Optimization over Polynomials with Moment Matrices and Sums of Squares: Introduction

Monique Laurent

CWI, Amsterdam

IMA Tutorial
Algebraic Algorithms in Optimization

Polynomial Optimization Problem (POP)

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

where

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

and $p, h_1, \ldots, h_m$ are real polynomials in n variables

Unconstrained Polynomial Minimization

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

Example 1: An $n \times n$ matrix M is copositive if $x^T M x \geq 0$ for all $x \in \mathbb{R}^n_+$. That is, $p_{\min} \geq 0$ for $p(x) = \sum_{i,j=1}^n M_{ij} x_i^2 x_j^2$ [i.e., $p(x)$ is nonnegative on $\mathbb{R}^n$]

$\rightsquigarrow$ co-NP-complete problem

Example 2: A sequence $a_1, \ldots, a_n \in \mathbb{N}$ can be partitioned if $p_{\min} = 0$ for $p(x) = (\sum_{i=1}^n a_i x_i)^2 + \sum_{i=1}^n (x_i^2 - 1)^2$

$\rightsquigarrow$ NP-complete problem

Example 3: A set of distances d_{ij} $(ij \in E, i, j = 1, \ldots, n)$ is realizable in $\mathbb{R}^k$ if $d_{ij} = \|x_i - x_j\|$ $(ij \in E)$ for some $x_1, \ldots, x_n \in \mathbb{R}^k$. That is,

$p_{\min} = 0$ for $p(x) = \sum_{ij \in E} (d_{ij}^2 - \sum_{h=1}^k (x_{ih} - x_{jh})^2)^2$

$\rightsquigarrow$ NP-complete problem, already for dimension $k = 1$

$\rightsquigarrow$ Unconstrained polynomial minimization is **NP-hard**, already for degree 4 polynomials

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:

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

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

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

Strategy

Approximate (POP) 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

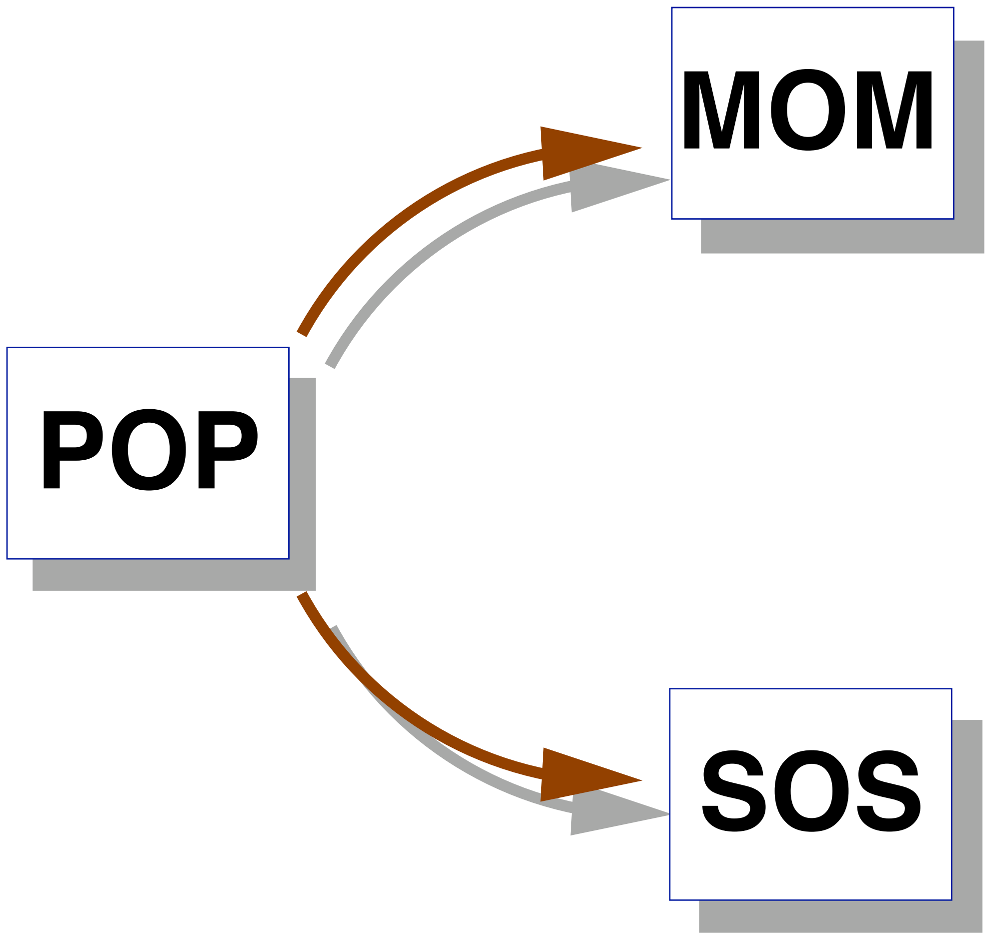
Underlying paradigm

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

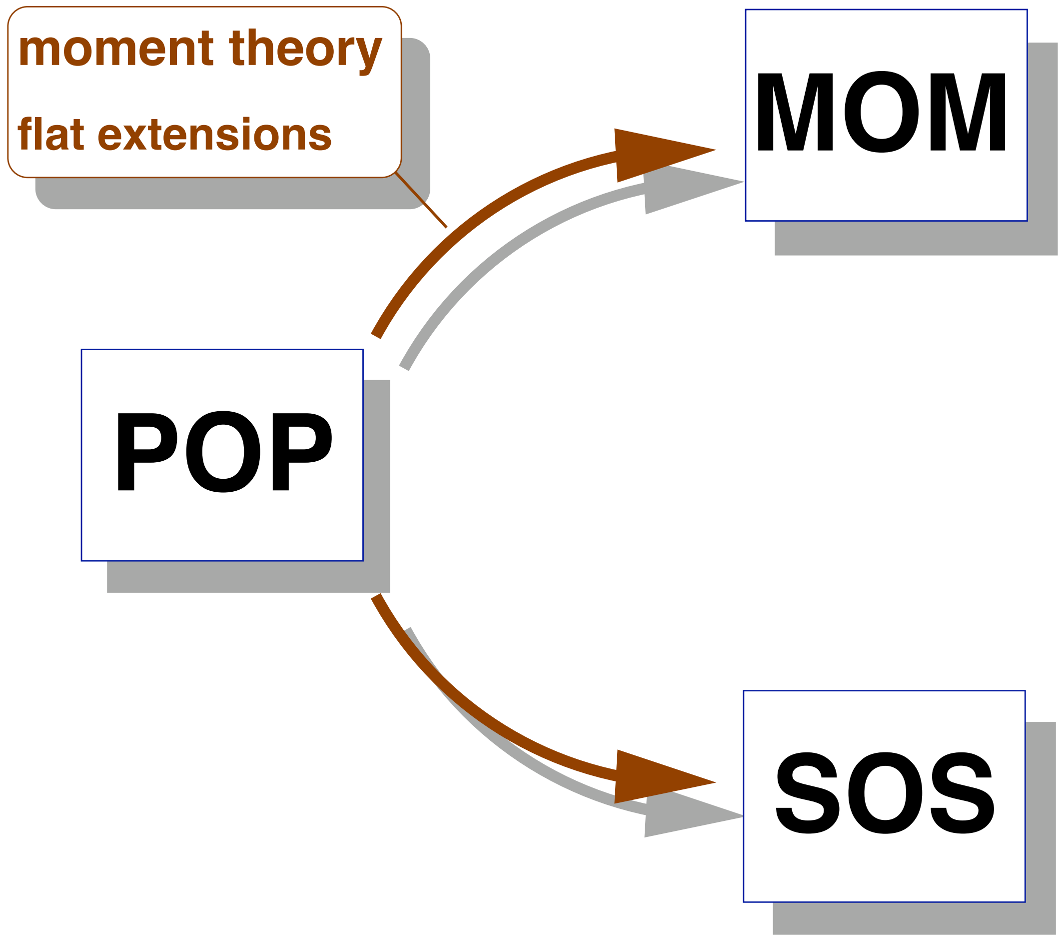
but

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

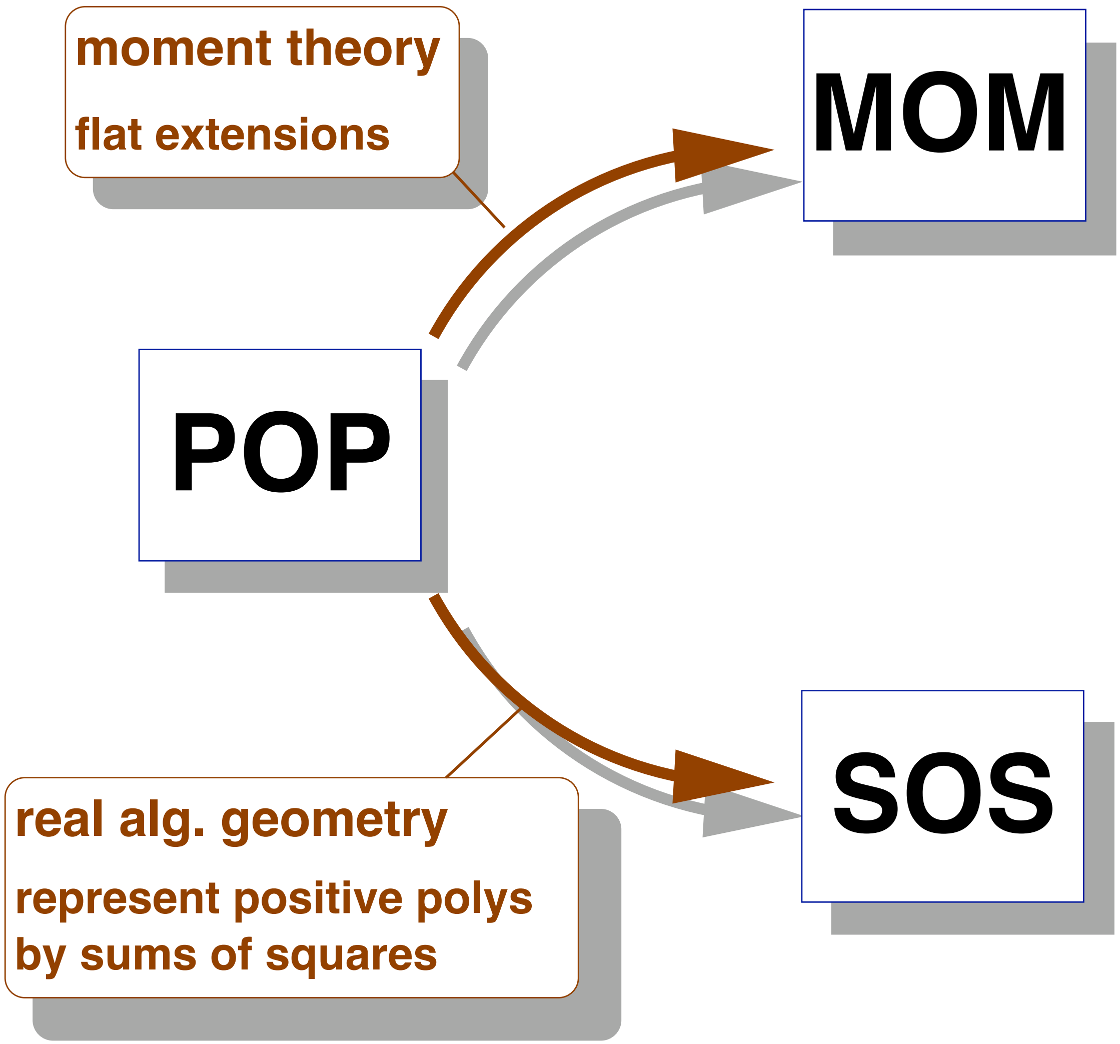
Overview



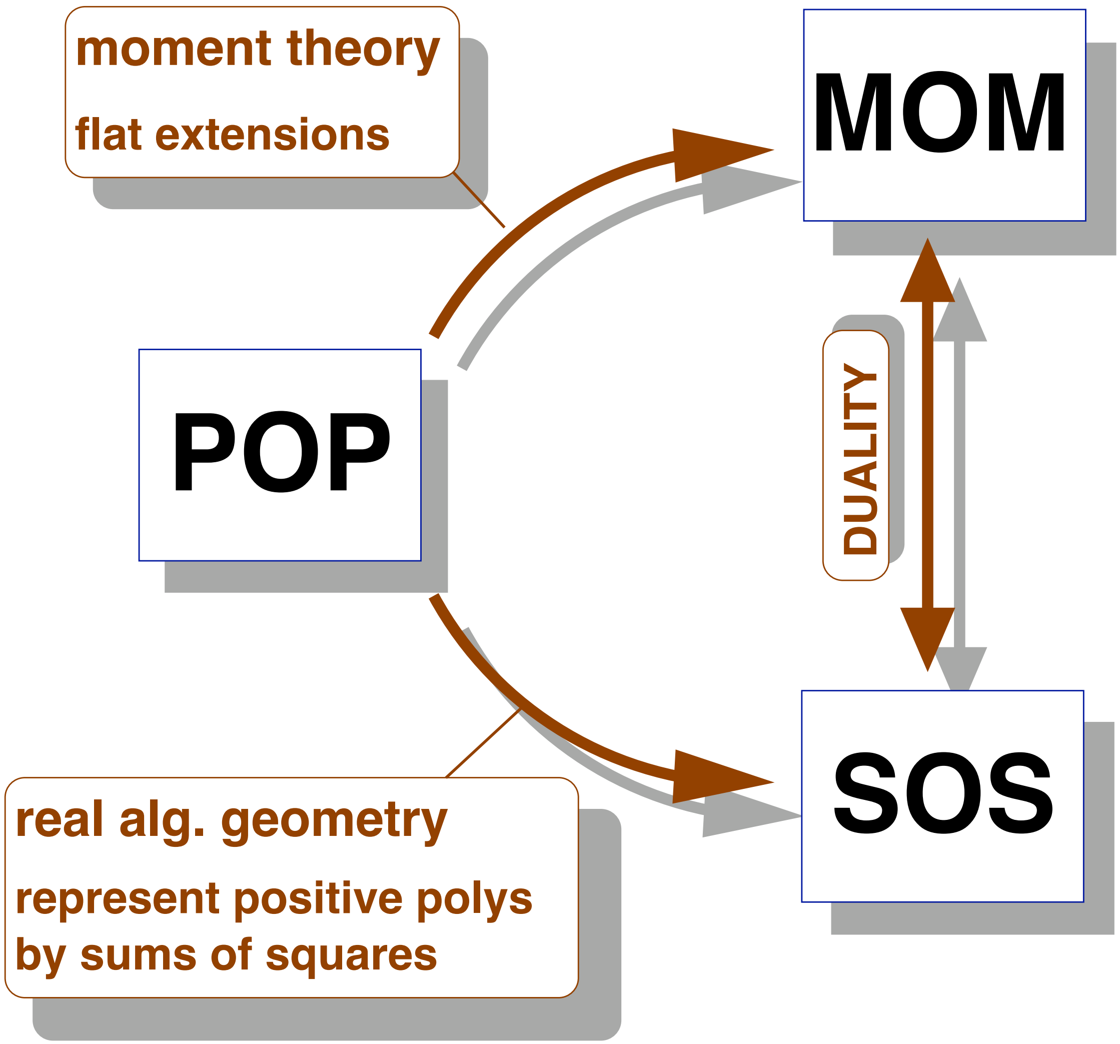
Overview



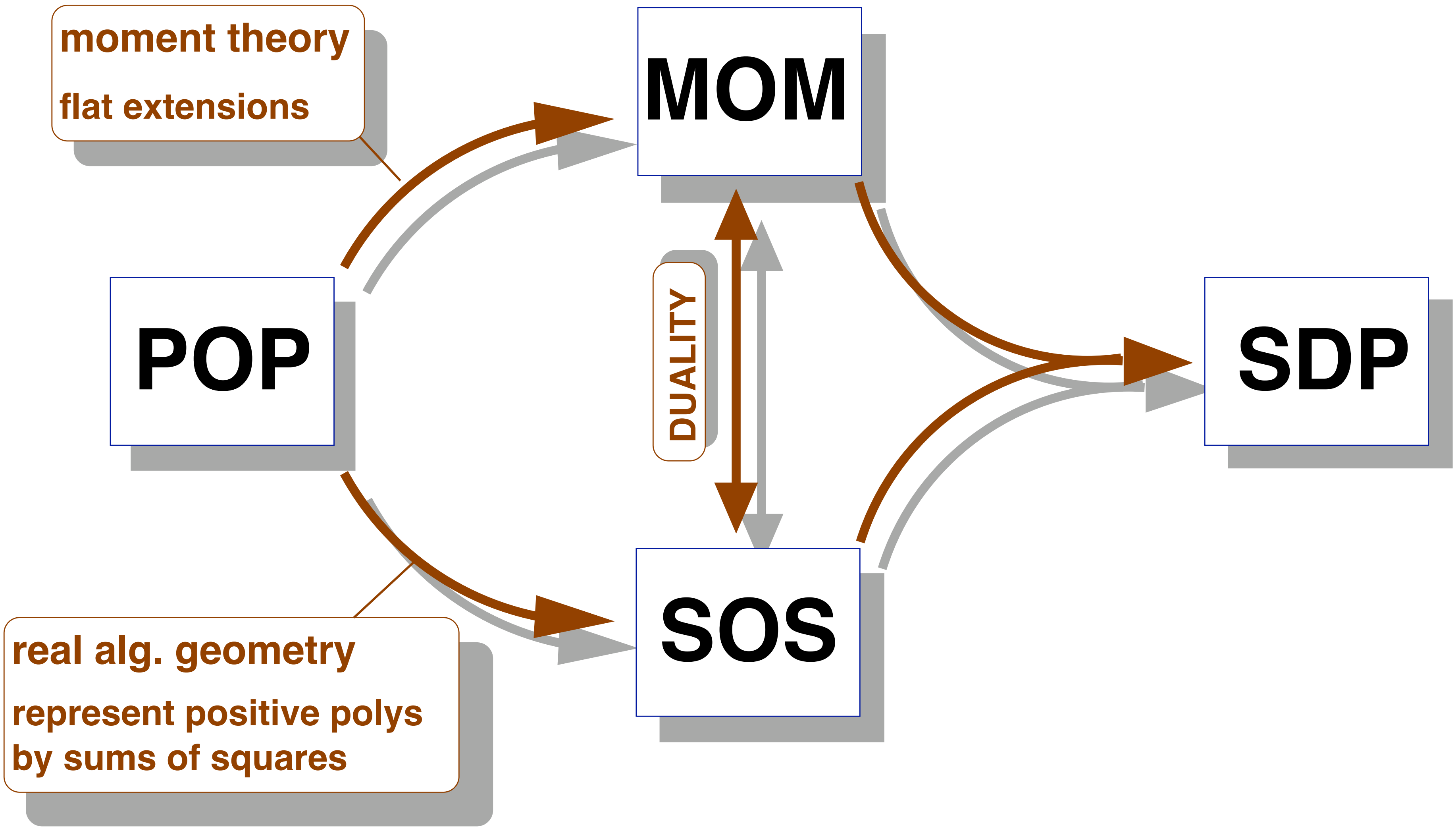
Overview


moment theory
flat extensions
MOM
POP
SOS
real alg. geometry
represent positive polys
by sums of squares

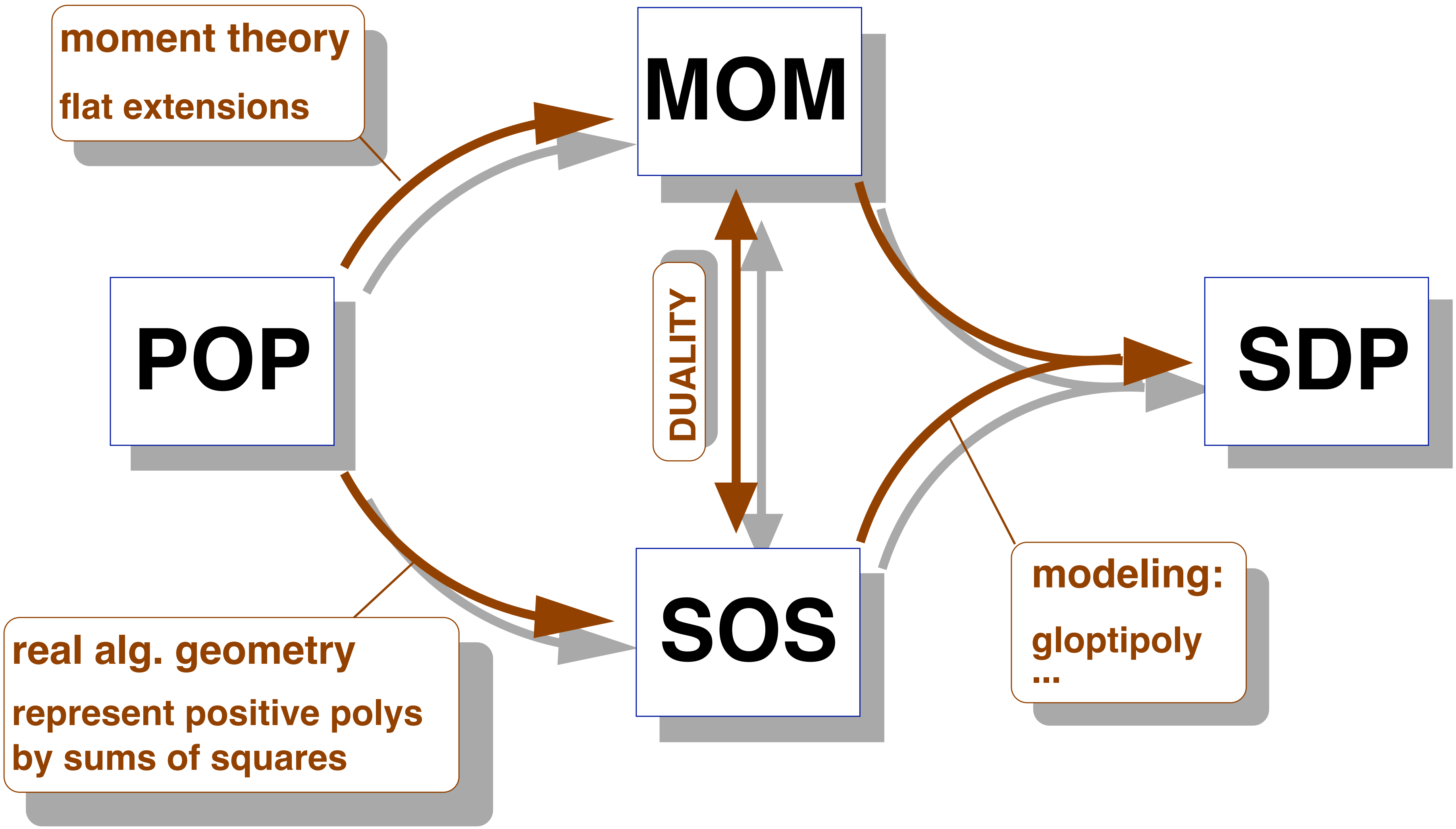
Overview


moment theory
flat extensions
MOM
POP
DUALITY
SOS
real alg. geometry
represent positive polys
by sums of squares

Overview

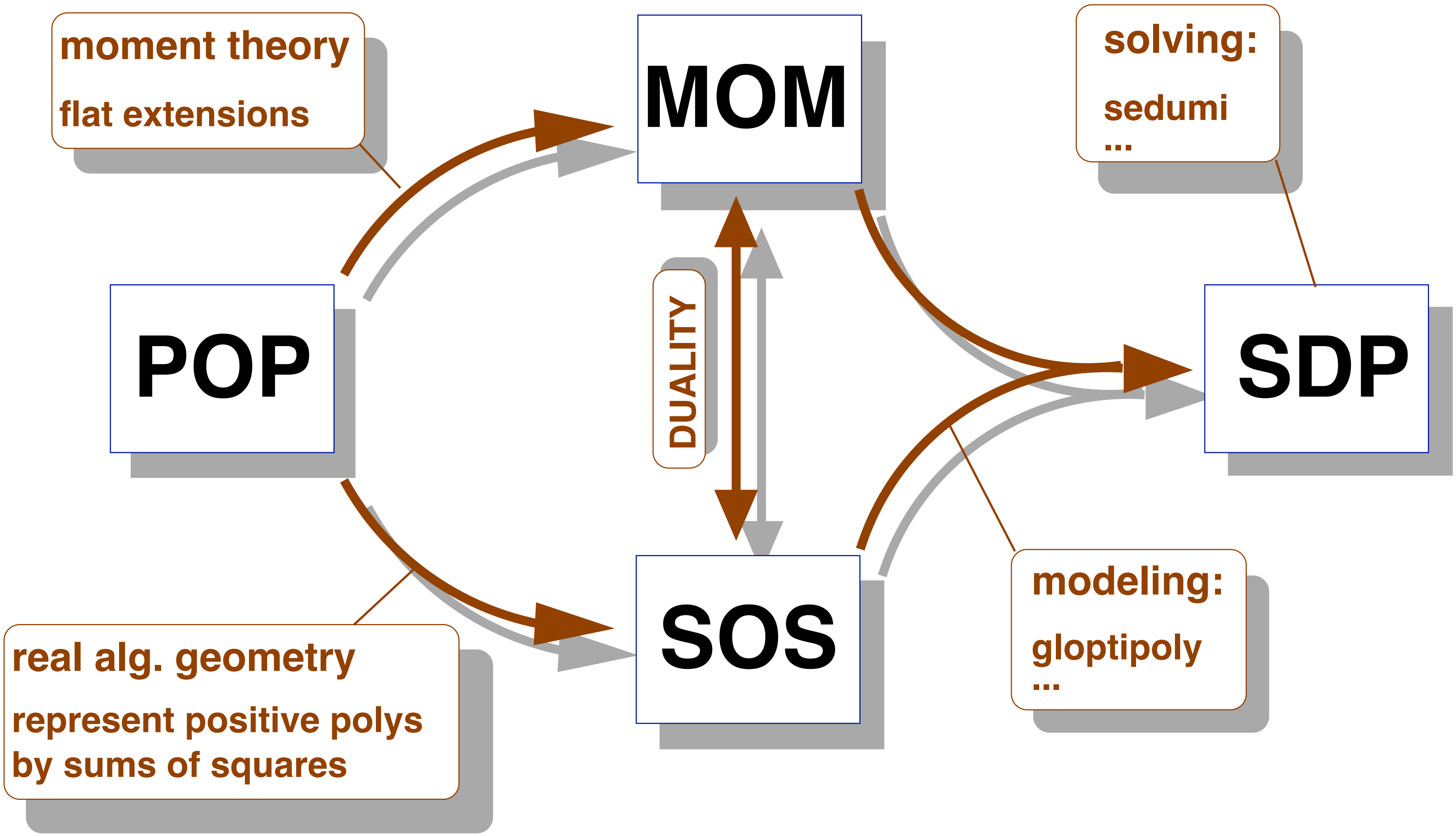

moment theory
flat extensions
MOM
POP
DUALITY
SDP
SOS
real alg. geometry
represent positive polys
by sums of squares

Overview



moment theory
flat extensions
MOM
POP
DUALITY
SDP
SOS
real alg. geometry
represent positive polys
by sums of squares
modeling:
gloptipoly
...

Overview


moment theory
flat extensions
MOM
solving:
sedumi
...
POP
DUALITY
SDP
SOS
modeling:
gloptipoly
...
real alg. geometry
represent positive polys
by sums of squares

Plan of lectures

- SOS/Moment relaxations for (POP): Introduction and main properties: **convergence, optimality criterion, extraction of global minimizers**

- Algebraic background about SOS and moment matrices

- Application 1: Unconstrained polynomial minimization

- Application 2: Find the **real** roots to polynomial equations

- Demo with GloptiPoly

- Exploiting sparsity, equations, symmetry, to reduce the problem size

Some notation about polynomials

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

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

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

- $\deg(p) = \max_{p_\alpha \neq 0} \deg(x^\alpha)$, $\deg(x^\alpha) = |\alpha| := \sum_{i=1}^n \alpha_i$

- p **homeogeneous** (**form**) if $p_\alpha = 0 \ \ \forall |\alpha| \neq d := \deg(p)$

$p \rightsquigarrow$ **form:** $\tilde{p}(x,t) := t^d p(x/t) = \sum_\alpha p_\alpha x^\alpha t^{d-|\alpha|} \in \mathbb{R}[x,t]$

$$p \geq 0 \text{ on } \mathbb{R}^n \iff \tilde{p} \geq 0 \text{ on } \mathbb{R}^{n+1}$$

Notation (continued)

- We **identify** a polynomial p with its sequence of coefficients $p = (p_\alpha)_\alpha$

Example: $p = 1 - x_1 + 3x_2 + 2x_1^2 + 7x_1x_2 - 2x_2^2 + x_1^3 - x_2^3$

$n = 2$ variables, degree $d = 3$

$\rightsquigarrow p = (1, -1, 3, 2, 7, -2, 1, 0, 0, -1, 0, \dots, 0)$

- M: matrix indexed by monomials of degree $\leq d$

Say that **the polynomial** $p \in \mathbb{R}[x]_d$ **lies in the kernel of** M if $Mp = 0$.

Notation (continued)

- p is a sum of squares of polynomials (SOS) if

$$p(x) = \sum_{j=1}^{m} [u_j(x)]^2 \quad \text{for some } u_1, \ldots, u_m \in \mathbb{R}[x]$$

Then, $\deg(p)$ is even and $\deg(u_i) \leq \deg(p)/2$

Example: $x^2 + y^2 + 2xy + z^6 = (x+y)^2 + (z^3)^2 \in \mathrm{SOS}_3$

- $p \in \mathrm{SOS}_n \iff$ its homogenization $\tilde{p} \in \mathrm{SOS}_{n+1}$

Linear/Semidefinite Programming: Short Recap

Given: $c, a_1, \ldots, a_m \in \mathbb{R}^n$, $b \in \mathbb{R}^m$

Variables: $x \in \mathbb{R}^n$ (primal variable), $y \in \mathbb{R}^m$ (dual variable)

Primal LP:

$$\boxed{p^* := \max\ c^T x \text{ s.t. } a_j^T x = b_j\ (j = 1, \ldots, m),\ x \geq 0}$$

Dual LP:

$$\boxed{d^* := \min\ y^T b \text{ s.t. } \sum_{j=1}^{m} y_j a_j - c \geq 0}$$

Strong duality: $\boxed{p^* = d^*}$

Semidefinite Programming

Semidefinite Programming (SDP) is the analogue of LP where:

vector variable		matrix variable
$x \in \mathbb{R}^n$	$\rightsquigarrow$	$X \in \mathrm{Sym}_n$
$x \geq 0$	$\rightsquigarrow$	$X \succeq 0$

$X \in \mathrm{Sym}_n$: X is a symmetric $n \times n$ matrix

$X \succeq 0$: X is a positive semidefinite matrix

$$
\begin{aligned}
X \succeq 0 \quad &\Longleftrightarrow u^T X u \geq 0 \ \text{ for all } u \in \mathbb{R}^n \\
&\Longleftrightarrow X = UU^T \ \text{ for some } U \in \mathbb{R}^{n \times n} \\
&\Longleftrightarrow \ \text{all eigenvalues are nonnegative}
\end{aligned}
$$

Semidefinite Programming (continued)

Given: $C, A_1, \ldots, A_m \in \mathrm{Sym}_n$, $b \in \mathbb{R}^m$
Variables: $X \in \mathrm{Sym}_n$ (primal), $y \in \mathbb{R}^m$ (dual)

Primal SDP:

$$\boxed{p^* := \sup \langle C, X \rangle \text{ s.t. } \langle A_j, X \rangle = b_j \; (j = 1, \ldots, m), \; X \succeq 0}$$

Dual SDP:

$$\boxed{d^* := \inf \sum_{j=1}^m b_j y_j \text{ s.t. } \sum_{j=1}^m y_j A_j - C \succeq 0}$$

- Weak duality: $\boxed{p^* \leq d^*}$
- Primal SDP strictly feasible ($\exists X \succ 0$ feasible) $\Longrightarrow p^* = d^*$

(no duality gap) and Dual SDP attains its infimum

Complexity

- LP can be solved in polynomial time
- SDP can be solved in polynomial time (to *arbitrary precision*)

$\rightsquigarrow$ (theoretically) with the ellipsoid method [Nemirovskii, Shor, ..] [since one can test in polynomial time whether $X \succeq 0$]

$\rightsquigarrow$ (practically) with interior-point methods [cf. Vandenberghe & Boyd, SIAM Review 1996; Handbook by Wolkowicz, Saigal & Vandenberghe 2000]

A geometric property of SDP

SDP: $\sup\langle C, X\rangle$ s.t. $\langle A_j, X\rangle = b_j \ (j = 1, \ldots, m), \ X \succeq 0$

Lemma: If X is an optimum solution with **maximum rank** then

$\mathrm{Ker} X \subseteq \mathrm{Ker} X'$ for any other optimum solution X'.

Proof: $X'' := \frac{1}{2}(X + X')$ is also an optimum solution with

$$\mathrm{Ker} X'' = \mathrm{Ker} X \cap \mathrm{Ker} X' \subseteq \mathrm{Ker} X$$
$$\Longrightarrow \mathrm{Ker} X = \mathrm{Ker} X'' \subseteq \mathrm{Ker} X'$$

Note: Interior-point algorithms return (*often*) maximum rank optimum solutions

Recognizing SOS of polynomials via SDP

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

i.e., $p(x) = \sum_j [u_j(x)]^2$, where $\deg(u_j) \le d$

$$\Updownarrow$$

$$p(x) = \zeta_{x,d}^T \underbrace{\Big(\sum_j u_j u_j^T\Big)}_{X \succeq 0} \zeta_{x,d},$$ setting $\zeta_{x,d} := (x^\beta)_{|\beta|\le d}$

$$\Updownarrow$$

The following semidefinite program is feasible:

$$\begin{cases} X \succeq 0 \\ \displaystyle\sum_{\substack{|\beta|,|\gamma|\le d \\ \beta+\gamma=\alpha}} X_{\beta,\gamma} = p_\alpha \quad (|\alpha| \le 2d) \end{cases}$$

$\rightsquigarrow$ system in matrix variable $\boldsymbol{X}$ of order $\binom{\boldsymbol{n+d}}{\boldsymbol{d}} \times \binom{\boldsymbol{n+d}}{\boldsymbol{d}}$ with $\binom{\boldsymbol{n+2d}}{\boldsymbol{2d}}$ equations

$\rightsquigarrow$ polynomial size **fixing either $\boldsymbol{n}$ or $\boldsymbol{d}$**

Note: $\mathbf{dim}\, \mathbb{R}[\boldsymbol{x}]_{\boldsymbol{d}} = \binom{\boldsymbol{n+d}}{\boldsymbol{d}}$

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

Solution: Try to write

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

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

- For $c = -1$, $\boldsymbol{X} = \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \\ 0 & 2 & 1 \end{pmatrix}$, giving the SOS

decomposition: $p = (x^2 + xy - y^2)^2 + (y^2 + 2xy)^2$

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

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

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:

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

$\mathrm{SOS}_{n,d} \subset \mathrm{POS}_{n,d}$ for all other (n, d)

Motzkin polynomial: $x^4y^2 + x^2y^4 - 3x^2y^2 + 1$ lies in
$\mathrm{POS}_{2,6} \setminus \mathrm{SOS}_{2,6}$

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:

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

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

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

How many nonnegative polynomials are sums of squares ?

- [Lasserre 2004]: **Many !**

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

If $p \in \mathbb{R}[x]$ is nonnegative on $\mathbb{R}^n$, then

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

Thus $\|p - p_{\epsilon,k}\|_1 \longrightarrow 0$ as $\epsilon \longrightarrow 0$

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

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

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

Sometimes, the common denominator is known:

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

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

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

Putinar's certificate for positivity on a semialgebraic set K

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

$$p \in M = \{s_0 + \sum_{j=1}^{m} s_j h_j \mid s_0, s_j \in \mathrm{SOS}_n\}$$

$$\Longrightarrow p \in T = \{ \sum_{e \in \{0,1\}^m} s_e h_1^{e_1} \cdots h_m^{e_m} \mid s_e \in \mathrm{SOS}_n\}$$

$$\Longrightarrow p \geq 0 \text{ on } K$$

Theorem: Assume K is compact.

Schmüdgen [1991]: $p > 0$ on $K \Longrightarrow p \in T$

Putinar [1993]: $p > 0$ on $K \Longrightarrow p \in M$ assuming

(A) $\exists f \in M$ s.t. $\{x \in \mathbb{R}^n \mid f(x) \geq 0\}$ is compact.

The assumption $p > 0$ cannot be relaxed to $p \geq 0$

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

$p = 1 - x^2 \geq 0$ on K, but $p \notin M$

Indeed, assume $1 - x^2 = s_0 + s_1(1 - x^2)^3$

where s_0, s_1 are SOS, $s_0 = \sum_l f_l^2$

At $x = \pm 1, \quad s_0(x) = 0 \Longrightarrow f_l(x) = 0 \ \ \forall l$

Hence, $1 - x^2$ divides each $f_l \Longrightarrow (1 - x^2)^2$ divides s_0

The Positivstellensatz [Stengle 1974] [Dubois 1969] [Risler 1970]

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

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

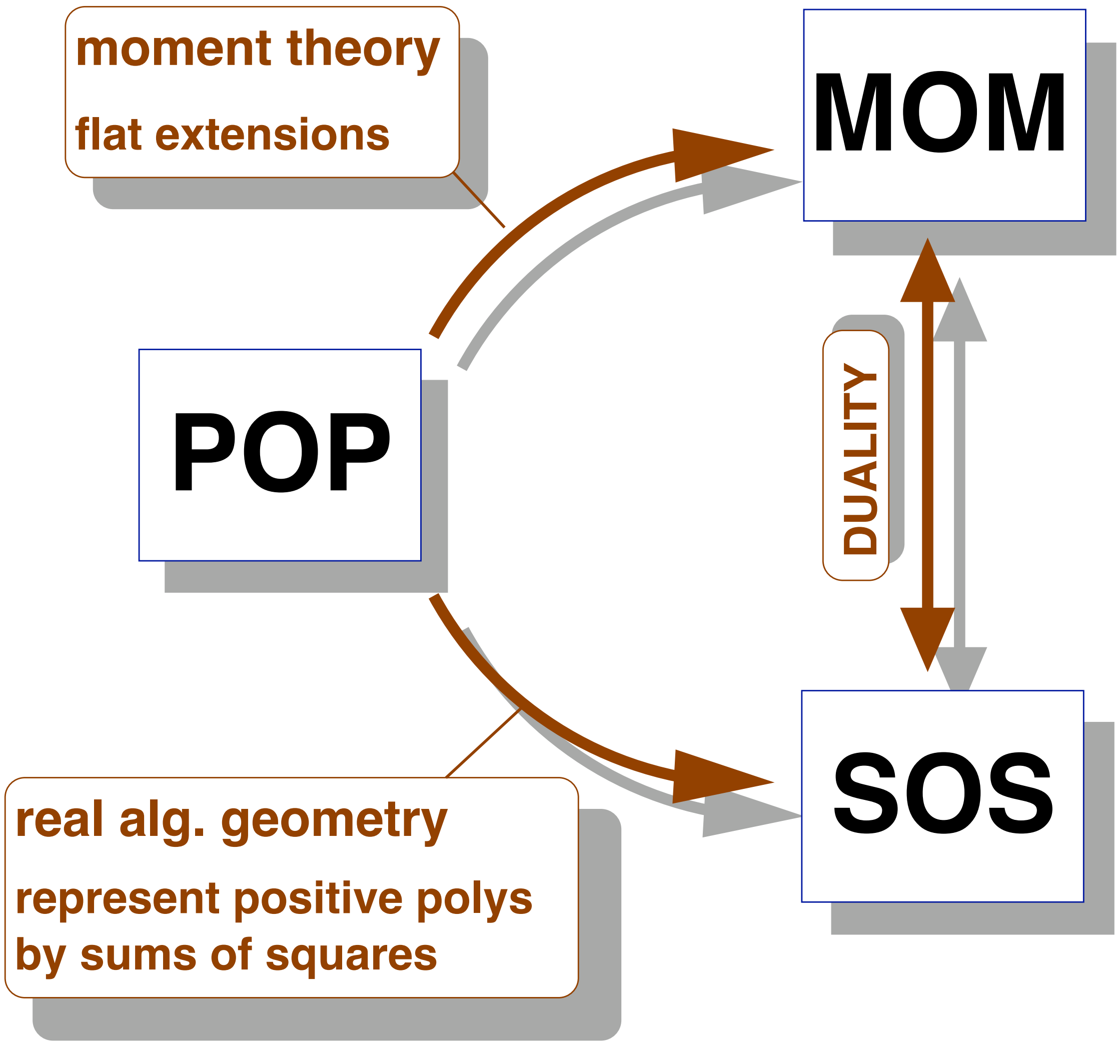
(2) $p \geq 0$ on $K \iff pf = f^{2m} + g$ for some $f, g \in T,\ m \in \mathbb{N}$

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

(2) $\Longrightarrow$ solution to **Hilbert's 17th problem**
(for $K = \mathbb{R}^n,\ T = \mathrm{SOS}_n$)

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

Overview


moment theory
flat extensions
MOM
POP
DUALITY
SOS
real alg. geometry
represent positive polys
by sums of squares

SOS relaxations for (POP) via Putinar's representation theorem

[following Lasserre (2001)]

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

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

Relax $p - \lambda \geq 0$ to $p - \lambda \in M$ $\rightsquigarrow$ relaxation (SOSt):

$$\boxed{p_t^{\text{sos}} := \sup \lambda \ \text{ s.t. } \quad \begin{array}{l} p - \lambda = s_0 + \sum_{j=1}^m s_j h_j \\ s_0, \ s_j \text{ are SOS} \\ \deg(s_0), \ \deg(s_j h_j) \leq 2t \end{array}}$$

$\boxed{p_t^{\text{sos}} \leq p_{t+1}^{\text{sos}} \leq p_{\min}}$ for $t \geq \lceil \deg(p)/2 \rceil, \ \lceil \deg(h_j)/2 \rceil$

Moment relaxations for (POP): Work in the (primal) space of variable x

Write $p(x) = p^T \zeta_x$, setting $\boxed{\zeta_x := (x^\alpha)_\alpha}$

$$
\begin{aligned}
p_{\min} &= \inf_{x \in K} p(x) \\
&= \inf_{x \in K} p^T \zeta_x = \inf_y p^T y \text{ s.t. } y \in \{\zeta_x \mid x \in K\} \\
&= \inf_y p^T y \text{ s.t. } y \in \operatorname{conv}\{\zeta_x \mid x \in K\} \\
&= \inf_y p^T y \text{ s.t. } y \text{ is the sequence of moments} \\
&\qquad\qquad\qquad \text{of a measure on } K
\end{aligned}
$$

Moment sequences and representing measures

Definition: Let μ be a probability measure on K. Then,

$\boxed{y_\alpha := \int_K x^\alpha \mu(dx)}$ is the **moment of order** α of μ

- $y = (y_\alpha)_\alpha$ is called the **sequence of moments** of μ
- μ is called a **representing measure** for y

Hence: $\boxed{p^T y = \int_K p(x)\mu(dx) \geq p_{\min}}$

Ex 1: $\mu = \delta_x$ (**Dirac measure** at x) $\rightsquigarrow y = \zeta_x$

Ex 2: $\mu = \sum_{i=1}^r \lambda_i \delta_{x_i}$ $(\lambda_i > 0)$: **finite atomic measure** with $\text{supp}(\mu) = \{x_1, \ldots, x_r\} \rightsquigarrow y = \sum_{i=1}^r \lambda_i \zeta_{x_i}$

Lemma: Assume y is the sequence of moments of a probability measure μ on K. Then,

$$\boxed{y_0 = 1}$$

$$\boxed{M_t(y) := (y_{\alpha+\beta})_{|\alpha|,|\beta|\leq t} \succeq 0} \quad \text{for } t \geq 0$$

moment matrix of order t

$$\boxed{M_t(h_j y) := (\textstyle\sum_\gamma (h_j)_\gamma y_{\alpha+\beta+\gamma})_{|\alpha|,|\beta|\leq t} \succeq 0} \quad \text{for } t \geq 0$$

localizing matrix

$$\boxed{\text{supp}(\mu) \subseteq \bigcap_{p \in \text{Ker} M_t(y)} V_{\mathbb{R}}(p) = V_{\mathbb{R}}(\text{Ker} M_t(y))}$$

$\text{supp}(\mu)$:= smallest closed set S with $\mu(\mathbb{R}^n \setminus S) = 0$

Proof: Let p be a polynomial with $\deg(p) \leq t$.

$$p^T M_t(y) p = \sum_{\alpha,\beta} p_\alpha p_\beta y_{\alpha+\beta} = \sum_{\alpha,\beta} p_\alpha p_\beta \int x^{\alpha+\beta} \mu(dx)$$
$$= \int p(x)^2 \mu(dx) \geq 0$$

If $p \in \text{Ker } M_t(y), \quad 0 = \int p(x)^2 \mu(dx) \implies \text{supp}(\mu) \subseteq V_{\mathbb{R}}(p$

$$p^T M_t(h_j y) p = \sum_{\alpha,\beta} p_\alpha p_\beta (h_j)_\gamma y_{\alpha+\beta+\gamma}$$
$$= \sum_{\alpha,\beta} p_\alpha p_\beta (h_j)_\gamma \int x^{\alpha+\beta+\gamma} \mu(dx) = \int h_j(x) p(x)^2 \mu(dx) \geq 0$$

Moment relaxation of order t

Relax the condition "y is the sequence of moments of a measure supported by K"
by "positive semidefiniteness of its (localizing) moment matrices":

(MOMt)

$$p_t^{\text{mom}} := \inf\ p^T y \ \text{ s.t. } \quad \begin{array}{l} y_0 = 1,\ M_t(y) \succeq 0 \\ M_{t-d_j}(h_j y) \succeq 0\ (j = 1, \ldots, m) \end{array}$$

for $t \geq \max(\lceil \deg(p)/2 \rceil,\ d_j = \lceil \deg(h_j)/2 \rceil)$

$$\boxed{p_t^* \leq p_{t+1}^* \leq p_{\min}}$$

Duality between the (SOSt) and (MOMt) relaxations

(SOSt) and **(MOMt)** are dual semidefinite programs.

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

No duality gap, e.g. if K has a nonempty interior [as (MOMt) is strictly feasible]

Example with ∞ duality gap: [Schweighofer 2005]

$$\min\ x_1x_2\ \text{ s.t. } -x_2^2 \geq 0,\ 1+x_1 \geq 0,\ 1-x_1 \geq 0$$

$$K = [-1,1] \times \{0\}$$

$$0 = p_{\min} = p_1^{\text{mom}} > p_1^{\text{sos}} = -\infty$$

Direct proof for weak duality using 'moment matrix calculus'

$$\boxed{y^T(f^2h) = f^T M_{t-d_h}(hy) f \quad \text{if } \deg(f^2h) \leq 2t}$$

Let y, (λ, s_j) be feasible for (MOMt), (SOSt), resp.

- $y_0 = 1$, $M_t(y) \succeq 0$, $M_{t-d_j}(h_j y) \succeq 0$
- $p - \lambda = s_0 + \sum_j s_j h_j$, $\deg(s_j h_j) \leq 2t$, $s_j = \sum_{i_j} u_{j,i_j}^2$

$$y^T(p-\lambda) = y^T\Big(\sum_{i_0} u_{0,i_0}^2 + \sum_j \sum_{i_j} u_{j,i_j}^2 h_j\Big)$$

$$= \sum_{i_0} {u_{0,i_0}}^T M_t(y) u_{0,i_0} + \sum_j \sum_{i_j} {u_{j,i_j}}^T M_{t-d_j}(h_j y) u_{j,i_j} \geq 0$$

$$\Longrightarrow y^T p \geq \lambda$$

Hence: $p_t^{\text{mom}} \geq p_t^{\text{sos}}$

Are the SOS/MOM bounds useful ?

Do they converge to $p_{\min}$?

Are they sometimes exact ?

Properties of the bounds

$$\boxed{p_t^{\text{sos}} \leq p_t^{\text{mom}} \leq p_{\min}}$$

- For fixed t, one can compute p_t^{sos}, p_t^{mom} in **polynomial time** (to any fixed precision)
- **Asymptotic (finite) convergence** to $p_{\min}$
via Putinar's representation theorem
- **Optimality certificate**
via theorems of Curto and Fialkow about moment matrices
- **Extracting global minimizers**
via the eigenvalue method for solving systems of polynomial equations

Asymptotic Convergence under Putinar's assumption

Putinar's representation theorem:
(cf. more elementary proof by [Schweighofer 2005])

If (A) $\exists u_j \in \mathrm{SOS}_n : \{x \mid \sum_j u_j(x)h_j(x) \geq 0\}$ is compact
then $\quad p > 0$ on $K \Longrightarrow p = s_0 + \sum_j s_j h_j$ with s_0, s_j SOS

Asymptotic convergence theorem: [Lasserre 2001]

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

Note: A representation result valid for "$p \geq 0$ on K" gives **finite convergence**: $\boxed{p_t^{\mathrm{sos}} = p_t^{\mathrm{mom}} = p_{\min} \quad \text{for some } t}$

Proof of the Convergence Theorem

For $\epsilon > 0$, $p - p_{\min} + \epsilon > 0$ on K

Putinar's theorem:

$$\exists s_0, s_j \in \mathrm{SOS}_n \ \text{ s.t. } p - p_{\min} + \epsilon = s_0 + \sum_j s_j h_j$$

Hence: $\exists t$ s.t. $p_{\min} - \epsilon$ is feasible for (SOSt)

$$\Longrightarrow \ p_t^{\text{sos}} \geq p_{\min} - \epsilon$$

Therefore:

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

Note: (A) $\Longrightarrow K$ compact

(A) holds, e.g.,

• if $\{x \mid h_j(x) \geq 0\}$ is compact for some j

• if the equations $x_i^2 = x_i$ are present in the description of K **(0/1 case)**

If $K \subseteq \{x \mid \sum_i x_i^2 \leq R^2\}$ for some known R

add the (redundant) constraint: $R^2 - \sum_i x_i^2 \geq 0$ to the description of K

then (A) holds.

Finite Convergence in the finite variety case

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

$$V = \{x \in \mathbb{C}^n \mid h_j(x) = 0 \ \ (j = 1 \ldots k)\}$$

$$V_{\mathbb{R}} = V \cap \mathbb{R}^n$$

Theorem: [La 2002] [Lasserre/La/Rostalski 2006]

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

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

How to recognize that a given relaxation is exact?

Let y be an optimum solution to (MOMt).

Easy stopping criteria:

(i) If $x^* = (y_{10\ldots0}, \ldots, y_{0\ldots01}) \in K$ and $p(x^*) = p_t^{\text{mom}}$, then (MOMt) is exact: $p_t^{\text{mom}} = p_{\min}$ and x^* is a global minimizer.

(ii) Same conclusion if $\text{rank} M_t(y) = 1$.

Proof of (ii): Indeed, $y = \zeta_{x^*}$ with $x^* \in K$ and

$$p_{\min} \geq p_t^{\text{mom}} = p^T y = p(x^*) \geq p_{\min}$$

A general stopping criterion

Stopping criterion: Let y be an optimum solution to (MOMt).

If y has a representing measure μ on K, then $p_{\min} = p_t^{\mathrm{mom}}$ and $\mathrm{supp}(\mu) \subseteq \{$global minimizers of p over $K\}$.

Proof: $p_{\min} \geq p_t^{\mathrm{mom}} = p^T y = \int_K p(x)\mu(dx) \geq p_{\min}$.

How to recognize when y has a representing measure ?

Optimality Certificate and Extracting Global Minimizers

Theorem: [Henrion-Lasserre 2005]
Let $d = \max_j d_j,\ t \geq \max(d, \lceil \deg(p)/2 \rceil)$. Let y be an optimum solution to program **(MOMt)** satisfying:

(RC) $\boxed{\operatorname{rank} M_t(y) = \operatorname{rank} M_{t-d}(y) =: r}$

(i) y has a representing measure μ on K
with $\operatorname{supp}(\mu) = V(\operatorname{Ker} M_t(y))$ and $|\operatorname{supp}(\mu)| = r$.

(ii) $p_t^{\text{mom}} = p_{\min}$

(iii) $V(\operatorname{Ker} M_t(y)) \subseteq \{$global minimizers of p over $K\}$,
with equality if $\operatorname{rank} M_t(y)$ is maximum.

Proof for (ii), (iii) assuming (i)

Set $X_p :=$ { global minimizers of p over K }

By (i) + stopping criterion: y has a measure μ on K

$$\Longrightarrow p_{\min} = p_t^{\text{mom}} \text{ and } V(\text{Ker}\, M_t(y)) = \text{supp}(\mu) \subseteq X_p$$

To show: $X_p \subseteq V(\text{Ker}\, M_t(y))$ if $\text{rank}\, M_t(y)$ is max.

$\forall x \in X_p,\ \zeta_x$ is an optimum solution to (MOMt)

$$\Downarrow \text{[maximality of } \text{rank}\, M_t(y)]$$

$\text{Ker}\, M_t(y) \subseteq \text{Ker}\, M_t(\zeta_x) \subseteq I(x)$ (polynomials vanishing at x)

$$\Longrightarrow x \in V(\text{Ker}\, M_t(y))$$

Some remarks

- Remains to prove **(i)** ($\rightsquigarrow$ algebraic study of moment matrices)

- (RC) holds at max. rank sol. $\Longrightarrow |X_p| = |\text{supp}(\mu)| < \infty$

But: $|X_p| < \infty \not\Longrightarrow$ (RC) !

Example: $p_{\min} := \min p(x)$ s.t. $\sum_{i=1}^n x_i^2 \leq 1$, where p is a form in $\text{POS}_n \setminus \text{SOS}_n$. Then, $p_t^{\text{mom}} = p_t^{\text{sos}} < p_{\min}$, since

$$p = s_0 + s_1(1 - \sum_i x_i^2) \text{ with } s_0, s_1 \in \text{SOS}_n \Longrightarrow p \in \text{SOS}_n$$

- **However:** (RC) holds in the finite variety case.

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

$$\begin{aligned}
\min \quad & p = -25(x_1-2)^2 - (x_2-2)^2 - (x_3-1)^2 \\
& \qquad -(x_4-4)^2 - (x_5-1)^2 - (x_6-4)^2 \\
\text{s.t.} \quad & (x_3-3)^2 + x_4 \geq 4, \ \ (x_5-3)^2 + x_6 \geq 4 \\
& x_1 - 3x_2 \leq 2, \ \ -x_1 + x_2 \leq 2, \ x_1 + x_2 \leq 6, \\
& x_1 + x_2 \geq 2, \ 1 \leq x_3 \leq 5, \ 0 \leq x_4 \leq 6, \\
& 1 \leq x_5 \leq 5, \ 0 \leq x_6 \leq 10, \ x_1, x_2 \geq 0
\end{aligned}$$

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

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

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

Motzkin form: $p = x_1^2 x_2^2 (x_1^2 + x_2^2 - 3x_3^2) + x_3^6$

$$p_{\min} = \inf_{x \in \mathbb{R}^3} p(x) = \inf_{x_1^2+x_2^2+x_3^2 \leq 1} p(x) = 0$$

Recall: $p \neq s_0 + s_1(1 - \sum_i x_i^2)$ with $s_0, s_1 \in \mathrm{SOS}_3$

order t	rank sequence	bound p_t^{mom}	solution extrac
3	1 4 9 13	-0.0045964	none
4	1 4 10 20 29	-0.00020329	none
5	1 4 10 20 34 44	$-2.8976\ 10^{5}$	none
6	1 4 10 20 34 56 84	$-6.8376\ 10^{-6}$	none
7	1 4 10 20 35 56 84 120	$-2.1569\ 10^{-6}$	none

$d = 3$

Application to Unconstrained Polynomial Minimization

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

where $\deg(p) = 2d$

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

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

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:

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

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

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

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

When p attains its minimum

The 'ball approach':

• Approximate p^*_{ball} with Lasserre's relaxation scheme:
Asymptotic convergence, as Putinar's assumption holds!
• Seems to work well if the radius of the ball is not too large ...

The 'gradient variety' approach:

Representation result: [Demmel, Nie, Sturmfels 2004]

$$\boxed{p > 0 \text{ on } V^{\mathbb{R}}_{\text{grad}} \Longrightarrow p = s_0 + \sum_{i=1}^n u_i \frac{\partial p}{\partial x_i}}$$

$$\boxed{p \geq 0 \text{ on } V^{\mathbb{R}}_{\text{grad}} \Longrightarrow p = s_0 + \sum_{i=1}^n u_i \frac{\partial p}{\partial x_i}} \text{ if } I_{\text{grad}} \text{ radical}$$

where $s_0 \in \mathrm{SOS}_n$, $u_i \in \mathbb{R}[x]$

Convergence Result [Demmel, Nie, Sturmfels 2004]

Asymptotic Convergence to p^*_{grad} of the SOS and moment bounds

Finite Convergence to p^*_{grad} when $I_{\text{grad}} = \langle \frac{\partial p}{\partial x_1}, \ldots, \frac{\partial p}{\partial x_n} \rangle$ is a **radical ideal**

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]

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

- p_ϵ has a minimum and $\boxed{\lim_{\epsilon \longrightarrow 0} (p_\epsilon)_{\min} = p_{\min}}$
- the gradient variety of p_ϵ is **finite**

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

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

Example: Perturb the Motzkin polynomial

Motzkin polynomial: $p = x^2y^2(x^2+y^2-3)+1$

Then, $p_{\min} = 0$, attained at $(\pm 1, \pm 1)$, and p is **not** SOS

Minimize the perturbed polynomial:

$$\inf_{x\in\mathbb{R}^2} p_\epsilon = p + \epsilon(x_1^8 + x_2^8)$$

ϵ	order t	rank sequence	$(p_\epsilon)_t^{\mathrm{mom}}$	extracted solutions
10^{-1}	4	1 3 4 4	0.1595	$(\pm 0.9453, \pm 0.9453)$
10^{-2}	4	1 3 4 4	0.0194	$(\pm 0.9935, \pm 0.9935)$
10^{-3}	5	1 3 4 4	0.0019	$(\pm 0.9993, \pm 0.9993)$

Example: Perturb the Motzkin polynomial (continued)

$$p = x^2y^2(x^2+y^2-3)+1, \quad p_\epsilon = p + \epsilon(x_1^8 + x_2^8)$$

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

ϵ	order t	rank sequence	p_t^{mom}	extracted solutions
10^{-1}	4	-	unbounded	
10^{-1}	5	1 3 4 4 4 10	0.0315	none
10^{-1}	7	1 3 4 4 4 4 -	0.0315	$(\pm 0.9453, \pm 0.9453)$
10^{-2}	4	1 3 6 10 15	-22.07	none
10^{-2}	5	1 3 4 4 4 10	0.0005	none
10^{-2}	7	1 3 4 4 4 4 4 -	0.0005	$(\pm 0.9935, \pm 0.9935)$
10^{-3}	5	1 3 6 10 15 21	-0.0279	none
10^{-3}	6	1 3 4 4 4 4 11	$5.3\ 10^{-6}$	none
10^{-3}	7	1 3 4 4 4 4 4 -	$5.3\ 10^{-6}$	$(\pm 0.9993, \pm 0.9993)$
10^{-4}	7	1 3 4 4 4 4 4 -	$5.3\ 10^{-8}$	$(\pm 0.9999, \pm 0.9999)$

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	$\pm(0.0643, 15.4849)$

Strategy 2: Minimize p over its 'gradient tentacle'

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

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

where

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

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

Schweighofer's Representation Result

$$\boxed{\begin{array}{l} f > 0 \text{ on } K_{\text{grad}} \Longrightarrow f = s_0 + s_1(1 - \|\nabla p(x)\|^2 \|x\|^2) \\ \text{for some } s_0, s_1 \in \text{SOS}_n \end{array}}$$

if p *has* only isolated singularities at infinity (*)

[e.g., $n = 2$ or K_{grad} compact]

(*): System $\nabla p_d(x) = 0$, $p_{d-1}(x) = 0$ has finitely many projective zeros, with $p = p_d + p_{d-1} + \ldots + p_0$, p_i homogeneous of degree i

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

What's Next ?

Details about the proof technique for the following results:

- Optimality certificate
- Extraction of global minimizers
- Finite convergence in the finite variety case

Tools:

- Algebraic facts about polynomial ideals
- The eigenvalue method for solving systems of polynomial equations
- Results about moment matrices of Curto and Fialkow