Moments for polynomial optimization - An illustrated tutorial

Draft lecture notes of October 10, 2023

Didier Henrion

LAAS-CNRS, Toulouse, France
Czech Tech. Univ. Prague, Czechia

These are lecture notes for a mini-course course given in October 2023 at Institut Henri Poincaré, Paris, in the scope of the research programme "Recent trends in computer algebra". I am especially grateful to Mohab Safey El Din for giving me the opportunity to present this material.

The objective of these lecture notes is to collect in an informal manner a certain number of elementary technical concepts and illustrative examples relevant for understanding the mathematical background of the moment-SOS hierarchy approach to polynomial optimization.

Polynomial optimization consists of minimizing a polynomial of many real variables subject to polynomial equality and inequality constraints. The moment-SOS hierarchy is an approach to polynomial optimization that solves it globally at the price of solving a family of convex optimization problems of increasing size.

I hope that the lecture notes can serve as a gentle introduction to a fascinating but sometimes technically difficult branch of applied mathematics.

Contents

Reader's guide

These lecture notes are conceived as an informal introduction to the moment-SOS hierarchy approach to polynomial optimization. Each key concept and idea is illustrated by an example and sometimes a picture. When a notion appears for the first time, it is emphasized in italic and indexed at the end of the document. Since priority is given to readability and accessibility, sometimes at the price of mathematical rigor and more complicated notions and notations, the reader interested in details and precise mathematical statements is referred to the notes and citations ending each chapter.

Acknowledgments

These notes benefited from feedback by Jérémy Berthomieu, Saroj Chhatoi, Marouan Handa, Carine Jauberthie, Jean Bernard Lasserre, Jakub Mareček, Simone Naldi, Edouard Pauwels, Mohab Safey El Din and Bruno Salvy.

Notations

We tried to stick as much as possible with the following conventions for mathematical notations: lowercase roman for scalars, lowercase bold for vectors, uppercase bold for matrices, math script for sets, except math blackboard for the set of real numbers, vectors and symmetric matrices.

Chapter 1

The moment cone and its dual

The purpose of this chapter is to introduce two convex cones in duality: the positive polynomial cone and the moment cone.

1.1 Positive polynomials

Let $\mathbb{R}[\mathbf{x}]_d$ denote the vector space of polynomials of degree at most d in the vector of indeterminates $\mathbf{x} \in \mathbb{R}^n$. Its dimension is the binomial coefficient $n_d := \frac{(n+d)!}{n!d!}$. The vector space $\mathbb{R}[\mathbf{x}]_d$ can be indexed by $\mathbb{N}^n_d := \{\mathbf{a} \in \mathbb{N}^n : \sum_{i=1}^n a_i \leq d\}$. Let $\mathbf{b}(\mathbf{x}) := (b_\mathbf{a}(\mathbf{x}))_{\mathbf{a}\in\mathbb{N}^n_d} \in \mathbb{R}[\mathbf{x}]^{n_d}$ denote a basis for this space, with the convention that $b_\mathbf{0}(\mathbf{x}) := 1$, so that every element $p \in \mathbb{R}[\mathbf{x}]_d$ can be expressed as a linear combination

$$p(\mathbf{x}) = \sum_{\mathbf{a}\in\mathbb{N}^n_d} p_\mathbf{a} b_\mathbf{a}(\mathbf{x}) = \mathbf{p}^T \mathbf{b}(\mathbf{x}) \tag{1.1}$$

with coefficient vector[1] $\mathbf{p} := (p_\mathbf{a})_{\mathbf{a}\in\mathbb{N}^n_d} \in \mathbb{R}^{n_d}$.

A notationally convenient choice of basis are *monomials*:

$$b_\mathbf{a}(\mathbf{x}) = \mathbf{x}^\mathbf{a} := \prod_{i=1,\dots,n} x_i^{a_i}.$$

Example 1.1. *Let $n = 1$, $d = 4$ and $p(x) = 3 - 2x^2 + x^4$. Then $n_d = 5$ and $p(x)$ has coefficient vector $\mathbf{p} = (p_0, p_1, p_2, p_3, p_4) = (3, 0, -2, 0, 1) \in \mathbb{R}^5$ in the monomial basis $\mathbf{b}(x) = (1, x, x^2, x^3, x^4) = (x^a)_{a=0,1,\dots,4}$.*

Example 1.2. *The monomial $p(x) = x^{10}$ has coefficient vector $\mathbf{p} = (p_a)_{a=0,1,\dots,10} = (126, 0, 210, 0, 120, 0, 45, 0, 10, 0, 1)/512 \in \mathbb{R}^{10}$ in the* Chebyshev polynomial *basis $\mathbf{b}(x) = (t_a(x))_{a=0,1,\dots,10}$ defined by $t_0(x) := 1$, $t_1(x) := x$, $t_{a+1}(x) := 2xt_a(x) - t_{a-1}(x)$, $a \geq 1$.*

Example 1.3. *The* Motzkin polynomial *$p(\mathbf{x}) := 1 + x_1^2x_2^2(x_1^2 + x_2^2 - 3)$ has $n = 2$ variables, degree $d = 6$ and nonzero coefficients $p_{00} = p_{42} = p_{24} = 1$, $p_{22} = -3$ in a monomial basis of dimension $n_d = 28$.*

[1] $\mathbf{p}^T$ denotes the row vector transpose of column vector $\mathbf{p}$.

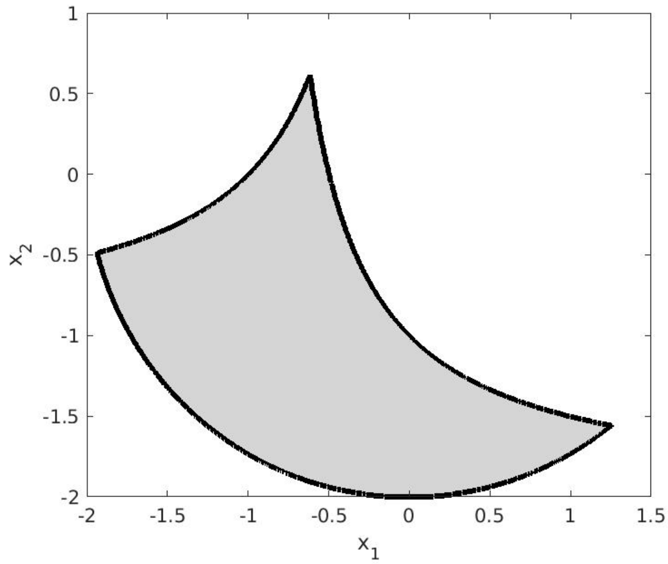


Figure 1.1: A basic semialgebraic set defined by quadratic inequalities

Given a vector $\mathbf{g}(\mathbf{x}) = (g_k(\mathbf{x}))_{k=1,\ldots,m} \in \mathbb{R}[\mathbf{x}]^m$, the set

$$\mathscr{X} := \{\mathbf{x} \in \mathbb{R}^n : g_k(\mathbf{x}) \geq 0, k = 1, \ldots, m\} \tag{1.2}$$

is called *basic semialgebraic*. It is defined as the intersection of finitely many sets defined by polynomial inequalities. More generally, a *semialgebraic set* is defined as a union of finitely many basic semialgebraic sets.

Note that when defining $\mathscr{X}$ in (1.2) we used polynomial inequalities. By choosing polynomials of opposite signs, e.g. $g_{k+1} := -g_k$ for some k, we can however also model polynomial equations of the kind $g_k(\mathbf{x}) = 0$. A set defined by finitely many polynomial equations on the real numbers is called a real *algebraic set* or real *algebraic variety*.

Example 1.4. *The Euclidean ball of radius R*

$$\mathscr{B}_R := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T\mathbf{x} \leq R^2\}$$

is an elementary example of basic semialgebraic set $\mathscr{X}$ that can be defined as in (1.2) *with $m = 1$ and $g_1(\mathbf{x}) = R^2 - \mathbf{x}^T\mathbf{x}$. Now if $m = 2$ and $g_1(\mathbf{x}) = -g_2(\mathbf{x}) = R^2 - \mathbf{x}^T\mathbf{x}$ the set $\mathscr{X} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T\mathbf{x} = R^2\}$ defined in* (1.2) *is the sphere, the boundary of the ball, an elementary example of a real algebraic variety.*

Example 1.5. *If $n = 2$, $m = 3$ with $g_1(\mathbf{x}) = 4-x_1^2-x_2^2$, $g_2(\mathbf{x}) = -1-2x_1-x_2-x_1x_2$, $g_3(\mathbf{x}) = 1 + x_1 + x_1x_2$, the set $\mathscr{X}$ defined in* (1.2) *is represented on Figure 1.1 in gray. Its boundary (thick black) consists of a circular arc (bottom) and two hyperbolic branches (top).*

Example 1.6. *If $n = 3$, $m = 2$, $g_1(\mathbf{x}) = 1 - x_1^2 - x_2^2 - x_3^2 + 2x_1x_2x_3$, $g_2(\mathbf{x}) = 3 - x_1^2 - x_2^2 - x_3^2$, the set $\mathscr{X}$ has the pillow shape of Figure 1.2. Note that by construction $\mathscr{X} \subset \mathscr{B}_{\sqrt{3}}$.*

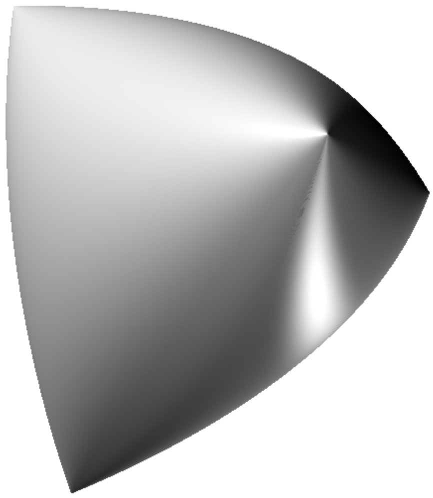


Figure 1.2: The pillow as a basic semialgebraic set

The set of polynomials of degree up to d that are positive[2] on $\mathscr{X}$ is denoted by

$$\mathscr{P}_d(\mathscr{X}) := \{p \in \mathbb{R}[\mathbf{x}]_d : p(\mathbf{x}) \geq 0, \ \forall \mathbf{x} \in \mathscr{X}\}.$$

It can be checked that this set is a *convex*[3] *cone*[4]. It is called the *positive polynomial cone*.

Example 1.7. *The Motzkin polynomial of Example 1.3 belongs to* $\mathscr{P}_6(\mathbb{R}^2)$*, i.e. it is globally positive, see Figure 1.3 for a logarithmic scale representation. Observe that the polynomial is zero at the 4 points* $x_1 = \pm 1, x_2 = \pm 1$*, and it tends to one from above along the 4 asymptotes* $x_1 = 0, x_2 \to \pm\infty$ *and* $x_1 \to \pm\infty, x_2 = 0$.

Example 1.8. *The polynomial* g_1 *defining the pillow* $\mathscr{X}$ *of Example 1.6 belongs to* $\mathscr{P}_6(\mathscr{X})$ *by construction, but it does not belong to* $\mathscr{P}_6(\mathscr{B}_{\sqrt{3}})$ *since it is negative e.g. at the point* $\mathbf{x} = (-1, 1, 1) \in \mathscr{B}_{\sqrt{3}}$.

1.2 Moments

Now consider a functional acting linearly on $\mathbb{R}[x]_d$. In basis $\mathbf{b}$, such a functional can be represented with a vector $\mathbf{y} \in \mathbb{R}^{n_d}$:

$$\begin{array}{rrcl} \ell_{\mathbf{y}} : & \mathbb{R}[\mathbf{x}]_d & \to & \mathbb{R} \\ & p(\mathbf{x}) = \mathbf{p}^T\mathbf{b}(\mathbf{x}) & \mapsto & \ell_{\mathbf{y}}(p) = \mathbf{p}^T\mathbf{y}. \end{array} \tag{1.3}$$

Informally, $\ell_{\mathbf{y}}$ linearizes polynomials. The space of vectors $\mathbf{y}$ is *dual*[5] to $\mathbb{R}[\mathbf{x}]_d$, and we denote it by $\mathbb{R}[\mathbf{x}]_d^\star$.

[2]By positive we mean non-negative, i.e. the value zero is allowed.
[3]A set is convex if it contains the line segment joining any two points.
[4]A set is a cone if is closed under positive scalar multiplication.
[5]The dual to a vector space is the set of its (bounded) linear functionals.

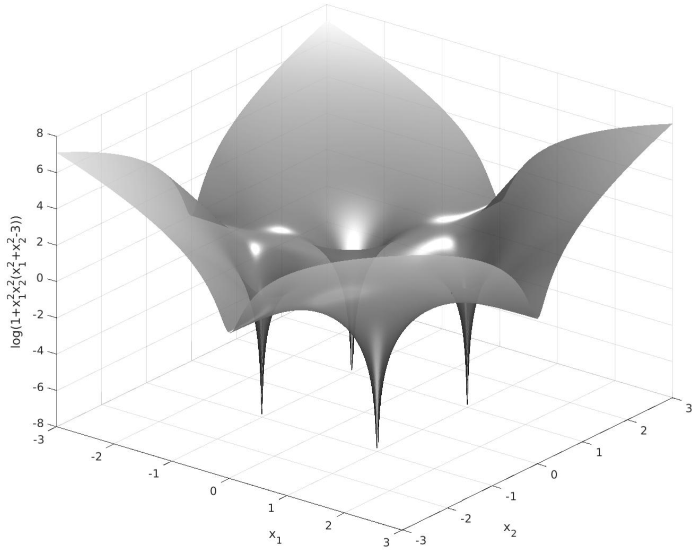


Figure 1.3: The logarithm of the Motzkin polynomial

Example 1.9. *In the monomial basis, the linear functional* $\ell_{\mathbf{y}}$ *acting on the polynomial* p *of Example 1.1 writes* $\ell_{\mathbf{y}}(p) = 3y_0 - 2y_2 + y_4$.

Example 1.10. *In the monomial basis, the linear functional* $\ell_{\mathbf{y}}$ *acting on the Motzkin polynomial* p *of Example 1.3 writes* $\ell_{\mathbf{y}}(p) = y_{00} - 3y_{22} + y_{42} + y_{24}$.

Now let us give some more examples of linear functionals.

Example 1.11. *Evaluation at a given point* $\mathbf{x}$*:* $p \mapsto p(\mathbf{x})$. *In the monomial basis, this functional writes* $\ell_{\mathbf{y}}(p) = \sum_{\mathbf{a}} p_{\mathbf{a}} \mathbf{x}^{\mathbf{a}}$, *i.e.* $\mathbf{y} := (\mathbf{x}^{\mathbf{a}})_{\mathbf{a}}$. *If* $n = 1$, $d = 4$ *and* $x = -2$, *we have* $\ell_{\mathbf{y}}(p) = p_0 - 2p_1 + 4p_2 - 8p_3 + 16p_4$.

Example 1.12. *Evaluation at several points* $(\mathbf{x}_k)_k$ *with weights* $(w_k)_k$*:* $p \mapsto \sum_k w_k p(\mathbf{x}_k)$. *The number of points is not necessarily finite. Sometimes we assume that the weights are non-negative and sum up to one, i.e.* $w_k \geq 0$, $\sum_k w_k = 1$.

Example 1.13. *Integration with respect to a given* measure[6] μ *on a given set* $\mathscr{X}$*:* $p \mapsto \int_{\mathscr{X}} p(\mathbf{x}) d\mu(\mathbf{x})$. *Informally, this linear functional can be interpreted as a continuous counterpart of the discrete weighted point evaluation of Example 1.12. Measure* μ *can be interpreted as a way to distribute the mass on set* $\mathscr{X}$. *Each measure* μ *can be identified to its linear functional*

$$\ell_{\mathbf{y}}(p) = \int_{\mathscr{X}} p(\mathbf{x}) d\mu(\mathbf{x}).$$

If the linear functional is normalized and positive, i.e. $\ell_{\mathbf{y}}(1) = 1$ *and* $\ell_{\mathbf{y}}(p) \geq 0$ *whenever* $p \in \mathscr{P}(\mathbb{R}^n)$, *it is referred to as a* probability measure.

[6] By measure we mean a non-negative Borel regular measure.

Example 1.14. *A classical choice of measure in Example 1.13 is the* Lebesgue *or* uniform measure*, which distributes the mass evenly on a given set $\mathscr{X}$. It is generally denoted* $d\mathbf{x}$ *and the value of the linear functional*

$$\ell_{\mathbf{y}}(p) = \int_{\mathscr{X}} p(\mathbf{x})d\mathbf{x}$$

is the average value or expectation of polynomial p on $\mathscr{X}$. The value $\ell_{\mathbf{y}}(1) = \int_{\mathscr{X}} d\mathbf{x}$ is called the Lebesgue measure or volume of $\mathscr{X}$.

Example 1.15. *Another classical choice of measure in Example 1.13 is the (standard)* Gaussian measure*, which distributes more weight near the origin and decays quickly out of the origin. The constant factor in the corresponding linear functional*

$$\ell_{\mathbf{y}}(p) = (2\pi)^{-\frac{1}{n}} \int_{\mathbb{R}^n} p(\mathbf{x})e^{-\frac{1}{2}\mathbf{x}^T\mathbf{x}}d\mathbf{x}$$

makes it a probability measure.

Example 1.16. *The linear functional of Example 1.12 also corresponds to a probability measure as soon as the weights are non-negative and sum up to one. It is called an* atomic measure*, a particular case of which is the linear functional of Example 1.11 which is called a* point *or* Dirac measure *at* $\mathbf{x}$*, denoted* $\delta_{\mathbf{x}}$.

Since a polynomial is a finite linear combination of basis elements, integration with respect to a measure as in Example 1.13 can be written as a finite sum

$$\ell_{\mathbf{y}}(p) = \int_{\mathscr{X}} p(\mathbf{x})d\mu(\mathbf{x}) = \int_{\mathscr{X}} \left(\sum_{\mathbf{a}} p_{\mathbf{a}} b_{\mathbf{a}}(\mathbf{x})\right) d\mu(\mathbf{x}) = \sum_{\mathbf{a}} p_{\mathbf{a}} \int_{\mathscr{X}} b_{\mathbf{a}}(\mathbf{x})d\mu(\mathbf{x})$$

that can be expressed as

$$\ell_{\mathbf{y}}(p) = \sum_{\mathbf{a}} p_{\mathbf{a}} \int_{\mathscr{X}} b_{\mathbf{a}}(\mathbf{x})d\mu(\mathbf{x}) = \sum_{\mathbf{a}} p_{\mathbf{a}} y_{\mathbf{a}} = \mathbf{p}^T\mathbf{y}$$

consistently with the representation (1.3) upon defining the *moment* of order $\mathbf{a} \in \mathbb{N}^n_d$

$$y_{\mathbf{a}} := \int_{\mathscr{X}} b_{\mathbf{a}}(\mathbf{x})d\mu(\mathbf{x}) = \ell_{\mathbf{y}}(b_a)$$

as the image of the basis element $b_{\mathbf{a}}$ through the linear functional. The *moment vector* is then

$$\mathbf{y} := \int_{\mathscr{X}} \mathbf{b}(\mathbf{x})d\mu(\mathbf{x}) \tag{1.4}$$

where vector integration is meant entrywise.

Example 1.17. *As a follow-up to Example 1.11, let us compute the moments of the Dirac measure at the point $z = -2$ in the monomial basis up to degree* 4*. It holds $y_a = \int x^a \delta_{-2}(dx) = (-2)^a$ so $\mathbf{y} = (y_0, y_1, y_2, y_3, y_4) = (1, -2, 4, -8, 16)$.*

Example 1.18. *As a follow-up to Example 1.14, let us compute the moments of the Lebesgue measure of* $\mathscr{X} = [-1,1]$ *in the monomial basis up to degree* 4*. It holds* $y_a = \int_{-1}^{1} x^a dx = \frac{1^{a+1}-(-1)^{a+1}}{a+1}$ *so* $\mathbf{y} = (y_0, y_1, y_2, y_3, y_4) = (2, 0, \frac{2}{3}, 0, \frac{2}{5})$*.*

Example 1.19. *Since a probability measure is a (non-negative normalized) weighted point evaluation, the* convex hull of a set $\mathscr{X}$*, i.e. the set of convex combinations of its points, can be expressed as the set of first degree moments of its probability measures:*

$$\begin{aligned} \textit{conv}\, \mathscr{X} &:= \left\{ \sum_k w_k x_k : x_k \in \mathscr{X},\ w_k \geq 0,\ \sum_k w_k = 1 \right\} \\ &= \left\{ \int_{\mathscr{X}} \mathbf{x}\, d\mu(\mathbf{x}) : \mu \textit{ probability measure on } \mathscr{X} \right\}. \end{aligned}$$

So far we have seen that integration of polynomials with respect to a given measure μ can be represented by its moment vector $\mathbf{y}$. Now we can ask the converse question: given a vector $\mathbf{y}$, does it represent a measure μ ? This is the celebrated (truncated) *moment problem.*

Moment problem. *Given* $\mathscr{X}$ *and* $\mathbf{y}$*, is there a measure* μ *such that* (1.4) *holds ?*

The set of moments of degree up to d of measures on $\mathscr{X}$ is denoted by

$$\mathscr{M}_d(\mathscr{X}) := \left\{ \mathbf{y} \in \mathbb{R}[\mathbf{x}]_d^\star : \mathbf{y} = \int_{\mathscr{X}} \mathbf{b}(\mathbf{x}) d\mu(\mathbf{x}) \text{ for some measure } \mu \text{ on } \mathscr{X} \right\}. \quad (1.5)$$

It can be checked that this set is a finite-dimensional convex cone. It is called the *moment cone.*

Hence $\mathbf{y} \in \mathscr{M}_d(\mathscr{X})$ if and only if $\mathbf{y}$ represents a measure on $\mathscr{X}$. It turns out that if $\mathscr{X}$ is *compact*[7], then cone $\mathscr{M}_d(\mathscr{X})$ is *dual*[8] to cone $\mathscr{P}_d(\mathscr{X})$.

Theorem 1.1. *If* $\mathscr{X}$ *is compact, it holds*

$$\mathscr{M}_d(\mathscr{X}) = \mathscr{P}_d(\mathscr{X})^\star := \{\mathbf{y} \in \mathbb{R}[\mathbf{x}]_d^\star : \ell_{\mathbf{y}}(p) \geq 0 \textit{ for all } p \in \mathscr{P}_d(\mathscr{X})\}.$$

So if there is a polynomial $p \in \mathscr{P}_d(\mathscr{X})$ such that $\ell_{\mathbf{y}}(p) < 0$ then $\mathbf{y}$ does not belong to $\mathscr{M}_d(\mathscr{X})$ i.e. it does not represent a measure on $\mathscr{X}$.

Example 1.20. *On Figure 1.4 we represent the set of moments* (y_1, y_2, y_3) *in the monomial basis* (x, x^2, x^3) *of all probability measures (i.e.* $y_0 = 1$*) on* $[-1,1]$*. It is the convex hull of the moment curve* $\{(x, x^2, x^3) : x \in [-1,1]\}$ *(thick black), see Example 1.19. The point* $(1,0,0)$ *does not belong to this set, and this can be checked with the polynomial* $p(x) = (1-x)^2 \in \mathscr{P}_2([-1,1])$ *which satisfies* $\ell_{\mathbf{y}}(p) = -1$ *for* $\mathbf{y} = (1,1,0,0)$*.*

[7]A finite-dimensional set is compact if it is bounded and closed.

[8]The dual of a cone is the set of all its (bounded) positive linear functionals.

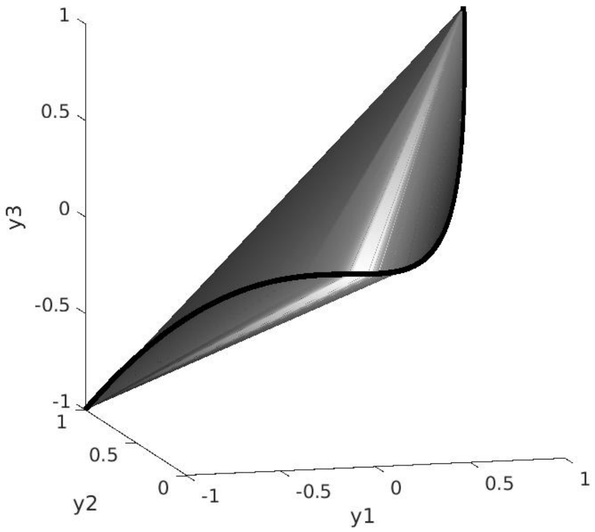


Figure 1.4: The set of moments of probability measures on $[-1,1]$ is the convex hull of the moment curve (thick black)

For compact $\mathscr{X}$, both $\mathscr{P}_d(\mathscr{X})$ and $\mathscr{M}_d(\mathscr{X})$ are convex and closed, and $\mathscr{M}_d(\mathscr{X}) = \mathscr{P}_d(\mathscr{X})^\star$ implies $\mathscr{M}_d(\mathscr{X})^\star = \mathscr{P}_d(\mathscr{X})^{\star\star} = \mathscr{P}_d(\mathscr{X})$. So the positive polynomial cone is dual to the moment cone.

1.3 Notes

Real algebraic geometry is concerned with polynomial equations and inequalities over the real numbers. It is exposed in [1] and [4] with a focus on computer algebra algorithms.

For an account of the numerical benefits of using Chebyshev polynomials over monomials in function approximation, see [38].

Background material on positive polynomials and moments and their dualities can be found in [23, 21]. More advanced material on positive polynomials and the moment problem can be found in [25, 36]. See in particular [36, Ch. 17-18] for the truncated moment problem.

Duality Theorem 1.1 is a finite-dimensional version of the Riesz-Haviland Theorem. A modern statement and proof can be found e.g. in [23, Theorem 5.13].

An accessible introduction to univariate measure theory can be found in [33, Part I]. More advanced topics on multivariate measure theory are covered in [33, Part III].

Chapter 2

Semidefinite relaxations of the moment cone

The solution to the moment problem reduces to finding whether a vector belongs to the moment cone. It turns out however that this problem is difficult in its full generality. In the sequel we describe an approximate solution to the moment problem, in the sense that we construct a family of outer approximations of the moment cone. The approximations are indexed by a relaxation order that controls tightness: the higher the order, the tighter the relaxation. Crucially, for a given order, testing whether a vector belongs to a semidefinite relaxation reduces to semidefinite optimization, a well-studied problem of convex optimization for which efficient algorithms are available.

2.1 Semidefinite cone

Let $\mathbb{S}^m$ denote the Euclidean space of symmetric matrices of $\mathbb{R}^{m\times m}$. Given a real quadratic form[1] $f : \mathbb{R}^m \to \mathbb{R}$, the (unique) matrix $\mathbf{X} \in \mathbb{S}^m$ such that $f(\mathbf{y}) = \mathbf{y}^T\mathbf{X}\mathbf{y}$ for all $\mathbf{y} \in \mathbb{R}^m$ is called the *Gram matrix* of f. We say that a matrix is *positive semidefinite* when it is the Gram matrix of a positive quadratic form. In other words, a matrix $\mathbf{X} \in \mathbb{S}^m$ is positive semidefinite, denoted by $\mathbf{X} \succeq 0$, if and only if $\mathbf{y}^T\mathbf{X}\mathbf{y} \geq 0$, $\forall \mathbf{y} \in \mathbb{R}^m$ or equivalently, if and only if the minimum eigenvalue of $\mathbf{X}$ is non-negative. This last statement makes sense since symmetric matrices have only real eigenvalues.

Example 2.1. *Note that $\mathbf{X} \succeq 0$ is not meant as a nonnegativity constraint on the individual entries of matrix $\mathbf{X}$. If $\mathbf{X}$ is positive semidefinite, then its diagonal entries are positive scalars, but there are positive semidefinite matrices with negative off-diagonal entries, i.e.*

$$\mathbf{X} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

and there are matrices with non-negative entries which are not positive semidefinite,

[1]A quadratic form is a polynomial with all terms of degree exactly two.

i.e.

$$\mathbf{X} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

The set of all positive quadratic forms, or equivalently, or all positive semidefinite matrices, is a cone that we will use systematically in our further developments. It is called the positive semidefinite cone, or just *semidefinite cone* for short, and it is denoted as $\{\mathbf{X} \in \mathbb{S}^m \ : \ \mathbf{X} \succeq 0\}$.

The intersection of the semidefinite cone with a linear subspace is a convex set called a *spectrahedron*. It can be expressed with a *linear matrix inequality* (LMI)

$$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}(\mathbf{x}) := \mathbf{A}_0 + \sum_{i=1}^{n} x_i \mathbf{A}_i \succeq 0\} \tag{2.1}$$

where matrices $\mathbf{A}_i \in \mathbb{S}^m$, $i = 0, 1, \ldots, n$ are given. Note that spectrahedron (2.1) is a cone if and only if $\mathbf{A}_0 = \mathbf{0}$.

Note that an LMI constraint is generally nonlinear, but it is always convex. To prove convexity, rewrite the matrix constraint as a scalar constraint

$$\mathbf{v}^T \mathbf{A}(\mathbf{x}) \mathbf{v} = (\mathbf{v}^T \mathbf{A}_0 \mathbf{v}) + \sum_{i=1}^{n} (\mathbf{v}^T \mathbf{A}_i \mathbf{v}) x_i \geq 0$$

which is linear on $\mathbf{x} \in \mathbb{R}^n$ for all $\mathbf{v} \in \mathbb{R}^m$, hence defining a convex set as an intersection of half-spaces.

Example 2.2. *If all matrices $\mathbf{A}_i$ are diagonal, then the LMI in* (2.1) *reduces to linear inequalities, and the corresponding spectrahedron reduces to a polyhedron.*

Example 2.3. *Consider the spectrahedron*

$$\{\mathbf{x} \in \mathbb{R}^2 : \mathbf{A}(\mathbf{x}) = \begin{pmatrix} 1 + x_1 & x_2 \\ x_2 & 1 - x_1 \end{pmatrix} \succeq 0\}.$$

Matrix $\mathbf{A}(\mathbf{x})$ is positive semidefinite if and only if its trace (sum of eigenvalues) and determinant (product of eigenvalues) are positive. Since trace $\mathbf{A}(\mathbf{x}) = 2$, *it follows that the spectrahedron is the unit ball* $\{\mathbf{x} \in \mathbb{R}^2 : \det \mathbf{A}(\mathbf{x}) = 1 - x_1^2 - x_2^2 \geq 0\} = \mathscr{B}_2$.

Example 2.4. *The spectrahedron defined as the set of vectors $\mathbf{x} \in \mathbb{R}^2$ satisfying*

$$\mathbf{A}(\mathbf{x}) = \begin{pmatrix} 2 - 2x_1 & x_2 & -1 + x_1 \\ x_2 & 1 + x_1 & 0 \\ -1 + x_1 & 0 & 1 \end{pmatrix} \succeq 0$$

is the convex set represented in gray on Figure 2.1. The cubic polynomial $\det \mathbf{A}(\mathbf{x}) = (1 + x_1)^2(1 - x_1) - x_2^2$ *vanishes on the spectrahedron boundary (thick black). Observe that* rank $\mathbf{A}(\mathbf{x}) = 2$ *along this curve, except at the point on the left where* rank $\mathbf{A}((-1, 0)) = 1$.

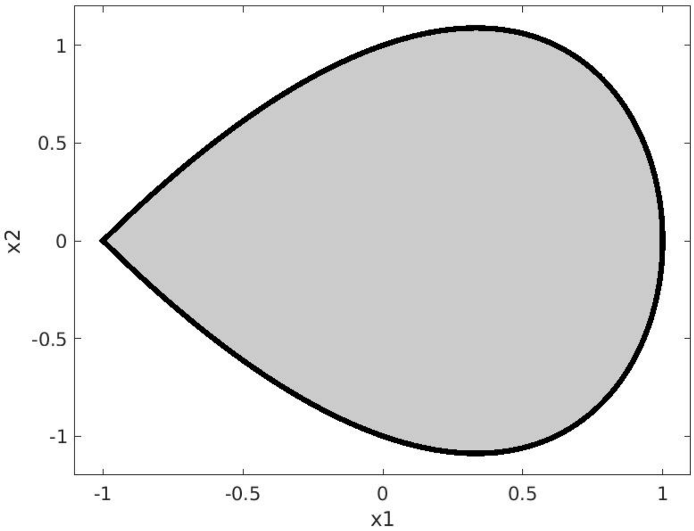


Figure 2.1: Cubic spectrahedron

Example 2.5. *The convex pillow of Example 1.6 is the spectrahedron defined by*

$$\mathbf{A}(\mathbf{x}) = \begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & x_3 \\ x_2 & x_3 & 1 \end{pmatrix} \succeq 0.$$

The cubic polynomial $\det \mathbf{A}(\mathbf{x}) = 1 - x_1^2 - x_2^2 - x_3^2 + 2x_1x_2x_3$ *vanishes on the boundary, where the rank of A is equal to 2, except at 4 distinguished points where it drops down to 1.*

Let us build the univariate polynomial

$$t \mapsto \det\left(t\mathbf{I}_m + \mathbf{A}(\mathbf{x})\right) = \sum_{k=0}^{m} g_{m-k}(\mathbf{x}) t^k \tag{2.2}$$

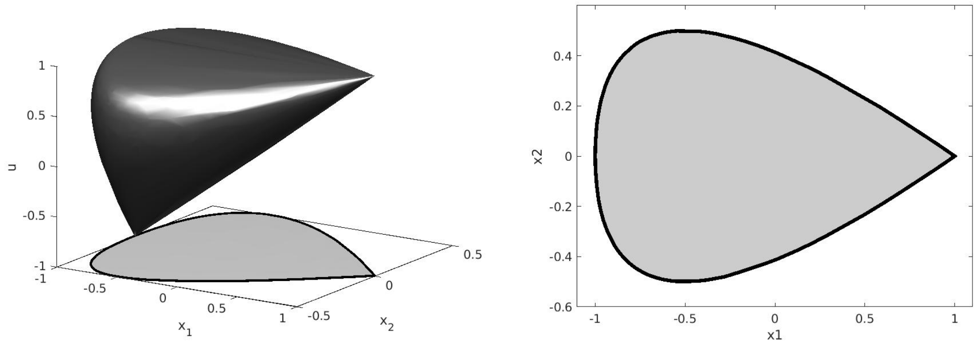
where $\mathbf{I}_m$ denotes the identity matrix of size m, hence $g_0(\mathbf{x}) = 1$. Coefficients $g_k \in \mathbb{R}[\mathbf{x}]$, $k = 1, \ldots, m$ are multivariate polynomials. They are elementary symmetric functions of the eigenvalues of $\mathbf{A}(\mathbf{x})$. For example $g_1(\mathbf{x}) = \operatorname{trace} \mathbf{A}(\mathbf{x})$ and $g_m(\mathbf{x}) = \det \mathbf{A}(\mathbf{x})$.

By construction, the roots of polynomial (2.2) are all real and equal to the eigenvalues of matrix $-\mathbf{A}(\mathbf{x})$. If this matrix has negative eigenvalues, then polynomial (2.2) has all negative roots, and hence all its coefficients $g_k(\mathbf{x})$, $k = 1, \ldots, m$ are positive. The following result follows from this observation.

Lemma 2.1. *A spectrahedron is a convex closed basic semialgebraic set:*

$$\{\mathbf{x} \in \mathbb{R}^n \;:\; \mathbf{A}(\mathbf{x}) \succeq 0\} \;=\; \{\mathbf{x} \in \mathbb{R}^n \;:\; g_k(\mathbf{x}) \geq 0,\; k = 1, \ldots, m\}.$$

Example 2.6. *Returning to Example 2.5, the elementary symmetric functions of the eigenvalues of* $\mathbf{A}(\mathbf{x})$ *are the polynomials* $g_1(\mathbf{x}) = 3$, $g_2(\mathbf{x}) = 3 - x_1^2 - x_2^2 - x_3^2$,



(a) Spectrahedron in space (x_1, x_2, u) (b) Shadow in plane (x_1, x_2)

Figure 2.2: A spectrahedron (left) and its shadow (right)

$g_3(\mathbf{x}) = 1 - x_1^2 - x_2^2 - x_3^2 + 2x_1x_2x_3$. So the convex pillow of Example 1.6 and Figure 1.2 is the intersection of the cubic semialgebraic set $\{\mathbf{x} \in \mathbb{R}^3 : g_3(\mathbf{x}) \geq 0\}$ with the ball $\mathscr{B}_{\sqrt{3}} = \{\mathbf{x} \in \mathbb{R}^3 : g_2(\mathbf{x}) \geq 0\}$.

Since all spectrahedra are convex closed basic semialgebraic, one may then wonder conversely whether all convex closed basic semialgebraic sets are spectrahedra. The answer is negative (for reasons that we do not explain here).

Example 2.7. *The TV-screen set $\{\mathbf{x} \in \mathbb{R}^2 : 1 - x_1^4 - x_2^4 \geq 0\}$ is convex closed basic semialgebraic but it is not a spectrahedron.*

Consequently, in order to represent convex closed basic semialgebraic sets, we have to go beyond spectrahedra. Let us consider *spectrahedral shadows*, which are projections of spectrahedra:

$$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}_0 + \sum_{i=1}^{n} x_i \mathbf{A}_i + \sum_{j=1}^{N} u_j \mathbf{B}_j \succeq 0 \text{ for some } \mathbf{u} \in \mathbb{R}^N\}. \tag{2.3}$$

Spectrahedral shadows are constructed by introducing new variables $\mathbf{u}$ called liftings.

Example 2.8. *On Figure 2.2 we represent a spectrahedron (left) and its planar shadow (right). The shadow is modeled with one lifting variable:*

$$\{\mathbf{x} \in \mathbb{R}^2 : \begin{pmatrix} 1 + x_1 & x_2 & 0 \\ x_2 & -x_1 + u & -x_2 \\ 0 & -x_2 & 1 - u \end{pmatrix} \succeq 0, \textit{ for some } u \in \mathbb{R}\}.$$

It can be shown that the shadow is the union of an ellipse and a triangle. Hence it is a non-basic semi-algebraic set, and as such it cannot be a spectrahedron.

Example 2.9. *A spectrahedral shadow representation of the TV-screen set of Ex-*

ample 2.7 is the set of vectors $\mathbf{x} \in \mathbb{R}^2$ *such that*

$$\left(\begin{array}{cc|cc|cc} 1+u_1 & u_2 & 0 & 0 & 0 & 0 \\ u_2 & 1-u_1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & x_1 & 0 & 0 \\ 0 & 0 & x_1 & u_1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & x_2 \\ 0 & 0 & 0 & 0 & x_2 & u_2 \end{array}\right) \succeq 0$$

for some $u \in \mathbb{R}^2$.

Finding whether a given vector belongs to a spectrahedral shadow, or the (slightly) more general problem of minimizing a linear function within a spectrahedral shadow, reduces to *semidefinite optimization*, a well-studied class of convex optimization problems for which efficient numerical algorithms based on interior-point methods are available.

Since spectrahedral shadows are convex closed semialgebraic sets, one may then wonder conversely whether all convex closed semialgebraic sets are spectrahedral shadows. Whereas this is true for planar sets ($n = 2$) and convex hulls of one-dimensional semi-algebraic sets in $\mathbb{R}^n$ for any n, in general the answer is negative (for reasons that we do not explain here).

Example 2.10. *The bivariate sextic moment cone* $\mathscr{M}_6(\mathbb{R}^2)$ *is not a spectrahedral shadow.*

This motivates us to study spectrahedral shadows as approximations to convex semialgebraic sets, rather than exact representations.

2.2 Polynomial sums of squares (SOS)

An even degree polynomial $p(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]_{2d}$ is a sum of squares (SOS) if it can be expressed as a finite sum $p(\mathbf{x}) = \sum_{i=1}^N q_i^2(\mathbf{x})$ for some distinct $q_i(\mathbf{x}) = \mathbf{q}_i^T \mathbf{b}(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]_d$, $i = 1, \ldots, N$. Hence

$$p(\mathbf{x}) = \sum_{i=1}^N q_i^2(\mathbf{x}) = \sum_{i=1}^N (\mathbf{q}_i^T \mathbf{b}(\mathbf{x}))^2 = \sum_{i=1}^N \mathbf{b}^T(\mathbf{x}) \mathbf{q}_i \mathbf{q}_i^T \mathbf{b}(\mathbf{x}) = \mathbf{b}^T(\mathbf{x}) \mathbf{P} \mathbf{b}(\mathbf{x}) \qquad (2.4)$$

where $\mathbf{P} := \sum_{i=1}^N \mathbf{q}_i \mathbf{q}_i^T \in \mathbb{S}^{n_d}$ is a positive semidefinite matrix of rank $N \leq n_d$, the Gram matrix of p expressed as a quadratic function of the basis $\mathbf{b}$. Given p, finding whether it is SOS amounts to finding whether it has a positive semidefinite Gram matrix $\mathbf{P}$ whose entries are linearly related to p. In other words, the set of SOS polynomials is a spectrahedral shadow.

Example 2.11. *Let* $n = 1$, $d = 1$, $p(x) = 1 + x + x^2$. *Then there is a unique Gram matrix representing* p *in the monomial basis, namely*

$$\mathbf{P} = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} = \frac{3}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}^T + \frac{1}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}^T$$

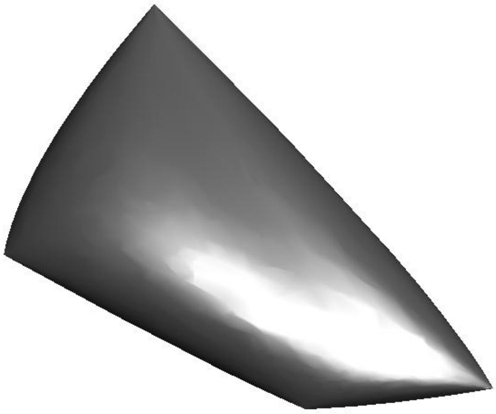


Figure 2.3: Gram spectrahedron of $1 + x + x^2 + x^3 + x^4 + x^5 + x^6$

from which it follows that $p(x) = (\frac{\sqrt{6}}{2}(1+x))^2 + (\frac{\sqrt{2}}{2}(1-x))^2 = q_1^2(x) + q_2^2(x)$.

Example 2.12. *Let* $n = 1$, $d = 2$, $p(x) = 1 + x + x^2 + x^3 + x^4$. *Then the set of Gram matrices representing* p *in the monomial basis is described by the LMI*

$$\mathbf{P}(u) = \begin{pmatrix} 1 & \frac{1}{2} & -u \\ \frac{1}{2} & 1+2u & \frac{1}{2} \\ -u & \frac{1}{2} & 1 \end{pmatrix} \succeq 0$$

where u *is a real lifting. It turns out that* $\mathbf{P}(u) \succeq 0$ *if and only if* u *belongs to the interval of eigenvalues of* $\mathbf{P}(u)$ *containing the origin.*

Example 2.13. *Let* $n = 1$, $d = 2$, $p(x) = 1 + x + x^2 + x^3 + x^4 + x^5 + x^6$. *Then the set of Gram matrices representing* p *in the monomial basis is described by the LMI*

$$\mathbf{P}(\mathbf{u}) = \begin{pmatrix} 1 & \frac{1}{2} & -u_1 & -u_2 \\ \frac{1}{2} & 1+2u_1 & \frac{1}{2}+u_2 & -u_3 \\ -u_1 & \frac{1}{2}+u_2 & 1+2u_3 & \frac{1}{2} \\ -u_2 & -u_3 & \frac{1}{2} & 1 \end{pmatrix} \succeq 0$$

where $\mathbf{u} \in \mathbb{R}^3$ *is a lifting vector. The corresponding spectrahedron is represented on Figure 2.3. In particular the choice* $\mathbf{u} = \frac{1}{2}(1, 2, 1)$ *corresponds to an SOS representation with* rank $\mathbf{P}(\mathbf{u}) = 2$ *terms.*

2.3 Approximations of the positive polynomial cone

Let us denote by $\Sigma[\mathbf{x}]_{2d} \subset \mathbb{R}[\mathbf{x}]_{2d}$ the cone of SOS polynomials of degree at most $2d$.

If a polynomial is SOS then clearly it is positive, i.e. $\Sigma[\mathbf{x}]_{2d} \subset \mathscr{P}(\mathbb{R}^n)_{2d}$. One may then wonder conversely whether all positive polynomials are SOS. The answer is negative, except in three special cases.

Theorem 2.1. $\Sigma[\mathbf{x}]_{2d} = \mathscr{P}(\mathbb{R}^n)_{2d}$ *if and only if* $n = 1$ *or* $d = 1$ *or* $n = d = 2$.

Example 2.14. *The most famous and simplest (*$n = 2$, $d = 3$*) example of a polynomial which is positive but not SOS is the Motzkin polynomial of Example 1.3.*

On a compact basic semialgebraic set

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

we can however approximate $\mathscr{P}(\mathscr{X})_d$ as closely as desired with the help of SOS cones of increasing degrees. Let us now describe this construction.

For convenience let $g_0(\mathbf{x}) := 1$ and without loss of generality, suppose that $g_1(\mathbf{x}) := R^2 - \mathbf{x}^T\mathbf{x}$ for R sufficienty large[2]. Let[3]

$$r_{\mathbf{g}} := \max_{k=0,1,\ldots,m} \left\lceil \frac{\deg g_k}{2} \right\rceil, \quad r_{\min} := \max\left(r_{\mathbf{g}}, \left\lceil \frac{\deg p}{2} \right\rceil\right) \tag{2.5}$$

and given $r \in \mathbb{N}$ such that $r \geq r_{\mathbf{g}}$, let

$$r_k := \left\lceil r - \frac{\deg g_k}{2} \right\rceil.$$

Define the *truncated quadratic module*

$$\mathscr{Q}_{d,r}(\mathbf{g}) := \{p \in \mathbb{R}[\mathbf{x}]_d : p = \sum_{k=0}^{m} s_k g_k, \;\; s_k \in \Sigma[\mathbf{x}]_{2r_k}, \; k = 0, 1, \ldots, m\}$$

a convex cone which is a spectrahedral shadow, with lifting variables the SOS polynomials s_k (of degree at most $2r_k$) and their corresponding Gram matrices.

On the one hand, observe that necessarily $2r \geq d$ since it is not possible to represent a degree d polynomial as a sum of polynomials of smaller degrees. On the other hand, observe that when $2r > d$ there is a cancellation of all terms of degrees $d+1, \ldots, 2r$ when summing up the polynomials $s_k g_k$.

Observe that the notation $\mathscr{Q}_{d,r}(\mathbf{g})$ emphasizes that the truncated quadratic module depends on the polynomials $\mathbf{g}$ defining $\mathscr{X}$, rather than on the intrinsic geometry of $\mathscr{X}$. If the same set is described by two different sets of polynomials, the respective truncated quadratic modules may differ.

[2]Since $\mathscr{X}$ is bounded, it is included in a ball of sufficiently large radius R.

[3]The notation $\lceil a \rceil$ stands for the smallest integer not smaller than $a \in \mathbb{R}$.

By construction it holds

$$\mathscr{Q}_{d,r}(\mathbf{g}) \subset \mathscr{Q}_{d,r+1}(\mathbf{g}) \subset \mathscr{P}_d(\mathscr{X}). \tag{2.6}$$

Moreover, we have the following approximation result which states that the positive polynomial cone can be approximated from inside arbitrarily well[4] with quadratic modules.

Theorem 2.2. *If $\mathscr{X}$ is compact, it holds*

$$\overline{\mathscr{Q}_{d,\infty}(\mathbf{g})} = \mathscr{P}_d(\mathscr{X}).$$

Example 2.15. *Consider again the Motzkin polynomial $p(\mathbf{x}) := 1+x_1^2x_2^2(x_1^2+x_2^2-3)$ of Examples 1.3 and 2.14. When $R = 1$, $m = 1$ with $g_1(\mathbf{x}) = 1 - x_1^2 - x_2^2$ i.e. $\mathscr{X} = \mathscr{B}_1$, it holds $p \in \mathscr{Q}_{6,3}(\mathbf{g})$ with the SOS liftings $s_0(\mathbf{x}) = (x_1(1-x_2^2))^2 + (x_2(1-x_1^2))^2 + (x_1x_2)^2$ and $s_1(\mathbf{x}) = 1$, i.e. $p(\mathbf{x}) = s_0(\mathbf{x}) + s_1(\mathbf{x})(1 - x_1^2 - x_2^2)$.*

2.4 Approximations of the moment cone

Since the truncated quadratic module $\mathscr{Q}_{d,r}(\mathbf{g})$ is an inner approximation of the positive polynomial cone $\mathscr{P}_d(\mathscr{X})$, by duality it follows that the dual truncated quadratic module

$$\mathscr{R}_{d,r}(\mathbf{g}) := \mathscr{Q}^\star_{d,r}(\mathbf{g})$$

is an outer approximation, or relaxation of the moment cone.

Dualizing inclusions (2.6) and recalling Theorem 1.1, we obtain embedded relaxations for the moment cone

$$\mathscr{R}_{d,r}(\mathbf{g}) \supset \mathscr{R}_{d,r+1}(\mathbf{g}) \supset \mathscr{M}_d(\mathscr{X}) \tag{2.7}$$

and the dual to Theorem 2.2 which states that the moment cone can be approximated from outside arbitrarily well with dual quadratic modules.

Theorem 2.3. *If $\mathscr{X}$ is compact, it holds*

$$\overline{\mathscr{R}_{d,\infty}(\mathbf{g})} = \mathscr{M}_d(\mathscr{X}).$$

Now we explicitly construct the dual truncated quadratic modules as spectrahedral shadows. By definition

$$\mathscr{R}_{d,r}(\mathbf{g}) = \{\mathbf{y} \in \mathbb{R}[\mathbf{x}]^\star_d : \ell_{\mathbf{y}}(p) \geq 0,\ \forall p \in \mathscr{Q}_{d,r}(\mathbf{g})\}$$

[4] The horizontal bar over a set means closure.

where the linear functional $\ell_{\mathbf{y}}(p)$ was defined in (1.3). Then it follows that

$$\begin{array}{rcl}\mathscr{R}_{d,r}(\mathbf{g}) & = & \{\mathbf{y}\in\mathbb{R}[\mathbf{x}]_d^\star : \ell_{\mathbf{y}}(\sum_{k=0}^m s_k g_k)\geq 0,\ \forall s_k\in\Sigma[\mathbf{x}]_{2r_k},\ k=0,1,\ldots,m\}\\ & = & \{\mathbf{y}\in\mathbb{R}[\mathbf{x}]_d^\star : \ell_{\mathbf{y}}(s_k g_k)\geq 0,\ \forall s_k\in\mathbb{R}[\mathbf{x}]_{2r_k},\ k=0,1,\ldots,m\}\\ & = & \{\mathbf{y}\in\mathbb{R}[\mathbf{x}]_d^\star : \ell_{\mathbf{y}}(q^2 g_k)\geq 0,\ \forall q\in\mathbb{R}[\mathbf{x}]_{r_k},\ k=0,1,\ldots,m\}\\ & = & \{\mathbf{y}\in\mathbb{R}[\mathbf{x}]_d^\star : \mathbf{M}_{r_k}(g_k\mathbf{y})\succeq 0,\ k=0,1,\ldots,m\}\end{array}$$

where matrix $\mathbf{M}_{r_k}(g_k\mathbf{y})$ represents the quadratic form $q\mapsto\ell_{\mathbf{y}}(g_k q^2)$ in basis $\mathbf{b}$, i.e.

$$\mathbf{M}_{r_k}(g_k\mathbf{y}):=\ell_{\mathbf{y}}(g_k\mathbf{b}\mathbf{b}^T) \tag{2.8}$$

with the linear functional acting entrywise on matrices. The notation $\mathbf{M}_{r_k}(g_k\mathbf{y})$ reflects the fact that the matrix depends linearly on g_k (for given $\mathbf{y}$) and also linearly on $\mathbf{y}$ (for given g_k).

The set of vectors $\mathbf{y}$ such that the symmetric linear matrix $\mathbf{M}_{r_k}(g_k\mathbf{y})$ is positive semidefinite describes a spectrahedron. This is also the case for

$$\mathscr{R}_{d,r}(\mathbf{g})=\{\mathbf{y}\in\mathbb{R}[\mathbf{x}]_d^\star : \mathbf{M}_{r_k}(g_k\mathbf{y})\succeq 0,\ k=0,1,\ldots,m\} \tag{2.9}$$

since the intersection of finitely many spectrahedra is a spectrahedron. Since this cone is an outer approximation of the moment cone, its elements are not necessarily moments of a measure, so they are sometimes called *pseudo-moments*.

Matrix $\mathbf{M}_{r_k}(g_k\mathbf{y})$ is called a *localizing matrix*. When $k=0$ and hence $r_0=r$, $g_0=1$, matrix $\mathbf{M}_r(\mathbf{y})$ is called the *moment matrix*.

Since the product of monomials is a monomial, i.e. $x^a x^b = x^{a+b}$ and $\ell_{\mathbf{y}}(x^a x^b) = \ell_{\mathbf{y}}(x^{a+b}) = y_{a+b}$, in the monomial basis the univariate moment matrices have Hankel structure, i.e. they have constant anti-diagonals.

Example 2.16. *If $n=1$ in the monomial basis, it holds*

$$\mathbf{M}_0(\mathbf{y})=y_0,\ \mathbf{M}_1(\mathbf{y})=\left(\begin{array}{cc} y_0 & y_1\\ y_1 & y_2\end{array}\right),\ \mathbf{M}_2(\mathbf{y})=\left(\begin{array}{ccc} y_0 & y_1 & y_2\\ y_1 & y_2 & y_3\\ y_2 & y_3 & y_4\end{array}\right).$$

In the multivariate case, the structure is more complicated.

Example 2.17. *If $n=2$ in the monomial basis, it holds*

$$\mathbf{M}_0(\mathbf{y})=y_{00},\ \mathbf{M}_1(\mathbf{y})=\left(\begin{array}{c|cc} y_{00} & y_{10} & y_{01}\\ \hline y_{10} & y_{20} & y_{11}\\ y_{01} & y_{11} & y_{02}\end{array}\right),\ \mathbf{M}_2(\mathbf{y})=\left(\begin{array}{c|cc|ccc} y_{00} & y_{10} & y_{01} & y_{20} & y_{11} & y_{02}\\ \hline y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12}\\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03}\\ \hline y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22}\\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13}\\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04}\end{array}\right).$$

In the Chebyshev polynomial basis, the univariate moment matrices have Hankel (i.e. constant anti-diagonals) plus Toeplitz (i.e. constant diagonals) structure. Indeed, the product of two Chebsyhev polynomials is the sum of two Chebyshev polynomials: $t_a(x)t_b(x)=\frac{1}{2}(t_{a+b}(x)+t_{|a-b|}(x))$.

Example 2.18. *If $n = 1$ in the Chebyshev basis, it holds*

$$\mathbf{M}_1(\mathbf{y}) = \frac{1}{2}\begin{pmatrix} y_0 & y_1 \\ y_1 & y_2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} y_0 & y_1 \\ y_1 & y_0 \end{pmatrix} = \begin{pmatrix} y_0 & y_1 \\ y_1 & \frac{1}{2}(y_0 + y_2) \end{pmatrix},$$

$$\mathbf{M}_2(\mathbf{y}) = \frac{1}{2}\begin{pmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} y_0 & y_1 & y_2 \\ y_1 & y_0 & y_1 \\ y_2 & y_1 & y_0 \end{pmatrix} = \begin{pmatrix} y_0 & y_1 & y_2 \\ y_1 & \frac{1}{2}(y_0 + y_2) & \frac{1}{2}(y_1 + y_3) \\ y_2 & \frac{1}{2}(y_1 + y_3) & \frac{1}{2}(y_0 + y_4) \end{pmatrix}.$$

The localizing matrices are weighted sums of shifted moment matrices. For notational convenience all the illustrative examples in the remainder of this document are expressed in the momonial basis.

Example 2.19. *If $n = 2$ and $g(x) = 1 - 2x_1 + 3x_2$ in the monomial basis:*

$$\mathbf{M}_1(g\mathbf{y}) = \left(\begin{array}{c|cc} y_{00} - 2y_{10} + 3y_{01} & y_{10} - 2y_{20} + 3y_{11} & y_{01} - 2y_{11} + 3y_{02} \\ \hline y_{10} - 2y_{20} + 3y_{11} & y_{20} - 2y_{30} + 3y_{21} & y_{11} - 2y_{21} + 3y_{12} \\ y_{01} - 2y_{11} + 3y_{02} & y_{11} - 2y_{21} + 3y_{12} & y_{02} - 2y_{12} + 3y_{03} \end{array}\right).$$

Example 2.20. *In Example 1.19 we saw that the convex hull of a set $\mathscr{X}$ are all first degree moments of probability measures on $\mathscr{X}$:*

$$\operatorname{conv} \mathscr{X} = \{(y_{\mathbf{a}})_{|\mathbf{a}|=1} : \mathbf{y} \in \mathscr{M}_1(\mathscr{X}),\ y_0 = 1\}.$$

Using moment and localization matrices as in (2.9)*, we can construct a family of outer approximations which are spectrahedral shadows:*

$$\mathscr{X}_r := \{(y_a)_{|a|=1} : \mathbf{y} \in \mathbf{M}_{r_k}(g_k\mathbf{y}) \succeq 0,\ y_0 = 1\}$$

so that inclusions (2.7) *and Theorem 2.3 imply*

$$\mathscr{X}_r \supset \mathscr{X}_{r+1} \supset \overline{\mathscr{X}_\infty} = \operatorname{conv} \mathscr{X}.$$

Example 2.21. *Let $n = 2$, $m = 3$ with $g_1(\mathbf{x}) = 4 - x_1^2 - x_2^2$, $g_2(\mathbf{x}) = -1 - 2x_1 - x_2 - x_1x_2$, $g_3(\mathbf{x}) = 1 + x_1 + x_1x_2$, as in Example 1.5. On Figure 2.4 are represented the outer approximations $\mathscr{X}_r$ for $r = 1, 2$ (dark gray) and the set $\mathscr{X}$ (light gray, thick black boundary). We observe that $\mathscr{X}_2 = \operatorname{conv} \mathscr{X}$.*

Example 2.22. *Let $n = m = 2$ with $g_1(\mathbf{x}) = 4 - x_1^2 - x_2^2$ and $g_2(\mathbf{x}) = x_1^3 + x_1x_2^2 - x_1^4 - x_1^2x_2^2 - x_2^4$. On Figure 2.5 are represented the outer approximations $\mathscr{X}_r$ for $r = 2, \ldots, 5$ (dark gray) and the set $\mathscr{X}$ (light gray, thick black boundary). We observe that consistently with Example 2.20, spectrahedral shadow approximations are embedded and tighter when r increases. It can however be shown that the singularity*[5] *of g_2 at the origin prevents the relaxations to be exact, i.e. there is no finite value of r for which $\mathscr{X}_r = \operatorname{conv} \mathscr{X}$ despite the fact that $\operatorname{conv} \mathscr{X} = \mathscr{X}$.*

[5] Singularity means that the (square-free) polynomial and its first partial derivatives are vanishing.

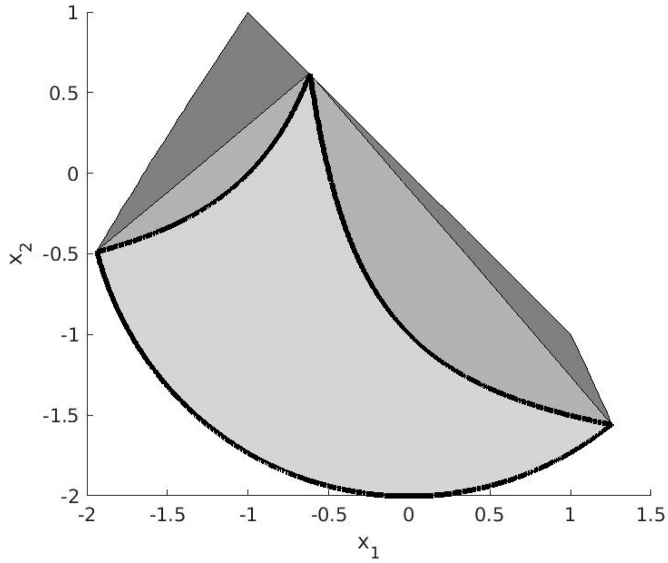


Figure 2.4: Embedded outer approximations (dark gray) of a non-convex set described by quadratic inequalities (light gray).

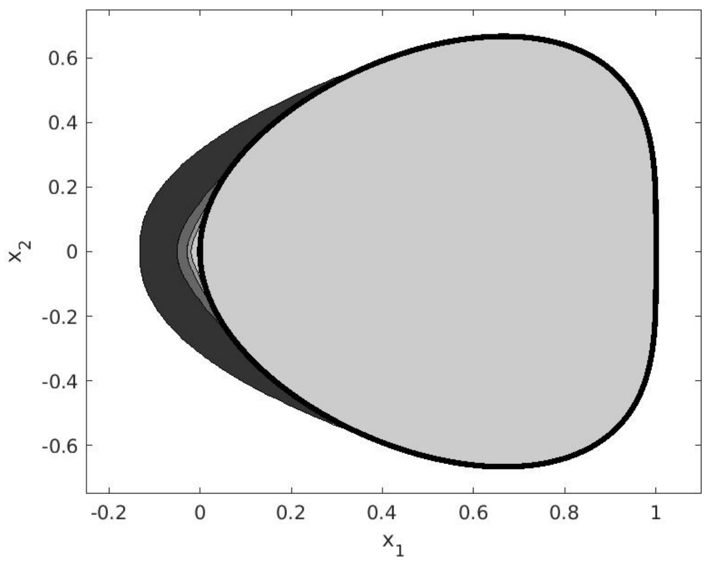


Figure 2.5: Embedded outer approximations (dark gray) of a convex set (light gray) whose boundary is the zero level set of a singular quartic (thick black).

After defining the basic semialgebraic set (1.2) with inequalities $g_k(\mathbf{x}) \geq 0$, we observed that we can also use equations $g_k(\mathbf{x}) = 0$. In the truncated quadratic module (2.9), the corresponding localizing matrix is then identically zero, i.e. $\mathbf{M}_{r_k}(g_k\mathbf{y}) = \mathbf{0}$. This generates a linear system of equations satisfied by the pseudo-moments.

Example 2.23. *Consider the discrete set*

$$\mathscr{X} = \{-1, +1\}^3 = \{\mathbf{x} \in \mathbb{R}^3 : x_k^2 - 1 = 0, k = 1, 2, 3\}$$

and the corresponding moment cone $\mathscr{M}_2(\mathscr{X})$. Then its first relaxation is

$$\mathscr{R}(\mathbf{g})_1 = \{\mathbf{y} \in \mathbb{R}^{10} : \left(\begin{array}{c|ccc} y_{000} & y_{100} & y_{010} & y_{001} \\ \hline y_{100} & y_{200} & y_{101} & y_{101} \\ y_{010} & y_{110} & y_{020} & y_{011} \\ y_{001} & y_{101} & y_{011} & y_{002} \end{array}\right) \succeq 0, \;\; \begin{array}{l} y_{200} - y_{000} = 0, \\ y_{020} - y_{000} = 0, \\ y_{002} - y_{000} = 0, \end{array} \;\; y_{000} = 1\}.$$

After substitution, we observe that in the space of second order moments $(y_{110}, y_{101}, y_{011})$ the relaxation

$$\begin{pmatrix} 1 & y_{101} & y_{101} \\ y_{110} & 1 & y_{011} \\ y_{101} & y_{011} & 1 \end{pmatrix} \succeq 0$$

corresponds to the pillow of Examples 1.6, 2.5 and 2.6. Note however that $\mathscr{X}_1 \neq$ $\operatorname{conv} \mathscr{X} = [-1, 1]^3$ because the boundary of the pillow is not flat.

2.5 Notes

Semidefinite optimization, also called semidefinite programming, or optimization over linear matrix inequalities (LMIs), has been a well-studied topic in applied mathematics since the mid 1990s. It is a versatile extension of linear and convex quadratic optimization for which efficient numerical algorithms can be designed [2]. A brief and elementary introduction to semidefinite optimization can be found in [3, Ch. 2].

The convex pillow of Example 2.5 is also called the elliptope. It corresponds to the set of correlation matrices arising in semidefinite relaxations of binary combinatorial optimization problems, as explained in Example 2.23. See e.g. [2, §4.3] for semidefinite relaxations of combinatorial optimization problems.

The TV-screen set of Example 2.7 is described in [6, Ex. 3.2.2]. Hyperbolic polynomials, also called real zero polynomials, are used in this reference to prove that this convex set cannot be a spectrahedron. The spectrahedral shadow of Example 2.9 is described in [8, §1.1]. See [32] for a tutorial on hyperbolic polynomials and their connection with semidefinite optimization. Lemma 2.1 on the basic semialgebraic formulation of spectrahedra can be found in [32, Theorem 20].

Spectrahedral shadows and their applications in convex optimization, are studied in [2] under the name of semidefinite representable sets, see [2, §4.2]. In the context of polynomial optimization, they are studied in [20, Ch. 11], [3, Ch. 6] or [21, Ch. 13].

The spectrahedral shadow of Example 2.8 was studied in [18, §2.3] in the context of LMIs for systems control.

It is shown in [34] that convex hulls of one-dimensional semialgebraic sets, as well as planar convex semialgebraic sets, are spectrahedral shadows. However, in higher dimensions, it is shown in [35] that there are convex closed semialgebraic sets that are not spectrahedral shadows. In particular, Example 2.10 follows from [35, Corollary 4.25].

Introductions to polynomial SOS can be found in [29], [23, §3], [3, Ch. 3] or [21, Ch. 2]. Theorem 2.1 is due to Hilbert in 1888, but the Motzkin polynomial of Example 1.3 was found only in 1967. Quadratic modules can be used to construct SOS representations of polynomials which are strictly positive on semialgebraic sets. Theorem 2.2 was proposed by Putinar in 1993 [31].

The semidefinite relaxations of the moment cone were introduced in 2001 by Lasserre in the context of polynomial optimization [19]. See [10] for an example of the use of the Chebyshev polynomial basis to manipulate moment problems arising in dynamical systems.

See [21, §13.2] for the use of moment relaxations to construct outer approximations of the convex hull of semialgebraic sets. In the three cases identified by Hilbert in Theorem 2.1, the convex hull of a rationally parametrized algebraic set is a spectrahedral shadow [9]. Example 2.21 was introduced in [13]. Example 2.22 was introduced in [7].

Chapter 3

Polynomial optimization and moment relaxations

In this chapter, we use the moment cone to reformulate a polynomial optimization problem as a moment problem. Then we use moment relaxations to deal numerically with the moment cone. This is the main idea behind the moment-SOS hierarchy.

3.1 Polynomial optimization

Given a vector $\mathbf{g} \in \mathbb{R}[\mathbf{x}]^m$, define as in (1.2) the basic semialgebraic set

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

and assume it is included in a unit ball $\mathscr{B}_R$ of sufficiently large radius R, so that without loss of generality we can enforce $g_1(\mathbf{x}) := R^2 - \mathbf{x}^T\mathbf{x}$.

Let $p \in \mathbb{R}[\mathbf{x}]$ be given and consider the *polynomial optimization problem* (POP) consisting of minimizing p over $\mathscr{X}$, namely

$$p^\star := \min_{\mathbf{x}} \; p(\mathbf{x}) \;\text{ s.t. }\; \mathbf{x} \in \mathscr{X} \tag{3.1}$$

Since $\mathscr{X}$ is compact and p is continuous, the minimum is attained at a given point $\mathbf{x}^\star \in \mathscr{X}$ called a minimizer. It is such that p achieves its minimum i.e. $p(x^\star) = p^\star$.

We do not have any convexity property on p or $\mathscr{X}$, so that problem (3.1) may feature several local minima, possibly several global minima, possibly infinitely many.

Example 3.1. *Let $p(x) = (1 - x^2)^2$ and $\mathscr{X} = \mathscr{B}_2$. Then POP (3.1) has two global minimizers $x^\star \in \{-1, +1\}$ with minimum $p^\star = 0$.*

Example 3.2. *Let $p(\mathbf{x}) = g_1(\mathbf{x}) = 1 - x_1^2 - x_2^2$, $g_2(\mathbf{x}) = x_1x_2$. Then POP (3.1) has infinitely many global minimizers $x^\star$ with the same minimum $p^\star = 0$ along the circular arcs $\{\mathbf{x} \in \mathbb{R}^2 : g_1(\mathbf{x}) = 0,\ g_2(\mathbf{x}) \geq 0\}$, see Figure 3.2.*

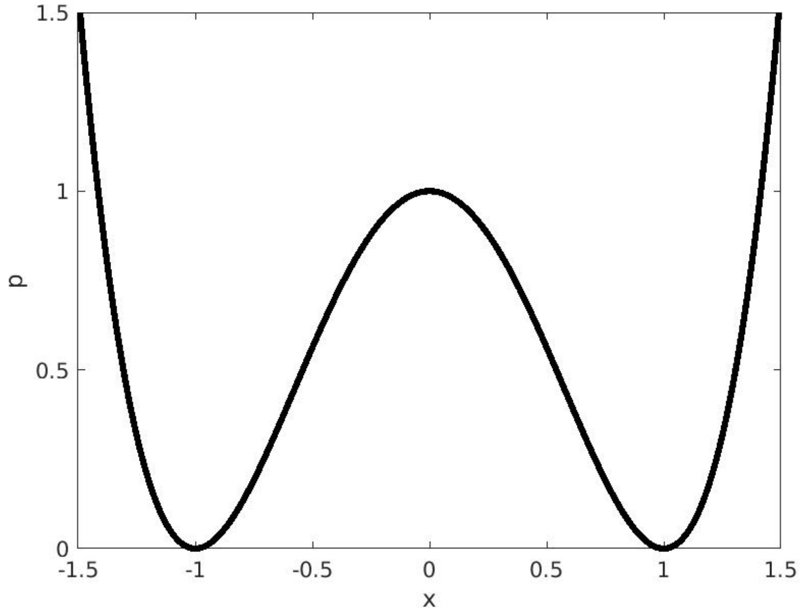


Figure 3.1: Polynomial $(1-x^2)^2$ with its two global minimizers at $x = \pm 1$.

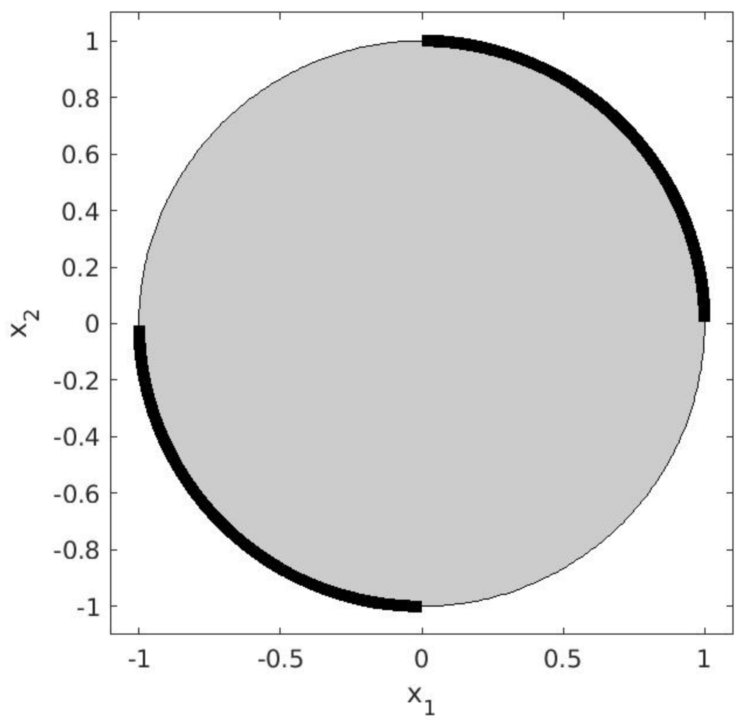


Figure 3.2: All the points on the circular arcs $\{\mathbf{x} \in \mathbb{R}^2 : 1 - x_1^2 - x_2^2 = 0,\, x_1 x_2 \geq 0\}$ (thick black) are optimal within unit ball $\mathscr{B}_1$ (gray).

3.2 Moment reformulation

Suppose we can generate distinct points $\mathbf{x}_k \in \mathscr{X}$, $k = 1, \ldots, N$, and we can evaluate p at these points. Obviously, we obtain an upper bound $\min_{k=1,\ldots,N} p(\mathbf{x}_k) \geq p^\star$ on the minimum. If N is large, the upper bound can be a good approximation to the minimum. Moreover, the point $\mathbf{x}_{k^\star}$ achieving the upper bound $\min_{k=1,\ldots,N} p(\mathbf{x}_k) = p(\mathbf{x}_{k^\star})$ can be a good approximation to the minimizer.

Given the points $\{\mathbf{x}_k\}_{k=1,\ldots,N} \subset \mathscr{X}$, consider the optimization problem

$$\min_{\mathbf{w}} \sum_{k=1}^{N} w_k p(\mathbf{x}_k) \text{ s.t. } w_k \geq 0, \sum_{k=1}^{N} w_k = 1 \tag{3.2}$$

which is linear in the unknown vector $\mathbf{w} = (w_k)_{k=1,\ldots,N}$ of non-negative weights summing up to one. Since the points $\mathbf{x}_k$ are distinct, the unique optimal solution $\mathbf{w}^\star$ to linear optimization problem (3.2) consists of putting the maximum weight on the point achieving the minimum, i.e. $w_k^\star = 1$ if $k = k^\star$ and $w_k^\star = 0$ if $k \neq k^\star$.

Now recall Examples 1.12 and 1.13 and the interpretation of a probability measure on $\mathscr{X}$ as a function distributing the mass on $\mathscr{X}$, such that the weight is everywhere non-negative and summing up to one. When $N \to \infty$, linear problem (3.2) becomes the optimization problem

$$p_M^\star := \min_{\mu} \int_{\mathscr{X}} p(\mathbf{x}) d\mu(\mathbf{x}) \text{ s.t. } \mu \text{ probability measure on } \mathscr{X} \tag{3.3}$$

which is linear in the unknown probability measure μ on $\mathscr{X}$.

Since p is a polynomial, it can be expressed as in (1.1) as a linear combination of basis elements, i.e.

$$p(\mathbf{x}) = \mathbf{p}^T \mathbf{b}(\mathbf{x})$$

and hence the objective function in linear problem (3.3) can be expressed as a linear function

$$\int_{\mathscr{X}} p(\mathbf{x}) d\mu(\mathbf{x}) = \mathbf{p}^T \int_{\mathscr{X}} \mathbf{b}(\mathbf{x}) d\mu(\mathbf{x}) = \mathbf{p}^T \mathbf{y}$$

of the vector of moments $\mathbf{y}$ of measure μ, as defined in (1.4). Overall, recalling the definition (1.5) of the moment cone $\mathscr{M}_d(\mathscr{X})$, linear measure problem (3.3) can be reformulated as the *moment optimization problem* (MOP)

$$p_M^\star = \min_{\mathbf{y}} \mathbf{p}^T \mathbf{y} \text{ s.t. } \mathbf{y} \in \mathscr{M}_d(\mathscr{X}),\ y_0 = 1 \tag{3.4}$$

which is linear in the unknown moment vector $\mathbf{y}$.

Theorem 3.1. *POP* (3.1) *and MOP* (3.4) *have the same value, i.e.* $p^\star = p_M^\star$. *Moreover, the moment vector* $\mathbf{y}^\star = \mathbf{b}(\mathbf{x}^\star)$ *is an optimal solution to MOP* (3.4) *whenever* $\mathbf{x}^\star$ *is an optimal solution to POP* (3.1).

The proof of this result is elementary: on the one hand, observe that if $\mathbf{x}^\star$ is optimal for POP, then $\mathbf{y}^\star = \mathbf{b}(\mathbf{x}^\star)$ is the vector of moments of the Dirac measure at $\mathbf{x}^\star$, recall Example 1.11, and hence it is admissible for MOP, but not necessarily optimal. Hence it yields a value $\mathbf{p}^T\mathbf{y}^\star = p(\mathbf{x}^\star) = p^\star$ which is an upper bound on the optimal value $p^\star_M$ of the MOM, i.e. $p^\star \geq p^\star_M$; on the other hand, from the definition of $p^\star$ it holds $p(\mathbf{x}) \geq p^\star$ for all $\mathbf{x} \in \mathscr{X}$, and hence for any probability measure μ on $\mathscr{X}$ we have $\int_{\mathscr{X}} p(\mathbf{x})d\mu(\mathbf{x}) \geq \int_{\mathscr{X}} p^\star d\mu(\mathbf{x}) = p^\star$ and therefore for any moment vector $\mathbf{y} \in \mathscr{M}_d(\mathscr{X})$ we have $\mathbf{p}^T\mathbf{y} \geq p^\star$. MOP consists of minimizing this linear function over all moment vectors, hence $p^\star_M \geq p^\star$.

Observe that if there are several global minima $(\mathbf{x}^\star_k)_{k=1,2,\ldots}$ to POP, then each moment vector $\mathbf{y}^\star_k = \mathbf{b}(\mathbf{x}^\star)$ is an optimal solution to MOP. Since MOP is linear, it follows that any convex combination $\sum_k w_k\mathbf{y}^\star_k$ with $w_k \geq 0$ and $\sum_k w_k = 1$ is also optimal for MOP.

3.3 Moment relaxations

In order to solve numerically MOP (3.4), we can now use the outer approximations described in Section 3.3. Given $r \in \mathbb{N}$ such that $r \geq r_{\min}$ as defined in (2.5), consider the semidefinite optimization problem

$$p^\star_r := \min_{\mathbf{y}} \ \mathbf{p}^T\mathbf{y} \ \text{ s.t. } \ \mathbf{y} \in \mathscr{R}_{d,r}(\mathbf{g}), \ y_0 = 1 \tag{3.5}$$

where $\mathscr{R}_{d,r}(\mathbf{g})$ is described in (2.9). Optimization problem (3.5) is called the *moment relaxation* of order r of the POP. Accordingly, integer r is called the *relaxation order*. The proof of the following fundamental result follows readily from the inclusion relations (2.7) and from Theorem 2.3.

Theorem 3.2. *Moment relaxations* (3.5) *generate a non-decreasing monotone sequence of lower bounds converging to the value of MOP* (3.4)*. It holds*

$$p^\star_r \leq p^\star_{r+1} \leq p^\star$$

and

$$p^\star_\infty = p^\star.$$

Recalling from Theorem 3.1 that the values of MOP (3.4) and POP (3.1) coincide, it means that we can solve the POP at the price of solving a family or hierarchy of semidefinite optimization problems. The family of moment relaxations (3.5) for increasing values of relaxation order r is called the *moment hierarchy*.

Moment problem (3.5) has a dual semidefinite optimization problem on the SOS cone that we do not describe here. It can be shown that there is no duality gap between the primal moment problem and the dual SOS problem. The family of primal moment and dual SOS problems is called the *moment-SOS hierarchy*.

Example 3.3. *Consider the POP of Example 3.1 where $n = 1$, $d = 4$, $m = 1$, $p(x) = 1 - 2x^2 + x^4$ and $g_1(x) := 4 - x^2$. The first relaxation in the moment hierarchy has order $r = r_{\min} = 2$ and it reads*

$$\begin{array}{rll} p_2^\star = & \min_{\mathbf{y}\in\mathbb{R}^5} & y_0 - 2y_2 + y_4 \\ & \text{s.t.} & \mathbf{M}_2(\mathbf{y}) = \begin{pmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{pmatrix} \succeq 0 \\ & & \mathbf{M}_1(g_1\mathbf{y}) = \begin{pmatrix} 4y_0 - y_2 & 4y_1 - y_3 \\ 4y_1 - y_3 & 4y_2 - y_4 \end{pmatrix} \succeq 0 \\ & & y_0 = 1. \end{array}$$

Solving numerically[1] *this semidefinite optimization problem yields a solution*[2] $\mathbf{y} = (1, 0, 1, 0, 1)$ *and the lower bound $p_2^\star = 0$ which is equal to the global minimum $p^\star = 0$.*

Example 3.4. *Consider the POP of Example 3.2 where $n = 2$, $d = 2$, $m = 2$, $p(\mathbf{x}) = g_1(\mathbf{x}) = 1 - x_1^2 - x_2^2$, $g_2(\mathbf{x}) = x_1x_2$. The first relaxation in the moment hierarchy has order $r = r_{\min} = 1$ and it reads*

$$\begin{array}{rll} p_1^\star = & \min_{\mathbf{y}\in\mathbb{R}^6} & y_{00} - y_{20} - y_{02} \\ & \text{s.t.} & \mathbf{M}_1(\mathbf{y}) = \left(\begin{array}{c|cc} y_{00} & y_{10} & y_{01} \\ \hline y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{array}\right) \succeq 0 \\ & & \mathbf{M}_0(g_1\mathbf{y}) = y_{00} - y_{20} - y_{02} \geq 0 \\ & & \mathbf{M}_0(g_2\mathbf{y}) = y_{11} \geq 0 \\ & & y_{00} = 1. \end{array}$$

Solving numerically this semidefinite optimization problem yields $y_{00} = 1$, $y_{10} = y_{01} = 0$, $y_{20} = y_{02} = \frac{1}{2}$, $y_{11} = \frac{1}{4}$ and the lower bound $p_1^\star = 0$ which is equal to the global minimum $p^\star = 0$.

Example 3.5. *Consider the problem consisting of minimizing $-x_2$ (or equivalently maximizing x_2) on the non-convex set $\mathscr{X}$ of Example 2.21, i.e. the POP*

$$\begin{array}{rll} p^* = & \min_{\mathbf{x}} & -x_2 \\ & \text{s.t.} & 4 - x_1^2 - x_2^2 \geq 0 \\ & & -1 - 2x_1 - x_2 - x_1x_2 \geq 0 \\ & & 1 + x_1 + x_1x_2 \geq 0. \end{array}$$

Its first moment relaxation ($r = r_{\min} = 1$) is

$$\begin{array}{rl} p_1^* = & \min_{\mathbf{y}\in\mathbb{R}^6} -y_{01} \\ \text{s.t.} & \left(\begin{array}{c|cc} y_{00} & y_{10} & y_{01} \\ \hline y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{array}\right) \succeq 0 \\ & 4y_{00} - y_{20} - y_{02} \geq 0 \\ & -y_{00} - 2y_{10} - y_{01} - y_{11} \geq 0 \\ & y_{00} + y_{10} + y_{11} \geq 0 \\ & y_{00} = 1 \end{array}$$

[1] We are using the interface GloptiPoly and the semidefinite solver SeDuMi under Matlab.

[2] Numerical outputs are rounded to the nearest integers or rationals, or to 4 significant digits.

and its second moment relaxation $(r = 2)$ *is*

$$\begin{array}{rcl} p_2^* & = & \min_{\mathbf{y}\in\mathbb{R}^{15}} -y_{01} \\ & \text{s.t.} & \left(\begin{array}{c|cc|ccc} y_{00} & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ \hline y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ \hline y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{array}\right) \succeq 0 \\ & & \left(\begin{array}{c|cc} 4y_{00}-y_{20}-y_{02} & 4y_{10}-y_{30}-y_{12} & 4y_{01}-y_{21}-y_{03} \\ \hline 4y_{10}-y_{30}-y_{12} & 4y_{20}-y_{40}-y_{22} & 4y_{11}-y_{31}-y_{13} \\ 4y_{01}-y_{21}-y_{03} & 4y_{11}-y_{31}-y_{13} & 4y_{02}-y_{22}-y_{04} \end{array}\right) \succeq 0 \\ & & \left(\begin{array}{c|cc} -y_{00}-2y_{10}-y_{01}-y_{11} & -y_{10}-2y_{20}-y_{11}-y_{21} & -y_{01}-2y_{11}-y_{02}-y_{12} \\ \hline -y_{10}-2y_{20}-y_{11}-y_{21} & -y_{20}-2y_{30}-y_{21}-y_{31} & -y_{11}-2y_{21}-y_{12}-y_{22} \\ -y_{01}-2y_{11}-y_{02}-y_{12} & -y_{11}-2y_{21}-y_{12}-y_{22} & -y_{02}-2y_{12}-y_{03}-y_{13} \end{array}\right) \succeq 0 \\ & & \left(\begin{array}{c|cc} y_{00}+y_{10}+y_{11} & y_{10}+y_{20}+y_{21} & y_{01}+y_{11}+y_{12} \\ \hline y_{10}+y_{20}+y_{21} & y_{20}+y_{30}+y_{31} & y_{11}+y_{21}+y_{22} \\ y_{01}+y_{11}+y_{12} & y_{11}+y_{21}+y_{22} & y_{02}+y_{12}+y_{13} \end{array}\right) \succeq 0 \\ & & y_{00} = 1. \end{array}$$

Numerically we obtain $p_1^\star \approx -1$ *and* $p_2^\star \approx -0.6180 \approx p^\star = -\frac{\sqrt{5}-1}{2}$. *So here the first moment relaxation gives a strict lower bound on the global maximum, and the second moment relaxation gives the global maximum. Notice that when the POP objective function has degree* $d = 1$*, i.e.* $p(\mathbf{x}) = \mathbf{p}^T\mathbf{x}$*, solving the moment relaxation* (3.5) *is equivalent to solving the semidefinite optimization problem*

$$p_r^\star = \min_{\mathbf{x}} \mathbf{p}^T\mathbf{x} \text{ s.t. } \mathbf{x} \in \mathscr{X}_r$$

on the outer approximation $\mathscr{X}_r$ *of the convex hull of* $\mathscr{X}$ *introduced in Example 2.20. Indeed, from Figure 2.4, we see that minimizing* $-x_2$ *on* $\mathscr{X}_1$ *yields* $x_2^\star = -1$*, whereas minimizing* $-x_2$ *on* $\mathscr{X}_2$ *yields* $x_2^\star \approx -0.6180$*. This is the global minimum since optimizing a linear function on a set is the same as optimizing the linear function on the convex hull of the set, and* $\mathscr{X}_2 = \operatorname{conv} \mathscr{X}$ *as observed in Example 2.4.*

Example 3.6. *Consider the POP consisting of minimizing* x_1 *on the convex set* $\mathscr{X}$ *of Example 2.22. As in Example 3.5 the polynomial to be minimized is linear and hence the moment relaxations are formulated on the outer approximations* $\mathscr{X}_r$ *of* $\mathscr{X}$*. From Figure 2.5 we see however that near the origin* $\mathbf{x} = (0, 0)$ *the outer approximations are not tight. Because of the singularity at the origin of the polynomial defining the boundary of* $\mathscr{X}$*, there will be no finite value of* r *for which* $p_r^\star = p^\star$*. Numerically, we obtain the following sequence*

$$(p_r^\star)_{r=2,3,4,5,\ldots} \approx (-0.1315, -0.04983, -0.02843, -0.02233, \ldots)$$

converging monotonically from below to $p^\star = 0$*. Now if instead of minimizing* x_1*, we minimize* $-x_1$*, we obtain immediately* $p_2^\star = p^\star = -1$ *at the first moment relaxation. Notice that there is no singularity of the polynomial defining the boundary at the point* $\mathbf{x} = (1, 0)$*.*

3.4 Notes

The moment-SOS hierarchy as presented in these notes was conceived by Lasserre in 2001 [19]. In this reference, duality between moments and SOS is emphasized, and convergence of the hierarchy is proven using Putinar's Theorem 2.2. The whole approach is explained in details in the survey [23]. Several books on polynomial optimization are based on these foundations: [20, 21, 11, 28].

GloptiPoly is a Matlab package that allows to model moment relaxations of POP as semidefinite optimization problems [12, 15]. The semidefinite optimization problems are then solved by a third-party conic solver such as SeDuMi [37].

Chapter 4

Solution recovery

When using the moment-SOS hierarchy, we would like to know whether a given moment relaxation of a polynomial optimization problem is exact, i.e. whether solving the relaxed problem solves the original problem. In this chapter we describe various strategies to detect exactness and recover the optimal solutions from the moments.

4.1 Finite convergence

In the previous chapter we saw that a monotically non-decreasing sequence of lower bounds $p_r^\star \leq p_{r+1}^\star \leq p^\star$ on the value $p^\star$ of a POP can be obtained by solving a hierarchy of moment relaxations of increasing size, ruled by the relaxation order r. The sequence converges asymptotically, i.e. $p_\infty^\star = p^\star$, but it may happen that this convergence is finite, i.e. there is a finite value of $r^\star$ such that $p_{r^\star}^\star = p^\star$. It turns out that this situation is generic in the following sense.

Theorem 4.1. *In the finite-dimensional space of coefficients of polynomials defining POP* (3.1)*, there is a low-dimensional algebraic variety which is such that if we choose an instance of POP* (3.1) *outside this variety, then the moment hierarchy has finite convergence, i.e. there exists a finite r^* such that $p_r^* = p^*$ for all $r \geq r^*$.*

Equivalently, finite convergence occurs under arbitrarily small perturbations of the POP data, and problems for which finite convergence does not occur are exceptional and degenerate in some sense, see e.g. the singular POP of Example 3.6.

4.2 Finitely many solutions

Theorem 4.1 does not tell us however a priori at which relaxation order finite convergence occurs. To detect finite convergence, we can use the following conditions. Let $\mathbf{y}^\star$ be the solution of moment relaxation (3.5) at a given relaxation order $r \geq r_{\min}$.

The first candidate for optimality is the vector of first degree moments.

Lemma 4.1. *Let* $\mathbf{x}^\star = (y^\star_{\mathbf{a}})_{|\mathbf{a}|=1}$. *If* $\mathbf{x}^\star \in \mathscr{X}$ *and* $p(\mathbf{x}^\star) = p^\star_r$ *then* $p^\star_r = p^\star$.

The proof is elementary. Every admissible vector for the POP yields an upper bound on the value of the POP. If this upper bound is a lower bound, it means that it is optimal.

Another useful property to check is whether the moment matrix has rank one.

Lemma 4.2. *If*

$$\operatorname{rank} \mathbf{M}_r(\mathbf{y}^\star) = 1. \tag{4.1}$$

then $p^\star_r = p^\star$.

The proof goes as follows. From the definition (2.8) of the moment matrix

$$\mathbf{M}_r(\mathbf{y}) = \ell_{\mathbf{y}}(\mathbf{b}\mathbf{b}^T)$$

if $\operatorname{rank} \mathbf{M}_r(\mathbf{y}^\star) = 1$ for some $\mathbf{y}^\star$ then

$$\mathbf{M}_r(\mathbf{y}^\star) = \mathbf{b}(\mathbf{x}^\star)\mathbf{b}(\mathbf{x}^\star)^T = \mathbf{y}^\star(\mathbf{y}^\star)^T$$

that is $\mathbf{y}^\star = \mathbf{b}(\mathbf{x}^\star)$ is the vector of moments of the Dirac measure at some $\mathbf{x}^\star$ as in Example 1.12. Since $\mathbf{y}^\star$ is admissible for moment relaxation (3.5), it follows that for all $k = 1, \ldots, m$, $\mathbf{M}_{r_k}(g_k\mathbf{y}^\star) = g_k(\mathbf{x}^\star)\mathbf{b}(\mathbf{x}^\star)\mathbf{b}(\mathbf{x}^\star)^T \succeq 0$ and hence $g_k(\mathbf{x}^\star) \geq 0$ which means that $\mathbf{x}^\star \in \mathscr{X}$ is admissible for POP (3.1). Since the value of the moment relaxation $p^\star_r = p(\mathbf{x}^\star)$ is a lower bound on $p^\star$ the value of the POP, and $\mathbf{x}^\star$ is admissible for the POP, it follows that $\mathbf{x}^\star$ is optimal and therefore $p^\star_r = p^\star$.

Theorem 4.2. *If for some* $r^* \geq r_{\min}$, *vector* $\mathbf{y}^*$ *is such that*

$$\operatorname{rank} \mathbf{M}_{r^*-r_{\mathbf{g}}}(\mathbf{y}^*) = \operatorname{rank} \mathbf{M}_{r^*}(\mathbf{y}^*) \tag{4.2}$$

then $p^*_{r^*} = p^*$.

Note that Lemma 4.1 is a particular case of Theorem 4.2 since if $\operatorname{rank} \mathbf{M}_{r^*}(\mathbf{y}^*) = 1$ for some r^* then $\operatorname{rank} \mathbf{M}_r(\mathbf{y}^*) = 1$ for all $r \in [r_{\min}, r^*]$. Similarly, Lemma 4.1 is a particular case of Theorem 4.2 since then the vector $\mathbf{y}^* = ((\mathbf{x}^*)^{\mathbf{a}})_{\mathbf{a}}$ is such that $\operatorname{rank} \mathbf{M}_r(\mathbf{y}^*) = 1$ for all $r \geq r_{\min}$.

If condition (4.2) is satisfied, then there are $N := \operatorname{rank} \mathbf{M}_{r^*}(\mathbf{y}^\star)$ weights $w_k \geq 0$, $\sum_{k=1}^N w_k = 1$ and points $\mathbf{x}^\star_k \in \mathscr{X}$ such that

$$\mathbf{M}_r(\mathbf{y}^\star) = \sum_{k=1}^N w_k \mathbf{b}(\mathbf{x}^\star_k)\mathbf{b}(\mathbf{x}^\star_k)^T \tag{4.3}$$

and $p^\star = p(\mathbf{x}^\star_k)$ for all $k = 1, \ldots, N$. We can apply numerical linear algebra algorithms (not described here) on $\mathbf{M}_r(\mathbf{y}^\star)$ to extract the N global optima $\mathbf{x}^\star_k$ for POP (3.1).

Example 4.1. *For the POP of Example 3.3, it holds* $r_{\mathbf{g}} = 1$ *and at the first relaxation* ($r = r_{\min} = 2$) *we obtain the moment matrix*

$$\mathbf{M}_2(\mathbf{y}^\star) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

whose submatrix

$$\mathbf{M}_1(\mathbf{y}^\star) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

is such that $\operatorname{rank} \mathbf{M}_1(\mathbf{y}^\star) = \operatorname{rank} \mathbf{M}_2(\mathbf{y}^\star) = 2$ *and the condition of Theorem 4.2 is satisfied for* $r^* = 2$. *Eigenvalue computations on the moment matrix yield the rank-one decomposition*

$$\mathbf{M}_2(\mathbf{y}^\star) = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}^T + \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}^T$$

which is equation (4.3) *with the expected two global minimizers* $\mathbf{x}_1^\star = 1$, $\mathbf{x}_2^\star = -1$ *and equal weights* $w_1 = w_2 = \frac{1}{2}$.

Example 4.2. *For the POP of Example 3.5, at the first relaxation we obtain the moment matrix*

$$\mathbf{M}_1(\mathbf{y}^\star) = \left(\begin{array}{r|rr} 1 & -1 & 1 \\ \hline -1 & 2 & 0 \\ 1 & 0 & 2 \end{array}\right)$$

which has rank 2, so we cannot apply Lemma 4.1 or Theorem 4.2. At the second relaxation we obtain the moment matrix

$$\mathbf{M}_2(\mathbf{y}^\star) \approx \left(\begin{array}{r|rr|rrr} 1.0000 & -0.6180 & 0.6180 & 0.3820 & -0.3820 & 0.3820 \\ \hline -0.6180 & 0.3820 & -0.3820 & -0.2361 & 0.2361 & -0.2361 \\ 0.6180 & -0.3820 & 0.3820 & 0.2361 & -0.2361 & 0.2361 \\ \hline 0.3820 & -0.2361 & 0.2361 & 0.1459 & -0.1459 & 0.1459 \\ -0.3820 & 0.2361 & -0.2361 & -0.1459 & 0.1459 & -0.1459 \\ 0.3820 & -0.2361 & 0.2361 & 0.1459 & -0.1459 & 0.1459 \end{array}\right)$$

which has numerically rank 1: it has a singular value approx. equal to 2.2016 *and its other singular values are of the order of* 10^{-8} *or less. According to Lemma 4.1, the first degree moments yield the global optimizer* $(y_{10}^\star, y_{01}^\star) \approx (-0.6180, 0.6180)$ *corresponding to the analytic solution* $\mathbf{x}^\star = (\frac{1-\sqrt{5}}{2}, \frac{1+\sqrt{5}}{2})$.

4.3 Solutions in a real algebraic variety

If the POP has infinitely many solutions, e.g. as in Example 3.4, we cannot use Lemma 4.1 and Theorem 4.2 because the rank of the moment matrix $\mathbf{M}_r(\mathbf{y}^\star)$ does not have to stabilize when r increases. Instead, we can approximate the set of solutions with the help of the eigenstructure of the moment matrix.

Let μ be a measure on $\mathscr{X}$. Given a set[1] $\mathscr{A} \subset \mathscr{X}$, the measure of this set is the real number $\mu(\mathscr{A}) := \int_{\mathscr{A}} d\mu(\mathbf{x})$. The support of measure μ, denoted $\operatorname{spt}\mu$, is the smallest closed set such that the measure of its complement is zero, i.e. $\mu(\mathscr{X}\backslash \operatorname{spt}\mu) = 0$.

Let $\mathscr{V}$ denote the smallest real algebraic variety which contains $\operatorname{spt}\mu$. It is the common zero set of finitely many polynomials. These polynomials generate an ideal[2] denoted by $\mathscr{I}$. As usual, let $\mathbf{b}$ denote a basis vector for the vector space of polynomials of degree at most d, and let $\mathbf{M}_d(\mathbf{y}) := \int_{\mathscr{X}} \mathbf{b}(\mathbf{x})\mathbf{b}(\mathbf{x})^T d\mu(\mathbf{x}) \in \mathbb{S}^{n_d}$ denote the moment matrix of order d of μ.

Lemma 4.3. *For all $d \in \mathbb{N}$, $n_d - \operatorname{rank} \mathbf{M}_d(\mathbf{y})$ is the dimension of the vector space of polynomials of degree up to d that vanish on variety $\mathscr{V}$. For d large enough, $\operatorname{rank} \mathbf{M}_d(\mathbf{y})$ is a polynomial[3] in d whose degree is the dimension of $\mathscr{V}$, and the kernel[4] $\ker \mathbf{M}_d(\mathbf{y})$ generates ideal $\mathscr{I}$.*

Example 4.3. *Returning to Example 3.4, the global minimum is already achieved at the first relaxation, but the corresponding moment matrix*

$$\mathbf{M}_1(\mathbf{y}^\star) = \left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & 0.5000 & 0.2500 \\ 0 & 0.2500 & 0.5000 \end{array}\right)$$

has full rank. The moment matrix obtained at the second relaxation is

$$\mathbf{M}_2(\mathbf{y}^\star) = \left(\begin{array}{c|cc|ccc} 1 & 0 & 0 & 0.5002 & 0.2542 & 0.4998 \\ \hline 0 & 0.5002 & 0.2542 & 0 & 0 & 0 \\ 0 & 0.2542 & 0.4998 & 0 & 0 & 0 \\ \hline 0.5002 & 0 & 0 & 0.3962 & 0.1271 & 0.1040 \\ 0.2542 & 0 & 0 & 0.1271 & 0.1040 & 0.1271 \\ 0.4998 & 0 & 0 & 0.1040 & 0.1271 & 0.3958 \end{array}\right)$$

and it has numerical rank 5, while its submatrix $\mathbf{M}_1$ has rank 3. The 1-dimensional kernel of $\mathbf{M}_2$ is spanned by the vector $\mathbf{p}_1 = (1, 0, 0, -1, 0, -1)$ corresponding to the polynomial $p_1(\mathbf{x}) = \mathbf{p}_1^T \mathbf{b}(\mathbf{x}) = 1 - x_1^2 - x_2^2$. At the third relaxation, $\mathbf{M}_1$ has rank 3, $\mathbf{M}_2$ has rank 5 and $\mathbf{M}_3$ has rank 7, a linear growth. The 3-dimensional kernel of M_3 is spanned by the polynomials $p_1(\mathbf{x}) = 1 - x_1^2 - x_2^2$, $p_2(\mathbf{x}) = (x_1 - x_2)p_1(\mathbf{x})$, $p_3(\mathbf{x}) = (x_1 + x_2)p_1(\mathbf{x})$. They belong to the ideal $\mathscr{I}$ generated by the polynomial $1 - x_1^2 - x_2^2$. These polynomials all vanish on the variety $\mathscr{V} := \{\mathbf{x} \in \mathbb{R}^2 : 1 - x_1^2 - x_2^2 = 0\}$, which is strictly larger than the set of globally optimal solutions to the POP.

Lemma 4.3 tells us that the set of polynomials vanishing on $\operatorname{spt}\mu$ can be constructed from the kernel of the moment matrix of μ. The common set of zeros of these polynomials is an algebraic variety. In POP however, because of the constraints, the set of global optimizers is not necessarily a variety. Example 4.3 illustrates that the set of global optimizers can be strictly smaller than the variety that contains them. In order to get tighter approximations on the set of global optimizers, we can further exploit knowledge from the eigenstructure of the moment matrix.

[1] More rigorously, set $\mathscr{A}$ should belong to the Borel sigma-algebra of $\mathscr{X}$.

[2] An ideal is a set of polynomials which is closed under linear combinations with polynomial coefficients.

[3] This polynomial is called the Hilbert function of the variety.

[4] Given a symmetric matrix $\mathbf{A} \in \mathbb{S}^m$, its null-space or kernel is $\ker \mathbf{A} := \{\mathbf{v} \in \mathbb{R}^m : \mathbf{A}\mathbf{v} = \mathbf{0}\}$.

4.4 Christoffel-Darboux polynomial

Since the moment matrix $\mathbf{M}_d$ is positive semi-definite, it has a spectral decomposition

$$\mathbf{M}_d = \mathbf{QEQ}^T$$

where $\mathbf{Q} \in \mathbb{R}^{n_d \times n_d}$ is an orthonormal[5] matrix with columns $\mathbf{q}_i$, $i = 1, \ldots, n_d$, and $\mathbf{E}$ is a diagonal matrix whose entries are the eigenvalues $e_{i+1} \geq e_i \geq 0$ of $\mathbf{M}_d$. By construction, each eigenvector $\mathbf{q}_i$ is the coefficient vector of a polynomial $q_i \in \mathbb{R}[\mathbf{x}]$ such that

$$\mathbf{q}_i^T \mathbf{M}_d \mathbf{q}_i = \mathbf{q}_i^T \left(\int_{\mathscr{X}} \mathbf{b}(\mathbf{x})\mathbf{b}(\mathbf{x})^T d\mu(\mathbf{x}) \right) \mathbf{q}_i = \int_{\mathscr{X}} q_i^2(\mathbf{x}) d\mu(\mathbf{x}) = e_i,\ i = 1, \ldots, n_d.$$

Lemma 4.3 tells us that if d is large enough, the ideal $\mathscr{I}$ of polynomials vanishing on sptμ consists of polynomials q_i such that $e_i = 0$, i.e. those in the kernel of the moment matrix. Let us now use the other polynomials to get tighter approximations of spt μ, and hence of the set of global optimizers of the POP.

Given a small parameter $\beta_d > 0$, let us construct the following Christoffel-Darboux polynomial

$$p_{\beta_d}(\mathbf{x}) := \sum_{i=1}^{n_d} \frac{q_i^2(\mathbf{x})}{e_i + \beta_d}. \tag{4.4}$$

Note that this polynomial is SOS of degree $2d$ and hence it is positive. Equation (4.4) can be written in matrix form as

$$p_{\beta_d}(\mathbf{x}) = \mathbf{b}(\mathbf{x})^T (\mathbf{M}_d + \beta_d \mathbf{I}_{n_d})^{-1} \mathbf{b}(\mathbf{x})$$

since an invertible symmetric matrix and its inverse share the same eigenvectors. Let us now argue that spt μ can be well approximated by the set of points $\mathbf{x}$ such that $p_{\beta_d}(\mathbf{x})$ is small.

Observe that if e_i is small, the term $q_i^2(\mathbf{x})$ takes small values on sptμ, and its relative weight $(e_i + \beta)^{-1}$ is large in the sum (4.4). Conversely, if e_i is large, $q_i^2(\mathbf{x})$ takes larger values on spt μ and its relative weight is smaller in the sum. Therefore, the sublevel set

$$\{\mathbf{x} \in \mathscr{X} : p_{\beta_d}(\mathbf{x}) \leq \alpha_d\}$$

for $\alpha_d > 0$ small is likely to be a good approximation to spt μ. The explicit choice of α_d, β_d as functions of d, and the corresponding convergence results depend on assumptions made on μ and the geometry of its support.

Let us mention here just one result, in the case μ is supported on the graph of a function

$$\begin{array}{rcccl} f & : & \mathscr{X} \subset \mathbf{R}^n & \to & \mathscr{F} \subset \mathbb{R} \\ & & \mathbf{x} & \mapsto & x_{n+1} = f(x) \end{array}$$

i.e. the vector of moments is given by

$$\mathbf{y} = \int_{\mathscr{X}} \int_{\mathscr{F}} \mathbf{b}(\mathbf{x}, x_{n+1})\, d\mu(\mathbf{x}, x_{n+1}) = \int_{\mathscr{X}} \mathbf{b}(\mathbf{x}, f(\mathbf{x}))\, d\mathbf{x}$$

[5]An orthonormal matrix has columns $\mathbf{q}_i$ satisfying $\mathbf{q}_i^T \mathbf{q}_i = 1$ and $\mathbf{q}_i^T \mathbf{q}_j = 0$ if $i \neq j$.

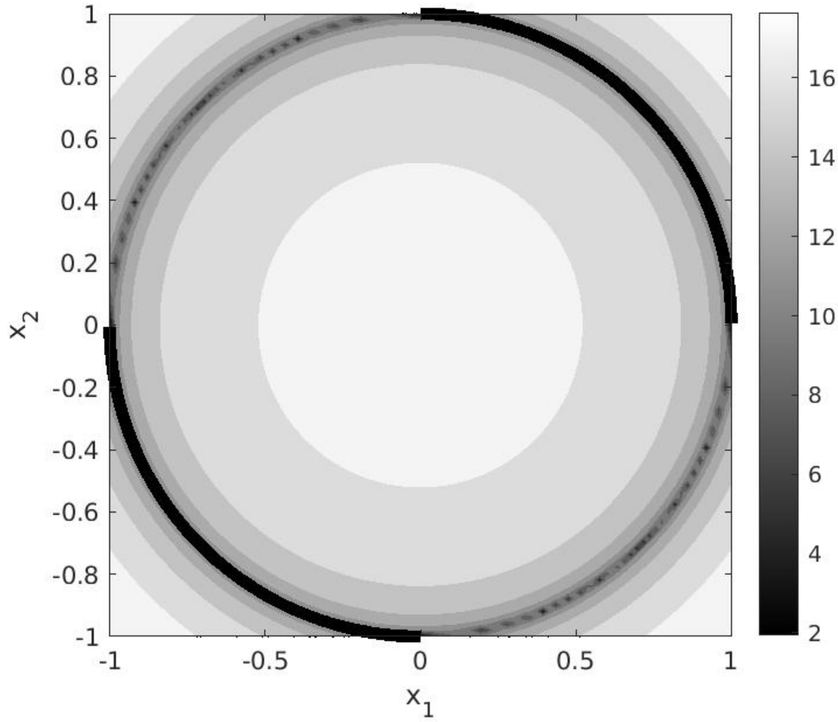


Figure 4.1: Logarithmic sublevel sets of the Christoffel-Darboux polynomial (gray) and argmin approximant (thick black).

for $\mathbf{b}$ a basis vector of polynomials of degree up to d in the joint variables $(\mathbf{x}, x_{n+1})$. Given $\mathbf{x} \in \mathscr{X}$, let us define[6] the Christoffel-Darboux approximant

$$f_{\beta_d}(\mathbf{x}) := \min\{\operatorname{argmin}_{x_{n+1}\in\mathscr{F}} p_{\beta_d}(\mathbf{x}, x_{n+1})\}. \tag{4.5}$$

Theorem 4.3. *With the choice $\beta_d = 2^{3-\sqrt{d}}$, if the set $\mathscr{S} \subset \mathscr{X}$ of continuity points of f is such that $\mathscr{X}\backslash\mathscr{S}$ has Lebesgue measure zero, then $\lim_{d\to\infty} f_{d,\beta_d}(\mathbf{x}) = f(\mathbf{x})$ for all $\mathbf{x} \in \mathscr{S}$ and $\lim_{d\to\infty} \int_{\mathscr{X}} |f(\mathbf{x}) - f_{d,\beta_d}(\mathbf{x})| d\mathbf{x} = 0$.*

Note that for each given $\mathbf{x} \in \mathscr{X}$, finding the argmin(s), i.e. the minimizer(s) $x^\star_{n+1} \in \mathscr{F}$ such that $p_{\beta_d}(\mathbf{x}, x^\star_{n+1}) = \min_{x_{n+1}\in\mathscr{F}} p_{\beta_d}(\mathbf{x}, x_{n+1})$, amounts to finding the roots of a univariate polynomial, a classical operation that boils down to numerical linear algebra.

Example 4.4. *The POP of Examples 3.2, 3.4 and 4.3 is such that the set of global minimizers is on the circular arcs $\{\mathbf{x} \in \mathbb{R}^2 : 1 - x_1^2 - x_2^2 = 0, x_1x_2 \geq 0\}$ which is also the graph $\{(x_1, x_2) : x_2 = f(x_1) \in [-1, 1], x_1 \in [-1, 1]\}$ of the function $f(x_1) := \operatorname{sign}(x_1)\sqrt{1 - x_1^2}$, see Figure 3.2. Consider the moment matrix $\mathbf{M}_2(\mathbf{y}^\star)$ of size 6 solving the second moment relaxation. With parameter $\beta_2 = 10^{-8}$, the sublevel sets of the Christoffel-Darboux polynomial p_{β_2} in (4.4) are represented in gray on Figure 4.1. We see that the small sublevel sets are concentrated on the unit circle. Also represented in thick black is the graph of function f_{β_2} in (4.5), an approximant of function f. The graph of f_{β_2} cannot be distinguished from the graph of f.*

[6]The set of minimizers is denoted by argmin.

4.5 Notes

The genericity result of Theorem 4.1 is from [27].

Theorem 4.2 is based on flat extension results of Curto and Fialkow in the 1990s [5], see [23, §5] or [20, §3.2.2].

The extraction algorithm to recover the solutions from the moments is described in [14] and explained in [23, §7.6]. It is implemented in the GloptiPoly software [12, 15].

For background material on polynomial ideals and algebraic varieties, see [4]. Lemma 4.3 is from [30, Prop. 1], see also [22, Ch. 5].

The Christoffel-Darboux polynomial and its applications are surveyed in [22]. Theorem 4.3 is from [26, Theorem 1], see also [22, §8.1].

Chapter 5

Limitations and extensions

Polynomial optimization problems (POPs) are difficult to solve in general, since they include as particular cases all combinatorial optimization problems. Still, the moment-SOS hierarchy performs well on most of the POPs encountered in practice. In particular, it behaves well on problems which have some specific features, such as convexity of the data, see e.g. [21, Ch. 13]. More generally, the relaxation order which is required to get a global optimality certificate can be interpreted as a measure of difficulty of a given POP.

The moment-SOS hierarchy is a numerical approach to POP: it relies on semidefinite optimization algorithms implemented in floating point arithmetic, and hence subject to conditioning problems and numerical inaccuracy. In some critical applications, this inaccuracy is not acceptable, and an exact solution is required. The solution to a POP with integer coefficients can be coded with algebraic numbers (i.e. roots of univariate polynomials with integer coefficients), and it can be obtained with computer algebra algorithms (e.g. Gröbner basis computations). Since the main computational ingredient in the moment-SOS hierarchy is semidefinite optimization, we can solve exactly a POP at the price of solving exactly semidefinite optimization problems. Computer algebra algorithms have been developed for that [16, 17], but they are currently limited to semidefinite optimization problems of small dimensions.

More generally, a main limitation of the moment-SOS hierarchy is the rapid growth of the size of the semidefinite optimization problems as a function of the relaxation order r. In the moment relaxation of order r of a POP with n variables, the number of moments is the binomial coefficient n_{2r}, and the size of the moment matrix is n_r. These numbers grow quickly with n and r. For example, if $n = 43$, the second ($r = 2$) moment relaxation of a quadratic POP features a moment matrix of size almost 1000 (and hence almost half a million entries) that depends on a vector of moments of size almost 200000, a challenging semidefinite optimization problem for a standard laptop at the time of writing these notes (2023).

For POPs with a moderate number of constraints, the computational complexity of solving the moment relaxation depends mostly on the size of the largest semidefinite matrix. So it is crucial to exploit as much as possible the problem structure to reduce this size, and decompose the large semidefinite matrices into sums of smaller semidefinite matrices. This can be achieved by exploiting symmetry and/or sparsity, see [23, §8], [21, Ch. 8] or [24].

Numerically it is generally preferable to use other bases than monomials to manipulate polynomials. In numerical linear algebra, Chebyshev polynomials of degrees by the thousands can routinely be used for function approximation [38]. Whether semidefinite solvers used in the moment-SOS hierarchy can benefit computationally from alternative polynomial bases remains to be seen.

Since the mid 2000s, there has been a lot of research activities to develop and apply the moment-SOS hierarchy beyond POPs. This includes e.g. polynomial optimal control and polynomial partial differential equations, but also many other problems in data science [11].

Index

Bibliography

[1] S. Basu, R. Pollack, M.-F. Roy. Algorithms in real algebraic geometry. Springer, 2006.

[2] A. Ben-Tal, A. Nemirovski. Lectures on modern convex optimization. SIAM, 2001.

[3] G. Blekerman, P. A. Parrilo, R. R. Thomas (Editors). Semidefinite optimization and convex algebraic geometry. SIAM, 2013.

[4] D. A. Cox, J. B. Little, D. O'Shea. Ideals, varieties, and algorithms. 3rd edition, Springer, 2007.

[5] R. Curto, L. Fialkow. Recursiveness, positivity, and truncated moment problems. Houston J. Math. 17:603-635, 1991.

[6] J. W. Helton, V. Vinnikov. Linear matrix inequality representation of sets. Comm. Pure Appl. Math. 60(5):654-674, 2007.

[7] D. Henrion. On semidefinite representations of plane quartics, LAAS-CNRS Research Report No. 08444, 2008.

[8] D. Henrion. Detecting rigid convexity of bivariate polynomials. Lin. Alg. Appl. 432:1218-1233, 2010.

[9] D. Henrion. Semidefinite representation of convex hulls of rational varieties. Acta Appl. Math. 115(3):319-327, 2011.

[10] D. Henrion. Semidefinite characterisation of invariant measures for one-dimensional discrete dynamical systems. Kybernetika 48(6):1089-1099, 2012.

[11] D. Henrion, M. Korda and J. B. Lasserre. The moment-SOS hierarchy. World Scientific, 2020.

[12] D. Henrion, J. B. Lasserre. GloptiPoly: Global optimization over polynomials with Matlab and SeDuMi. ACM Trans. Math. Software 29(2):165-194, 2003.

[13] D. Henrion, J. B. Lasserre. Convergent LMI relaxations for non-convex optimization over polynomials in control. IEEE Control Systems Magazine 24(3):72-83, 2004.

[14] D. Henrion, J. B. Lasserre. Detecting global optimality and extracting solutions in GloptiPoly. In D. Henrion, A. Garulli (Editors). Positive polynomials in control. LNCIS, Springer, 2005.

[15] D. Henrion, J. B. Lasserre, J. Löfberg. GloptiPoly 3: moments, optimization and semidefinite programming. Optim. Methods and Software 24(4-5):761-779, 2009.

[16] D. Henrion, S. Naldi, M. Safey El Din. Exact algorithms for linear matrix inequalities. SIAM J. Optim. 26(4):2512-2539, 2016.

[17] D. Henrion, S. Naldi, M. Safey El Din. SPECTRA - a Maple library for solving linear matrix inequalities in exact arithmetic. Optim. Methods and Software 34(1):62-78, 2019.

[18] D. Henrion, M. Šebek, V. Kučera. Positive polynomials and robust stabilization with fixed-order controllers. IEEE Trans. Autom. Control 48(7):1178-1186, 2003.

[19] J. B. Lasserre. Global optimization with polynomials and the problem of moments. SIAM J. Optim. 11(3):796–817, 2001.

[20] J. B. Lasserre. Moments, positive polynomials and their applications. Imperial College Press, 2009.

[21] J. B. Lasserre. An Introduction to polynomial and semi-algebraic optimization. Cambridge Univ. Press, 2015.

[22] J. B. Lasserre, E. Pauwels, M. Putinar. The Christoffel-Darboux kernel for data analysis. Cambridge Univ. Press, 2022.

[23] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In M. Putinar, S. Sullivant (Editors). Emerging applications of algebraic geometry. IMA Vol. Math. Appl., Springer, 2009.

[24] V. Magron, J. Wang. Sparse polynomial optimization. World Scientific, 2023.

[25] M. Marshall. Positive polynomials and sums of squares. AMS, 2008.

[26] S. Marx, E. Pauwels, T. Weisser, D. Henrion, J. B. Lasserre. Tractable semi-algebraic approximation using Christoffel-Darboux kernel. Constructive Approx. 54(3):391-429, 2021.

[27] J. Nie. Optimality conditions and finite convergence of Lasserre's hierarchy. Math. Prog. 146:97-121, 2014.

[28] J. Nie. Moments and polynomial optimization. SIAM, 2023.

[29] P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. Math. Prog. 96:293-320, 2003.

[30] E. Pauwels, M. Putinar, J. B. Lasserre. Data analysis from empirical moments and the Christoffel function. Found. Comput. Math. 21(1):243-273, 2021.

[31] M. Putinar. Positive polynomials on compact semi-algebraic sets. Indiana Univ. Math. J. 42:969-984, 1993.

[32] J. Renegar. Hyperbolic programs, and their derivative relaxations. Found. Comput. Math. 59–79, 2005.

[33] H. Royden, P. Fitzpatrick. Real analysis. 4th edition. Prentice Hall, 2010.

[34] C. Scheiderer. Semidefinite representation for convex hulls of real algebraic curves SIAM J. Appl. Algebra Geom. 2(1):1-25, 2018.

[35] C. Scheiderer. Spectrahedral shadows. SIAM J. Appl. Algebra Geom. 2(1):26-44, 2018.

[36] K. Schmüdgen. The moment problem. Springer, 2017.

[37] J. F. Sturm. Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. Optim. Methods and Software 11-12:625-633, 1999.

[38] L. N. Trefethen. Approximation theory and approximation practice, SIAM, 2013