Didier Henrion

The moment-SOS hierarchy

Outline

1. Polynomial optimization (POP)

2. Linear conic reformulation

3. Polynomial sums of squares (SOS) and moments

4. The moment-SOS aka Lasserre hierarchy

5. Extensions

1 - Polynomial optimization

POP (Polynomial Optimization Problem)

Given polynomials $p, p_1, \ldots p_m \in \mathbb{R}[x]$ of the indeterminate $x \in \mathbb{R}^n$, consider the **nonlinear nonconvex** global optimization problem

$$p^* = \min_{x \in X} p(x)$$

defined on the bounded basic semialgebraic set

$$X := \{x \in \mathbb{R}^n : p_1(x) \geq 0, \ \ldots \ p_m(x) \geq 0\}$$

In general POP can be very challenging:

- p can be nonconvex
- X can be nonconvex and/or disconnected and/or discrete
- there can be several global optimizers, maybe infinitly many

2 - Linear conic reformulation

Primal linear reformulation

Instead of the POP

$$p^* = \min_{x \in X} p(x)$$

over vectors in X, consider the **linear** problem (LP)

$$p^*_M = \min_{\mu} \int_X p(x) d\mu(x)$$

over probability measures (normalized bounded linear functionals on continuous functions) on X

Lemma: $p^*_M = p^*$ and the LP has for optimal solution the Dirac measure at any optimal solution of the POP

Dual linear reformulation

The Lagrange dual to the LP on probability measures

$$\min_{\mu} \int_X p(x) d\mu(x)$$

reads

$$\max_{p_L} \; p_L \text{ s.t. } p(x) \geq p_L \;\; \forall x \in X$$

which can be rephrased as an LP on **positive polynomials**

$$\max_{p_L} \; p_L \text{ s.t. } p(x) - p_L \in P(X)_d$$

where $P(X)_d$ denotes the convex cone of polynomials of degree up to d that are non-negative on X

Moments

Let $(b_a(x))_{a\in\mathbb{N}^n_d}$ denote a basis of the vector space of n-variate polynomials of degree at most d of dimension $\binom{n+d}{n}$, indexed in $\mathbb{N}^n_d := \{a \in \mathbb{N}^n : \sum_{k=1}^n a_k \leq d\}$

The polynomial p can then be written as

$$p(x) = \sum_{a\in\mathbb{N}^n_d} p_a b_a(x)$$

and the objective function can be written as

$$\int_X p(x)d\mu(x) = \sum_{a\in\mathbb{N}^n_d} p_a y_a$$

which is a linear function of the **moments** of measure μ

$$y_a = \int_X b_a(x)d\mu(x)$$

Moments and positive polynomials

The LP on probability measures

$$\min_{\mu} \int_X p(x) d\mu(x)$$

becomes an LP on **moments**

$$\min_{y} \sum_a p_a y_a \text{ s.t. } y_0 = 1,\ y \in P(X)'_d$$

which is dual to the LP on **positive polynomials**

$$\max_{p_L} \ p_L \text{ s.t. } p(x) - p_L \in P(X)_d$$

since

- measures on compact X are uniquely determined by moments
- the constraint $y_0 = 1$ corresponds to the normalization
- by the Riesz-Haviland Theorem, the cone of moments is dual to the cone of positive polynomials

Wonderful but ...

Challenging convex cones

Testing whether $p \in P(X)_d$ or $y \in P(X)'_d$ is difficult

Not much is known about the geometry of these cones

No efficient barrier function is known

... so we will content ourselves with **approximations**

3 - Polynomial sums of squares and moments

Approximating positive polynomials

The cone of positive polynomials $P(X)_d$ on the compact set

$$X := \{x \in \mathbb{R}^n : p_k(x) \geq 0, k = 1, \ldots, m\}$$

is generally **intractable**, so we will approximate it.

Denoting $p_0(x) := 1$ and enforcing (without loss of generality) $p_1(x) := R^2 - \sum_{i=1}^n x_i^2$ for R large enough, consider for given r

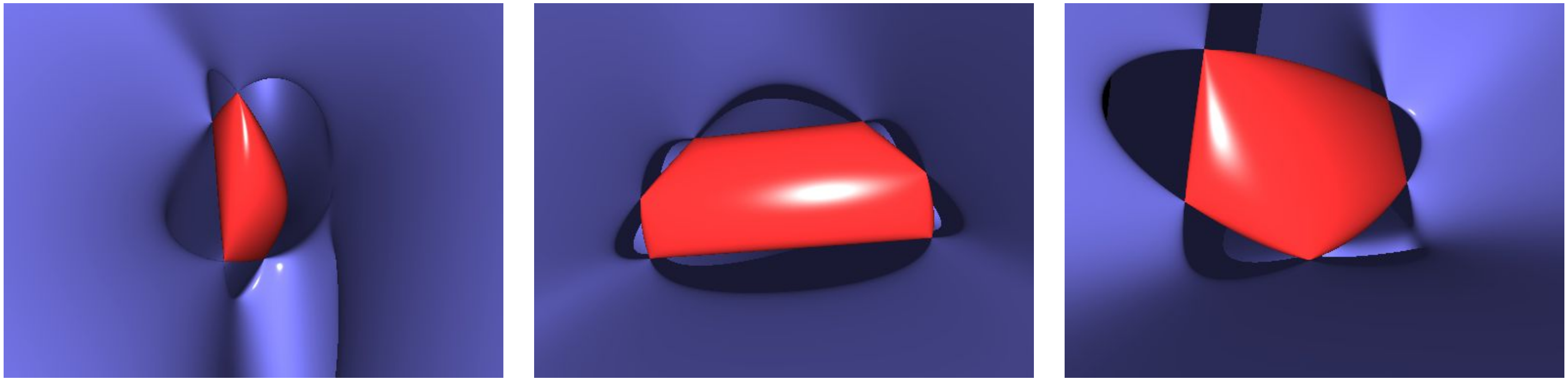
$$\Sigma(X)_r := \{p \in \mathbb{R}[x]_d : p = \sum_{k=0}^m s_k p_k, \;\; s_k \in \Sigma_{2\lceil r - \deg p_k/2 \rceil}\}$$

where Σ_{2d} denotes the cone of **sums of squares** (SOS) of polynomials of degree at most d

Lemma: By construction $\Sigma(X)_r \subset \Sigma(X)_{r+1} \subset P(X)_d$

Polynomial SOS

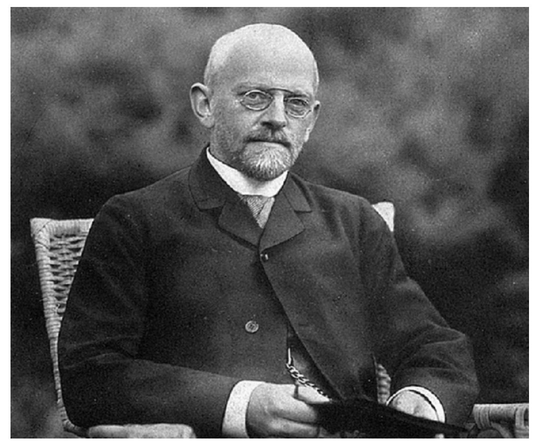
Lemma: Deciding whether a polynomial is SOS reduces to semidefinite programming



Semidefinite programs can be solved efficiently with primal-dual interior-point methods

SOS and positivity

Theorem (Hilbert 1888): $\Sigma(\mathbb{R}^n)_{2d} = P(\mathbb{R}^n)_{2d}$ if and only if $n = 1$ or $d = 1$ or $n = d = 2$



Hilbert's 17th problem at ICM Paris 1900

Motzkin's 1965 example

$$p = 1 - 3x_1^2x_2^2 + x_1^4x_2^2 + x_1^2x_2^4 \in P(\mathbb{R}^2)_6 \backslash \Sigma(\mathbb{R}^2)_6$$

Moment relaxations

Hence we have a hierarchy of tractable **inner approximations** for the cone of positive polynomials

$$\Sigma(X)_r \subset \Sigma(X)_{r+1} \subset P(X)_d$$

Using convex duality, we also have a hierarchy of tractable **outer approximations** for the cone of moments

$$\Sigma(X)'_r \supset \Sigma(X)'_{r+1} \supset P(X)'_d$$

Elements of $\Sigma(X)'_r$ are sometimes called pseudo-expectations or pseudo-moments, since some of them are not moments

We also say that $\Sigma(X)'_r$ is a **relaxation** of $P(X)'_d$

4 - The Lasserre aka moment-SOS hierarchy

The Lasserre hierarchy for POP

Replace the intractable problems

$$p^* = \min_y \sum_a p_a y_a \text{ s.t. } y_0 = 1, \; y \in P(X)'_d$$

$$p^* = \max_{p_L} \; p_L \text{ s.t. } p(x) - p_L \in P(X)_d$$

with the hierarchy of **semidefinite** problems

$$p_r^* = \min_y \sum_a p_a y_a \text{ s.t. } y_0 = 1, \; y \in \Sigma(X)'_r$$

$$p_r^* = \max_{p_L} \; p_L \text{ s.t. } p(x) - p_L \in \Sigma(X)_r$$

for increasing values of r

Convergence

Integer r is called the **relaxation order**

Since $\Sigma(X)_r \subset \Sigma(X)_{r+1} \subset P(X)_d$, we have a monotone non-decreasing sequence of lower bounds on the POP value:

$$p_r^* \leq p_{r+1}^* \leq p^*$$

Theorem (Putinar 1993): $\overline{\Sigma(X)_\infty} = P(X)_d$

Theorem (Lasserre 2001): $p_\infty^* = p^*$

Finite convergence

Theorem (Nie 2014): Generically $\exists r < \infty$ such that $p_r^* = p^*$

In other words, a vanishing small random perturbation of the input data of a given POP ensures **finite convergence** of the Lasserre hierarchy

We have linear algebra conditions to ensure finite convergence, **certify** global optimality and **extract** minimizers

We can also use polynomial kernels to retrieve approximately and sometimes exactly the **variety** of global minimizers

Extracting finitely many atoms

Define the **moment matrix** of degree $2d$

$$M_d(y) := \int_X a(x)a(x)^\top d\mu(x)$$

as a symmetric matrix linear function of the moments y of μ

Theorem (Flat Extension, Curto & Fialkow 1991): If the rank of $M_r(y)$ does not increase when r increases, then the moment relaxation is exact

Global solutions extracted by linear algebra, as implemented in our Matlab interface GloptiPoly **(H & Lasserre 2003)**

Approximating the support

Since the moment matrix is positive semidefinite, it has a spectral decomposition

$$M_d(y) = QEQ^\top$$

where Q is an orthonormal matrix whose columns are denoted q_i, $i = 1, 2, \ldots$ and E is a diagonal matrix with diagonal entries are eigenvalues $e_{i+1} \geq e_i \geq 0$ of the moment matrix

Each column q_i is the vector of coefficients in basis a of a polynomial $q_i(x)$, so that

$$q_i^\top M_d(y) q_i = \int q_i^2(x) d\mu(x) = e_i$$

Let $r \in \mathbb{N}$ and $\beta > 0$, let

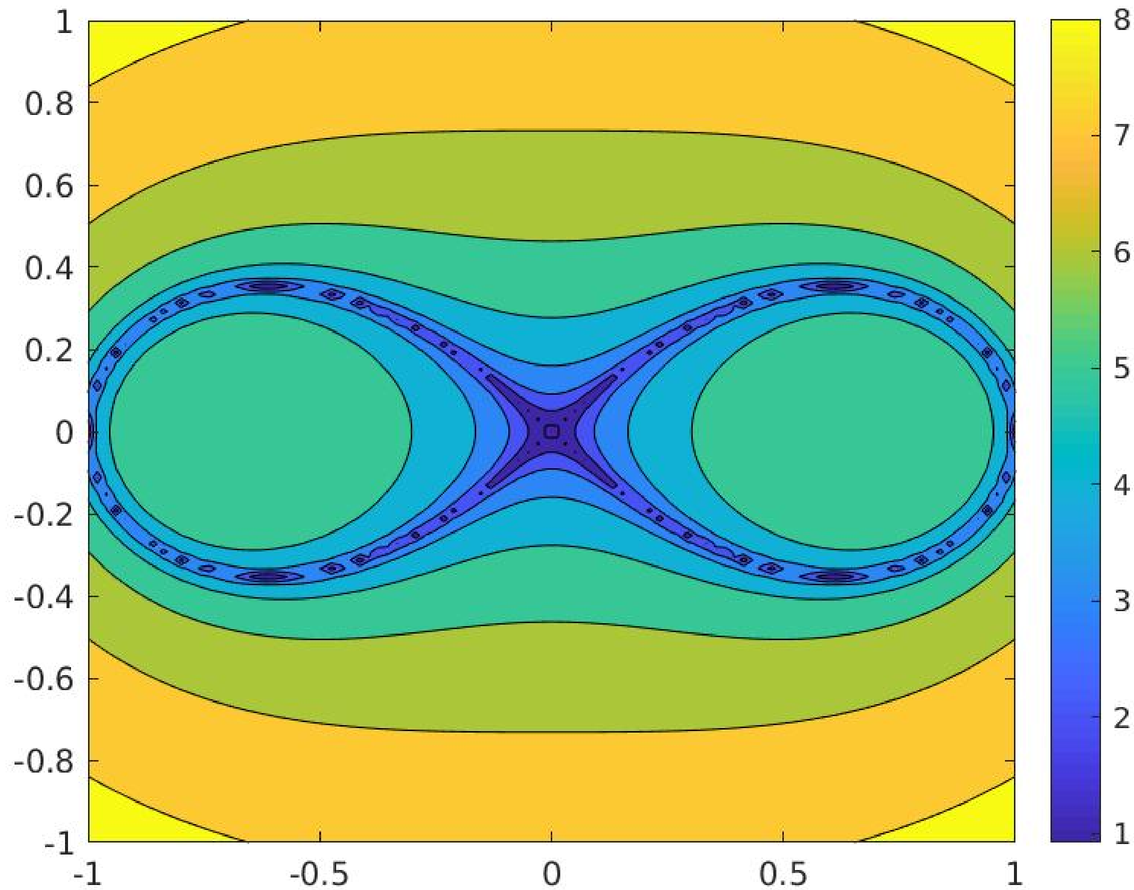
$$\gamma = \frac{\sum_{i=1}^{r} e_i}{\beta}$$

and follow **(Lasserre & Pauwels 2019)** to define the **Christoffel-Darboux** SOS polynomial

$$p_{\mathrm{cd}}(x) := \sum_{i=1}^{r} q_i^2(x)$$

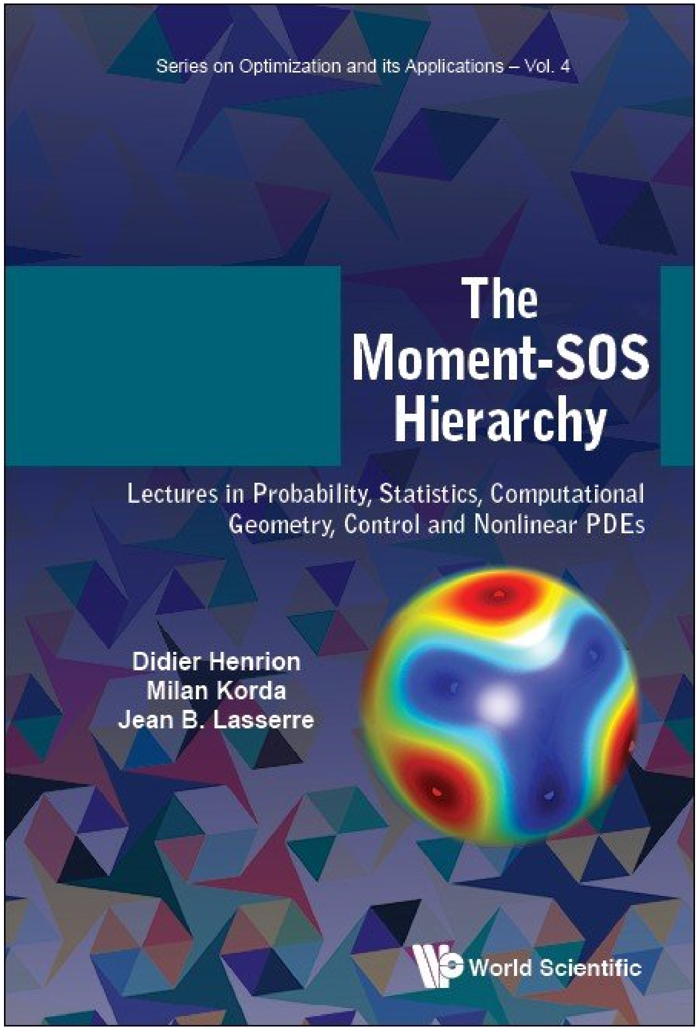
Lemma (concentration): $\mu(\{x \in X : p_{\mathrm{cd}}(x) \leq \gamma\}) \geq 1 - \beta$

We can retrieve approximately the support of a measure from sublevel sets of its Christoffel-Darboux polynomial



Moment matrix of order 4 and size 15 for the POP $\min_{x\in\mathbb{R}^2}(x_1^2+x_2^2)^2-x_1^2+x_2^2$

5 - Extensions



Polynomial Optimization, Efficiency through Moments and Algebra

poema-network.eu

tenors-network.eu

Tensor modeling, geometry and optimization

Tutorial at homepages.laas.fr/henrion/courses/moments.pdf