Optimization over Polynomials with Sums of Squares and Moment Matrices

Monique Laurent

Centrum Wiskunde & Informatica (CWI), Amsterdam
and University of Tilburg

Positivity, Valuations and Quadratic Forms

Konstanz, October 2009

Polynomial optimization problem

(P) *Minimize a polynomial function* p *over a basic closed semi-algebraic set* K

$$p_{\min} := \inf_{x \in K} p(x)$$

where

$$K := \{x \in \mathbb{R}^n \mid h_1(x) \geq 0, \ldots, h_m(x) \geq 0\}$$

$p, h_1, \ldots, h_m \in \mathbb{R}[x]$ are multivariate polynomials

$$p_{\min} := \inf_{x \in \mathbb{R}^n} p(x)$$

$$p_{\min} \geq 0 \iff p \geq 0 \text{ on } \mathbb{R}^n$$

Example: The partition problem.
A sequence $a_1, \ldots, a_n \in \mathbb{N}$ can be partitioned
if $\sum_{i \in I} a_i = \sum_{i \in [n] \setminus I} a_i$ for some $I \subseteq [n]$, i.e. if
$p_{\min} = 0$, where $p(x) = (\sum_{i=1}^{n} a_i x_i)^2 + \sum_{i=1}^{n} (x_i^2 - 1)^2$

E.g., the sequence **1, 1, 2, 2, 3, 4, 5** can be partitioned.

$\rightsquigarrow$ **NP-complete problem**

Example: Testing matrix copositivity

$M \in \mathbb{R}^{n\times n}$ is copositive if $\ x^T M x \geq 0 \quad \forall x \in \mathbb{R}^n_+$

i.e. if $p_{\min} = 0$, where $\ p(x) = \sum_{i,j=1}^n M_{ij} x_i^2 x_j^2$

$\rightsquigarrow$ **co-NP-complete problem**

$$\begin{pmatrix} 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \text{ is copositive}$$

$$\begin{pmatrix} 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 \end{pmatrix} \text{ is copositive}$$

0/1 Linear programming

$$\min\ c^T x \ \text{s.t.}\ Ax \le b,\ x_i^2 = x_i\ \ (i = 1, \ldots, n)$$

Example: The **stability number** $\alpha(G)$ of a graph $G = (V, E)$ can be computed via any of the programs:

$$\alpha(G) = \max \sum_{i \in V} x_i \ \text{s.t.}\ x_i + x_j \le 1\ (ij \in E),\ x_i^2 = x_i\ (i \in V)$$

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

$\rightsquigarrow$ (P) is **NP-hard** for linear objective and quadratic constraints, or for quadratic objective and linear constraints

Strategy

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

Shor (1987), Nesterov, Lasserre, Parrilo (2000-)

Such relaxations can be constructed using

representations of nonnegative polynomials as sums of squares of polynomials

and

the dual theory of moments

Underlying paradigm

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

but

one can test whether p is a sum of squares of polynomials
efficiently via semidefinite programming

Plan of the talk

- Role of semidefinite programming in sums of squares

- SOS/Moment relaxations for (P)

- Main properties:

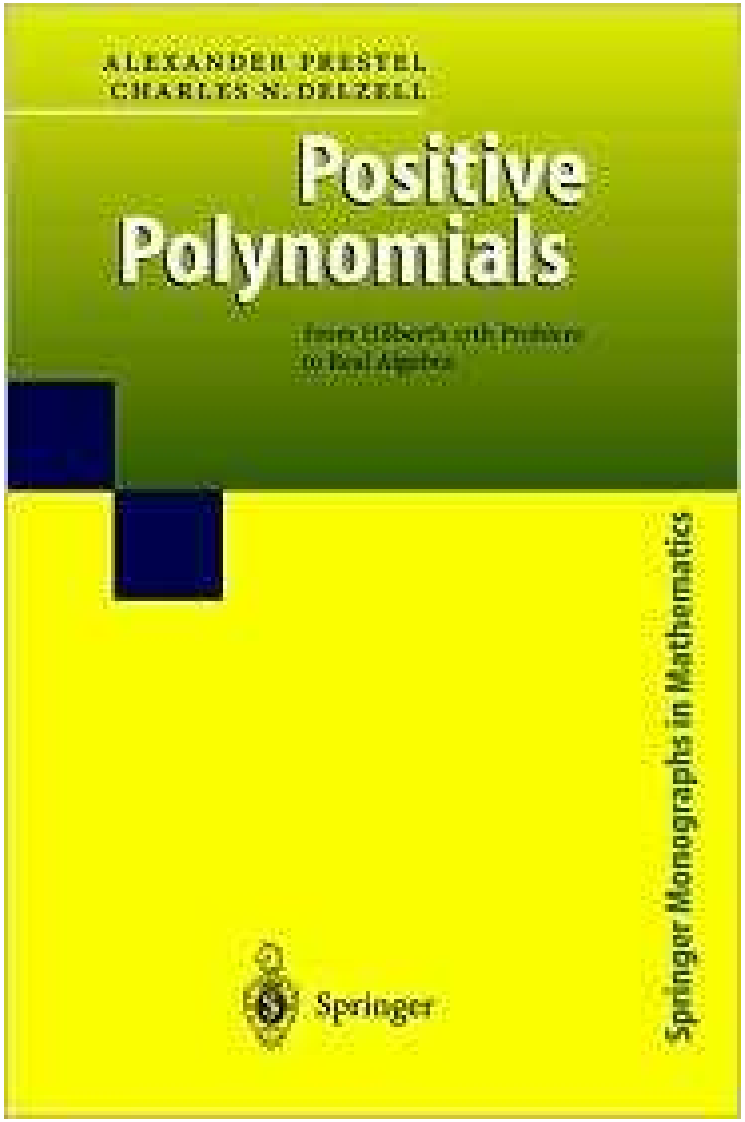
(1) Asymptotic/finite convergence
via SOS representation results for positive polynomials

(2) Optimality criterion
via results for the moment problem

(3) Extract global minimizers
by solving polynomial equations

- Application to unconstrained polynomial optimization

A beautiful monograph about positive polynomials ...



Alexander Prestel
Charles N. Delzell

Positive Polynomials

From Hilbert's 17th Problem to Real Algebra

Some notation

- $\mathbb{R}[x] = \mathbb{R}[x_1, \ldots, x_n]$: ring of polynomials in n variables

- $\mathbb{R}[x]_d$: all polynomials with degree $\leq d$

$$p \in \mathbb{R}[x]_d \quad \rightsquigarrow p(x) = \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| \leq d}} p_\alpha \underbrace{x_1^{\alpha_1} \cdots x_n^{\alpha_n}}_{x^\alpha} = \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| \leq d}} p_\alpha x^\alpha$$

$$\rightsquigarrow \; p(x) = \vec{p}^{\,T} [x]_d$$

after setting $\vec{p} = (p_\alpha)_\alpha$: vector of coefficients

and $[x]_d = (x^\alpha)_\alpha$: vector of monomials

What is semidefinite programming?

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

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

$$\begin{array}{ll} \sup_{\boldsymbol{X}} & \langle \boldsymbol{C}, \boldsymbol{X} \rangle \\ \text{s.t.} & \langle \boldsymbol{A}_{\boldsymbol{j}}, \boldsymbol{X} \rangle = \boldsymbol{b}_{\boldsymbol{j}} \quad (\boldsymbol{j} = \mathbf{1}, \ldots, \boldsymbol{m}) \\ & \color{red}{\boldsymbol{X} \succeq \mathbf{0}} \end{array}$$

There are efficient algorithms to solve semidefinite programs

A small example of SDP

$$\max\ (X_{13} + X_{31})/2 \quad \text{such that} \quad \begin{aligned} & X \succeq 0,\ X \in \mathbb{R}^{3\times 3} \\ & X_{11} = 1,\ X_{12} = 1 \\ & X_{23} = 1,\ X_{33} = 2 \\ & 2X_{13} + X_{22} = 3 \end{aligned}$$

$$\max\ c \quad \text{such that } X = \begin{pmatrix} 1 & 1 & c \\ 1 & 3-2c & 1 \\ c & 1 & 2 \end{pmatrix} \succeq 0$$

One can check that $\max\ c = 1$ and $\min\ c = -1$

The Gram-matrix method to recognize sums of squares [cf. Powers-Wörmann 1998]

Write $p(x) = \sum_{|\alpha| \leq 2d} p_\alpha x^\alpha \in \mathbb{R}[x]_{2d}$ as a sum of squares:

$$p(x) = \sum_{j=1}^{k} (u_j(x))^2 = \sum_{j=1}^{k} [x]_d^T \vec{u_j} \, \vec{u_j}^T [x]_d$$

$$= [x]_d \underbrace{\left(\sum_{j=1}^{k} \vec{u_j} \vec{u_j}^T \right)}_{=: U \succeq 0} [x]_d = \sum_{|\beta|, |\gamma| \leq d} x^\beta x^\gamma U_{\beta,\gamma}$$

$$= \sum_{|\alpha| \leq 2d} x^\alpha \underbrace{\left(\sum_{\substack{|\beta|, |\gamma| \leq d \\ \beta + \gamma = \alpha}} U_{\beta,\gamma} \right)}_{= p_\alpha}$$

Recognize sums of squares via SDP

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

$$\Updownarrow$$

The following semidefinite program is feasible:

$$\begin{cases} U \succeq 0 \\ \sum_{\substack{|\beta|,|\gamma| \leq d \\ \beta+\gamma=\alpha}} U_{\beta,\gamma} = p_\alpha \quad (|\alpha| \leq 2d) \end{cases}$$

Example: Is $p = x^4 + 2x^3y + 3x^2y^2 + 2xy^3 + 2y^4$ SOS ?

Solution: Try to write

$$p(x,y) \equiv (x^2 \; xy \; y^2) \underbrace{\begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}}_{U} \begin{pmatrix} x^2 \\ xy \\ y^2 \end{pmatrix} \quad \text{with } U \succeq 0$$

Equating coefficients:

$$\begin{array}{ll} x^4 = x^2 \cdot x^2 & 1 = a \\ x^3y = x^2 \cdot xy & 2 = 2b \\ x^2y^2 = xy \cdot xy = x^2 \cdot y^2 & 3 = d + 2c \\ xy^3 = xy \cdot y^2 & 2 = 2e \\ y^4 = y^2 \cdot y^2 & 2 = f \end{array}$$

Example continued

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

- For $c = -1$, $U = \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \\ 0 & 2 & 1 \end{pmatrix}$

$\rightsquigarrow p = (x^2 + xy - y^2)^2 + (y^2 + 2xy)^2$

- For $c = 0$, $U = \begin{pmatrix} 1 & 0 & 0 \\ 1 & \sqrt{\frac{3}{2}} & \sqrt{\frac{1}{2}} \\ 0 & \sqrt{\frac{3}{2}} & -\sqrt{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & \sqrt{\frac{3}{2}} & \sqrt{\frac{3}{2}} \\ 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{pmatrix}$

$\rightsquigarrow p = (x^2 + xy)^2 + \frac{3}{2}(xy + y^2)^2 + \frac{1}{2}(xy - y^2)^2$

Which nonnegative polynomials are SOS ?

Hilbert [1888] classified the pairs (n, d) for which
every nonnegative polynomial of degree d in n variables is SOS:

- $n = 1$
- $d = 2$
- $n = 2,\ d = 4$

$\Sigma_{n,d} \subset \mathcal{P}_{n,d}$ for all other (n, d)

Motzkin polynomial: $x^4y^2 + x^2y^4 - 3x^2y^2 + 1$ lies in
$\mathcal{P}_{2,6} \setminus \Sigma_{2,6}$

How many nonnegative polynomials are sums of squares ?

[Blekherman 2003]: **Very few !**

At **fixed degree** $2d$ and large number n of variables, there are significantly more nonnegative polynomials than sums of squares:

$$c \cdot n^{\frac{d-1}{2}} \leq \left(\frac{\mathrm{vol}(\widehat{\mathcal{P}}_{n,2d})}{\mathrm{vol}(\widehat{\Sigma}_{n,2d})} \right)^{\frac{1}{D}} \leq C \cdot n^{\frac{d-1}{2}}$$

$$\widehat{\mathcal{P}}_{n,2d} := \{p \in \mathcal{P}_{n,2d} \mid \ p \text{ homogeneous, } \deg(p) = 2d, \int_{S^{n-1}} p(x)\mu(dx) = 1\}$$

$D := \binom{n+2d-1}{2d} - 1$, the dimension of the ambient space

How many nonnegative polynomials are sums of squares ?

Many !

The SOS cone is dense in the cone of nonnegative polynomials (allowing **variable degrees**):

[Lasserre 2004]: If $p \geq 0$ on $\mathbb{R}^n$, then

$$\forall \epsilon > 0 \;\; \exists k \in \mathbb{N} \;\; \text{s.t.} \;\; p + \epsilon \left(\sum_{h=0}^{k} \sum_{i=1}^{n} \frac{x_i^{2h}}{h!} \right) \;\; \text{is SOS}$$

[Lasserre-Netzer 2006]: If $p \geq 0$ on $[-1, 1]^n$, then

$$\forall \epsilon > 0 \;\; \exists k \in \mathbb{N} \;\; \text{s.t.} \;\; p + \epsilon \left(1 + \sum_{i=1}^{n} x_i^{2k} \right) \;\; \text{is SOS}$$

Artin [1927] solved Hilbert's 17th problem [1900]

$$p \geq 0 \text{ on } \mathbb{R}^n \Longrightarrow p = \sum_i \left(\frac{p_i}{q_i}\right)^2, \text{ where } p_i, q_i \in \mathbb{R}[x]$$

That is, $p \cdot q^2$ is SOS for some $q \in \mathbb{R}[x]$

Sometimes, the shape of the common denominator is known:

Pólya [1928] + Reznick [1995]: For p *homogeneous*

$$p > 0 \text{ on } \mathbb{R}^n \setminus \{0\} \Longrightarrow p \cdot \left(\sum_{i=1}^n x_i^2\right)^r \text{ SOS for some } r \in \mathbb{N}$$

An example [Parrilo 2000]

$$M := \begin{pmatrix} 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 \end{pmatrix}$$

$$p := \sum_{i,j=1}^{5} M_{ij} x_i^2 x_j^2$$

p is **not SOS**

But $(\sum_{i=1}^{5} x_i^2)p$ **is SOS**

This is a **certificate** that $p \geq 0$ on $\mathbb{R}^5$, i.e., that M is copositive

SOS certificates for positivity on a semi-algebraic set K

Let $K = \{x \in \mathbb{R}^n \mid h_1(x) \geq 0, \ldots, h_m(x) \geq 0\}$

Set $h_0 := 1$

Quadratic module: $M(h) := \left\{ \sum_{j=0}^{m} s_j h_j \mid s_j \in \Sigma_n \right\}$

Preordering: $T(h) := \left\{ \sum_{e \in \{0,1\}^m} s_e h_1^{e_1} \cdots h_m^{e_m} \mid s_e \in \Sigma_n \right\}$

$$p \in M(h) \Longrightarrow p \in T(h) \Longrightarrow p \geq 0 \text{ on } K$$

The Positivstellensatz [Krivine 1964] [Stengle 1974]

Not an equivalence:

$K = \{x \in \mathbb{R} \mid (1-x^2)^3 \geq 0\}$

$p = 1 - x^2$

Then, $p \geq 0$ on K, but $p \notin T(h)$

The Positivstellensatz characterizes equivalence:

$p \geq 0$ on $K \iff pf = p^{2m} + g$ for some $f, g \in T(h)$ and some $m \in \mathbb{N}$

But this does not yield tractable relaxations for (P)!

The Positivstellensatz [Krivine 1964] [Stengle 1974]

$$K = \{x \in \mathbb{R}^n \mid h_1(x) \geq 0, \ldots, h_m(x) \geq 0\}$$

(1) $p > 0$ on $K \iff pf = 1 + g$ for some $f, g \in T(h)$

(2) $p \geq 0$ on $K \iff pf = p^{2m} + g$ for some $f, g \in T(h)$, $m \in \mathbb{N}$

(3) $p = 0$ on $K \iff -p^{2m} \in T(h)$ for some $m \in \mathbb{N}$

(2) $\Longrightarrow$ solution to **Hilbert's 17th problem**

(3) $\Longrightarrow$ **Real Nullstellensatz**

But this does not yield tractable relaxations for (P)!!!

SOS certificates of positivity on K compact

Schmüdgen [1991]: Assume K is compact.

$$p > 0 \text{ on } K \implies p \in T(h)$$

Putinar [1993]:

Assume the following *Archimedean condition* holds:

(A) $\forall p \in \mathbb{R}[x] \;\; \exists N \in \mathbb{N} \;\; N \pm p \in M(h)$

Equivalently: $\exists \; N \in \mathbb{N} \quad N - \sum_{i=1}^{n} x_i^2 \in M(h)$

$$p > 0 \text{ on } K \implies p \in M(h)$$

SOS relaxations for (P) [Lasserre 2001, Parrilo 2000]

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

Relax	$p - \lambda \geq 0$ on K	[hard condition]
by	$p - \lambda \in M(h)$	[SOS but unbounded degrees ...]
by	$p - \lambda \in M(h)_{2t}$	[tractable SDP!]

$$M(h)_{2t} := \left\{ \sum_{j=0}^{m} s_j h_j \mid s_j \in \Sigma_n,\ \deg(s_j h_j) \leq 2t \right\}$$

$\rightsquigarrow$ Relaxation (SOSt):

$$p_t^{\text{sos}} := \sup \lambda \text{ s.t. } p - \lambda \in M(h)_{2t} \leq p_{t+1}^{\text{sos}} \leq p_{\min}$$

Asymptotic convergence

If (A) holds for K, then $\lim_{t\to\infty} p_t^{\text{sos}} = p_{\min}$

Proof: $p - p_{\min} + \epsilon > 0$ on K
$\Longrightarrow \exists t \;\; p - p_{\min} + \epsilon \in M(h)_{2t}$
$\Longrightarrow p_t^{\text{sos}} \geq p_{\min} - \epsilon$

Note: A representation result valid for "$p \geq 0$ on K" gives

finite convergence: $p_t^{\text{sos}} = p_{\min}$ for some t

[Nie-Schweighofer 2007]: $p_{\min} - p_t^{\text{sos}} \leq \frac{c'}{\sqrt[c]{\log(t/c)}}$ for t big,

where $c = c(h)$ and $c' = c(p, h)$

Dual moment relaxations for (P) [Lasserre 2001]

$$p_{\min} = \inf_{x\in K} p(x) = \inf_{\substack{\mu \text{ probability} \\ \text{measure on } K}} \int_K p(x)d\mu(x)$$

$$= \inf_{L\in\mathbb{R}[x]^*} L(p) \;\text{ s.t. } L \text{ comes from a probability measure } \mu \text{ on } K$$

$$= \inf_{L\in\mathbb{R}[x]^*} L(p) \;\text{ s.t. } L(f) \geq 0 \;\forall f \geq 0 \text{ on } K$$

[Haviland] The following are equivalent for $L \in \mathbb{R}[x]^*$:

- L *comes from a nonnegative measure* on K, i.e.,

  $L(f) = \int_K f(x)d\mu(x) \;\; \forall f \in \mathbb{R}[x]$

- $L(p) \geq 0$ if $p \geq 0$ on K

Dual moment relaxations for (P) [continued]

$$p_{\min} = \inf_{L \in \mathbb{R}[x]^*} L(p) \text{ s.t. } L(f) \geq 0 \ \forall f \geq 0 \text{ on } K$$

Relax $L(f) \geq 0 \quad \forall f \geq 0 \text{ on } K$

by $L(f) \geq 0 \quad \forall f \in M(h)$

by $L(f) \geq 0 \quad \forall f \in M(h)_{2t}$

$\rightsquigarrow$ Relaxation (MOMt):

$$p_t^{\text{mom}} := \inf_{L \in \mathbb{R}[x]_{2t}^*} L(p) \text{ s.t. } L \geq 0 \text{ on } M(h)_{2t}$$

Weak duality: $p_t^{\text{sos}} \leq p_t^{\text{mom}} \leq p_{\min}$

Equality: $p_t^{\text{sos}} = p_t^{\text{mom}}$ e.g. if $\text{int}(K) \neq \emptyset$

The dual relaxation (MOMt) is again an SDP

$L \in \mathbb{R}[x]^*_{2t} \rightsquigarrow M_t(L) := (L(x^\alpha x^\beta))_{|\alpha|,|\beta| \le t}$

$M_t(L)$ is the *moment matrix* of L (of order t)

Lemma: $L(f^2) \ge 0 \ \forall f \in \mathbb{R}[x]_t \iff M_t(L) \succeq 0$

Proof: $L(f^2) = \vec{f}^T M_t(L) \vec{f}$

$\rightsquigarrow$ Can express $L \ge 0$ on $M(h)_{2t}$,
i.e., $L(f^2 h_j) \ge 0 \ \forall f \ \forall j$ with $\deg(f^2 h_j) \le 2t$
as SDP conditions

Optimality criterion

Observation: Let L be an optimum solution to (MOMt).
If L comes from a probability measure μ on K,
then (MOMt) is exact: $p_t^{\text{mom}} = p_{\min}$.

Question: How to recognize whether L has a representing measure on K ?

Next: Sufficient condition of Curto-Fialkow for the moment problem

A sufficient condition for the moment problem

Theorem [Curto-Fialkow 1996] Let $L \in \mathbb{R}[x]^*_{2t}$.

> If $M_t(L) \succeq 0$ and (RC) $\text{rank} M_t(L) = \text{rank} M_{t-1}(L)$,
> then L has a representing measure.

Corollary [Curto-Fialkow 2000] + [Henrion-Lasserre 2005]

Set $d := \max_j \lfloor \deg(h_j)/2 \rfloor$.

Let L be an optimum solution to (MOMt) satisfying

$\text{rank} M_t(L) = \text{rank} M_{t-d}(L)$. Then, $p_t^{\text{mom}} = p_{\min}$ and

> $V(\text{Ker} M_t(L)) \subseteq \{$ global minimizers of p on $K\}$

with **equality** if $\text{rank} M_t(L)$ is **maximum**.

Remarks

- Compute $V(\mathrm{Ker} M_t(L))$ with the *eigenvalue method*.

- If the rank condition holds at a maximum rank solution, then (P) has *finitely many global minimizers*.

But the reverse is not true !

- The rank condition *holds* in the **finite variety case**:

$$K = \{x \in \mathbb{R}^n \mid \underbrace{h_1(x) = 0, \ldots, h_k(x) = 0}_{\text{ideal } I}, h_{k+1}(x) \geq 0, \ldots\}$$

with $|V_{\mathbb{R}}(I)| < \infty$

Finite convergence in the finite variety case

$$K = \{x \in \mathbb{R}^n \mid \underbrace{h_1(x) = 0, \ldots, h_k(x) = 0}_{\text{ideal } I}, h_{k+1}(x) \geq 0, \ldots\}$$

Theorem: [La 2002] [Lasserre/La/Rostalski 2007]

(i) If $|V_{\mathbb{C}}(I)| < \infty$, $\boxed{p_t^{\text{sos}} = p_t^{\text{mom}} = p_{\min} \quad \text{for some } t}$

(ii) If $|V_{\mathbb{R}}(I)| < \infty$, $\boxed{p_t^{\text{mom}} = p_{\min} \quad \text{for some } t}$

The flat extension theorem

Theorem [Curto-Fialkow 1996] Let $L \in \mathbb{R}[x]^*_{2t}$.

> If $\text{rank } M_t(L) = \text{rank } M_{t-1}(L)$,
> then there is an extension $\tilde{L} \in \mathbb{R}[x]^*$ of L
> for which $\text{rank } M(\tilde{L}) = \text{rank } M_t(L)$.

Main tool: $\text{Ker} M(\tilde{L})$ is an ideal.

[La-Mourrain 2009] The flat extension theorem can be generalized to matrices indexed by a set $\mathcal{C}$ of monomials (*connected to 1*) and its *closure* $\mathcal{C}^+ = \mathcal{C} \cup x_1\mathcal{C} \cup \ldots \cup x_n\mathcal{C}$, satisfying $\text{rank } M_{\mathcal{C}} = \text{rank } M_{\mathcal{C}^+}$.

The finite rank moment matrix theorem

Theorem: [Curto-Fialkow 1996] Let $L \in \mathbb{R}[x]^*$.

$M(L) \succeq 0$ and $\text{rank}\, M(L) =: r < \infty$

$\Longleftrightarrow$ L has a (unique) r-atomic representation measure μ.

[La 2005] **Simple proof for** $\Longrightarrow$:

- $I := \text{Ker} M(L)$ is a real radical ideal
- I is 0-dimensional, as $\dim \mathbb{R}[x]/I = r$

Hence: $V(I) = \{x_1, \ldots, x_r\} \subseteq \mathbb{R}^n$

Verify: L is represented by $\mu = \sum_{i=1}^r L(p_i^2)\delta_{x_i}$, where the p_i's are interpolation polynomials at the x_i's

Implementations of the SOS/moment relaxation method

GloptiPoly by Henrion, Lasserre
(incorporates the optimality stopping criterion and the extraction procedure for global minimizers)

SOSTOOLS by Prajna, Papachristodoulou, Seiler, Parrilo

YALMIP by Löfberg

SparsePOP by Waki, Kim, Kojima, Muramatsu

Example 1

$$\min \quad p = -25(x_1-2)^2 - (x_2-2)^2 - (x_3-1)^2 - (x_4-4)^2 - (x_5-1)^2 - (x_6-4)^2$$

$$\text{s.t.} \quad (x_3-3)^2 + x_4 \geq 4, \ \ (x_5-3)^2 + x_6 \geq 4$$

$$x_1 - 3x_2 \leq 2, \ \ -x_1 + x_2 \leq 2, \ x_1 + x_2 \leq 6,$$

$$x_1 + x_2 \geq 2, \ 1 \leq x_3 \leq 5, \ 0 \leq x_4 \leq 6,$$

$$1 \leq x_5 \leq 5, \ 0 \leq x_6 \leq 10, \ x_1, x_2 \geq 0$$

order t	rank sequence	bound p_t^{mom}	solution extracted
1	1 7	unbounded	none
2	1 1 21	-310	$(5,1,5,0,5,10)$

$d = 1$

The global minimum is found at the relaxation of order $t = 2$

Example 2

$$\begin{array}{ll} \min & p = -x_1 - x_2 \\ \text{s.t.} & x_2 \le 2x_1^4 - 8x_1^3 + 8x_1^2 + 2 \\ & x_2 \le 4x_1^4 - 32x_1^3 + 88x_1^2 - 96x_1 + 36 \\ & 0 \le x_1 \le 3,\ 0 \le x_2 \le 4 \end{array}$$

order t	rank sequence	bound p_t^{mom}	solution extracted
2	1 1 4	-7	none
3	1 2 2 4	-6.6667	none
4	1 1 1 1 6	-5.5080	$(2.3295, 3.1785)$

$d = 2$

The global minimum is found at the relaxation of order $t = 4$

An example where (RC) cannot hold

Perturbed Motzkin form:

$$p = x_1^2 x_2^2 (x_1^2 + x_2^2 - 3x_3^2) + x_3^6 + \epsilon(x_1^6 + x_2^6 + x_3^6)$$

$K = \{x \mid \sum_{i=1}^3 x_i^2 \leq 1\} \rightsquigarrow (0,0)$ is the *unique minimizer*

But (RC) never holds

as $p \notin M(h)$ and $p_t^{\text{sos}} = p_t^{\text{mom}} < p_{\min} = 0$

order t	rank sequence	bound p_t^{mom}	val. moment vect.
3	1 4 9 13	$-2.11\ 10^{-5}$	$1.67\ 10^{-44}$
4	1 4 10 20 35	$-1.92\ 10^{-9}$	$4.47\ 10^{-60}$
5	1 4 10 20 35 56	$2.94\ 10^{-12}$	$1.26\ 10^{-44}$
6	1 4 10 20 35 56 84	$3.54\ 10^{-12}$	$1.5\ 10^{-44}$
7	1 4 10 20 35 56 84 120	$4.09\ 10^{-12}$	$2.83\ 10^{-43}$

$d = 3$, $\epsilon = 0.01$

Application to Unconstrained Polynomial Minimization

$$p_{\min} = \inf_{x\in\mathbb{R}^n} p(x)$$

where $\deg(p) = 2d$

As there is **no constraint**, the relaxation scheme just gives **one bound**:

$$p_t^{\text{sos}} = p_t^{\text{mom}} = p_d^{\text{sos}} = p_d^{\text{mom}} \leq p_{\min} \quad \text{for all } t \geq d$$

with equality iff $p(x) - p_{\min}$ is SOS

How to get better bounds ?

Idea: Transform the *Unconstrained* Problem into a *Constrained* Problem

If p has a minimum:

$$p_{\min} = p^*_{\text{grad}} := \inf_{x \in V^{\mathbb{R}}_{grad}} p(x)$$

where $$V^{\mathbb{R}}_{\text{grad}} := \{x \in \mathbb{R}^n \mid \nabla p(x) = 0 \ (i = 1, \ldots, n)\}$$

If, moreover, a bound R is known on the norm of a global minimizer:

$$p_{\min} = p^*_{\text{ball}} := \inf_{R^2 - \sum_i x_i^2 \geq 0} p(x)$$

When p attains its minimum

The 'ball approach':

- Convergence of the SOS/MOM bounds to $p_{\min} = p^*_{\text{ball}}$

The 'gradient variety' approach:

[Demmel, Nie, Sturmfels 2004]:

$$\boxed{p > 0 \text{ on } V^{\mathbb{R}}_{\text{grad}} \Longrightarrow p \in M_{\text{grad}}}$$

$$\boxed{p \geq 0 \text{ on } V^{\mathbb{R}}_{\text{grad}} \Longrightarrow p \in M_{\text{grad}}} \text{ if } I_{\text{grad}} \text{ radical}$$

$$M_{\text{grad}} := M(\pm \partial p/\partial x_i) = \Sigma_n + \underbrace{\sum_{i=1}^n \mathbb{R}[x]\partial p/\partial x_i}_{I_{\text{grad}}}$$

Convergence Result [Demmel, Nie, Sturmfels 2004]

Asymptotic convergence of the SOS/MOM bounds to p^*_{grad}

Finite Convergence to p^*_{grad} when the gradient ideal I_{grad} is **radical**

Hence: When p attains its minimum, we have a converging hierarchy of SDP bounds to $p_{\min}$

Example: $p = x^2 + (xy - 1)^2$ does **not** attain its minimum

$p_{\min} = 0 < p^*_{grad} = 1$

What if p is not known to have a minimum?

Strategy 1: Perturb the polynomial p

[Hanzon-Jibetean 2003] [Jibetean-Laurent 2004]

$$p_\epsilon(x) := p(x) + \epsilon \left(\sum_{i=1}^{n} x_i^{2d+2} \right)$$ for small $\epsilon > 0$

- p_ϵ has a minimum and $$\lim_{\epsilon \to 0} (p_\epsilon)_{\min} = p_{\min}$$

- The global minimizers of p_ϵ converge to global minimizers of p as $\epsilon \longrightarrow 0$

- The gradient variety of p_ϵ is **finite**

$\rightsquigarrow$ finite convergence of $(p_\epsilon)_t^{\mathrm{sos}}$, $(p_\epsilon)_t^{\mathrm{mom}}$ to $(p_\epsilon)_{\min}$

Example: Perturb the polynomial $p = (xy - 1)^2 + x^2$

$$\inf p(x) \text{ s.t. } \partial p_\epsilon / \partial x_1 = 0, \; \partial p_\epsilon / \partial x_2 = 0$$

ϵ	order t	rank sequence	p_t^{mom}	extracted solutions
10^{-2}	3	2 6 8	0.00062169	
10^{-2}	4	2 2 2 7	0.33846	
10^{-2}	5	2 2 2 2 -	0.33846	$\pm(0.4729, 1.3981)$
10^{-3}	5	2 2 2 2 -	0.20824	$\pm(0.4060, 1.9499)$
10^{-4}	5	2 2 2 2 -	0.12323	$\pm(0.3287, 2.6674)$
10^{-5}	5	2 2 2 2 -	0.07132	$\pm(0.2574, 3.6085)$
10^{-6}	5	2 2 2 2 -	0.040761	$\pm(0.1977, 4.8511)$
10^{-7}	5	2 2 2 2 -	0.023131	$\pm(0.1503, 6.4986)$
10^{-8}	5	2 2 2 2 -	0.013074	$\pm(0.1136, 8.6882)$
10^{-9}	5	2 2 2 2 -	0.0073735	$\pm(0.0856, 11.6026)$
10^{-10}	5	2 2 2 2 -	0.0041551	$\pm(0.0643, 15.4849)$

When p does not have a minimum:

Algebraic/analytical approach of Schweighofer [2005]

Strategy 2: Minimize p over its 'gradient tentacle'

If $p_{\min} > -\infty$, then

$$p_{\min} = \inf_{x \in K_{\text{grad}}} p(x)$$

where

$$K_{\text{grad}} := \{x \in \mathbb{R}^n \mid \|\nabla p(x)\|^2 \|x\|^2 \leq 1\} \supseteq V_{\text{grad}}^{\mathbb{R}}$$

$$\nabla p = (\partial p / \partial x_i)_{i=1}^n$$

Representation result on the gradient tentacle K_{grad}

[Schweighofer 2005]: Assume $p_{\min} > -\infty$ and p has only isolated singularities at infinity (*) (e.g. $n = 2$). Then,

$$p \geq 0 \text{ on } \mathbb{R}^n \iff p \geq 0 \text{ on } K_{\text{grad}}$$
$$\iff \forall \epsilon > 0 \quad p + \epsilon \in M(1 - \|\nabla p(x)\|^2 \|x\|^2)$$

$\rightsquigarrow$ Convergent SOS/moment bounds to $p_{\min} = \inf_{K_{\text{grad}}} p(x)$

(*): The system $\nabla p_d(x) = 0$, $p_{d-1}(x) = 0$ has finitely many projective zeros, where $p = p_d + p_{d-1} + \ldots + p_0$ and p_i is the homogeneous component of degree i

Tools: Algebra (extension of Schmüdgen's theorem) + analysis (Parusinski's results on behaviour of polynomials at infinity)

When p does not have a minimum:

The 'tangency variety' approach of Vui-Son [2008]

Strategy 3: Minimize p over its 'truncated tangency variety':

$$\Gamma_p := \left\{x \in \mathbb{R}^n \mid \text{rank} \begin{pmatrix} \nabla p(x) \\ x \end{pmatrix} \leq 1\right\}$$

$$= \{x \mid g_{ij} := x_i \, \partial p/\partial x_j - x_j \, \partial p/\partial x_i = 0 \; \forall i, j \leq n\}$$

$$\Gamma_p^0 := \Gamma_p \cap \{x \mid p(x) \leq p(0)\}$$

[Vui-Son 2008]: For $p \in \mathbb{R}[x]$ such that $p_{\min} > -\infty$

$$p \geq 0 \text{ on } \mathbb{R}^n \Longleftrightarrow p \geq 0 \text{ on } \Gamma_p^0$$
$$\Longleftrightarrow \forall \epsilon > 0 \quad p + \epsilon \in M(p(0) - p, \pm g_{ij})$$

$\rightsquigarrow$ Convergent SOS/moment bounds to $p_{\min} = \inf_{\Gamma_p^0} p(x)$

Further directions

• Exploit structure (equations, sparsity, symmetry, convexity, ...) to get smaller SDP programs
[Kojima, Grimm, Helton, Lasserre, Netzer, Nie, Riener, Schweighofer, Theobald, Parrilo, ...]

• Application to the generalized problem of moments, to approximating integrals over semi-algebraic sets, ...
[Henrion, Lasserre, Savorgnan, ...]

• Extension to NC variables
[Helton, Klep, McCullough, Schmüdgen, Schweighofer,...]