An Overview of Convergence Rates for Sum of Squares Hierarchies in Polynomial Optimization

Monique Laurent⬤ **and Lucas Slot**⬤

Abstract In this chapter we consider polynomial optimization problems, asking to minimize a polynomial function over a compact semialgebraic set, defined by polynomial inequalities. This models a great variety of (in general, nonlinear nonconvex) optimization problems. Various hierarchies of (lower and upper) bounds have been introduced, having the remarkable property that they converge asymptotically to the global minimum. These bounds exploit algebraic representations of positive polynomials in terms of sums of squares and can be computed using semidefinite optimization. Our focus in this chapter lies in the performance analysis of these hierarchies of bounds, namely, in how far the bounds are from the global minimum as the degrees of the sums of squares they involve tend to infinity. We present the main state-of-the-art results and offer a gentle introductory overview over the various techniques that have been recently developed to establish them, stemming from the theory of orthogonal polynomials, approximation theory, Fourier analysis, and more.

1 Introduction

This paper offers a gentle introduction and overview over the *design and performance analysis of approximation hierarchies for polynomial optimization*. In this section we will first introduce polynomial optimization, and its use for modeling hard optimization problems arising within a broad range of fields and application domains. After that we will explain how to design hierarchies of approximations using sums of squares of polynomials as a tractable surrogate for polynomial nonnegativity, and how sums of squares of polynomials can be modeled using semidefinite optimization.

M. Laurent (✉)
CWI Amsterdam, Amsterdam, XG, The Netherlands
e-mail: M.Laurent@cwi.nl

Tilburg University, Tilburg, AB, The Netherlands

L. Slot
ETH, Zürich, Switzerland
e-mail: lucas.slot@inf.ethz.ch

1.1 Polynomial Optimization

Throughout we use the following notation. We let $\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, \ldots, x_n]$ denote the ring of n-variate real polynomials in the variables $\mathbf{x} = (x_1, \ldots, x_n)$. For a multi-index $\alpha \in \mathbb{N}^n$, we let $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ denote the associated monomial, whose degree is $|\alpha| := \|\alpha\|_1 = \alpha_1 + \ldots + \alpha_n$. We let $\mathbb{N}_d^n$ denote the set of sequences $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq d$. Any polynomial $f \in \mathbb{R}[\mathbf{x}]$ can be expressed in the monomial basis as $f = \sum_\alpha f_\alpha x^\alpha$ for some coefficients $f_\alpha \in \mathbb{R}$. Here, the sum is finite and the largest value of $|\alpha|$ for which $f_\alpha \neq 0$ is the *degree* of f, denoted $\deg(f)$. For an integer $d \geq 0$, we let $\mathbb{R}[\mathbf{x}]_d$ denote the set of polynomials with degree at most d. It is a vector space, with $[x]_d = \{x^\alpha : \alpha \in \mathbb{N}_d^n\}$ as its standard monomial basis. As we will see later, it is sometimes convenient to use other polynomial bases, that are orthogonal with respect to some inner product induced by a selected measure on the space $\mathbb{R}^n$.

The general setting of polynomial optimization is as follows. We are given polynomials $f, g_1, \ldots, g_m \in \mathbb{R}[\mathbf{x}]$ and the goal is to find the minimum value that f takes on the feasible region defined by the polynomials g_j. In other words, the goal is to solve the following optimization problem.

The polynomial optimization problem

$$f_{\min} = \min\{f(\mathbf{x}) : g_1(\mathbf{x}) \geq 0, \ldots, g_m(\mathbf{x}) \geq 0\}$$
$$= \min\{f(\mathbf{x}) : \mathbf{x} \in \mathbf{X}\}, \tag{1}$$

after setting

$$\mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_1(\mathbf{x}) \geq 0, \ldots, g_m(\mathbf{x}) \geq 0\}. \tag{2}$$

We also sometimes use the notation $f_{\min,\mathbf{X}}$ for the minimum value of (1) and the notation $\mathrm{Se}(\mathbf{g})$ for the set in (2).

Such a set $\mathbf{X}$ is known as a basic closed semialgebraic set. Typical instances include the unit sphere $\mathbb{S}^{n-1}$, the standard simplex Δ^n, the unit ball B^n, the box $[-1, 1]^n$ or $[0, 1]^n$, and the binary cube $\{\pm 1\}^n$ or $\{0, 1\}^n$, where

$$\mathbb{S}^{n-1} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\}, \ B^n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq 1\},$$

$$\Delta^n = \left\{\mathbf{x} \in \mathbb{R}^n : x_i \geq 0 \ (i \in [n]), \sum_{i=1}^n x_i = 1\right\},$$

and $\|\mathbf{x}\|^2 = \sum_{i=1}^n x_i^2$ denotes the squared Euclidean norm. Throughout we assume $\mathbf{X}$ is compact. So, problem (1) always has a global minimizer.

The polynomial optimization problem (1) contains linear programming as a (very) special case, when all polynomials $f, g_1, \ldots, g_m$ are linear (i.e., of degree at most 1). In this case, efficient optimization algorithms exist (see, e.g., Roos et al. (2005)). But,

it also models a much broader range of problems, which are generally nonlinear and nonconvex. This includes well-known NP-hard problems, already when restricting to seemingly simple feasible regions $\mathbf{X}$ in (2), such as the simplex, the box, or the sphere as we illustrate on some examples in the next section.

Polynomial optimization has received growing research interest in the past decades, when it was realized that algebraic and geometric properties of polynomials could be exploited to design dedicated methods, able to capture the *global minimum*, in contrast to general nonlinear optimization methods where one can often only gain information about local minima. In a nutshell, this research direction builds on combining real algebraic geometry results (about sums of squares of polynomials) and functional analytical results (about moments of measures) with semidefinite optimization. It roots in foundational works, in particular, by Shor (1987), Nesterov (2000), Lasserre (2001) and Parrilo (2000, 2003). The field has substantially grown[1] and has a broad literature. We mention some books and overviews that can serve as introduction to the topic and give further references to many applications and additional aspects that are not mentioned in the present paper; in particular, by Lasserre (2009), Lasserre (2015), Laurent (2009), Blekherman et al. (2012), Henrion et al. (2020), Magron and Wang (2023) and Nie (2023).

The present paper will focus on the performance analysis of various hierarchies of bounds that have been introduced for the polynomial optimization problem (1), based on using tailored sums of squares representations for positive polynomials. We will recall the definition of these hierarchies of (upper and lower) bounds and discuss the main state-of-the-art results that have been shown in recent years about their quantitative convergence properties (these are not covered in the literature mentioned above). We focus on offering a gentle overview of the main techniques that are needed to prove these quantitative results.

1.2 Examples and Applications

We begin with mentioning a few instances of the polynomial optimization problem (1) that capture well-known hard combinatorial optimization problems.

Consider a graph $G = (V, E)$, where $V = [n] = \{1, \ldots, n\}$ is the set of vertices and the pairs in $E \subseteq V \times V$ correspond to the edges of G. A set $I \subseteq V$ is said to be independent (or stable) if it contains no edge and a fundamental combinatorial problem is determining the largest cardinality of an independent set, denoted $\alpha(G)$, a well-known NP-hard problem (see Garey and Johnson (1979)). Interestingly, this problem admits several equivalent reformulations as instances of polynomial optimization over the boolean cube, the box, the simplex, the sphere, respectively:

---

[1] This is witnessed, e.g., by the fact that it has received its own Mathematics Subject Classification number 90C23.

$$\alpha(G) = \max \left\{ \sum_{i \in V} x_i : x_i x_j = 0 \text{ for } \{i, j\} \in E, x_i^2 = x_i \text{ for } i \in V \right\},$$

$$\alpha(G) = \max \left\{ \sum_{i \in V} x_i - \sum_{\{i,j\} \in E} x_i x_j : x \in [0, 1]^n \right\},$$

$$\frac{1}{\alpha(G)} = \min \left\{ x^\mathsf{T}(I + A_G)x : x \in \mathbb{R}_+^n, \sum_{i \in V} x_i = 1 \right\},$$

$$\frac{1}{\alpha(G)} = \min \left\{ (x^{\circ 2})^\mathsf{T}(I + A_G)x^{\circ 2} : x \in \mathbb{R}^n, \sum_{i \in V} x_i^2 = 1 \right\},$$

where we set $x^{\circ 2} = (x_1^2, \ldots, x_n^2)^\mathsf{T}$ and A_G is the adjacency matrix of graph G. An additional reformulation is in terms of linear optimization over copositive matrices:

$$\alpha(G) = \min\{\lambda : \lambda(I + A_G) - J \text{ is copositive}\},$$

where J denotes the all-ones matrix. We refer to de Klerk and Pasechnik (2002) for details about the above formulations.

Another fundamental combinatorial problem is the maximum cut problem, asking for the largest cardinality of a cut in a graph $G = (V, E)$, denoted as mc(G), that can be formulated as any of the following polynomial optimization problems:

$$\mathrm{mc}(G) = \max \left\{ \sum_{\{i,j\} \in E} \frac{1}{2}(1 - x_i x_j) : x \in \{\pm 1\}^n \right\},$$

$$\mathrm{mc}(G) = \max \left\{ \frac{1}{4} x^\mathsf{T} L_G x : x \in [-1, 1]^n \right\},$$

where $L_G = D - A_G$ is the Laplacian matrix, with D the diagonal matrix having the degrees of the vertices as diagonal entries.

A fundamental question in analysis that can be formulated in terms of polynomial optimization is testing whether an n-variate polynomial f is convex. Indeed, f is convex if and only if its Hessian matrix $H_f(\mathbf{x}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \right)_{i,j=1}^n$ is positive semidefinite at any $\mathbf{x} \in \mathbb{R}^n$ or, equivalently, if the $2n$-variate polynomial $F(\mathbf{x}, \mathbf{y}) = \mathbf{y}^\mathsf{T} H_f(\mathbf{x})\mathbf{y}$ is nonnegative over $\mathbb{R}^n \times \mathbb{R}^n$ (i.e., its global minimum is 0). It has been shown by Ahmadi et al. (2013) that testing whether a quartic polynomial is convex is an NP-hard problem. This implies hardness of testing global nonnegativity of a quartic polynomial.

So, the above examples show that computing the global minimum of a polynomial over simple regions such as the simplex, the sphere, the box, the full space, or the boolean box, are computationally hard already when restricting to small degree (at most 4). Another notable application which can be cast as instance of polynomial optimization problem involving only quadratic polynomials, is the optimal power flow problem in energy (see, e.g., Zohrizadeh et al. (2020)). Polynomial optimization has a broad modeling power and permits to capture problems from various areas,

such as probablity, mathematical finance, control, game theory, for which we refer to the exposition by Lasserre (2009).

1.3 Nonnegative Polynomials and Sums of Squares

A polynomial $f \in \mathbb{R}[\mathbf{x}]_{2d}$ is called a *sum of squares* (abbreviated as *sos*) if it can be written as a sum of squares of other polynomials, i.e., $f = \sum_{j=1}^{k} q_j^2$ for some $q_j \in \mathbb{R}[\mathbf{x}]$ and $k \geq 1$. Then, each q_j has degree at most d and can be assumed to be homogeneous if f is homogeneous. We let Σ denote the set of sums of squares of polynomials and set $\Sigma_{2d} = \Sigma \cap \mathbb{R}[\mathbf{x}]_{2d}$. We may use the notation $\Sigma[\mathbf{x}]$ to stress which variables are used.

Let $\mathcal{P}$ denote the set of polynomials f that are globally nonnegative (i.e., $f(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$), and $\mathcal{P}(\mathbf{X})$ the set of polynomials that are nonnegative over a given set $\mathbf{X} \subseteq \mathbb{R}^n$.

Clearly, every sum of squares of polynomials is globally nonnegative, i.e., the inclusion $\Sigma \subseteq \mathcal{P}$ holds. As is well-known, this inclusion is in general strict. By a result of Hilbert (1888) we know that any globally nonnegative n-variate polynomial with degree $2d$ is a sum of squares of polynomials only in the following three cases: $n = 1$ (univariate), $d = 1$ (quadratic), and the exceptional case ($n = 2, d = 2$) (i.e., quartic in two variables).

A nonnegative polynomial that is not sos

The first explicit example of a nonnegative polynomial that is not sos was found by Motzkin in 1967, it is bivariate with degree 6 and reads

$$f(x, y) = x^4 y^2 + x^2 y^4 - 3x^2 y^2 + 1. \tag{3}$$

It is depicted in Fig. 1a. To see global nonnegativity one may use the arithmetic-geometric mean inequality, and use 'brute force' to show f is not sos (write f as a sum of squares of polynomials and compare coefficients). Also the Robinson polynomial

$$f(x, y, z) = x^6 + y^6 + z^6 - x^2 y^2 (x^2 + y^2) - x^2 z^2 (x^2 + z^2) - y^2 z^2 (y^2 + z^2) + 3x^2 y^2 z^2 \tag{4}$$

is globally nonnegative but not sos. We refer to Reznick (1996) and Powers (2025) (and references therein) for a nice historic discussion and many more examples.

At the 1900 International Congress of Mathematicians in Paris Hilbert asked whether every nonnegative polynomial can be written as a sum of squares of *rational* functions (known as *Hilbert's seventeenth problem*). This question was settled in the affirmative by Artin (1927), a result that started the flourishing field of real
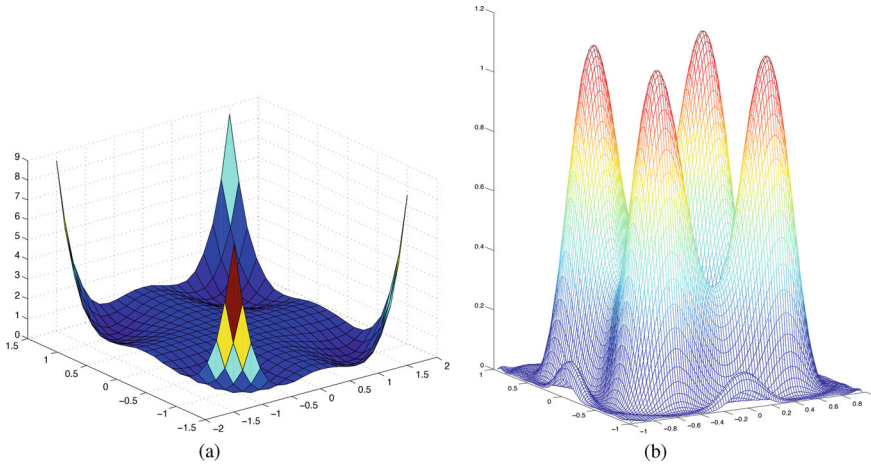
Fig. 1 **a** On the left: the Motzkin polynomial (3); **b** on the right: the optimal degree 16 sum-of-squares polynomial σ in program (14)

algebraic geometry (see, e.g., Prestel and Delzell (2001) and Marshall (2008)). A landmark result is the Positivstellensatz[2] by Krivine (1964) and Stengle (1996) that a.o. characterizes the polynomials that are nonnegative on a semialgebraic set **X** as in (2). To cite this result we need the notion of *preordering* $\mathcal{T}(\mathbf{g})$ generated by the polynomials $\mathbf{g} = \{g_1, \ldots, g_m\}$ entering the algebraic description of the set **X**:

$$\mathcal{T}(\mathbf{g}) = \Big\{ \sum_{J \subseteq [m]} \sigma_J \prod_{j \in J} g_j : \sigma_J \in \Sigma \text{ for } J \subseteq [m] \Big\} \tag{5}$$

(setting $g_\emptyset = 1$). Clearly, $\mathcal{T}(\mathbf{g}) \subseteq \mathcal{P}(\mathbf{X})$. Krivine-Stengle show that a polynomial $f \in \mathbb{R}[x]$ is nonnegative on **X** if and only if $p_1 f = f^{2k} + p_2$ for some polynomials $p_1, p_2 \in \mathcal{T}(\mathbf{g})$ and some integer $k \in \mathbb{N}$. This leads to a sos-type decomposition of f as $f = f^{2k}/p_1 + p_2/p_1$, thus 'with a denominator'.

Simpler sos-type decompositions have been shown later under more restrictive assumptions, typically assuming *strict positivity* of f. In particular, Reznick (1995) shows a sharper result for *homogeneous* polynomials: If f is homogeneous and satisfies $f(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$, then there exists an integer $r \geq 0$ such that $(\sum_{i=1}^n x_i^2)^r f(\mathbf{x}) \in \Sigma$.

When the set **X** is compact, Schmüdgen (1991) shows membership in the pre-ordering under strict positivity:

---

[2] The terminology of *Positivtsellensatz* refers to Hilbert's celebrated Nullstellenzatz that characterizes the polynomials vanishing at a given complex variety.

Schmüdgen's Positivstellensatz

If $\mathbf{X}$ is compact and f is strictly positive on $\mathbf{X}$, then $f \in \mathcal{T}(\mathbf{g})$.

Under the Archimedean condition, Putinar (1993) shows membership in the *quadratic module* $Q(\mathbf{g})$, which is defined as

$$Q(\mathbf{g}) = \Big\{ \sum_{j=0}^{m} \sigma_j g_j : \sigma_j \in \Sigma \text{ for } j \in \{0, 1, \ldots, m\} \Big\} \qquad (6)$$

(setting $g_0 = 1$). So, $Q(\mathbf{g}) \subseteq \mathcal{T}(\mathbf{g}) \subseteq \mathcal{P}(\mathbf{X})$. The advantage of the quadratic module $Q(\mathbf{g})$ over the preordering $\mathcal{T}(\mathbf{g})$ is that it involves less terms: $m + 1$ sos polynomials for $Q(\mathbf{g})$, instead of 2^m for $\mathcal{T}(\mathbf{g})$.

The Archimedean condition asks that $R - \sum_{i=1}^{n} x_i^2 \in Q(\mathbf{g})$ for some $R > 0$ and can be seen as an algebraic certificate of compactness. It is a property of the algebraic description of $\mathbf{X}$ rather than $\mathbf{X}$ itself. It is, however, easy to satisfy, simply by adding the inequality of a ball containing $\mathbf{X}$ to its description. The Archimedean condition holds for most sets $\mathbf{X}$ of interest in applications such as the sphere, box, simplex, etc.

Putinar's Positivstellensatz

If $Q(\mathbf{g})$ is Archimedean and f is strictly positive on $\mathbf{X}$, then $f \in Q(\mathbf{g})$.

So, sums of squares belong to a classical topic in real algebraic geometry, with a rich history going back to early work by Hilbert. It is only recently that their relevance to optimization has been fully appreciated, starting with ground works by Shor (1987), Nesterov (2000), Lasserre (2001), and Parrilo (2000, 2003). A key ingredient for this link to optimization is the fact that sums of squares can be modeled using semidefinite programs, which makes them amenable to numerical algorithms. Next, we explain how to model sos polynomials using semidefinite programming, and thereafter how to use sos polynomials to define hierarchies of (lower and upper) bounds for the original polynomial optimization problem (1).

1.4 Sums of Squares and Semidefinite Optimization

We begin with a quick recap on semidefinite optimization, and refer, e.g., to de Klerk (2002) for a detailed treatment. A semidefinite program (in primal form) reads

$$p^* = \sup\{\langle C, X \rangle : \langle A_j, X \rangle = b_j \ (j \in [m]), X \in \mathcal{S}_+^N\}. \qquad (7)$$

Here, $C, A_1, \ldots, A_m \in \mathcal{S}^N$ are symmetric $N \times N$ matrices and $b \in \mathbb{R}^m$–these are the data of the problem– and $X \in \mathcal{S}_+^N$ is the matrix variable, required to be symmetric

and positive semidefinite (also written as $X \succeq 0$). So, the program (7) is a linear program over the cone $\mathcal{S}_+^N$ of positive semidefinite matrices. Its dual program reads

$$d^* = \inf \Big\{ \sum_{j=1}^m b_j y_j : \sum_{j=1}^m y_j A_j - C \in \mathcal{S}_+^N \Big\}. \tag{8}$$

Weak duality holds: $p^* \leq d^*$, and strong duality $p^* = d^*$ holds under some Slater-type conditions. Semidefinite programs contain linear programs as a special case (when all data matrices C, A_j are diagonal). The crucial fact is that semidefinite programs can be solved efficiently up to any given precision, under some assumptions (like knowing a small ball inside the feasible region of (7) and a ball enclosing it).

As we now see, sums of squares can be modeled with semidefinite programs. As a warm-up observe that a quadratic form $f = x^T M x$ is a sum of squares precisely when the matrix M is positive semidefinite. Consider now a polynomial $f = \sum_{\alpha \in \mathbb{N}_{2d}^n} f_\alpha x^\alpha$ of even degree $2d$ for which we wish to decide whether $f \in \Sigma$. Assume $f \in \Sigma$, i.e., $f = q_1^2 + \cdots + q_k^2$ for some polynomials q_j. Then, each $q_j \in \mathbb{R}[\mathbf{x}]_d$ can be written in the monomial basis as $q_j = [\mathbf{x}]_d^\mathsf{T} \mathbf{a}_j$, where $\mathbf{a}_j = ((\mathbf{a}_j)_\beta)_{\beta \in \mathbb{N}_d^n}$ is the vector of coefficients of the polynomial q_j in the monomial basis $[\mathbf{x}]_d$. Thus, we obtain

$$f = \sum_{j=1}^k q_j^2 = \sum_{j=1}^m [\mathbf{x}]_d^\mathsf{T} \mathbf{a}_j \mathbf{a}_j^\mathsf{T} [\mathbf{x}]_d = [\mathbf{x}]_d \Big(\sum_{j=1}^k \mathbf{a}_j \mathbf{a}_j^\mathsf{T} \Big)[\mathbf{x}]_d = [\mathbf{x}]_d^\mathsf{T} Q [\mathbf{x}]_d,$$

setting $Q := \sum_{j=1}^k \mathbf{a}_j \mathbf{a}_j^\mathsf{T}$. By construction the matrix Q is positive semidefinite and, by equating coefficients at both sides of the above polynomial identity, we arrive at the following characterization.

Modeling sums of squares with semidefinite programs
A polynomial $f = \sum_{\alpha \in \mathbb{N}_{2d}^n} f_\alpha x^\alpha \in \mathbb{R}[\mathbf{x}]_{2d}$ is a sum of squares of polynomials if and only if the following semidefinite program has a feasible solution:

$$Q \in \mathcal{S}_+^{N_d}, \qquad \sum_{\beta, \gamma \in \mathbb{N}_d^n : \beta + \gamma = \alpha} Q_{\beta, \gamma} = f_\alpha \ \text{ for all } \alpha \in \mathbb{N}_{2d}^n, \tag{9}$$

where the matrix variable Q is indexed by $\mathbb{N}_d^n$ and $N_d = |\mathbb{N}_d^n| = \binom{n+d}{d}$.

Note that the choice of the monomial basis is not important in the above derivation, it would work mutatis mutandis using any other polynomial basis of the polynomial space (this fact will be used in Sect. 2.2).

Now that we know how to express sos polynomials using semidefinite programs we can also express polynomials in $Q(\mathbf{g})_{2r}$, the quadratic module truncated at a given degree $2r$, defined by

$$Q(\mathbf{g})_{2r} = \Big\{ \sum_{j=0}^{m} \sigma_j g_j : \sigma_j \in \Sigma, \deg(\sigma_j g_j) \le 2r \text{ for } j = 0, 1, \ldots, m \Big\}, \qquad (10)$$

or in the truncated preordering $\mathcal{T}(\mathbf{g})_{2r}$ defined by

$$\mathcal{T}(\mathbf{g})_{2r} = \Big\{ \sum_{J \subseteq [m]} \sigma_J \prod_{j \i J} g_j : \sigma_J \in \Sigma, \deg\Big(\sigma_J \prod_{j \in J} g_j\Big) \le 2r \text{ for } J \subseteq [m] \Big\}. \quad (11)$$

For this, set $d_j = \lceil \deg(g_j)/2 \rceil$, so that polynomials in $Q(\mathbf{g})_{2r}$ can be modeled as $\sum_{j=0}^{m} g_j [\mathbf{x}]_{r-d_j}^{\mathsf{T}} Q_j [\mathbf{x}]_{r-d_j}$ for some positive semidefinite matrices $Q_j \in \mathcal{S}^{N_{d_j}}$ (for $j = 0, \ldots, m$). In other words, membership in the truncated quadratic module can be modeled as a semidefinite program. The same holds of course for membership in the truncated preordering.

We now have all tools in hands to define the hierarchies of upper bounds and lower bounds for the polynomial optimization problem (1).

1.5 Upper Bounds

To define upper bounds on $f_{\min}$ the starting point is to observe that problem (1) can be reformulated as a linear optimization problem over the set $\mathcal{M}(\mathbf{X})$ of positive Borel measures supported within the set $\mathbf{X}$:

$$f_{\min} = \min \Big\{ \int_{\mathbf{X}} f(\mathbf{x}) d\nu(\mathbf{x}) : \nu \in \mathcal{M}(\mathbf{X}), \int_{\mathbf{X}} d\nu(\mathbf{x}) = 1 \Big\}. \qquad (12)$$

The argument is simple: On the one hand, $\int_{\mathbf{X}} f d\nu \ge f_{\min}$ since ν is a probability measure on $\mathbf{X}$. On the other hand, the Dirac delta $\nu = \delta_{\mathbf{x}^*}$ at a global minimizer $\mathbf{x}^*$ of f in $\mathbf{X}$ provides a feasible solution to (12) with value $f(\mathbf{x}^*) = f_{\min}$.

Let is now fix a (reference) measure $\mu \in \mathcal{M}(\mathbf{X})$ whose support is equal to $\mathbf{X}$. The next step is that one can restrict the optimization in (12) to the measures ν that have a sos density with respect to this given measure μ:

$$f_{\min} = \min \Big\{ \int_{\mathbf{X}} f(\mathbf{x}) \sigma(\mathbf{x}) d\mu(\mathbf{x}) : \sigma \in \Sigma, \int_{\mathbf{X}} \sigma(\mathbf{x}) d\mu(\mathbf{x}) = 1 \Big\} \qquad (13)$$

(Lasserre (2010)). Intuitively, this relies on the fact that the Dirac delta $\delta_{\mathbf{x}^*}$ can be well approximated by sos polynomials. The upper bounds on $f_{\min}$ are then obtained by restricting the optimization to sos polynomials of a given degree.

The upper bounds on $f_{\min}$

For any $r \in \mathbb{N}$, define the parameter

$$\mathrm{ub}(f, \mathbf{X}, \mu)_r = \min \left\{ \int_{\mathbf{X}} f(\mathbf{x})\sigma(\mathbf{x})d\mu(\mathbf{x}) : \sigma \in \Sigma_{2r}, \int_{\mathbf{X}} \sigma(\mathbf{x})d\mu(\mathbf{x}) = 1 \right\}. \tag{14}$$

We have $f_{\min} \leq \mathrm{ub}(f, \mathbf{X}, \mu)_{r+1} \leq \mathrm{ub}(f, \mathbf{X}, \mu)_r$.

As an illustration, we show in Fig. 1a the Motzkin polynomial and in (b) the optimal sos density σ for program (14) at order $2r = 16$ (which approximates well the sum of Dirac functions at the four minimizers $(\pm 1, \pm 1)$).

So, the parameter (14) depends on the choice of the reference measure μ on $\mathbf{X}$ and it can be computed via a semidefinite program (in fact, as an eigenvalue problem as we see in Sect. 2.2). In view of (13), the bounds $\mathrm{ub}(f, \mathbf{X}, \mu)_r$ converge asymptotically to $f_{\min}$ as $r \to \infty$. We return to these bounds in Sect. 2, where we will discuss in detail the convergence rate of the error range $\mathrm{ub}(f, \mathbf{X}, \mu)_r - f_{\min}$ as the relaxation order r grows.

For now, we summarize in Table 1 the known results about this convergence rate for various classes of compact sets X (and reference measures μ). As we see there, one can show a convergence rate in $O(1/r^2)$ (up to a $\log^2(r)$ factor) for a large class of compact sets. For instance, one can show a rate in $O(\log^2(r)/r^2)$ for any convex body $\mathbf{X}$, and in $O(1/r^2)$ when $\mathbf{X}$ is 'round', which means that there are tangent inner and outer balls at any point on its boundary. The convergence rate in $O(1/r^2)$ is essentially optimal, as can be seen for the case $\mathbf{X} = [-1, 1]$ (due to the explicit link to extremal roots of orthogonal polynomials, see relation (23) in Sect. 2.3), or for the case $\mathbf{X} = \mathbb{S}^{n-1}$ (due to a link to cubature rules, see Sect. 1.7).

1.6 Lower Bounds

To define lower bounds on $f_{\min}$ the starting point is to observe that the minimum value taken by the polynomial f over the set $\mathbf{X}$ is equal to the largest scalar λ for which the polynomial $f - \lambda$ is nonnegative over $\mathbf{X}$:

$$f_{\min} = \sup\{\lambda : \lambda \in \mathbb{R}, \ f(\mathbf{x}) - \lambda \geq 0 \ \text{ for all } \mathbf{x} \in \mathbf{X}\}. \tag{15}$$

Now, we obtain bounds by replacing the positivity condition by membership in either the truncated quadratic module $Q(\mathbf{g})_{2r}$ or preordering $\mathcal{T}(\mathbf{g})_{2r}$ (see (10), (11)).

Table 1 Overview of known results on the asymptotic error $\text{ub}(f, \mathbf{X}, \mu)_r - f_{\min}$ of Lasserre's hierarchy of upper bounds

$\mathbf{X}$ (compact)	Error	Measure μ	References
Geometric assumption	$O(1/\sqrt{r})$	Lebesgue	de Klerk et al. (2017)
Convex body	$O(1/r)$	Lebesgue	de Klerk and Laurent (2018)
Semialgebraic with dense interior, convex body	$O(\log^2(r)/r^2)$	Lebesgue	Slot and Laurent (2021)
$\mathbb{S}^{n-1}$	$O(1/r)$	Uniform	Doherty and Wehner (2013)
$\mathbb{S}^{n-1}$	$O(1/r^2)$	Uniform	de Klerk and Laurent (2022)
$[-1, 1]^n$	$O(1/r^2)$	$\prod_i (1 - \mathbf{x}_i)^\lambda d\mathbf{x}$ ($\lambda = -\frac{1}{2}$)	de Klerk and Laurent (2020)
$[-1, 1]^n$	$O(1/r^2)$	$\prod_i (1 - \mathbf{x}_i)^\lambda d\mathbf{x}$ ($\lambda \geq -\frac{1}{2}$)	Slot and Laurent (2022d)
'Round' convex body	$O(1/r^2)$	Lebesgue	Slot and Laurent (2022d)
B^n	$O(1/r^2)$	$(1 - \|\mathbf{x}\|^2)^\lambda d\mathbf{x}$ ($\lambda \geq 0$)	Slot and Laurent (2022d)
Δ^n	$O(1/r^2)$	Lebesgue	Slot and Laurent (2022d)

The lower bounds on $f_{\min}$

For any integer $r \geq \deg(f)/2$, define the parameters

$$\text{lb}(f, Q(\mathbf{g}))_r = \sup\{\lambda : \lambda \in \mathbb{R}, \ f - \lambda \in Q(\mathbf{g})_{2r}\}, \qquad (16)$$

$$\text{lb}(f, \mathcal{T}(\mathbf{g}))_r = \sup\{\lambda : \lambda \in \mathbb{R}, \ f - \lambda \in \mathcal{T}(\mathbf{g})_{2r}\}. \qquad (17)$$

By definition, we have $\text{lb}(f, Q(\mathbf{g}))_r \leq \text{lb}(f, \mathcal{T}(\mathbf{g}))_r$ for all r. Furthermore, we have $\text{lb}(f, Q(\mathbf{g}))_r \leq \text{lb}(f, Q(\mathbf{g}))_{r+1} \leq f_{\min}$, and $\text{lb}(f, \mathcal{T}(\mathbf{g}))_r \leq \text{lb}(f, \mathcal{T}(\mathbf{g}))_{r+1} \leq f_{\min}$.

As an application of the earlier mentioned Positivstellensätze of Putinar and Schmüdgen, if $Q(\mathbf{g})$ is Archimedean (resp., $\mathbf{X}$ is compact), then the bounds $\text{lb}(f, Q(\mathbf{g}))_r$ (resp., $\text{lb}(f, \mathcal{T}(\mathbf{g}))_r$) converge to $f_{\min}$ as $r \to \infty$ (Lasserre (2001)). We return to these bounds in Sect. 3, where we will discuss their asymptotic convergence rates as the relaxation order r grows. We summarize the main known results in Table 2 below. As we see, the results depend on the algebraic structure of the semialgebraic set $\mathbf{X}$.

Table 2 Overview of known results on the asymptotic error $f_{\min} - \text{lb}(f, \mathbf{X})_r$ of Lasserre's hierarchies of lower bounds. ($\star$ The result on the sphere $\mathbb{S}^{n-1}$ is shown in Fang and Fawzi (2021) only for homogeneous polynomials f, but it can be extended to general polynomials Blomenhofer et al. (2025))

$\mathbf{X}$ (compact)	Error	Certificate	References
Archimedean	$O(1/\log(r)^c)$	$Q(\mathbf{g})$	Nie and Schweighofer (2007)
Archimedean	$O(1/r^c)$	$Q(\mathbf{g})$	Baldi and Mourrain (2023), and Baldi et al. (2025)
General	$O(1/r^c)$	$\mathcal{T}(\mathbf{g})$	Schweighofer (2004)
$[-1, 1]^n$	$O(1/r)$	$Q(\mathbf{g})$	Baldi and Slot (2024)
$\mathbb{S}^{n-1}$	$O(1/r^2)^\star$	$Q(\mathbf{g}) (= \mathcal{T}(\mathbf{g}))$	Fang and Fawzi (2021)
B^n	$O(1/r^2)$	$Q(\mathbf{g}) (= \mathcal{T}(\mathbf{g}))$	Slot (2022c)
Δ^n	$O(1/r^2)$	$\mathcal{T}(\mathbf{g})$	Slot (2022c)
$[-1, 1]^n$	$O(1/r^2)$	$\mathcal{T}(\mathbf{g})$	Laurent and Slot (2023)

1.7 Link to Cubature Rules

A natural approach to design upper bounds on the minimum value of f over $\mathbf{X}$ is by minimizing f over a well-chosen finite set of points $\mathbf{X}_r \subseteq \mathbf{X}$. For example, for a set $\mathbf{X} \subseteq [0, 1]^n$, one may take for $\mathbf{X}_r$ the set of rational points in $\mathbf{X}$ that have a given denominator $r \geq 1$. This clearly gives a hierarchy of upper bounds that converge asymptotically to $f_{\min}$ when the denominator r grows. For the simplex $\mathbf{X} = \Delta^n$, we have $|\mathbf{X}_r| = \binom{n+r}{r}$, thus polynomial in n for any fixed r. In fact, the parameters $f_{\min, \mathbf{X}_r}$ lead to a polynomial-time approximation scheme with convergence rate $\min_{\mathbf{x} \in \mathbf{X}_r} f(\mathbf{x}) - f_{\min} = O(1/r)$ (de Klerk et al. (2006)). However, for the hypercube $\mathbf{X} = [0, 1]^n$, we have $|\mathbf{X}_r| = (r + 1)^n$, which is thus *exponential* in the dimension n. We refer to Martinez et al. (2019) for related work for well-chosen finite meshes on $\mathbf{X}$. Let us mention the following simple link to cubature rules.

Let $\mu \in \mathcal{M}(\mathbf{X})$ with support $\mathbf{X}$. Assume that $\mathbf{X}_r = \{x^{(i)} : i \in [N]\} \subseteq \mathbf{X}$, together with positive weights $w = (w_i)_{i \in [N]}$, provides a positive cubature rule for $(\mathbf{X}, \mu)$ that is exact at degree $\deg(f) + 2r$. That is, for any polynomial p with $\deg(p) \leq \deg(f) + 2r$ we have $\int_{\mathbf{X}} p \, d\mu = \sum_{i=1}^{N} w_i p(x^{(i)})$. Then, one easily sees that

$$\text{ub}(f, \mathbf{X}, \mu)_r \geq \min_{\mathbf{x} \in \mathbf{X}_r} f(\mathbf{x}) = f_{\min, \mathbf{X}_r} \geq f_{\min, \mathbf{X}}.$$

Sometimes, this permits to show tightness of the convergence rate for the upper bounds $\text{ub}(f, \mathbf{X}, \mu)_r$ from information about cubature rules. This is done, e.g., for the sphere $\mathbf{X} = \mathbb{S}^{n-1}$ (equipped with the uniform Haar measure) by de Klerk and Laurent (2022), who show $\text{ub}(f, \mathbf{X}, \mu)_r - f_{\min} = \Omega(1/r^2)$ for linear f.

2 Performance Analysis of the Upper Bounds

Recall the set up: $\mathbf{X} \subseteq \mathbb{R}^n$ is a compact set and μ is a reference measure supported by $\mathbf{X}$. Consider the bilinear form $\langle \cdot, \cdot \rangle_\mu$ on $\mathbb{R}[\mathbf{x}]$ induced by μ, defined by

$$\langle p, q \rangle_\mu = \int_{\mathbf{X}} p(\mathbf{x})q(\mathbf{x})d\mu(\mathbf{x}) \ \text{ for } p, q \in \mathbb{R}[\mathbf{x}]. \tag{18}$$

If $\mathbf{X}$ has a nonempty interior, then this defines an inner product on $\mathbb{R}[\mathbf{x}]$. Otherwise, if $\mathbf{X}$ has an empty interior (e.g., $\mathbf{X} = \mathbb{S}^{n-1}$ is the unit sphere), then this defines an inner product on the space $\mathcal{R}(\mathbf{X})$ of polynomials on $\mathbf{X}$. Let $\mathcal{B}_\mu = \{P_\alpha : \alpha \in \mathbb{N}^n\}$ be an orthonormal basis of $\mathcal{R}(\mathbf{X})$ with respect to this inner product, with the property that the set

$$\mathcal{B}_{\mu,d} = \{P_\alpha : \alpha \in \mathbb{N}_d^n\}, \ \text{ sometimes also denoted as the vector } [P_\alpha]_d, \tag{19}$$

is a basis of $\mathcal{R}(\mathbf{X})_d$ (the set of polynomials that agree on $\mathbf{X}$ with a degree d polynomial) for each $d \in \mathbb{N}$.

Our objective here is to give a (rough) overview of the main ideas used to show the results in Table 1, where rates in $O(1/r^2)$ and in $O(\log^2(r)/r^2)$ are presented.

The starting point for showing a $O(r^2)$ rate is establishing an eigenvalue reformulation for the parameter $\mathrm{ub}(f, \mathbf{X}, \mu)_r$ (Sect. 2.2) and a link to extremal roots of orthogonal polynomials in the univariate case (Sect. 2.3). In addition, one uses some simple 'tricks' to reduce the analysis to (at most) quadratic polynomials and to simpler sets and measures (Sect. 2.1).

To show the (slightly weaker) rate $O(\log^2(r)/r^2)$ for much more general sets $\mathbf{X}$ one again follows a univariate strategy, now reducing the search in problem (14) to sos polynomials $\sigma(\mathbf{x}) = s(f(\mathbf{x}))$, where s is a *univariate* sos polynomial (see Sect. 2.4).

2.1 Some Useful Tricks for the Analysis

We group here some simple facts, useful for the analysis of the upper bounds.

Lemma 1 (de Klerk et al. (2017)) *Let* $f, g \in \mathbb{R}[\mathbf{x}]$ *satisfy (i)* $f(\mathbf{x}) \leq g(\mathbf{x})$ *for all* $\mathbf{x} \in \mathbf{X}$, *and (ii)* f *and* g *take the same minimum value on* $\mathbf{X}$, *i.e.,* $f_{min,\mathbf{X}} = g_{min,\mathbf{X}}$. *Then, for any* $r \in \mathbb{N}$, *we have* $\mathrm{ub}(f, \mathbf{X}, \mu)_r - f_{min,\mathbf{X}} \leq \mathrm{ub}(g, \mathbf{X}, \mu)_r - g_{min,\mathbf{X}}$.

Using Taylor's expansion, one can see that any polynomial f admits an upper estimator g on $\mathbf{X}$ that satisfies the assumptions (i), (ii) of Lemma 1 and that is (linear or) quadratic.

The next lemma permits to reduce the analysis for a pair $(\mathbf{X}, w)$, where $\mathbf{X}$ is equipped with the absolutely continuous measure $w(\mathbf{x})d\mathbf{x}$, to a possibly simpler pair $(\widehat{\mathbf{X}}, \widehat{w})$, where $\mathbf{X} \subseteq \widehat{\mathbf{X}}$ and $w \leq \widehat{w}$ 'look the same' around a minimizer.

Lemma 2 (Slot and Laurent (2022d)) *Let $\mathbf{X} \subseteq \widehat{\mathbf{X}}$ be compact sets, where $\mathbf{X}$ (resp., $\widehat{\mathbf{X}}$) is equipped with an absolutely continuous measure $w(\mathbf{x})d\mathbf{x}$ (resp., $\widehat{w}(\mathbf{x})d\mathbf{x}$). Let $x^* \in \mathbf{X}$ be a minimizer of f in $\mathbf{X}$. Assume the following conditions hold:*

- *(i) $\mathbf{X}, \widehat{\mathbf{X}}$ are 'locally similar' at $\mathbf{x}^*$: $\mathbf{X} \cap B^n(\mathbf{x}^*, \epsilon) = \widehat{\mathbf{X}} \cap B^n(\mathbf{x}^*, \epsilon)$ for some $\epsilon > 0$ (with $B^n(\mathbf{x}^*, \epsilon)$ the ball centered at $\mathbf{x}^*$ with radius ϵ).*
- *(ii) $w(\mathbf{x}) \leq \widehat{w}(\mathbf{x})$ for all $\mathbf{x} \in int(\mathbf{X})$.*
- *(iii) $w, \widehat{w}$ are 'locally comparable' at $\mathbf{x}^*$: $C \cdot \widehat{w}(\mathbf{x}) \leq w(\mathbf{x})$ for $\mathbf{x} \in int(K) \cap B^n(\mathbf{x}^*, \epsilon)$ for some $\epsilon > 0$ and $C > 0$.*

Then, there exists a (linear or) quadratic polynomial g such that $f(\mathbf{x}) \leq g(\mathbf{x})$ for all $\mathbf{x} \in \widehat{\mathbf{X}}$, $f_{min,\mathbf{X}} = g_{min,\mathbf{X}}$, and $ub(f, \mathbf{X}, w)_r - f_{min,\mathbf{X}} \leq \frac{2}{C}(ub(g, \widehat{\mathbf{X}}, \widehat{w})_r - g_{min,\widehat{\mathbf{X}}})$.

The up-shot of these two lemmas can be summarized as follows.

> **Recipe for analyzing the upper bounds**
> It suffices to analyze the convergence rate of the range $ub(f, \mathbf{X}, \mu)_r - f_{min}$ for *(linear or) quadratic f* and 'simple' sets $\mathbf{X}$ (like the box $[-1, 1]^n$ or the ball B^n).

2.2 Reformulation as an Eigenvalue Problem

We begin with giving an eigenvalue reformulation for the parameter $ub(f, \mathbf{X}, \mu)_r$ from (14). For this, we express the polynomial $\sigma \in \Sigma_{2r}$ entering the definition of the parameter $ub(f, \mathbf{X}, \mu)_r$ using the orthonormal basis $\mathcal{B}_{\mu,r}$, ordered as the vector $[P_\alpha]_r$ (as in (19)). So, we can write $\sigma = \langle Q, [P_\alpha]_r [P_\alpha]_r^\mathsf{T} \rangle$, where Q is the matrix variable (indexed by $\mathcal{B}_{\mu,r}$). Then, we have $\int_\mathbf{X} f\sigma d\mu = \langle Q, M_{\mu,r}(f) \rangle$ and $\int_\mathbf{X} \sigma d\mu = \text{Tr}(Q)$, after defining the matrix

$$M_{\mu,r}(f) = \left(\int_\mathbf{X} f P_\alpha P_\beta d\mu \right)_{\alpha, \beta \in \mathbb{N}_r^n}.$$

In this way we arrive at the following *eigenvalue reformulation*:

$$ub(f, \mathbf{X}, \mu)_r = \min\{\langle M_{\mu,r}(f), Q \rangle : Q \succeq 0, \ \text{Tr}(Q) = 1\} = \lambda_{\min}(M_{\mu,r}(f)), \tag{20}$$

showing that the parameter $ub(f, \mathbf{X}, \mu)_r$ is equal to the smallest eigenvalue of the matrix $M_{\mu,r}(f)$. Some remarks are in order here: this computation relies on the matrix $M_{\mu,r}(f)$, which in turn requires to be able to integrate a polynomial on the

set $\mathbf{X}$ w.r.t. the measure μ. Thus, for practical computation one needs to restrict to some relatively easy sets equipped with well-understood measures.

Estimating this smallest eigenvalue remains a difficult problem in general. However, there is a situation where it is very well-understood: in the univariate case for the interval $\mathbf{X} = [-1, 1]$ equipped with a 'nice' measure. We first consider this case, which will form the basis for understanding the general multivariate case.

2.3 Univariate Case: Links to Roots of Orthogonal Polynomials

Here, we consider the univariate case $n = 1$ and the interval $\mathbf{X} = [-1, 1]$ equipped with a measure μ supported on $[-1, 1]$. Consider as above the orthonormal basis $\mathcal{B}_\mu = \{P_k : k \geq 0\}$ of $\mathbb{R}[x]$ w.r.t. the inner product $\langle \cdot, \cdot \rangle_\mu$. Then, the polynomials P_k satisfy the well-known *3-term recurrence*: there exist scalars a_k, b_k ($k \geq 0$) such that

$$x P_k = a_{k-1} P_{k-1} + b_k P_k + a_k P_{k+1} \text{ for } k \geq 0 \quad (\text{setting } a_{-1} = 0). \quad (21)$$

Observe that for the polynomial $f = x$, the matrix $M_{\mu,r}(x) = \left(\int_{-1}^{1} x P_i P_j d\mu \right)_{i,j=0}^{r}$ (also known as the *Jacobi matrix*) is tri-diagonal and its eigenvalues are the roots of the degree $r + 1$ orthogonal polynomial P_{r+1}. In particular,

$$\lambda_{\min}(M_{\mu,r}(x)) = \text{ smallest root of the orthogonal polynomial } P_{r+1} \text{ (w.r.t. } \mu). \quad (22)$$

We refer, e.g., to Dunkl and Xu (2014) for background on orthogonal polynomials.

From now on, we will assume that the reference measure μ is of *Jacobi-type*.

Jacobi-type measures on $[-1, 1]$
The Jacobi-type measure is $d\mu(x) = (1 - x)^\lambda (1 + x)^{\lambda'} dx$, where $\lambda, \lambda' > -1$; the associated orthogonal polynomials are known as the Jacobi polynomials. The case when $\lambda = \lambda' = -1/2$ is of special interest; then, the 3-term recurrence (21) reads $x P_k = (P_{k-1} + P_{k+1})/2$ and the associated orthogonal polynomials are the Chebychev polynomials.
The reason for restricting to Jacobi-type measures is that the behaviour of the extremal roots of their associated orthogonal polynomials is well-understood: The smallest root of P_r is $-1 + \Theta(1/r^2)$ (Dimitrov and Nikolov (2010), and Driver and Jordaan (2012)).

As observed earlier, it suffices to analyze the upper bounds for *linear* or *quadratic* f.

For *linear* $f = x$, since the smallest root of P_{r+1} is $-1 + \Theta(1/r^2)$, by combining with relations (20) and (22), we directly obtain the error analysis

$$\mathrm{ub}(x, [-1, 1], \mu)_r - f_{\min} = \lambda_{\min}(M_{\mu,r}(x)) + 1 = \Theta\left(\frac{1}{r^2}\right) \qquad (23)$$

(from de Klerk and Laurent (2020)).

We now consider the case when f is *quadratic*. We distinguish two cases, depending whether f has a minimizer in the boundary or in the interior of $[-1, 1]$. In the former case, f admits a linear upper estimator; then we can apply Lemma 1 and use (23) to conclude that $\mathrm{ub}(f, [-1, 1], \mu)_r - f_{\min} = O(1/r^2)$.

So, assume now $f = x^2 + cx$ with $a \in (-2, 2)$, so that f attains its minimum value in $(-1, 1)$. Then, one needs to estimate the smallest eigenvalue of the matrix $M_{\mu,r}(f)$, a difficult task in general. Indeed, $M_{\mu,r}(f)$ is a 5-diagonal matrix, with its entries depending on the parameters a_k, b_k in the 3-term recurrence (21) and the parameter c in the definition of f. Estimating this smallest eigenvalue is, however, easier in the Chebyshev case.

In the case when the measure μ is of Chebyshev-type, i.e., $d\mu = (1 - x^2)^{-1/2}dx$, the matrix $M_{\mu,r}(f)$ is 'almost' a circulant matrix: After modifying its first two rows and columns it can be made a circulant matrix, whose eigenvalues can be explicitly computed. Combining with an interlacing argument, one can show again that $\mathrm{ub}(f, [-1, 1], \mu)_r - f_{\min} = O(1/r^2)$ (de Klerk and Laurent (2020)). Combining this with using Lemma 2 one can show the same result for a general measure μ with weight $(1 - x^2)^\lambda$ when $\lambda \geq -1/2$ (Slot and Laurent (2022d)).

In summary, we have $\mathrm{ub}(f, [-1, 1], \mu)_r - f_{\min} = O(1/r^2)$ for any polynomial f, in the case when $[-1, 1]$ is equipped with a measure μ with Jacobi weight $(1 - x^2)^\lambda$ and $\lambda \geq -1/2$. This error estimate in $O(1/r^2)$ extends then to the box $[-1, 1]^n$, equipped with a product of such measures.

Similarly, to establish the rate $O(1/r^2)$ for the other sets in Table 1, one uses the result (23) for the univariate case, combined with the results of Lemmas 1 and 2 (see Slot and Laurent (2022d) for the ball, simplex, and round convex bodies), and possibly some 'integration trick' (see de Klerk and Laurent (2022) for the sphere).

2.4 Another Analysis Technique Using Needle Polynomials

At this point there remains to explain how to show the (slightly weaker) rate $O(\log^2(r)/r^2)$ presented in Table 1 for general compact sets $\mathbf{X}$ like semialgebraic sets (with dense interior) and convex bodies. For this the key idea is again to follow a univariate approach, as explained by Slot and Laurent (2021). Instead of searching over all (multivariate) sos polynomials $\sigma \in \mathbb{R}[\mathbf{x}]$ in program (14), the idea is to restrict the search to *univariate* sos polynomials $s \in \mathbb{R}[x]$ and then set $\sigma(\mathbf{x}) = s(f(\mathbf{x}))$. In this way one arrives at the following (weaker) bounds (introduced by Lasserre (2021)):

$$\text{ub}_\#(f, \mathbf{X}, \mu)_r = \min \left\{ \int_{\mathbf{X}} f(\mathbf{x}) s(f(\mathbf{x})) d\mu(\mathbf{x}) : s \in \mathbb{R}[x]_{2r} \text{ sos}, \int_{\mathbf{X}} s(f(\mathbf{x})) d\mu(\mathbf{x}) = 1 \right\}$$
$$= \min \left\{ \int_{f(\mathbf{X})} x s(x) df_\# \mu(x) : s \in \mathbb{R}[x]_{2r} \text{ sos}, \int_{f(\mathbf{X})} s(x) df_\# \mu(x) = 1 \right\}, \quad (24)$$

where $f_\# \mu$ denotes the univariate measure obtained by taking the push-forward of μ by f. Then, we have $\text{ub}(f, \mathbf{X}, \mu)_{rd} \leq \text{ub}_\#(f, \mathbf{X}, \mu)_r$ if $d = \deg(f)$.

In view of relations (20) and (22), the analysis of the bounds $\text{ub}_\#(f, \mathbf{X}, \mu)_r$ relies on the smallest roots of the orthogonal polynomials w.r.t. the push-forward measure $f_\# \mu$. However, these orthogonal polynomials are not well understood in general, so another approach is needed. Up to translation one may assume that $f_{\min} = 0$. Then, the idea is to find a univariate sos polynomial s that approximates well the Dirac delta at the origin. For this, one can employ the so-called *needle polynomials* (from Kroó (2015)) that are widely used in the literature of approximation theory. We refer to Slot and Laurent (2021) for the technical details and extension to compact sets satisfying a suitable geometric assumption.

3 Performance Analysis of the Lower Bounds

We turn now to the lower bounds. Let $\mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_1(\mathbf{x}) \geq 0, \ldots, g_m(\mathbf{x}) \geq 0\}$ be as in (2). We assume that the quadratic module $Q(\mathbf{g})$ is Archimedean. Recall that this implies that, for any f, the bounds $\text{lb}(f, Q(\mathbf{g}))_r$ and $\text{lb}(f, \mathcal{T}(\mathbf{g}))_r$ converge to $f_{\min}$ as $r \to \infty$ by Putinar's and Schmüdgen's Positivstellensätze, respectively. In this section, we outline the main ideas used to obtain the convergence rates presented in Table 2. For ease of writing, we will use the letter C to refer to either $Q(\mathbf{g})$ or $\mathcal{T}(\mathbf{g})$. A useful observation is the following.

Observation 1 *Let* $f \in \mathbb{R}[\mathbf{x}]$. *Then, for any* $\epsilon > 0$ *and* $r \in \mathbb{N}$*, we have*

$$f_{min} - lb(f, C)_r \leq \epsilon \iff f - f_{min} + \epsilon \in C_{2r}.$$

Thus, in order to prove convergence rates for Lasserre's hierarchies of lower bounds, it suffices to find low-degree sos-representations of $f - f_{\min} + \epsilon$. In what follows, we discuss two methods of obtaining such representations. We give particular attention to the so-called *polynomial kernel method* in Sect. 3.1 below, as it reveals an interesting connection between the analysis of the *lower* bounds and the *upper* bounds discussed in Sect. 2.

Throughout this section, we set $d = \deg(f)$, which should be thought of as being fixed, while the relaxation order r will grow.

3.1 Analysis via the Polynomial Kernel Method

The *polynomial kernel method* (PKM) has been successful for proving strong convergence rates for optimization over certain distinguished sets $\mathbf{X}$, including $\mathbf{X} = \mathbb{S}^{n-1}, B^n, \Delta^n, [-1, 1]^n$. It was initially described[3] in the special case $\mathbf{X} = \mathbb{S}^{n-1}$ by Fang and Fawzi (2021), who use it to show a convergence rate in $O(1/r^2)$ in that setting. It was later applied by Laurent and Slot (2023) and Slot (2022c), to prove rates in $O(1/r^2)$ for optimization over $[-1, 1]^n$ and Δ^n, B^n, respectively. The latter work by Slot (2022c) is the first to describe the technique in full generality, and we follow its exposition here.

Suppose we were able to construct a linear operator $\mathbf{K} : \mathbb{R}[\mathbf{x}] \to \mathbb{R}[\mathbf{x}]$ with the following three properties:

$$\mathbf{K}(1) = 1, \tag{P1}$$

$$\mathbf{K}p \in C_{2r} \quad \text{for all } p \in \mathcal{P}(\mathbf{X}), \tag{P2}$$

$$\max_{\mathbf{x} \in \mathbf{X}} |\mathbf{K}^{-1} f(\mathbf{x}) - f(\mathbf{x})| \leq \epsilon. \tag{P3}$$

Then, we claim that $(f - f_{\min}) + \epsilon \in C_{2r}$ (which shows that $f_{\min} - \text{lb}(f, C)_r \leq \epsilon$ by Observation 1). Indeed, by (P3), we have $\mathbf{K}^{-1} f(\mathbf{x}) \geq f(\mathbf{x}) - \epsilon \geq f_{\min} - \epsilon$ on $\mathbf{X}$. Using (P1), we obtain $\mathbf{K}^{-1}(f - f_{\min} + \epsilon) = \mathbf{K}^{-1} f - f_{\min} + \epsilon \in \mathcal{P}(\mathbf{X})$. By (P2), we may then conclude that

$$f - f_{\min} + \epsilon = \mathbf{K}\left[\mathbf{K}^{-1}\left(f - f_{\min} + \epsilon\right)\right] \in C_{2r}.$$

3.1.1 Constructing Linear Operators

It remains to construct operators $\mathbf{K}$ that enjoy these special properties. For this, we rely on the theory of (polynomial) *reproducing kernels*. Let μ be a (sufficiently nice) measure supported on $\mathbf{X}$. Then, any polynomial $\mathcal{K} \in \mathbb{R}[\mathbf{x}, \mathbf{y}]$ induces a linear operator $\mathbf{K}$ on $\mathbb{R}[\mathbf{x}]$ via convolution:

$$\mathbf{K}p(\mathbf{x}) := \int_{\mathbf{X}} \mathcal{K}(\mathbf{x}, \mathbf{y}) p(\mathbf{y}) d\mu(\mathbf{y}) \quad (p \in \mathbb{R}[\mathbf{x}]). \tag{25}$$

---

[3] Earlier, weaker analyses of the lower bounds on $\mathbb{S}^{n-1}$ due to Reznick (1995) and Doherty and Wehner (2013) already relied on the PKM *implicitly*.

The goal is to choose $\mathcal{K}$ in such a way that $\mathbf{K}$ satisfies (P1)-(P3). For (P2), it suffices to choose $\mathcal{K}$ so that the polynomial $\mathbf{x} \mapsto \mathcal{K}(\mathbf{x}, \mathbf{y})$ lies in C_{2r} for all *fixed* $\mathbf{y} \in \mathbf{X}$. This is a consequence of Tchakaloff's Theorem, which allows us to write the integral (25) as (positively weighted) finite sum over a cubature rule (see Laurent and Slot (2023)). To establish (P1) and (P3), it turns out that it is enough to control the *eigenvalues* of $\mathbf{K}$. The eigenvalues of $\mathbf{K}$ can be related to the polynomial $\mathcal{K}$ directly if we make the assumption that it is of the form

$$\mathcal{K}(\mathbf{x}, \mathbf{y}) = \sum_{\|\alpha\|_1 \leq 2r} \lambda_\alpha P_\alpha(\mathbf{x}) P_\alpha(\mathbf{y}) \quad (\lambda_\alpha \in \mathbb{R}), \tag{26}$$

where $\{P_\alpha : \alpha \in \mathbb{N}^n\}$ is an orthonormal basis of $\mathbb{R}[\mathbf{x}]$ w.r.t. μ, ordered so that $\deg(P_\alpha) = \|\alpha\|_1$ for all α. Indeed, by orthonormality, the eigenvalues of $\mathbf{K}$ are then given by the coefficients λ_α in (26). If we set $\lambda_\alpha = 1$ for all α, the resulting polynomial is called the *reproducing kernel* for $\mathbb{R}[\mathbf{X}]_{2r}$ w.r.t. μ. Its associated operator $\mathbf{K}$ acts as the identity on $\mathbb{R}[\mathbf{x}]_{2r}$. It thus satisfies (P1) and (P3) (for $\epsilon = 0$). However, it does not have an sos-representation in general, and so it does not satisfy (P2). The idea is to *perturb* the reproducing kernel slightly, choosing $\lambda_\alpha \approx 1$ (for all $|\alpha| \leq d$), in such a way that all three properties hold. The following lemma makes the relation between the λ_α and properties (P1), (P3) precise.

Lemma 3 *Let $\mathcal{K}$ be as in (26), with $\lambda_0 = 1$ and $\lambda_\alpha \in (1/2, 1]$ for all $\|\alpha\|_1 \leq d$. Then, its associated linear operator $\mathbf{K}$ satisfies (P1), and*

$$\max_{\mathbf{x} \in \mathbf{X}} |\mathbf{K}^{-1} f(\mathbf{x}) - f(\mathbf{x})| \leq \gamma \cdot \sum_{|\alpha| \leq d} (1 - \lambda_\alpha) \cdot (f_{\max} - f_{\min}).$$

for some constant $\gamma > 0$ depending only on $\mathbf{X}$ and d (but not on f).

The inequality in Lemma 3 above can be improved in certain special cases (such as $\mathbf{X} = \mathbb{S}^{n-1}$). In particular, the constant γ can be bounded nontrivially, but this is beyond the scope of this survey.

3.1.2 A Connection to the Upper Bounds

In general, it is not obvious how to choose the λ_α such that (P2) holds. However, for the aforementioned special choices of $\mathbf{X}$, one can rely on special structure of the reproducing kernel to make this problem tractable. For instance, in the original application of the PKM by Fang and Fawzi (2021) for the sphere $\mathbb{S}^{n-1}$, the classical *Funk-Hecke formula* tells us that, for any $k \in \mathbb{N}$,

$$\sum_{\|\alpha\|_1 = k} P_\alpha(\mathbf{x}) P_\alpha(\mathbf{y}) = \mathcal{G}_k(\mathbf{x} \cdot \mathbf{y}) \quad (\mathbf{x}, \mathbf{y} \in \mathbb{S}^{n-1}),$$

where $\mathcal{G}_k$ is the *Gegenbauer polynomial* of degree k. These are the (univariate) orthogonal polynomials on the interval $[-1, 1]$ w.r.t to weight $w(t) = (1 - t^2)^{\frac{n-3}{2}}$. Assuming that the λ_α in (26) depend only on $|\alpha|$, the polynomial $\mathcal{K}$ then equals

$$\mathcal{K}(\mathbf{x}, \mathbf{y}) = \sum_{k=0}^{2r} \lambda_k \mathcal{G}_k(\mathbf{x} \cdot \mathbf{y}).$$

Write $q(t) := \sum_{k=0}^{2r} \lambda_k \mathcal{G}_k(t)$. Note that if $q \in \Sigma[t]_{2r}$ is a (univariate) sum of squares, then $\mathcal{K}(\mathbf{x}, \mathbf{y})$ lies in the quadratic module of the sphere for fixed $\mathbf{y} \in \mathbb{S}^{n-1}$, and so **K** will satisfy (P2). In light of Lemma 3, we thus wish to select the λ_k so that $\lambda_0 = 1$ (P1); $q \in \Sigma[t]_{2r}$ (P2); and $\sum_{k=1}^{d}(1 - \lambda_k)$ is as small as possible (P3). Remarkably, the optimal selection of the λ_k corresponds to a particular instance of Lasserre's *upper* bounds. This correspondence allows us to transport the analysis of these bounds in Sect. 2 to our present setting. Indeed, by orthogonality, and after choosing the right normalization of the Gegenbauer polynomials $\mathcal{G}_k$, we have that $\lambda_k = \int_{-1}^{1} \mathcal{G}_k(t)q(t)w(t)dt$. Now set $g(t) = d - \sum_{k=0}^{d} \mathcal{G}_k(t)$. Then, we have that

$$\int_{-1}^{1} g(t)q(t)w(t)dt = \sum_{k=1}^{d}(1 - \lambda_k).$$

Thus, choosing the λ_k optimally reduces to solving the optimization problem

$$\text{opt} := \inf_{q \in \Sigma[t]_{2r}} \left\{ \underbrace{\int_{-1}^{1} \sum_{k=1}^{d} g(t)q(t)w(t)dt}_{\sum_{k=1}^{d}(1-\lambda_k)} : \underbrace{\int_{-1}^{1} q(t)w(t)dt = 1}_{\lambda_0} \right\}.$$

We recognize this as the program (14) that defines the upper bound $\text{ub}(g, [-1, 1])_r$ for the minimization of g on $[-1, 1]$. As $g_{\min} = 0$ (attained at $t = 1$), we may conclude that $\text{opt} = O(1/r^2)$, and a convergence rate of the *lower* bounds on $\mathbb{S}^{n-1}$ of the same order follows. We refer to Fang and Fawzi (2021) (and the exposition in Slot (2022b)) for details.

For $\mathbf{X} \in \{B^n, \Delta^n\}$, one has (more complicated) analogs of the Funk-Hecke formula. This allows one to establish a correspondence between upper and lower bounds similar to the above. For $\mathbf{X} = [-1, 1]^n$, there is no such formula. However, the reproducing kernel has a product structure in that case, which can be exploited to obtain representations of $\mathcal{K}$ in the truncated preordering. We refer to Slot (2022c), Slot (2022b) and Laurent and Slot (2023) for details.

Recipe for analyzing the lower bounds (I): Polynomial kernel method
Convergence rates for Lasserre's hierarchy of lower bounds may be obtained
from sum-of-squares representations of (perturbed) reproducing kernels on **X**.
For distinguished semialgebraic sets **X**, such representations can be found via
the hierarchy of *upper* bounds.

3.2 Analysis via Algebro-Geometric Reduction

We now discuss a technique employed first by Schweighofer (2004), Nie and
Schweighofer (2007), Baldi and Mourrain (2023) and Baldi et al. (2025), to prove
convergence rates for Lasserre's lower bounds on general (compact, Archimedean)
semialgebraic sets $\mathbf{X} = \mathrm{Se}(\mathbf{g})$ (as in (2)). With respect to the polynomial kernel
method, this technique yields weaker guarantees, but it does not rely on special
structure of the set **X**. We follow the exposition of Baldi and Mourrain (2023).

Suppose that, for a given polynomial f, we wish to show that $f - f_{\min} + \epsilon \in Q(\mathbf{g})_{2r}$ for some small $\epsilon > 0$. The idea is to embed the (potentially complicated)
set **X** in a simple semialgebraic set $\mathbf{Y} = \mathrm{Se}(\mathbf{h}) \supseteq \mathrm{Se}(\mathbf{g}) = \mathbf{X}$. For instance, **Y** might
be a (scaled) ball, simplex or box. Following Baldi and Mourrain (2023), let us fix
$\mathrm{Se}(\mathbf{h}) = [-1, 1]^n$ for concreteness. Using specialized results on the convergence of
Lasserre's hierachy on $[-1, 1]^n$ (e.g., those obtained via the PKM), we know that,
for some $\epsilon = O(1/r^2)$,

$$f - f_{\min,\mathbf{Y}} + \epsilon \in \mathcal{T}(\mathbf{h})_{2r}. \tag{27}$$

This fact is not immediately useful to us, for two reasons. First, we wish to obtain a
representation in $Q(\mathbf{g})$, not $\mathcal{T}(\mathbf{h})$. Second, the minimum $f_{\min,\mathbf{Y}}$ of f on **Y** might be
(much) smaller than $f_{\min} = f_{\min,\mathbf{X}}$. The first issue is resolved by proving an inclusion
$\mathcal{T}(\mathbf{h})_{2r} \subseteq Q(\mathbf{g})_{2r+\ell}$. Note that the existence of such an inclusion is not surprising,
as $\mathbf{Y} \supseteq \mathbf{X}$, and thus $\mathcal{P}(\mathbf{X}) \supseteq \mathcal{P}(\mathbf{Y})$. Nonetheless, careful arguments are needed to
control the additional degree term ℓ, but we do not discuss these here. The second
issue is more serious. It is resolved as follows: Suppose we were able to construct
a regularizing polynomial $q \in Q(\mathbf{g})_{\ell'}$ with the property that $f_{\min,\mathbf{X}} \leq (f - q)_{\min,\mathbf{Y}}$.
Then, setting $\tilde{f} = f - q$, we could apply (27) to $\tilde{f}$, and use the inclusion of $\mathcal{T}(\mathbf{h})$
in $Q(\mathbf{g})$ to obtain, for some $\tilde{\epsilon} = O(1/r^2)$,

$$\tilde{f} - \tilde{f}_{\min,\mathbf{Y}} + \tilde{\epsilon} \in \mathcal{T}(\mathbf{h})_{2r} \subseteq Q(\mathbf{g})_{2r+\ell}.$$

Using the identity $f - f_{\min,\mathbf{X}} + \tilde{\epsilon} = (\tilde{f} - \tilde{f}_{\min,\mathbf{Y}} + \tilde{\epsilon}) + q + \tilde{f}_{\min,\mathbf{Y}} - f_{\min,\mathbf{X}}$, com-
bined with $\tilde{f}_{\min,\mathbf{Y}} \geq f_{\min,\mathbf{X}}$ and $q \in Q(\mathbf{g})_{\ell'}$, this gives us the representation

$$f - f_{\min,\mathbf{X}} + \tilde{\epsilon} \in Q(\mathbf{g})_{\max\{2r+\ell, \ell'\}}.$$

Note that $\tilde{\epsilon}$ depends on the degree of q, which may be substantially larger than d. Therefore, $\tilde{\epsilon} \gg \epsilon$ in general. In addition to the extra degree terms ℓ, ℓ', this is what explains the weaker rates obtained from this technique compared to the PKM.

3.2.1   Construction of the Regularizing Polynomial

It remains to construct the polynomial $q \in Q(\mathbf{g})$ described above. As mentioned, this construction has a large impact on the quality of the resulting convergence guarantees. In fact, it is arguably the construction of q that sets apart the results of Nie and Schweighofer (2007), Baldi and Mourrain (2023) and Baldi et al. (2025), leading to stronger rates in the latter works. We give a sketch of the construction in Baldi and Mourrain (2023).

Recall that after setting $\tilde{f} = f - q$, we want that $\tilde{f}_{\min,\mathbf{Y}} \geq f_{\min,\mathbf{X}}$. Thus, we wish that $q \ll 0$ on $\mathbf{Y}\backslash\mathbf{X}$, and $q \approx 0$ on $\mathbf{X}$ (as $q \in Q(\mathbf{g})$, we know that $q \geq 0$ on $\mathbf{X}$). Without loss of generality, assume that the polynomials g_j defining $\mathbf{X}$ satisfy $-1 \leq g_j(\mathbf{x}) \leq 1$ for all $\mathbf{x} \in \mathbf{Y}$ and $j \in [m]$. The idea is to first construct a *univariate* sum-of-squares polynomial $h \in \Sigma[t]$, with the property that, for some small $\eta > 0$, we have $h(t) \approx 0$ for $t \in [\eta, 1]$ and $h(x) \approx 1$ for $t \in [-1, -\eta]$. This can be done via Chebyshev approximation, see Baldi and Mourrain (2023, Sect. 2.3). Then, we set $q(\mathbf{x}) = M \cdot \sum_{j=1}^{m} g_j(\mathbf{x})h(g_j(\mathbf{x}))$, with $M > 0$. Note that $q \in Q(\mathbf{g})$ by definition. Note further that, for $\mathbf{x} \notin \mathbf{X}$ sufficiently far from the boundary bd$(\mathbf{X})$, at least one $g_j(\mathbf{x}) < -\eta$, whence $g_i(\mathbf{x})h(g_i(\mathbf{x})) \ll 0$, whereas for $\mathbf{x} \in \mathbf{X}$ sufficiently far from bd$(\mathbf{X})$, all $g_j(\mathbf{X}) \in [\eta, 1]$, whence $q(\mathbf{x}) \approx 0$. It remains to analyze the situation for $\mathbf{x}$ close to bd$(\mathbf{X})$, where (some of the) constraints $g_j(\mathbf{x})$ are close to 0. There, a careful comparison of the gradients ∇f of f and ∇g_j of the constraints g_j is required to finish the argument. This comparison involves the so-called Łojasiewicz-constant of $\mathbf{g}$, which appears in the exponent of the final convergence rate. See Baldi and Mourrain (2023) for details.

> **Recipe for analyzing the lower bounds (II): Algebro-geometric reduction**
> Convergence rates for Lasserre's hierarchy of lower bounds on general semi-algebraic sets may be derived from rates on special sets (such as the simplex or hypercube), through delicate algebraic and geometric arguments.

3.3   Extensions

We end this section with some recent extensions of the techniques described above.

3.3.1 Algebro-Geometric Reduction for the Hypercube

Baldi and Slot (2024) use a variation of the reduction technique described above to prove convergence rates in $O(1/r)$ for the Putinar-type lower bounds on $[-1, 1]^n$, based on the quadratic module (recall that Laurent and Slot (2023) only show rates for the stronger Schmüdgen-type bounds, based on the preordering). This extension shows that the algebro-geometric method can also be useful for analyses on specific, rather than general sets $\mathbf{X}$. The main difference w.r.t. the above is that the 'simple' set $\mathbf{Y} \supseteq [-1, 1]^n$ used in Baldi and Slot (2024) depends on f, r (while it was static before). In fact, somewhat remarkably, $\mathbf{Y} = \text{Se}(\mathbf{h}) = [1 - \eta, 1 + \eta]^n$ is itself just a scaled hypercube, with $\eta = \eta(f, r) > 0$. The use of a dynamic $\mathbf{Y}$ eliminates the need for the regularizing polynomial q, permitting to prove stronger rates than those in Baldi and Mourrain (2023) and Baldi et al. (2025) (which would give a rate in $O(1/\sqrt[10]{r})$ in this setting). On the other hand, proving an inclusion $\mathcal{T}(\mathbf{h}) \subseteq Q(\mathbf{g})$ with proper degree bounds is more complicated. See Baldi and Slot (2024) for details.

3.3.2 Sparse Polynomial Optimization

As we have seen in Sect. 1.4, the semidefinite programs used to model Lasserre's hierarchies in n variables at degree r involve matrices of size $N_r = \binom{n+r}{r}$. These SDPs are thus intractable already for moderately large values of n. To address this issue, variants of Lasserre's hierarchy of lower bounds that exploit *sparsity* of the underlying polynomial optimization problem have been proposed in the literature, see Magron and Wang (2023) for an overview. A recent work by Korda et al. (2024) extends both of the techniques discussed in Sect. 3 to this setting, yielding a performance analysis for the sparse bounds. Furthermore, the convergence rates achieved there are stronger than those of Sect. 3 for sufficiently sparse POPs (relative to the size of the resulting SDPs).

3.3.3 Generalized Moment Problems

Polynomial optimization is a special case of the *generalized problem of moments* (GPM), which asks to minimize a linear function over the cone of positive Borel measures on $\mathbb{R}^n$ under some linear conditions on the moments. Lasserre's approximation hierarchies naturally extend to the GPM, see Lasserre (2009) and de Klerk and Laurent (2019). Convergence rates for these extended hierarchies have been established, especially in the context of dynamical systems, optimal control and volume estimation, by Korda et al. (2017), Korda and Henrion (2018), Schlosser et al. (2024a) and Schlosser and Tacchi (2024b). The analyses in the latter two recent works rely directly on (a combination of) existing convergence results for polynomial optimization discussed above.

4   Discussion

Tightness of the Performance Analysis

As we have seen, the convergence rates for the hierarchy of *upper* bounds presented in Sect. 2 are essentially tight (up to log-factors). For the *lower* bounds, the situation is much less clear. The literature on 'worst-case' examples for the lower bounds is mostly *qualitative* in nature, see, e.g., Scheiderer (2005) and Powers and Scheiderer (2001). On the quantitative side, Baldi and Slot (2024) recently showed that the Putinar-type lower bounds on $\mathbf{X} = [-1, 1]^n$ converge at a rate no better than $\Omega(1/r^8)$. Note that this negative result is still rather far away from the best-known positive result in that setting, which is in $O(1/r)$. It is an interesting research direction to prove negative results on the convergence of the lower bounds that either match the positive results more closely; or apply to more general $\mathbf{X}$; or apply to the Schmüdgen-type bounds.

Exponential Convergence Under Local Optimality Conditions

A common feature of the convergence rates in Table 2 is that they are *sub*exponential, in particular no better than $O(1/r^c)$ for some constant $c > 0$ depending on the set $\mathbf{X}$ and the type of certificate. In light of the negative results mentioned above, these are likely 'optimal' (up to improving the exponent c). On the other hand, Bach and Rudi (2023) show an *exponential* convergence rate for the Schmüdgen-type bounds on $\mathbf{X} = [-1, 1]^n$, which holds under an additional assumption on the objective function f. Roughly speaking, f should have a (strictly) positive definite Hessian at its global minimizer $\mathbf{x}^* \in [-1, 1]^n$, see Bach and Rudi (2023) for details.[4] Under this assumption, $f - f_{\min}$ can be written as a sum of squares of *smooth* functions (not necessarily polynomials). In turn, these smooth functions can be approximated by polynomials via (truncated) expansion in the Fourier basis, leading to a sum-of-squares representation of $f - f_{\min}$. The exponential convergence rate for Lasserre's lower bounds then follows from the fact that the Fourier coefficients of a smooth function decay exponentially quickly. As the approach of Bach and Rudi relies primarily on Fourier analysis, it seem likely that it could be extended to other distinguished $\mathbf{X}$, such as $\mathbf{X} = \mathbb{S}^{n-1}$. More unclear is whether it could also be used to prove guarantees for general semialgebraic sets $\mathbf{X}$, which we believe is an interesting open question.

---

[4] Note that this assumption holds *generically*. Under a similar assumption, Nie (2014) showed earlier that the hierarchy of lower bounds has finite convergence for general $\mathbf{X} = \mathrm{Se}(\mathbf{g})$, i.e., $f - f_{\min}$ belongs to $Q(\mathbf{g})_r$ for some $r \in \mathbb{N}$. However, his result gives no quantitative information on r.

Optimization Over Finite Semialgebraic Sets

In this survey, we have focused on the setting where $\mathbf{X} \subseteq \mathbb{R}^n$ is an infinite set. Sum-of-squares hierarchies have been extensively studied for optimization over finite semialgebraic $\mathbf{X}$ as well, particularly for (subsets of) the boolean hypercube: $\mathbf{X} \subseteq \{-1, 1\}^n$ (or $\mathbf{X} \subseteq \{0, 1\}^n$). A key difference in the finite setting is that the hierarchies always converge in a finite number of steps (under a minor condition on the description of $\mathbf{X}$), see Nie (2013). In fact, they converge in $\Theta(n)$ steps for a semialgebraic subset of the boolean cube, see Laurent (2023) and Fawzi et al. (2016). For this reason, asymptotic analysis as $r \to \infty$ does not make sense. Rather, one often fixes the level $r \in \mathbb{N}$ of the hierarchy, and lets the number of variables n tend to infinity. There is a large literature on this regime in the theoretical computer science community, see, e.g., Barak and Steurer (2014) and references therein. One could also consider a 'hybrid' regime, where $r = c \cdot n$ for some constant $c > 0$ and $n, r \to \infty$ simultaneously. In this regime, the polynomial kernel method (see Sect. 3.1) yields error guarantees for the upper and lower bounds on $\mathbf{X} = \{-1, 1\}^n$, see Slot and Laurent (2022a) for details.

References

Ahmadi, A.A., Olshevsky, A., Parrilo, P.A., Tsitsiklis, J.N.: NP-hardness of deciding convexity of quartic polynomials and related problems. Math. Program. **137**(1–2), 453–476 (2013)

Artin, E.: Über die Zerlegung definiter Funktionen in Quadrate. Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg **5**, (1927)

Bach, F., Rudi, A.: Exponential convergence of sum-of-squares hierarchies for trigonometric polynomials. SIAM J. Optim. **33**(3), 2137–2159 (2023)

Baldi, L., Mourrain, B.: On moment approximation and the effective Putinar's Positivstellensatz. Math. Program. **200**, 71–103 (2023)

Baldi, L., Slot, L.: Degree bounds for Putinar's Positivstellensatz on the hypercube. SIAM J. Appl. Algebr. Geom. **8**(1), 1–25 (2024)

Baldi, L., Mourrain, B., Parusinski, A.: On Łojasiewicz inequalities and the effective Putinar's Positivstellensatz. J. Algebra **662**, 741–767 (2025)

Barak, B., Steurer, D.: Sum-of-squares proofs and the quest toward optimal algorithms. In: Proceedings of the International Congress of Mathematicians, Seoul, vol. 4, pp. 509–533 (2014)

Blekherman, G., Parrilo, P.A., Thomas, R.R.: Semidefinite Optimization and Convex Algebraic Geometry. MOS-SIAM Series on Optimization, vol. 13 (2012)

Blomenhofer, A.T., Laurent, M., Slot, L.: In preparation

de Klerk, E., Laurent, M.: A survey of semidefinite programming approaches to the generalized problem of moments and their error analysis. In: Araujo, C., Benkart, G., Praeger, C., Tanbay, B. (eds.) World Women in Mathematics 2018. Association for Women in Mathematics Series, vol. 20. Springer, Cham (2019)

de Klerk, E.: Aspects of Semidefinite Programming–Interior Point Algorithms and Selected Applications. Kluwer (2002)

de Klerk, E., Laurent, M.: Comparison of Lasserre's measure-based bounds for polynomial optimization to bounds obtained by simulated annealing. Math. Oper. Res. **43**, 1317–1325 (2018)

de Klerk, E., Laurent, M.: Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere. Math. Program. **193**, 665–685 (2022)

de Klerk, E., Pasechnik, D.: Approximation of the stability number of a graph via copositive programming. SIAM J. Optim. **12**, 875–892 (2002)

Dimitrov, D.K., Nikolov, G.P.: Sharp bounds for the extreme zeros of classical orthogonal polynomials. J. Approx. Theory **162**, 1793–1804 (2010)

Doherty, C., Wehner, S.: Convergence of SDP hierarchies for polynomial optimization on the hypersphere (2013. arXiv:1210.5048

Driver, K., Jordaan, K.: Bounds for extreme zeros of some classical orthogonal polynomials. J. Approx. Theory **164**, 1200–1204 (2012)

Dunkl, C.F., Xu, Y.: Orthogonal Polynomials of Several Variables. Encyclopedia of Mathematics and its Applications. Cambridge University Press (2014)

Fang, K., Fawzi, H.: The sum-of-squares hierarchy on the sphere, and applications in quantum information theory. Math. Program. Ser. A **190**, 331–360 (2021)

Fawzi, H., Saunderson, J., Parrilo, P.A.: Sparse sums of squares on Abelian groups and improved semidefinite liftings. Math. Program. **160**, 149–191 (2016)

Garey, M.R., Johnson, D.S.: Computers and Intractability. A Guide to the Theory of NP-Completeness. Freeman, San Francisco (1979)

Henrion, D., Korda, M., Lasserre, J.B.: The Moment-SOS Hierarchy—Lectures in Probability, Statistics. Computational Geometry, Control and Nonlinear PDEs. World Scientific (2020)

Hilbert, D.: Über die Darstellung definiter Formen als Summe von Formenquadraten. Math. Ann. **32**, 342–350 (1888)

Klerk, E., Laurent, M.: Worst-case examples for Lasserre's measure-based hierarchy for polynomial optimization on the hypercube. Math. Oper. Res. **45**(1), 86–98 (2020)

Klerk, E., Laurent, M., Parrilo, P.: A PTAS for the minimization of polynomials of fixed degree over the simplex. Theor. Comput. Sci. **361**(2–3), 210–225 (2006)

Klerk, E., Laurent, M., Sun, Z.: Convergence analysis for Lasserre's measure-based hierarchy of upper bounds for polynomial optimization. Math. Program. Ser. A **162**(1), 363–392 (2017)

Korda, M., Magron, V., Rios-Zertuche, R.: Convergence rates for sums-of-squares hierarchies with correlative sparsity. Math. Program. (2024). https://doi.org/10.1007/s10107-024-02071-6

Korda, M., Henrion, D.: Convergence rates of moment-sum-of-squares hierarchies for volume approximation of semialgebraic sets. Optim. Lett. **12**(3), 435–442 (2018)

Korda, M., Henrion, D., Jones, C.N.: Convergence rates of moment-sum-of squares hierarchies for optimal control problems. Syst. & Control Lett. **100**, 1–5 (2017)

Krivine, J.-L.: Anneaux préordonnés. Journal d' Analyse Mathématique **12**, 307–326 (1964)

Kroó, A.: Multivariate needle polynomials with application to norming sets and cubature formulas. Acta Math. Hung. **147**(1), 46–72 (2015)

Lasserre, J.B.: Moments, Positive Polynomials and their Applications. World Scientific, Singapore (2009)

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

Lasserre, J.B.: A new look at nonnegativity on closed sets and polynomial optimization. SIAM J. Optim. **21**(3), 864–885 (2010)

Lasserre, J.B.: Introduction to Polynomial and Semi-Algebraic Optimization. Cambridge University Press, Cambridge, UK (2015)

Lasserre, J.B.: Connecting optimization with spectral analysis of tri-diagonal Hankel matrices. Math. Program. **190**, 795–809 (2021)

Laurent, M.: Sums of squares, moment matrices and optimization over polynomials. In: Putinar, M., Sullivant, S. (eds.) Emerging Applications of Algebraic Geometry. IMA Volumes in Mathematics and its Applications, vol. 149, pp. 157–270. Springer (2009)

Laurent, M.: A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0–1 programming. Math. Oper. Res. **28**(3), 470–496 (2023)

Laurent, M., Slot, L.: An effective version of Schmüdgen's Positivstellensatz for the hypercube. Optim. Lett. **17**, 515–530 (2023)

Magron, V., Wang, J.: Sparse Polynomial Optimization. Series on Optimization and its Applications. World Scientific (Europe) (2023)

Marshall, M.: Positive Polynomials and Sums of Squares. Mathematical Surveys and Monographs, vol. 146. American Mathematical Society, Providence, Rhode Island (2008)

Martinez, A., Piazzon, F., Sommariva, A., Vianello, M.: Quadrature-based polynomial optimization. Optim. Lett. **14**(5), 1027–1036 (2019)

Nesterov, Y.: Squared functional systems and optimization problems. In: Frenk, J.B.G., Roos, C., Terlaky, T., Zhang, S. (eds.) High Performance Optimization. Applied Optimization, vol. 33, pp. 405–440. Springer, Boston, Massachussets (2000)

Nie, J.: Moment and Polynomial Optimization. MOS-SIAM Series on Optimization (2023)

Nie, J.: Polynomial optimization with real varieties. SIAM J. Optim. **23**(3), 1634–1646 (2013)

Nie, J.: Optimality conditions and finite convergence of Lasserre's hierarchy. Math. Program. **146**(1), 97–121 (2014)

Nie, J., Schweighofer, M.: On the complexity of Putinar's Positivstellensatz. J. Complex. **23**(1), 135–150 (2007)

Parrilo, P.A.: Semidefinite programming relaxations for semialgebraic problems. Math. Program. Ser. B **96**, 293–320 (2003)

Parrilo, P.A.: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D. thesis, California Institute of Technology (2000)

Pólya, G.: Uber positive Darstellung von Polynomen, Vierteljahresschrift der Naturforschenden Gesellschaft in Zürich **73**, 14–145 (1928). Reprinted. In: Collected Papers, vol. 2, pp. 309–313. MIT Press, Cambridge (1974)

Powers, V.: Certificates of Positivity for Real Polynomials: Theory, Practice, and Applications. Developments in Mathematics, vol. 69 (2021)

Powers, V., Scheiderer, C.: The moment problem for non-compact semialgebraic sets. Adv. Geom. **1**(1), 71–88 (2001)

Prestel, A., Delzell, C.N.: Positive Polynomials—From Hilbert's 17th Problem to Real Algebra. Springer Monographs in Mathematics. Springer, Berlin (2001)

Putinar, M.: Positive polynomials on compact semi-algebraic sets. Indiana Univ. Math. J. **42**, 969–984 (1993)

Reznick, B.: Some concrete aspects of Hilbert's 17th problem. In: Real Algebraic Geometry and Ordered Structures (Baton Rouge, Louisiana, 1996; Delzell, C.N., Madden, J.J. (eds.)) Contemporary Mathematics, vol. 253, pp. 251–272. American Mathematical Society, Providence, Rhode Island (2000)

Reznick, B.: Uniform denominators in Hilbert's seventeenth problem. Math. Z. **220**, 75–97 (1995)

Roos, C., Terlay, T., Vial, J.-P.: Interior Point Methods for Linear Optimization. Springer (2005)

Scheiderer, C.: Non-existence of degree bounds for weighted sums of squares representations. J. Complex. **21**(6), 823–844 (2005)

Schlosser, C., Tacchi, M., Lazarev, A.: Convergence rates for the moment-SoS hierarchy (2024). arXiv:2402.00436

Schlosser, C., Tacchi, M.: Specialized effective Positivstellensätze for improved convergence rates of the moment-SOS hierarchy (2024). arXiv:2403.05934

Schmüdgen, K.: The K-moment problem for compact semi-algebraic sets. Math. Ann. **289**(2), 203–206 (1991)

Schweighofer, M.: On the complexity of Schmüdgen's Positivstellensatz. J. Complex. **20**(4), 529–543 (2004)

Shor, N.Z.: An approach to obtaining global extremums in polynomial mathematical programming problems. Kibernetika **5**, 102–106 (1987)

Slot, L., Laurent, M.: Sum-of-squares hierarchies for binary polynomial optimization. Math. Program. **197**, 621–660 (2022a); Singh, M., Williamson, D.P. (eds.) Integer Programming and Combinatorial Optimization (IPCO 2021). Lecture Notes in Computer Science, vol. 12707, pp. 43–57. Springer, Cham

Slot, L.: Asymptotic analysis of semidefinite bounds for polynomial optimization and independent sets in geometric hypergraphs. Ph.D. thesis, Tilburg University (2022b)

Slot, L.: Sum-of-squares hierarchies for polynomial optimization and the Christoffel-Darboux kernel. SIAM J. Optim. **32**(4), 2612–2635 (2022c)

Slot, L., Laurent, M.: Near-optimal analysis of Lasserre's univariate measure-based bounds for multivariate polynomial optimization. Math. Program. **188**, 443–460 (2021)

Slot, L., Laurent, M.: Improved convergence analysis of Lasserre's measure-based upper bounds for polynomial minimization on compact sets. Math. Program. **193**, 831–871 (2022d)

Stengle, G.: Complexity estimates for the Schmüdgen Positivstellensatz. J. Complex. **12**, 167–174 (1996)

Szegö, G.: Orthogonal Polynomials. American Mathematical Society Colloquium Publications, vol. 23. American Mathematical Society (1959)

Zohrizadeh, F., Josz, C., Jin, M., Madani, R., Lavaei, J., Sojoudi, S.: A survey on conic relaxations of optimal power flow problem. Eur. J. Oper. Res. **287**, 391–409 (2020)