

Systematic construction of stronger bounds?

Use *higher order* semidefinite relaxations

General approach to polynomial optimization

Strategy

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

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

Such relaxations can be constructed using

sums of squares of polynomials

and

the dual theory of moments

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

Sums of squares

Strategy (use sums of squares)

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

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

Use SDP to find sums of squares

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

Gram matrix method:

$$f(x,y) = (x^2 \; xy \; y^2) \underbrace{\begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}}_{M \succeq 0 \rightsquigarrow M = \sum_i p_i p_i^T \rightsquigarrow f = \sum_i p_i(x)^2} \begin{pmatrix} x^2 \\ xy \\ y^2 \end{pmatrix}$$

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: $f = 2$

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

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

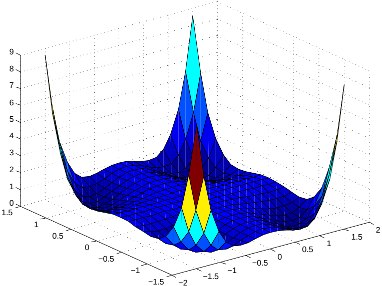
$$c = 0 \rightsquigarrow f = (x^2 + xy)^2 + \tfrac{3}{2}(xy + y^2)^2 + \tfrac{1}{2}(xy - y^2)^2$$

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 is 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,

not a sum of squares, but

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

Are many nonnegative polynomials sums of squares?

Theorem: [Blekherman 2003]

Few SOS polynomials

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

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

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

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

Positivity certificates over K

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

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

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

Theorem: Assume K compact.

- [Schmüdgen 1991] $f > 0$ on $K \implies f \in P(g)$.
- [Putinar 1993] **Archimedean condition:** $\exists R: \ R - \sum_i x_i^2 \in Q(g)$.
  $f > 0$ on $K \implies f \in Q(g)$.

Positivstellensatz: [Krivine 1964, Stengle 1974]

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

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

SOS relaxations for (P)

Truncated quadratic module:

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

Replace

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

by

(SOSt) $f_t^{\text{sos}} = \sup \lambda \text{ s.t. } f - \lambda \in Q(g)_t$

- Each bound f_t^{sos} can be computed with SDP.
- $f_t^{\text{sos}} \leq f_{t+1}^{\text{sos}} \leq f_{\min}$.
- **Asymptotic convergence:** $\lim_{t \to \infty} f_t^{\text{sos}} = f_{\min}$. [Lasserre 2001]

Degree bounds? Convergence?

• For the general **Positivstellensatz**, recent degree bounds by [Lombardi-Perruci-Roy 2014] (depending on n, m and the maximum degree, exponential with 5 towers)

• For the **Positivstellensatz of Schmüdgen**, degree bounds and error estimates by [Schweighofer 2004]: There exist constant $c, c' > 0$ such that, for any polynomial f of degree d:

(i) If $f > 0$ on K then $f \in P_t(g)$ for $t \leq cd^2 \left(1 + \left(d^2 n^d \frac{L(f)}{f_{\min}}\right)^c\right)$.

(ii) $f_{\min} - f_t^{\text{sos}} \leq c' L(f) \frac{d^4 n^{2d}}{\sqrt[c']{t}}$ for any $t \geq c' d^{c'} n^{c'd}$.

Sharper estimates for $K = [0,1]^n$ by [De Klerk-L 2010], roughly $c = c' = 1$. Also for the standard simplex by [DeKlerk-Parrilo-L 2006] (convergence in $1/t$).

• For **Putinar's Positivstellensatz**, results by [Nie-Schweighofer 2007]:

If $f > 0$ on K then $f \in Q_t(g)$ for $t \leq c'' \exp\{(2d^2 n^d)^{c''}\}$.

• When $K \subseteq \{0,1\}^n$, convergence in n steps.

For $K = \{0,1\}^n$, there are linear lower bounds [L 2003].

Moments

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

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

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 **necessary condition**:

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

$$L(p^2) \geq 0 \;\; \forall p, \quad \text{i.e.,} \quad M(L) = (L(x^{\alpha+\beta}))_{\alpha,\beta \in \mathbb{N}^n} \succeq 0$$

and $$L(g_j p^2) \geq 0 \;\; \forall p, \quad \text{i.e.,} \quad M(g_j L) = (L(g_j(x) x^{\alpha+\beta}))_{\alpha,\beta \in \mathbb{N}^n} \succeq 0$$

$M(L)$ is a **moment matrix** and $M(g_j L)$ are **localizing moment matrices**

Moment relaxations for (P)

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

Truncate at degree $2t$:

(MOMt) $f_t^{\text{mom}} = \inf_{L\in\mathbb{R}[x]_{2t}^*} L(f)$ s.t. $L(1) = 1$, $L \geq 0$ on $Q(g)_t$

i.e., $M_t(L) \succeq 0$, $M_{t-d_j}(g_j L) \succeq 0 \ \ \forall j$

(SOSt) $f_t^{\text{sos}} = \sup \lambda$ s.t. $f - \lambda \in Q(g)_t$

$$f_t^{\text{sos}} \leq f_t^{\text{mom}} \leq f_{\min}$$

Sufficient condition for representing measure

$$M_t(L) = \begin{array}{|c|c|}\hline M_{t-1}(L) & \\ \hline & \\ \hline \end{array}$$

Theorem [Curto-Fialkow 1996 - L 2005: short algebraic proof]

If $M_t(L) \succeq 0$ and $\mathrm{rank}\ M_t(L) = \mathrm{rank}\ M_{t-1}(L)$,

then L has a representing measure μ: $L(f) = \int f(x)\mu(dx)$ for $\deg(f) \leq 2t$

- Extend L to $L \in \mathbb{R}[x]^*$ with $\mathrm{rank}\ M(L) = \mathrm{rank} M_t(L) =: r$
- $M(L) \succeq 0$ with finite rank $r \implies L$ has a r-atomic measure μ
  1. the kernel $I = \mathrm{Ker} M(L)$ is a **real radical ideal** in $\mathbb{R}[x]$
  2. the quotient algebra $\mathbb{R}[x]/I$ has **finite dimension r**
  3. the variety $V(I)$ has **r points in** $\mathbb{C}^{\mathbf{n}}$, in fact $\mathbb{R}^{\mathbf{n}}$ $\rightsquigarrow$ **support of** μ
  4. it can be computed with the **eigenvalue method**
  5. define the **positive measure** μ (using interpolation polynomials at the points of $V(I)$)

Optimality criterion for moment relaxation (MOMt)

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

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

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

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

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

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

$V(\mathrm{Ker} M_s(L)) \subseteq \{$ global minimizers of f on $K\}$,

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

Properties

- Interior point algs for SDP give a **maximum rank** optimal solution

- Algorithm for computing the (finitely many) **real roots** of polynomial equations (and real radical ideals)
  [Lasserre-L-Rostalski 2008,2009]
  [Lasserre-L-Mourrain-Rostalski-Trebuchet 2013]

- **Finite convergence** holds in the **finite variety case** [L 2007]
  in the **convex case** [Lasserre 2009, de Klerk-L 2011]

- **Finite convergence** holds **generically** [Nie 2013]

- Several implementations: GloptiPoly [Henrion-Lasserre], SOSTOOLS [Prajna et al.], SparsePOP [Waki et al.], YALMIP [Löfberg]

Applications to $\alpha(G)$ and $\chi(G)$

Semidefinite hierarchies for $\alpha(G)$, $\chi(G)$

$$\alpha(G) = \max \sum_{i \in V} x_i \ \text{ s.t. } \ x_i x_j = 0 \ (ij \in E), \ x_i^2 = x_i \ (i \in V)$$

$$\text{las}_t(G) = \max \ L(\sum_{i \in V} x_i) \ \text{ s.t. } \ M_t(L) \succeq 0, \ L(1) = 1, \ L(x_i x_j) = 0 \ (ij \in E),$$
$$L = 0 \text{ on truncated ideal } (x_i^2 - x_i : i \in [n])_{2t}$$

$$\chi(G) = \min \ c \ \text{ s.t. } \ \alpha(G \square K_c) = |V(G)|$$

$$\text{Las}_t(G) = \min \ c \ \text{ s.t. } \ \text{las}_t(G \square K_c) = |V(G)|$$

- $\alpha(G) \leq \text{las}_t(G)$, with equality if $t = \alpha(G)$.
- $\text{Las}_t(G) \leq \chi(G)$, with equality if $t = n$. [Gvozdenović-L 2008]
- At $t = 1$, get theta number: $\text{las}_1(G) = \vartheta(G)$, $\text{Las}_1(G) = \vartheta(\overline{G})$.
- These hierarchies refine other known hierarchies [L 2003] [GL 2008]

Applications

- Coding problem: Find the maximum cardinality of an error correcting binary code of length N
- Compute the stability number $\alpha(G_N)$ of a Hamming graph G_N: $V = \{0,1\}^N$, edges = pairs $\{u,v\}$ at small Hamming distance.
- Big graph ($|V| = 2^N$), but with a large automorphism group.
- $\mathrm{las}_t(G_N)$: SDP of size $O(N^{2^t-1})$ instead of $O(|V|^t = 2^{tN})$

  [L 2007]
- Best known upper bounds for $\alpha(G_N)$ obtained by computing (variations of) the Lasserre bound of order $t = 2$.

  Also bounds for $\chi(G_N)$ (and other symmetric graphs).

  [Gijswijt, Gvozdenović, L, Mittelman, Regts, Schrijver, Vallentin...]
- Extensions to geometric problems: kissing number problem [Bachoc-Vallentin 2008], coloring of the Euclidean space [Bachoc-Nebe-Oliveira-Vallentin 2009] (α, χ in infinite graphs)

Concluding remarks

- SDP hierarchies are used for **approximation algorithms** in TCS
- SDP hierarchies are used for **noncommutative** polynomials
  - Evaluate polynomials at matrices or operators (instead of scalars)
  - **[Helton 2002]** $f(X_1, \ldots, X_n) \succeq 0$ for all $X_1, \ldots, X_n \in \mathcal{S}^d$ $(\forall d \geq 1)$
    $\Longleftrightarrow f$ is a sum of hermitian squares
  - SDP hierarchies for quantum analogues of $\alpha(G)$ and $\chi(G)$?
- $M_G = \alpha(G)(I + A_G) - J \quad \rightsquigarrow \quad f_G = \sum_{i,j \in V} x_i^2 x_j^2 (M_G)_{ij} \ \geq 0$

Conjecture: $f_G \left(\sum_{i \in V} x_i^2\right)^{\alpha(G)-1}$ is a sum of squares

[de Klerk-Pasechnik 2002] Partial proof [Gvozdenović-L 2007]

- **Question** [DeKlerk-L 2010] What is the smallest constant C_n such that $x_1 \cdots x_n + C_n \in Q_n(x_i^2 - x_i : i \in [n])$? $C_n = \frac{1}{n(n+2)}$?

Thank you

Example

Compute the real roots of the polynomial equations:

$$
\begin{aligned}
g_1 &= 5x_1^9 - 6x_1^5 x_2 + x_1 x_2^4 + 2x_1 x_3, \\
g_2 &= -2x_1^6 x_2 + 2x_1^2 x_2^3 + 2x_2 x_3, \\
g_3 &= x_1^2 + x_2^2 - 0.265625.
\end{aligned}
$$

order t	rank sequence	accuracy
5	1 4 8 16 25 34	—
6	1 3 9 15 22 26 32	—
7	1 3 8 10 12 16 20 24	—
8	1 4 8 8 8 12 16 20 24	4.6789e-5

8 real roots (and 20 complex roots).