# Moment Matrices and Optimization over Polynomials

Monique Laurent

CWI, Amsterdam

ADONET Doctoral school

Optimization over Polynomials and Semidefinite Programming

Klagenfurt

September 12-16, 2005

# Polynomial Optimization Problem (P)

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

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

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

# Unconstrained Polynomial Minimization

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

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

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

Hence, unconstrained polynomial minimization is **NP-hard** for degree 4 polynomials

## $0/1$ Linear Programming

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

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

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

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

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

# Some notation about polynomials

- Multivariate polynomial:

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

- $\deg(x^\alpha) = |\alpha| := \sum_{i=1}^n \alpha_i$

- $\deg(p) = \max \deg(x^\alpha)$ s.t. $p_\alpha \neq 0$

- We identify a polynomial $p(x)$ with its sequence of coefficients $p = (p_\alpha)_{|\alpha| \leq d}$, if $\deg(p) = d$

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

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

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

**Example:** $x^2 + y^2 + 2xy + z^6 = (x+y)^2 + (z^3)^2$ is SOS

# Recognizing SOS of polynomials via SDP

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

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

$$\Updownarrow$$

$$p(x) = z^T \underbrace{\left(\sum_j u_j u_j^T\right)}_{X \succeq 0} z, \quad \text{setting } z := (x^\beta)_{|\beta| \le d}$$

$$\Updownarrow$$

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 $X$ of order $\binom{n+d}{d} \times \binom{n+d}{d}$, with $\binom{n+2d}{2d}$ equations

$\rightsquigarrow$ polynomial size *fixing either* $n$ *or* $d$

# When does a polynomial have a decomposition as a sum of squares?

Hilbert [1888] classified the pairs $(n, d)$ for which *every nonnegative polynomial of degree* $d$ *in* $n$ *variables is a sum of squares of polynomials*:

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

E.g., the **Motzkin polynomial:** $x^4y^2 + x^2y^4 - 3x^2y^2 + 1$ is nonnegative but *not* a SOS

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

$$\boxed{p \geq 0 \text{ on } \mathbb{R}^n \Longrightarrow p = \sum_i \left(\frac{p_i}{q_i}\right)^2, \text{ where } p_i, q_i \in \mathbb{R}[x_1, \ldots, x_n]}$$

That is,

$$\boxed{p \cdot q^2 \text{ is SOS for some } q \in \mathbb{R}[x_1, \ldots, x_n]}$$

Sometimes, the common denominator is known:

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

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

## (Dual) SOS approximations for (P)

following [Lasserre 2001]

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

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

Replacing the *nonnegativity* condition by a stronger *SOS* condition: formulate **(SOSt)**

$$\begin{array}{ll} \rho_t^* := \sup \ \rho \ \text{ s.t.} & p(x) - \rho = s_0(x) + \sum_{j=1}^m s_j(x) h_j(x) \\ & s_0, \ s_j \text{ are SOS} \\ & \deg(s_0), \ \deg(s_j h_j) \leq 2t \end{array}$$

for any $t \geq \lceil \deg(p)/2 \rceil, \ \lceil \deg(h_j)/2 \rceil$

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

**Example:** $K = \mathbb{R}^n$ (unconstrained case), $\deg(p) = 2d$
$\rho_t^* = \rho_d^* \leq p_{\min}$, with equality iff $p(x) - p_{\min}$ is SOS

Equality in the univariate case (Shor [1987], Nesterov [2000])

# (Primal) Moment Approximations

$$\boxed{p_{\min} = \inf_{x \in K} \ p(x) = \inf_{\mu} \int p(x)\mu(dx) = \inf_{y} p^T y}$$

where $\mu$ is a probability measure supported by $K$ and $y$ is the sequence of moments of a prob. measure supported by $K$

**Note:** $$\int p(x)\mu(dx) = \sum_\alpha p_\alpha \underbrace{\int x^\alpha \mu(dx)}_{:= y_\alpha} = p^T y$$

**Definition:** For a measure $\mu$, $y_\alpha := \int x^\alpha \mu(dx)$ is called the moment of order $\alpha$ of the measure $\mu$; $y = (y_\alpha)_\alpha$ is the sequence of moments of $\mu$, and $\mu$ is a representing measure for $y$.

## Lemma: Assume $y$ is the sequence of moments of $\mu \in \mathcal{P}_K$. 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*

**Moreover:** $\boxed{\text{supp}(\mu) \subseteq \text{Zeros}(p)} \quad \forall p \in \text{Ker} M_t(y)$

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

## Proof: For $p = (p_\alpha)_{|\alpha| \le t}$

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

$$= \int p(x)^2 \mu(dx) \ge 0$$

$$p^T M_t(h_j * y) p = \sum_{\alpha,\beta} p_\alpha p_\beta (h_j)_\gamma y_{\alpha+\beta+\gamma}$$

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

For $p \in \text{Ker } M_t(y), \quad 0 = p^T M_t(y) p = \int p(x)^2 \mu(dx)$

Implying: $\text{supp}(\mu) \subseteq \text{Zeros}(p)$

If we relax the condition "$y$ is the sequence of moments of a measure supported by $K$" by requiring positive semidefiniteness of its moment matrix, we obtain a hierarchy of tractable SDP relaxations:

**(MOMt)**

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

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

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

## Lemma:

The semidefinite programs **(SOSt)** and **(MOMt)** are dual of each other.

**Weak duality:** $\boxed{\rho_t^* \leq p_t^*}$

No duality gap, e.g., if $K$ has a nonempty interior

## Properties of the bounds

$$\boxed{\rho_t^* \le p_t^* \le p_{\min}}$$

- For *fixed* $t$, one can compute $\rho_t^*$, $p_t^*$ in **polynomial time** (to any fixed precision)
- **Asymptotic (finite) convergence** to $p_{\min}$
via some representation theorems for positive polynomials from real algebraic geometry
- **Optimality certificate**
via some theorems of Curto and Fialkow about moment matrices
- **Extracting global minimizers**
via the eigenvalue method for solving systems of polynomial equations

## Asymptotic Convergence

**Representation Theorem:** [Putinar 1993]
(more elementary proof by [Schweighofer 2005])

If $\quad \exists u_j$ SOS : $\{x \mid \sum_j u_j(x)h_j(x) \geq 0\}$ is compact $\qquad (*)$

then $\quad p > 0$ on $K \Longrightarrow p = s_0 + \sum_j s_j h_j$ with $s_0, s_j$ SOS

**Convergence Theorem:** [Lasserre 2001]

If $(*)$ then $\boxed{\lim_{t\to\infty} \rho_t^* = \lim_{t\to\infty} p_t^* = p_{\min}}$

## Proof of the Convergence Theorem

Fix $\epsilon > 0$

The polynomial $p - p_{\min} + \epsilon$ is positive on $K$

**By Putinar's Theorem,** $p - p_{\min} + \epsilon = s_0 + \sum_j s_j h_j$ for some polynomials $s_0, s_j$ that are SOS

**Thus,** $p_{\min} - \epsilon$ is feasible for (SOS$t$) for some $t$, implying $\rho_t^* \geq p_{\min} - \epsilon$

**Therefore:** $\forall \epsilon > 0 \; \exists t$ for which $\rho_t^* \geq p_{\min} - \epsilon$

**This shows:**

$$\lim_{t \to \infty} \rho_t^* = p_{\min}$$

**Note:** $(*)$ implies $K$ compact. Moreover, $(*)$ 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 the radius $R$ of a ball containing $K$ is known, then add the (redundant) constraint $R^2 - \sum_i x_i^2 \geq 0$ to the description of $K$

**Note:** A representation result valid for "$p \geq 0$ on $K$" would give a *finite* convergence result

# Finite Convergence: Finite Variety

Assume the equations: $h_1(x) = 0 \ldots h_k(x) = 0$ are present in the description of $K$ and

$$V := \{x \in \mathbb{C}^n \mid h_1(x) = 0 \ldots h_k(x) = 0\} \text{ is } \textit{finite}$$

**Then:** $\boxed{p_t^* = p_{\min} \quad \text{for } t \text{ large enough}}$ [Laurent 2002]

**Moreover:** $\boxed{\rho_t^* = p_t^* = p_{\min} \quad \text{for } t \text{ large enough}}$

if $h_1, \ldots, h_k$ form a *Groebner basis* [Laurent 2002]
or if they generate a *radical ideal* [Parrilo 2002]

**Representation result in radical case:** [Parrilo 2002]
$p \geq 0$ on $K \Longrightarrow p = s_0 + \sum_j s_j h_j$ ($s_0, s_j$ SOS)

# Optimality Certificate and Extracting Global Minimizers

**Theorem:** [Henrion-Lasserre 2005]
If $y$ is an optimum solution to the program **(MOMt)** satisfying:

$$\boxed{\mathrm{rank} M_t(y) = \mathrm{rank} M_{t-d}(y)}$$

where $d = \max_j \lceil \deg(h_j)/2 \rceil$

then: $\boxed{p_t^* = p_{\min}}$

Moreover, the solutions to the system of polynomial equations:

$$\boxed{p(x) = 0 \quad \text{for } p \in \mathrm{Ker}\ M_t(y)}$$

are global minimizers of $p$ over $K$

$\rightsquigarrow$ they can be found using the *eigenvalue method*

# 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, Parrilo

## 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 $t$ | bound $p_t^*$ | solution extracted |
|---|---|---|
| 1 | unbounded | none |
| 2 | -310 | (5,1,5,0,5,10) |

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 $t$ | bound $p_t^*$ | solution extracted |
|---|---|---|
| 2 | -7 | none |
| 3 | -6.6667 | none |
| 4 | -5.5080 | (2.3295,3.1785) |

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

## What's Next ?

**Details about the proof technique for the following results:**

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

**Tools:**

- Algebraic facts about polynomial ideals
- Results about moment matrices of Curto and Fialkow

**After that:**

- Some more representation results yielding approximation schemes for unconstrained polynomial minimization, matrix copositivity, stable set problem, etc.