Four Lectures on Polynomial Optimization for Control

Lecture 1: The Moment-Sum-of-Squares Hierarchy

Joe Warrington

General optimization problem

Consider the problem

$$
\begin{aligned}
f^\star = \quad & \inf_{x \in \mathbb{R}^n} \quad f(x) \\
& \text{s. t.} \quad x \in \Omega
\end{aligned}
$$

where $\Omega := \{x : g_j(x) \geq 0, \quad j = 1, \ldots, m\} \subseteq \mathbb{R}^n$ for some functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1 \ldots, m$. This can be rewritten in two ways:

$$
\begin{aligned}
(1) \ f^\star = \inf_{\mu \in \mathcal{M}_+} & \int_\Omega f(x)\mu(x)\, \mathrm{d}x \\
\text{s. t.} & \int_\Omega \mu(x)\, \mathrm{d}x = 1
\end{aligned}
$$

Interpretation: $\mu^\star$ is the probability measure supported on Ω over which the expectation of $f(x)$ is minimized.

$$
\begin{aligned}
(2) \ f^\star = \sup_{\lambda \in \mathbb{R}} & \ \lambda \\
\text{s. t.} & \ f(x) - \lambda \geq 0, \ \forall x \in \Omega
\end{aligned}
$$

Interpretation: $\lambda^\star$ is the largest number that can be subtracted from $f(x)$ such that the result stays positive on Ω.

In fact these two are duals of each other (they are infinite-dimensional linear programs), and strong duality often holds. The only problem is that (1) has infinitely many decision variables and (2) has infinitely many constraints...

Polynomial optimization problem

What if f, $g_1, \dots, g_m$ are polynomial?

$$f^\star = \inf_{x \in \mathbb{R}^n} \quad f(x)$$
$$\text{s. t.} \quad x \in \Omega$$

Then tools are available to make the problem finite-dimensional. Define the notation

- $\mathbb{R}[x]$ as the *ring*[1] *of polynomials* on $x = [x_1, \dots, x_n]^\top$.
- x^α for $x \in \mathbb{R}^n$ and vector $\alpha \in \mathbb{N}^n$ means $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$, i.e., shorthand notation for a monomial.

Example

For example if $n = 2$ and $\alpha = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, then $x^\alpha = x_1^2 x_2$.

- f_α is the coefficient that multiplies the monomial x^α

$$\Rightarrow \text{we can write} \quad f(x) = \sum_{\alpha \in \mathbb{N}^n} f_\alpha x^\alpha$$

[1] The "ring" part just means polynomials come with operators for commutative addition, multiplication, and scalar multiplication.

Bounded-degree polynomial

For a polynomial whose maximum total degree amongst all monomials is d, define the following:

- $\mathbb{R}[x]_d$ as the *ring of polynomials of degree at most* d.
- $\mathbb{N}^n_d$ as the set of integer vectors summing to no more than d:

$$\mathbb{N}^n_d := \left\{ \alpha \in \mathbb{N}^n \, : \, \sum_i \alpha_i \leq d \right\}$$

Example: $n = 2$, $d = 2$

We have $\mathbb{N}^2_2 = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right\}$ so the different possible values of x^α for $\alpha \in \mathbb{N}^2_2$ are the monomials 1, x_1, x_2, x_1^2, $x_1 x_2$, and x_2^2.

There are $|\mathbb{N}^n_d| = s(d) := \binom{n+d}{n}$ monomials in an n-dimensional, degree-d polynomial $f \in \mathbb{R}[x]_d$. So f is described by $s(d)$ scalar coefficients f_α:

$$f(x) = \sum_{\alpha \in \mathbb{N}^n_d} f_\alpha x^\alpha$$

Expectation of a polynomial function

The α-moment of a measure μ is defined by $y_\alpha := \int x^\alpha \mu(x)\,\mathrm{d}x$. The expectation of a polynomial $f(x)$ with respect to a measure μ with moments $\{y_\alpha\}_{\alpha\in\mathbb{N}^n}$ is therefore

$$\begin{aligned}\int f(x)\mu(x)\,\mathrm{d}x &= \int \sum_{\alpha\in\mathbb{N}^n} f_\alpha x^\alpha \mu(x)\,\mathrm{d}x \\ &= \sum_{\alpha\in\mathbb{N}^n} (f_\alpha \underbrace{\int x^\alpha \mu(x)\,\mathrm{d}x)}_{\text{Definition of } \alpha\text{-moment of } \mu} = \sum_{\alpha\in\mathbb{N}^n} f_\alpha y_\alpha\end{aligned}$$

The expectation depends on the same moments as the monomials present in f.

Riesz Functional

For *any* sequence of moments $\{y_\alpha\}$, the **Riesz Functional** $L_\mathbf{y} : \mathbb{R}[x] \to \mathbb{R}$ is defined as

$$L_\mathbf{y}(f) = \sum_{\alpha\in\mathbb{N}^n} f_\alpha y_\alpha$$

Moment matrix $\mathbf{M}_d(\mathbf{y})$

The **moment matrix** $\mathbf{M}_d(\mathbf{y})$ is an $s(d) \times s(d)$ real, symmetric matrix with rows and columns indexed by $\mathbb{N}^n_d$. Its (α, β) entry is equal to $y_{\alpha+\beta}$.

Thus, populating $\mathbf{M}_d(\mathbf{y})$ requires a $\mathbf{y}$ containing moments up to degree $2d$.
It has the following useful property:

Theorem

"Pseudo-moment vector $\mathbf{y}$ *has a representing measure* μ*"* $\Rightarrow \mathbf{M}_d(\mathbf{y}) \succeq 0$

Proof.

If $\mathbf{y}$ has a representing measure μ then $\int f(x)^2 \mu(x) dx \geq 0$ for any polynomial $f \in \mathbb{R}[x]_d$, as neither f^2 nor μ are negative anywhere. It can be shown that

$$\int f(x)^2 \mu(x)\,\mathrm{d}x \geq 0 \;\forall f \in \mathbb{R}[x]_d \Leftrightarrow \sum_{\alpha \in \mathbb{N}^n_d} \sum_{\beta \in \mathbb{N}^n_d} f_\alpha f_\beta y_{\alpha+\beta} \geq 0 \;\forall f \in \mathbb{R}[x]_d \,.$$

The right-hand side is equivalent to $\mathbf{f}^\top \mathbf{M}_d(\mathbf{y}) \mathbf{f} \geq 0$ for all $\mathbf{f} \in \mathbb{R}^{s(d)}$, which is the definition of $\mathbf{M}_d(\mathbf{y}) \succeq 0$. □

Localizing matrix $\mathbf{M}_d(g\mathbf{y})$

Given a polynomial $g(x) = \sum_\gamma g_\gamma x^\gamma$, the **localizing matrix** $\mathbf{M}_d(g\mathbf{y})$ is an $s(d) \times s(d)$ real, symmetric matrix with rows and columns indexed by $\mathbb{N}^n_d$. Its (α, β) entry is equal to $\sum_\gamma g_\gamma y_{\alpha+\beta+\gamma}$. Populating $\mathbf{M}_d(g\mathbf{y})$ requires a $\mathbf{y}$ containing moments up to degree $2d + \deg(g)$.[2] It has the following useful property:

Theorem

"Pseudo-moment vector $\mathbf{y}$ has a representing measure μ whose support is contained in the set $\{x : g(x) \geq 0\}$" $\Rightarrow \mathbf{M}_d(g\mathbf{y}) \succeq 0$

Proof.

If $\mathbf{y}$ has a representing measure μ supported on $\{x : g(x) \geq 0\}$, then $\int f(x)^2 g(x)\mu(x)dx \geq 0$ for any polynomial $f \in \mathbb{R}[x]_d$, as neither f^2, nor g, nor μ are negative. It can be shown that

$$\int f(x)^2 g(x)\mu(x)\,\mathrm{d}x \geq 0 \;\forall f \in \mathbb{R}[x]_d \Leftrightarrow \sum_{\alpha,\beta,\gamma} f_\alpha f_\beta g_\gamma y_{\alpha+\beta+\gamma} \geq 0 \;\forall f \in \mathbb{R}[x]_d \,.$$

The right-hand side is equivalent to $\mathbf{f}^\top \mathbf{M}_d(g\mathbf{y})\mathbf{f} \geq 0 \;\forall \mathbf{f} \in \mathbb{R}^{s(d)}$, i.e. $\mathbf{M}_d(g\mathbf{y}) \succeq 0$. □

The moment matrix is just a special case of the localizing matrix, with $g(x) \equiv 1$.

[2] The common convention is to use $\mathbf{M}_{d-\lceil \deg(g)/2 \rceil}(g\mathbf{y})$ instead, to limit moments of $\mathbf{y}$ to degree $2d$.

Moment relaxation of problem (1)

We rewrote $f^\star = \inf_{x\in\mathbb{R}^n} f(x)$ s. t. $x \in \Omega$ as

$$(1) \qquad f^\star = \inf_{\mu\in\mathcal{M}_+} \int_\Omega f(x)\mu(x)\,\mathrm{d}x \quad \text{s. t.} \int_\Omega \mu(x)\,\mathrm{d}x = 1$$

The level-d **moment relaxation** of (1)

$$(\mathbf{P}_d) \qquad \rho_d = \inf_{\mathbf{y}} L_{\mathbf{y}}(f)$$

$$\text{s. t. } y_0 = 1, \qquad \text{(A)}$$

$$\mathbf{M}_d(\mathbf{y}) \succeq 0, \qquad \text{(B)}$$

$$\mathbf{M}_{d-d_j}(g_j\mathbf{y}) \succeq 0, \quad j = 1, \ldots, m \qquad \text{(C)}$$

where $\mathbf{y} = \{y_\alpha\}_{\alpha\in\mathbb{N}^n_{2d}}$, $d_j := \lceil \deg(g_j)/2 \rceil$, and $d \geq \max\{\deg(f), \deg(g_1), \ldots, \deg(g_m)\}$.

(A) is necessary for $\mathbf{y}$ to correspond to a probability measure, i.e., integrating to 1.

(B) is necessary for the vector of *pseudo-moments* $\mathbf{y}$ to have a representing measure μ. That is, that there can exist a measure μ whose moments are $\{y_\alpha\}$.

(C) lists the so-called *localizing* constraints that are necessary for μ to be supported on Ω. Constraints (A) to (C) are necessary but not sufficient for $\int_\Omega \mu(x)\,\mathrm{d}x = 1$.

$\Rightarrow$ **Problem $(\mathbf{P}_d)$ is a tractable SDP relaxation of (1).**

Sum-of-squares polynomials

When is a polynomial $f \in \mathbb{R}[x]_d$ non-negative on all of $\mathbb{R}^n$?

Clearly it is **sufficient** if one can write it as a sum of squared polynomials:

$$f(x) = \sum_i f_i(x)^2 \implies f(x) \geq 0 \; \forall x \in \mathbb{R}^n$$

Sum-of-squares polynomial

The condition $f(x) = \sum_i f_i(x)^2$ is equivalent to

$$\exists \mathbf{F} \in \mathbb{R}^{s(d) \times s(d)} \text{ s. t. } \mathbf{F} \succeq 0 \text{ and } f = \mathbf{x}^\top \mathbf{F} \mathbf{x} \qquad \text{(SOS)}$$

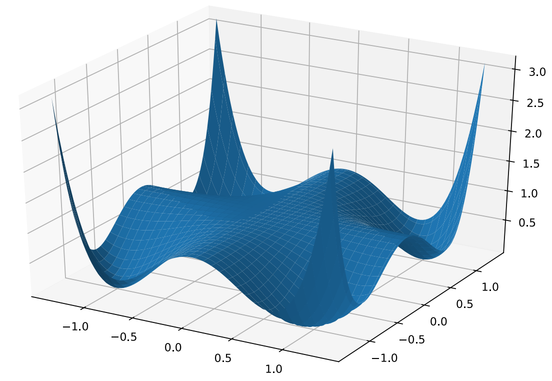
where $\mathbf{x}$ is an $s(d)$-dimensional vector containing the monomials x^α for all $\alpha \in \mathbb{N}_d^n$. All the coefficient data appears in the matrix $\mathbf{F}$. We use $\Sigma[x]_d$ to denote the set of all degree-d polynomials satisfying (SOS).

Sum-of-squares polynomials

But this condition is in general **not necessary**, i.e., $f(x) \in \Sigma[x]_d \not\Leftarrow f(x) \geq 0 \; \forall x \in \mathbb{R}^n$.

Example (Motzkin 1967)

The polynomial in $n = 2$, $d = 6$, $f(x) = x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 + 1$, is non-negative on all of $\mathbb{R}^2$ but has **no SOS representation**.



Sum-of-squares polynomials

Although $f(x) \in \Sigma[x]_d \not\Leftarrow f(x) \geq 0 \ \forall x \in \mathbb{R}^n$ in general, the reverse implication **does** turn out to hold in the following special cases:[3]

The only cases where non-negativity also implies SOS

- $n = 1$, $d \geq 0$: univariate polynomials of any degree;
- $n \geq 1$, $d = 2$: quadratic polynomials of any dimension;
- $n = 2$, $d = 4$: quartic polynomials in 2 dimensions

Obviously, Motzkin's polynomial does not fall into any of these categories.

[3] J.-B. Lasserre, "An Introduction to Polynomial and Semi-Algebraic Optimization", Chapter 2, Cambridge University Press, 2015

Quadratic module

For the dual problem (2), we have to check the condition $f(x) - \lambda \geq 0,\ \forall x \in \Omega$. Each point in Ω defines a constraint $\implies$ intractable!

We instead try to find a polynomial that can be certified as non-negative on Ω.[4]

Quadratic module

For $g := (g_1, g_2, \dots, g_m)$, the **quadratic module** is defined as

$$Q(g) := \left\{ \sigma_0 + \sum_{i=1}^{m} \sigma_j g_j \;\middle|\; \sigma_j \in \Sigma[x],\ j = 0, \dots, m \right\}.$$

Truncated quadratic module

The **truncated quadratic module** is defined, for polynomial degree k, as

$$Q_k(g) := \left\{ \sigma_0 + \sum_{i=1}^{m} \sigma_j g_j \;\middle|\; \sigma_0 \in \Sigma[x]_{k},\ \sigma_j \in \Sigma[x]_{k-d_j},\ j = 1, \dots, m \right\},$$

where $d_j := \lceil \deg(g_j)/2 \rceil$.

[4] Note that non-negativity on Ω is much less restrictive than non-negativity on $\mathbb{R}^n$, so simply enforcing $f(x) - \lambda \in \Sigma[x]$ is generally way too conservative.

Putinar's *Positivstellensatz*

We require one more technicality, namely an algebraic guarantee that set Ω is compact:

Definition (Lasserre 2018, Def. 1)

Archimedean condition: The quadratic module $Q(g)$ associated with Ω is said to be *Archimedean* if there exists $M > 0$ such that the quadratic polynomial $M - \|x\|^2$ satisfies

$$M - \|x\|^2 \in Q_k(g)$$

for some k.

Then the following result holds:

Theorem (Putinar 1993)

For $\Omega = \{x \mid g_j(x) \geq 0,\ j = 1, \ldots, m\} \subset \mathbb{R}^n$ with associated Archimedean quadratic module $Q(g)$:

(a) *If a polynomial $f \in \mathbb{R}[x]$ is strictly positive on Ω then $f \in Q(g)$.*

(b) *The pseudo-moments $\{y_\alpha\}_{\alpha \in \mathbb{N}^n}$ have a representing measure on Ω if and only if $\mathbf{M}_d(\mathbf{y}) \succeq 0$ and $\mathbf{M}_d(g_j \mathbf{y}) \succeq 0$ for $j = 1, \ldots, m$, and for all $d \in \mathbb{N}$.*

Sum-of-squares restriction of problem (2)

We rewrote $f^\star = \inf_{x\in\mathbb{R}^n} f(x)$ s. t. $x \in \Omega$ as

$$(2) \qquad f^\star = \sup_{\lambda\in\mathbb{R}} \lambda \quad \text{s. t. } f(x) - \lambda \geq 0,\ \forall x \in \Omega$$

The level-d **SOS restriction of problem** (2)

$$(\mathbf{D}_d) \qquad \delta_d = \sup_{\lambda\in\mathbb{R},\sigma_0,\sigma_1,\ldots,\sigma_m} \lambda$$

$$\text{s. t.} \quad f - \lambda = \sigma_0 + \sum_{j=1}^{m} \sigma_j g_j, \qquad \text{(D)}$$

$$\sigma_0 \in \Sigma[x]_d \qquad \text{(E)}$$

$$\sigma_j \in \Sigma[x]_{d-d_j} \quad j = 1, \ldots, m \qquad \text{(F)}$$

where $d_j = \lceil \deg(g_j)/2 \rceil$.

(D) is enforced as $s(2d)$ scalar constraints equating monomial coefficients on either side. The variable λ contributes to the "1" monomial for this purpose.

(E) and (F) ensure each multiplier polynomial is SOS; implemented as LMIs.

$\Rightarrow$ **Problem $(\mathbf{D}_d)$ is a tractable SDP restriction of** (2).

Properties of problems $(\mathbf{P}_d)$ and $(\mathbf{D}_d)$

- Both problems are LMIs and are therefore convex and in principle tractable e.g. with MOSEK, Sedumi, etc.
- They form a primal-dual pair for given d.
- From weak duality we know $\delta_d \leq \rho_d$, and we also know $\rho_d \leq f^\star$.

Theorem (Lasserre 2000)

Let Ω be compact and the associated $Q(g)$ Archimedean. Then the following hold:

(i) *As $d \to \infty$, $\rho_d \nearrow f^\star$ and $\delta_d \nearrow f^\star$.*

(ii) *If, for some d, the primal-optimal solution $\mathbf{y}^d$ satisfies*

$$\operatorname{rank}(\mathbf{M}_d(\mathbf{y}^d)) = \operatorname{rank}(\mathbf{M}_{d-s}(\mathbf{y}^d)),$$

where $s = \max_j d_j = \max_j \lceil \deg(g_j)/2 \rceil$, then $\rho_d = f^\star$. If t is the rank obtained in the above, there are t global minimizers $x_1^\star, \dots, x_t^\star \in \Omega$.

The global solutions $x_1^\star, \dots, x_t^\star$ can be extracted from $\mathbf{y}^d$ using a linear algebra routine.

Properties of problems $(\mathbf{P}_d)$ and $(\mathbf{D}_d)$

- Lasserre's theorem does not guarantee that the rank condition is ever satisfied for any finite d. But in practice, convergence at a "small" value of d is typical.

Theorem (Nie 2014)

Assume Ω is compact, the associated $Q(g)$ Archimedean, and that for each global solution $x^\star$ the following technical conditions hold:

(i) *The gradients $\nabla g_1(x^\star), \ldots, \nabla g_m(x^\star)$ are linearly independent;*

(ii) *Strict complementarity holds: $g_j(x^\star) = 0 \implies \lambda_j^\star > 0$;*

(iii) *Hessian of the Lagrangian is strictly positive definite:*

$$u^\top \left[\nabla_x^2 \left(f(x^\star) - \sum_{j=1}^m \lambda_j^\star g_j(x^\star) \right) \right] u > 0$$

for all $0 \neq u \in \nabla(f(x^\star) - \sum_{j=1}^m \lambda_j^\star g_j(x^\star))^\perp$.

Then $f - f^\star \in Q(g)$, i.e., there exists a hierarchy level $d^\star \in \mathbb{N}$, and associated SOS polynomials $\sigma_0^\star \in \Sigma[x]_{d^\star}$, $\sigma_1^\star \in \Sigma[x]_{d^\star - d_1}$, ..., $\sigma_m^\star \in \Sigma[x]_{d^\star - d_m}$ such that

$$f(x) - f^\star = \sigma_0^\star(x) + \sum_{j=1}^m \sigma_j^\star(x) g_j(x) \quad \forall x \in \mathbb{R}^n.$$

Further properties of the Moment-SOS hierarchy

- Nie's conditions can be seen as **analogous to the KKT conditions** from convex optimization, but extend to non-convex polynomial problems.
  - The SOS polynomials $\sigma_j(x)$ play the role of the usual Lagrange multipliers λ_j for the constraints $-g_j(x) \leq 0$.
  - Non-trivial σ_j indicates constraint j makes a difference to the value of $f^\star$, even if not active at $x^\star$! This cannot happen in convex optimization with conventional scalar multipliers λ_j.
- Polynomial problems of degree d can be described[5] by a point in $\mathbb{R}^{(m+1)s(d)}$. Problems satisfying Nie's conditions are **dense** in this space.
- For **SOS-convex problems**[6], the Moment-SOS hierarchy attains $f^\star$ at the first legal value of d.
  - Attractive, because otherwise the method would be disadvantageous for "easy" problems.
  - If convex but not SOS-convex, then convergence is still finite – as long as $\nabla^2 f(x^\star) \succ 0$ for every global minimizer $x^\star$ – but not guaranteed to occur at the first legal d.

[5] This is because each function $f, g_1, \ldots, g_m$ is fully described by $s(d)$ monomial coefficients.

[6] In an SOS-convex problem, the functions $f, -g_1, \ldots, -g_m$ are SOS-convex polynomials, meaning $\nabla^2 f(x) = L(x)L(x)^\top$ for some $L \in \mathbb{R}[x]^{n \times p}$.

Notes and references

Toolboxes

YALMIP functionality for SOS and Moment Relaxations:
https://yalmip.github.io/tutorial/sumofsquaresprogramming/
https://yalmip.github.io/tutorial/momentrelaxations/

Gloptipoly: http://www.laas.fr/~henrion/software/gloptipoly

Literature

J.-B. Lasserre "The Moment-SOS Hierarchy," 2018, https://arxiv.org/abs/1808.03446. This lecture covered Sections 1 and 2.

J.-B. Lasserre, "An Introduction to Polynomial and Semi-Algebraic Optimization," *Cambridge University Press*, 2015

M. Putinar, "Positive polynomials on compact semi-algebraic sets," *Indiana Univ. Math. J.*, vol. 42, 1993.

J. Nie, "Optimality Conditions and Finite Convergence of Lasserre's Hierarchy," *Math. Program. Ser. A*, vol. 146, 2014.