Density, Determinacy, Duality
and a Regularized Moment-SOS Hierarchy

Didier Henrion[*]

December 4, 2025

Abstract

The standard moment-sum-of-squares (SOS) hierarchy is a powerful method for solving global polynomial optimization problems. However, its convergence relies on Putinar's Positivstellensatz, which requires the feasible set to satisfy the algebraic Archimedean property. In this paper, we introduce a regularized moment-SOS hierarchy capable of handling problems on unbounded sets or bounded sets violating the Archimedean property. Adopting a functional analysis viewpoint, we rely on the multivariate Carleman condition for measure determinacy rather than algebraic compactness. We prove that finite degree projections of the quadratic module are dense in the cone of positive polynomials with respect to the square norm induced by the measure. Based on these density results, we prove the convergence of a regularized hierarchy without invoking any Positivstellensatz. Furthermore, we propose a penalized formulation of the hierarchy which, combined with Bernstein-Markov inequalities, provides a monotonically non-decreasing sequence of certified lower bounds on the global minimum. The approach is illustrated on several benchmark problems known to be difficult or ill-posed for the standard hierarchy.

The *moment-SOS hierarchy* is a powerful approach for solving globally difficult non-linear non-convex problems. It was designed originally for polynomial optimization [5], for recent overviews and more references see e.g. [4, 13, 14]. Its convergence proof relies on Putinar's Positivstellensatz (Psatz) [16], which is a distinguished sum-of-squares (SOS) representation of positive polynomials on semialgebraic sets of the form

$$K := \{x \in \mathbb{R}^n : g_j(x) \geq 0, \ j = 1, \ldots, m\}$$

where $g_j \in \mathbb{R}[x]$ are given real polynomials of the indeterminate $x \in \mathbb{R}^n$. Let $g_0 := 1$ for notational convenience. Convergence of the hierarchy is guaranteed under a specific algebraic compactness assumption on the data which is called the *Archimedean property*: the quadratic module

$$Q(g) := \left\{ \sum_{j=0}^m s_j g_j : s_j \in \Sigma[x], \ j = 0, 1, \ldots, m \right\}$$

must contain the polynomial

$$R^2 - x^T x$$

for some number R, where $\Sigma[x]$ is the cone of SOS polynomials, see e.g. [15, 11]. This property implies that K is compact, and this restricts the application of the hierarchy in its native form.

---

[*]LAAS-CNRS, University of Toulouse, France and Faculty of Electrical Engineering, Czech Technical University in Prague, Czechia.

Indeed there are compact semialgebraic sets which violate the Archimedean property, one of the simplest being

$$\{x \in \mathbb{R}^2 : 2x_1 - 1 \geq 0, \ 2x_2 - 1 \geq 0, \ 1 - x_1 x_2 \geq 0\}$$

see [15, Ex. 6.3.1] and [12, Ex 6.5]. As surveyed e.g. in [13], the compactness assumption can be circumvented by using various reformulation strategies (change of coordinates, homogenization) so that the problem eventually boils down to a modified POP satisfying the Archimedean property. The compactness assumption was relaxed to a certain extent by introducing the multivariate Carleman condition in the context of polynomial optimization [6]. The Carleman condition ensures measure *determinacy*, i.e. the property that a measure is uniquely specified by its moments. This allowed to construct simple high degree perturbation to approximate positive polynomials with SOS [9], and this culminated in an elegant treatment in [8] of the problem of moments on unbounded semialgebraic sets, a problem dual to the SOS representation of positive polynomials, see e.g. [17] for a comprehensive account.

In this paper, we introduce a *regularized* moment-SOS hierarchy that can deal with POPs on unbounded sets, or bounded sets defined by polynomials violating the Archimedean property. The convergence proof of our regularized hierarchy does not rely on Putinar's Psatz. In contrast with [8] which develops a coefficient-space program in a weighted topology of discrete moment sequences, our treatment follows a function-space viewpoint. Similarly to [8], our approach relies on the Carleman condition. We formulate our results as *density* and separation properties in the topology of the Hilbert space $L^2(\mu)$ where μ is a measure satisfying the Carleman condition.

Our contributions are as follows. After the preliminary Sections 1 and 2, we provide in Section 3 a direct functional analytic proof that SOS polynomials are uniformly dense in positive polynomials. Without assuming the Archimedean property of $Q(g)$ or the compactness of K, we prove in Section 4 that finite degree projections of the quadratic module $Q(g)$ are uniformly dense in finite degree positive polynomials. In Section 5 we introduce a regularized version of the moment-SOS hierarchy and prove its convergence with our density results. In particular our convergence proof does not rely on any Psatz. We also describe an alternative penalized version of the regularized moment-SOS hierarchy that provides a monotonically non-decreasing sequence of lower bounds on the global minimum on compact sets. We illustrate our regularized moment-SOS hierarchy on various degenerate benchmark POPs in Section 6.

1   Carleman Determinacy

In the following, μ is a finite positive Borel measure. The Lebesgue space $L^1(\mu)$ consists of measurable integrable functions $f : \mathbb{R}^n \to \mathbb{R}$. This is a Banach space with the norm

$$\|f\|_{L^1(\mu)} := \int_{\mathbb{R}^n} |f(x)| \, d\mu(x).$$

The Lebesgue space $L^2(\mu)$ consists of measurable square integrable functions $f : \mathbb{R}^n \to \mathbb{R}$. This is a Hilbert space with inner product

$$\langle f, g \rangle_{L^2(\mu)} := \int f(x) g(x) \, d\mu(x)$$

and norm

$$\|f\|_{L^2(\mu)} := \Big(\int f^2(x) \, d\mu(x) \Big)^{1/2}.$$

Since μ is finite, $L^2(\mu) \subset L^1(\mu)$ and $\|f\|_{L^1(\mu)} \leq \mu(\mathbb{R}^n)^{1/2} \|f\|_{L^2(\mu)}$.

Definition 1. *Given an integer vector $\alpha \in \mathbb{N}^n$, the* moment *of order α of the measure μ is the real number*

$$y_\alpha = \int x^\alpha d\mu(x) := \int x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} d\mu(x_1, x_2, \ldots, x_n).$$

The moments of μ are the infinite countable sequence $(y_\alpha)_{\alpha \in \mathbb{N}^n} \subset \mathbb{R}$.

Definition 2. *A measure with given moments is called* determinate *if it is the unique measure with these moments.*

Definition 3. *For $i = 1, \ldots, n$, define the even marginal moments*

$$y_{i,2k} := \int_{\mathbb{R}^n} x_i^{2k} \, d\mu(x_1, \ldots, x_n), \qquad k \in \mathbb{N}.$$

We say that μ satisfies the multivariate Carleman condition, or satisfies Carleman for short, if for each $i = 1, \ldots, n$,

$$\sum_{k=1}^\infty y_{i,2k}^{-1/(2k)} = +\infty.$$

Theorem 1. *If μ satisfies Carleman, then it is determinate.*

For a statement and references to the original proof see e.g. [7, Prop. 3.5] and [17, Sec. 14.4].

Lemma 1. *If μ satisfies Carleman, then all its moments are finite.*

Proof. Even marginal moments are finite. Let $i \in \{1, \ldots, n\}$ be given. If $y_{i,2k_0} = +\infty$ for some k_0, then on a finite measure space $L^p \subset L^q$ for $p > q$, hence $|x_i| \notin L^{2k}$ for every $k \geq k_0$ and thus $y_{i,2k} = +\infty$ for all $k \geq k_0$. The Carleman series would then have only finitely many positive terms and converge, contradicting the hypothesis. Hence $y_{i,2k} < \infty$ for all $k \in \mathbb{N}$ and each i.

Marginal moments are finite. Let $i \in 1, \ldots, n$ and $r \geq 0$ be given. Choose $k \in \mathbb{N}$ with $2k \geq r$. Since $|x_i| \in L^{2k}(\mu)$ as above and $\mu(\mathbb{R}^n) < \infty$, the embeddings on finite measure spaces give

$$\int_{\mathbb{R}^n} |x_i|^r \, d\mu = \| |x_i| \|_{L^r(\mu)}^r \leq \mu(\mathbb{R}^n)^{1 - \frac{r}{2k}} \| |x_i| \|_{L^{2k}(\mu)}^r < \infty. \tag{1}$$

Absolute mixed moments are finite. Let $\alpha \in \mathbb{N}^n$ and $m := |\alpha| = \sum_{i=1}^n \alpha_i$. For $m \geq 1$, apply the weighted arithmetic-geometric inequality to $a_i := |x_i|^m$ with weights $w_i := \alpha_i/m$ (so $\sum_i w_i = 1$):

$$|x^\alpha| = \prod_{i=1}^n |x_i|^{\alpha_i} = \prod_{i=1}^n a_i^{w_i} \leq \sum_{i=1}^n w_i a_i \leq \sum_{i=1}^n |x_i|^m.$$

Integrating and using (1) with $r = m$ yields

$$\int_{\mathbb{R}^n} |x^\alpha| \, d\mu \leq \sum_{i=1}^n \int_{\mathbb{R}^n} |x_i|^m \, d\mu < \infty.$$

Signed moments are finite. By the triangle inequality,

$$\left| \int_{\mathbb{R}^n} x^\alpha \, d\mu \right| \leq \int_{\mathbb{R}^n} |x^\alpha| \, d\mu < \infty,$$

so $\int x^\alpha \, d\mu$ is finite for every α. $\qquad \square$

Lemma 2. *If μ has compact support, then it satisfies Carleman.*

Proof. Since the support of μ is compact, for each $i = 1, \ldots, n$ there exists $R_i > 0$ such that $|x_i| \leq R_i$ μ-a.e. Let $M := \mu(\mathbb{R}^n) \in (0, \infty)$. Then $y_{i,2k} \leq R_i^{2k} M$, $k \in \mathbb{N}$, so $y_{i,2k}^{-1/(2k)} \geq R_i^{-1} M^{-1/(2k)}$. Since $M^{-1/(2k)} \to 1$ as $k \to \infty$, there is k_0 such that $y_{i,2k}^{-1/(2k)} \geq \frac{1}{2} R_i^{-1}$ for all $k \geq k_0$. Hence $\sum_{k=1}^{\infty} y_{i,2k}^{-1/(2k)} \geq \sum_{k=k_0}^{\infty} \frac{1}{2} R_i^{-1} = +\infty$. As this holds for each coordinate i, the multivariate Carleman condition is satisfied. $\qquad\square$

Lemma 3. *Let $p \in \mathbb{R}[x]$ be a coercive polynomial, i.e., $\lim_{\|x\| \to \infty} p(x) = +\infty$. Then the Gibbs measure $\mu(dx) = e^{-p(x)} dx$ satisfies Carleman.*

Proof. If $p(x)$ is a coercive polynomial of degree d, then d is an even integer and the highest-degree part is positive definite. Hence there exist constants $a > 0$ and $b \in \mathbb{R}$ such that for all $x \in \mathbb{R}^n$:

$$p(x) \geq a\|x\|^d - b.$$

Therefore, the measure μ is bounded above by a simpler radial measure:

$$d\mu(x) = e^{-p(x)} dx \leq e^b e^{-a\|x\|^d} dx.$$

We must check Carleman's condition for each one-dimensional marginal. Let's bound the moments $y_{i,2k} = \int_{\mathbb{R}^n} x_i^{2k} d\mu(x)$ for a fixed coordinate i. Since $|x_i| \leq \|x\|$, we have $x_i^{2k} \leq \|x\|^{2k}$ and

$$y_{i,2k} = \int_{\mathbb{R}^n} x_i^{2k} e^{-p(x)} dx \leq e^b \int_{\mathbb{R}^n} \|x\|^{2k} e^{-a\|x\|^d} dx.$$

To evaluate the integral, let us use spherical coordinates. Let $r = \|x\|$ and let S_{n-1} be the surface area of the unit $(n-1)$-sphere.

$$\int_{\mathbb{R}^n} \|x\|^{2k} e^{-a\|x\|^d} dx = S_{n-1} \int_0^{\infty} r^{2k} e^{-ar^d} r^{n-1} dr = S_{n-1} \int_0^{\infty} r^{2k+n-1} e^{-ar^d} dr.$$

This integral evaluates in terms of the Gamma function to:

$$\frac{S_{n-1}}{d} a^{-\frac{2k+n}{d}} \Gamma\left(\frac{2k+n}{d}\right).$$

Thus, we have an upper bound on the moments where c is a constant independent of k:

$$y_{i,2k} \leq c a^{-\frac{2k}{d}} \Gamma\left(\frac{2k+n}{d}\right).$$

Hence

$$y_{i,2k}^{-1/(2k)} \geq c^{-1/(2k)} a^{1/d} \Gamma\left(\frac{2k+n}{d}\right)^{-1/(2k)}.$$

Using Stirling's formula for the Gamma function and the fact that $d \geq 2$, one obtains for some constant $c' > 0$ and all large k,

$$y_{i,2k}^{-1/(2k)} \geq c' k^{-1/d}.$$

Since $1/d \leq 1/2 < 1$, the series

$$\sum_{k=1}^{\infty} y_{i,2k}^{-1/(2k)} \geq c' \sum_{k=1}^{\infty} \frac{1}{k^{1/d}}$$

diverges. This holds for every coordinate i, so μ satisfies the multivariate Carleman condition. $\qquad\square$

In [3, Thm. 3.2.17] it is shown that if μ satisfies the condition

$$\int e^{c|x|} d\mu(x) < \infty \tag{2}$$

for some $c > 0$ then μ satisfies Carleman.

2 Polynomial Density

Lemma 4. *If μ satisfies Carleman, every polynomial $p \in \mathbb{R}[x]$ belongs to $L^2(\mu)$.*

Proof. The squared norm $\|p\|_{L^2(\mu)}^2 = \int p(x)^2 \, d\mu(x)$ is a finite linear combination of moments and it is therefore finite by Lemma 1. $\qquad\square$

Lemma 5. *A subspace is dense in $L^2(\mu)$ iff its orthogonal complement is trivial.*

Proof. For a linear subspace $V \subset L^2(\mu)$, its orthogonal complement is

$$V^\perp := \{\, h \in L^2(\mu) : \langle h, v\rangle_{L^2(\mu)} = 0 \text{ for all } v \in V \,\}.$$

It is a closed subspace of $L^2(\mu)$ and

$$\overline{V} \;=\; (V^\perp)^\perp. \tag{3}$$

To see this, observe first that $V \subset (V^\perp)^\perp$ by definition, hence $\overline{V} \subset (V^\perp)^\perp$ since $(V^\perp)^\perp$ is closed. For the reverse inclusion, let $P := \overline{V}$. The orthogonal decomposition theorem says every $h \in L^2(\mu)$ can be written uniquely as $h = p + q$ with $p \in P$ and $q \in P^\perp$. If $h \in (V^\perp)^\perp$, then h is orthogonal to $V^\perp$, hence to $P^\perp$ (because $V^\perp \supset P^\perp$). Thus $0 = \langle h, q\rangle = \|q\|_2^2$, so $q = 0$ and $h \in P = \overline{V}$. Therefore $(V^\perp)^\perp \subset \overline{V}$.

Now if $V^\perp = \{0\}$ then $(V^\perp)^\perp = L^2(\mu)$, and from relation (3) it holds $\overline{V} = L^2(\mu)$, i.e. V is dense. Conversely, if V is dense, then $\overline{V} = (V^\perp)^\perp = L^2(\mu)$, which forces $V^\perp = \{0\}$. $\qquad\square$

Lemma 6. *Let μ satisfy Carleman. Let ν be a signed measure which is absolutely continuous w.r.t. μ with density $f \in L^2(\mu)$, i.e. $d\nu = f \, d\mu$. Let $\nu = \nu^+ - \nu^-$ be its Jordan decomposition, so that $d\nu^\pm = f_\pm \, d\mu$ with $f_+ := \max\{f, 0\}$ and $f_- := \max\{-f, 0\}$. Then each $\nu^\pm$ satisfies Carleman.*

Proof. Given $i \in \{1, \ldots, n\}$, $k \in \mathbb{N}$, consider the marginal moments $y_{i,2k}$ as in Definition 3. Since μ satisfies Carleman, Lemma 1 implies $y_{i,2k} < \infty$ for all k, and the Carleman series

$$\sum_{k=1}^\infty y_{i,2k}^{-1/(2k)} = +\infty.$$

Let $f_\pm$ be as above, so $d\nu^\pm = f_\pm \, d\mu$ and $f_\pm \in L^2(\mu)$, $f_\pm \geq 0$. Define the even marginal moments of $\nu^\pm$ by

$$z_{2k}^\pm := \int_{\mathbb{R}^n} x_i^{2k} \, d\nu^\pm(x) = \int_{\mathbb{R}^n} x_i^{2k} f_\pm(x) \, d\mu(x), \qquad k \in \mathbb{N}.$$

Step 1: finiteness of the moments of $\nu^\pm$. By Cauchy-Schwarz,

$$z_{2k}^\pm \leq \|f_\pm\|_{L^2(\mu)} \left(\int_{\mathbb{R}^n} x_i^{4k} \, d\mu(x) \right)^{1/2},$$

and the right-hand side is finite since all moments of μ are finite (Lemma 1 with $r = 4k$). Thus $z_{2k}^\pm < \infty$ for all k.

Step 2: a lower bound on the Carleman terms. From the inequality above we obtain, for $k \geq 1$,

$$\left(z_{2k}^\pm \right)^{-1/(2k)} \;\geq\; \|f_\pm\|_{L^2(\mu)}^{-1/(2k)} \left(\int_{\mathbb{R}^n} x_i^{4k} \, d\mu(x) \right)^{-1/(4k)}.$$

Let

$$c_k := y_{i,2k}^{-1/(2k)}.$$

5

Then
$$\left(z_{2k}^{\pm}\right)^{-1/(2k)} \geq \|f_{\pm}\|_{L^2(\mu)}^{-1/(2k)} c_{2k}.$$

Since $\|f_{\pm}\|_{L^2(\mu)}$ is fixed, we have $\|f_{\pm}\|_{L^2(\mu)}^{-1/(2k)} \to 1$ as $k \to \infty$. Thus there exist constants $\gamma \in (0,1)$ and $k_0 \in \mathbb{N}$ such that

$$\left(z_{2k}^{\pm}\right)^{-1/(2k)} \geq \gamma\, c_{2k} \qquad \text{for all } k \geq k_0. \tag{4}$$

Step 3: divergence of the subsequence c_{2k}. We now use only that $c_k = y_{i,2k}^{-1/(2k)}$ comes from moments of a positive measure. Define $m_r := \left(\int |x_i|^r \, d\mu\right)^{1/r}$ for $r > 0$. By Hölder's inequality, the map $r \mapsto m_r$ is nondecreasing: if $0 < r < s$ then $m_r \leq m_s$. In particular the sequence

$$m_{2k} = \left(y_{i,2k}\right)^{1/(2k)}, \qquad k \geq 1,$$

is nondecreasing, hence

$$c_k = y_{i,2k}^{-1/(2k)} = m_{2k}^{-1}$$

is nonincreasing in k.

We know that the full Carleman series $\sum_{k=1}^{\infty} c_k$ diverges and that (c_k) is nonincreasing and nonnegative. For such a sequence, any even index subseries also diverges. Indeed, for each N,

$$\sum_{k=1}^{N} c_{2k} \geq \frac{1}{2} \sum_{k=1}^{N} (c_{2k} + c_{2k+1}) = \frac{1}{2}\Big(\sum_{j=2}^{2N+1} c_j\Big) = \frac{1}{2}\Big(\sum_{j=1}^{2N+1} c_j - c_1\Big).$$

As $N \to \infty$, the sum $\sum_{j=1}^{2N+1} c_j$ tends to $+\infty$, hence $\sum_{k=1}^{\infty} c_{2k}$ also diverges.

Step 4: Carleman for $\nu^{\pm}$. Combining (4) with the divergence of $\sum c_{2k}$ we obtain

$$\sum_{k=1}^{\infty} \left(z_{2k}^{\pm}\right)^{-1/(2k)} \geq \sum_{k=k_0}^{\infty} \left(z_{2k}^{\pm}\right)^{-1/(2k)} \geq \gamma \sum_{k=k_0}^{\infty} c_{2k} = +\infty.$$

Thus the Carleman series for the i-th marginal of $\nu^{\pm}$ diverges. $\qquad\square$

Theorem 2. *If μ satisfies Carleman, then polynomials are dense in $L^2(\mu)$. Equivalently, for every $f \in L^2(\mu)$ and every $\varepsilon > 0$, there exists $p \in \mathbb{R}[x]$ such that $\|f - p\|_{L^2(\mu)} < \varepsilon$.*

Proof. Assume $f \in L^2(\mu)$ is orthogonal to all polynomials:

$$\int_{\mathbb{R}^n} f(x)\, p(x)\, d\mu(x) = 0 \qquad \forall p \in \mathbb{R}[x].$$

Define a finite signed measure ν by $d\nu := f\, d\mu$. Finiteness holds because μ is finite and $\int |f|\, d\mu \leq \mu(\mathbb{R}^n)^{1/2} \|f\|_2 < \infty$.

For any multi-index $\alpha \in \mathbb{N}^n$, Cauchy-Schwarz and Lemma 4 give

$$\int_{\mathbb{R}^n} |x^{\alpha}|\, d|\nu|(x) = \int_{\mathbb{R}^n} |x^{\alpha}|\, |f(x)|\, d\mu(x) \leq \left(\int_{\mathbb{R}^n} x^{2\alpha}\, d\mu\right)^{1/2} \|f\|_{L^2(\mu)} < \infty.$$

Hence the α-moment of ν is finite, and by orthogonality

$$\int_{\mathbb{R}^n} x^{\alpha}\, d\nu(x) = \int_{\mathbb{R}^n} f(x)\, x^{\alpha}\, d\mu(x) = 0 \qquad \forall \alpha \in \mathbb{N}^n.$$

6

Write the Jordan decomposition $\nu = \nu^+ - \nu^-$ with $\nu^\pm$ finite positive measures. Since $|\nu|$ has all moments finite and $\nu^\pm \leq |\nu|$ as measures, each $\nu^\pm$ also has all (absolute) moments finite. Moreover,

$$\int_{\mathbb{R}^n} x^\alpha \, d\nu(x) = \int_{\mathbb{R}^n} x^\alpha \, d\nu^+(x) - \int_{\mathbb{R}^n} x^\alpha \, d\nu^-(x) = 0 \qquad \forall \alpha \in \mathbb{N}^n,$$

so

$$\int x^\alpha \, d\nu^+(x) = \int x^\alpha \, d\nu^-(x) \qquad \forall \alpha \in \mathbb{N}^n.$$

Since μ satisfies Carleman, by Lemma 6, $\nu^\pm$ satisfy Carleman as well. By Theorem 1 they are determinate. Since they share all moments they be must be equal: $\nu^+ = \nu^-$ and therefore $\nu = 0$.

Since $\nu = f\,\mu = 0$ as a signed measure, we have $\int_E f \, d\mu = 0$ for all measurable E. Taking $E = \{f > 0\}$ and $E = \{f < 0\}$ yields $\mu(\{f \neq 0\}) = 0$, i.e., $f = 0$ in $L^2(\mu)$.

Thus the orthogonal complement of the polynomials is $\{0\}$, so polynomials are dense in $L^2(\mu)$.
$\square$

It is recalled in [3, Sec. 3.2.3, p. 69] or [17, Sec. 14.1] that in the univariate case, if μ is determinate then polynomials are dense in $L^2(\mu)$, but this is not true in the multivariate case. If μ satisfies (2), it is shown in [3, Thm. 3.2.18] that polynomials are dense in $L^2(\mu)$. In [17, Def. 14.1] a distinction is made between determinate, strictly determinate and strongly determinate measures, depending on L^2 density of polynomials. In [17, Thm. 14.19] it is shown that if a measure satisfies Carleman, then it is strongly determinate. From [17, Thm. 14.2] a strongly determinate measure is strictly determinate, which implies $L^2(\mu)$ density of polynomials. So our Theorem 2 can be alternatively proven with [17, Thm. 14.2] and [17, Thm. 14.19].

3   SOS Density

Let

$$P := \{p \in \mathbb{R}[x] : \ p(x) \geq 0, \ \forall x \in \mathbb{R}^n\}$$

denote the cone of *positive polynomials*, and let

$$\Sigma := \left\{ \sum_k q_k^2 : \ q_k \in \mathbb{R}[x] \right\} \subset P$$

denote the cone of *sums of squares (SOS)* of polynomials.

We assume in the remainder of the paper that μ satisfies Carleman, so that the density Theorem 2 applies. The following instrumental result was stated and proved in [8, Theorem 2.2].

Lemma 7. *If $h \in L^2(\mu)$ satisfies $\int q(x)^2 h(x) \, d\mu(x) \geq 0$ for all $q \in \mathbb{R}[x]$, then $h \geq 0$ μ-a.e.*

Theorem 3. *If μ satisfies Carleman, then the cone Σ is dense in P w.r.t. $L^2(\mu)$. Equivalently, for every $p \in P$ and every $\varepsilon > 0$, there exists $s \in \Sigma$ such that $\|p - s\|_{L^2(\mu)} < \varepsilon$.*

Proof. Let C denote the $L^2(\mu)$ closure of the SOS cone Σ, a nonempty, closed, convex cone in the Hilbert space $L^2(\mu)$. Fix $p \in P$ and let us prove that $p \in C$. Assume by contradiction that $p \notin C$. By the Separating Hyperplane Theorem in Hilbert space – see e.g. [2, Thm. 3.38] – there exist $h \in L^2(\mu) \setminus \{0\}$ such that

$$\langle h, p \rangle_{L^2(\mu)} \ < \ 0 \ \leq \ \langle h, s \rangle_{L^2(\mu)} \quad \forall s \in C. \tag{5}$$

Since Σ contains q^2 for every polynomial q, the first inequality in (5) gives $\int q^2 \, h \, d\mu \geq 0$ for all $q \in \mathbb{R}[x]$. By Lemma 7 this implies $h \geq 0$ μ-a.e. Because $p \in P$ is nonnegative everywhere and $h \geq 0$ a.e., we have $\langle h, p \rangle_{L^2(\mu)} = \int p \, h \, d\mu \geq 0$, contradicting (5). Therefore $p \in C$. $\qquad\square$

Let $\mathbb{R}[x]_d$ denote the vector space of polynomials of degree up to d. Let $b = (b_\alpha)_{|\alpha| \leq d}$ denote a basis of $\mathbb{R}[x]_d$ orthonormal w.r.t. μ. In this basis any polynomial $p = \sum_{|\alpha| \leq d} p_\alpha b_\alpha$ can be expressed with the coefficient vector $(p_\alpha)_{|\alpha| \leq d}$, and by orthonormality and bilinearity,

$$\|p\|^2_{L^2(\mu)} = \Big\langle \sum_\alpha p_\alpha b_\alpha, \ \sum_\beta p_\beta b_\beta \Big\rangle_{L^2(\mu)} = \sum_{\alpha,\beta} p_\alpha p_\beta \, \langle b_\alpha, b_\beta \rangle_{L^2(\mu)} = \sum_\alpha p_\alpha^2 =: \|p\|_2^2. \tag{6}$$

In other words, the $L^2(\mu)$ norm of a polynomial is the Euclidean norm of its coefficient vector in an orthonormal basis of $L^2(\mu)$. Note that we use the notation p indifferently for the polynomial function or the vector of its coefficients in basis b.

The $L^2(\mu)$ orthogonal *projection* onto $\mathbb{R}[x]_d$ is the linear operator

$$\Pi_d : L^2(\mu) \to \mathbb{R}[x]_d \subset L^2(\mu)$$

defined as follows: for any $f \in L^2(\mu)$, the polynomial $\Pi_d f \in \mathbb{R}[x]_d$ is the unique minimizer

$$\Pi_d f \ = \ \arg \min_{p \in \mathbb{R}[x]_d} \|f - p\|_{L^2(\mu)}.$$

Equivalently, the projection is characterized by the orthogonality conditions

$$\int \big(f(x) - \Pi_d f(x) \big) q(x) \, d\mu(x) \ = \ 0 \quad \text{for all } q \in \mathbb{R}[x]_d$$

or the orthogonal basis expression

$$\Pi_d f(x) \ = \ \sum_{|\alpha| \leq d} \langle f, b_\alpha \rangle_{L^2(\mu)} \, b_\alpha(x).$$

Theorem 4. *If μ satisfies Carleman, then the cone $\Pi_{2d}(\Sigma)$ is dense in $P \cap \mathbb{R}[x]_{2d}$ w.r.t $L^2(\mu)$. Equivalently, for every $p \in P \cap \mathbb{R}[x]_{2d}$ and every $\varepsilon > 0$, there exists $s \in \Sigma$ such that $\|p - \Pi_{2d}s\|_{L^2(\mu)} < \varepsilon$.*

Proof. By Theorem 3, Σ is dense in P w.r.t. $L^2(\mu)$. Given $p \in P \cap \mathbb{R}[x]_{2d}$ and $\varepsilon > 0$, pick $s \in \Sigma$ with $\|p - s\|_{L^2(\mu)} < \varepsilon$. Since $p \in \mathbb{R}[x]_{2d}$ and Π_{2d} is the $L^2(\mu)$ orthogonal projection,

$$\| p - \Pi_{2d}s \|_{L^2(\mu)} \ \leq \ \| p - s \|_{L^2(\mu)} \ < \ \varepsilon,$$

which proves the density of $\Pi_{2d}(\Sigma)$ in $P \cap \mathbb{R}[x]_{2d}$. $\qquad\square$

Remark 1. *An $L^2(\mu)$ orthogonal projection does not preserve positivity (or SOS) in general. For a simple example, take for μ the uniform probability measure on $[-1, 1]$ and the SOS polynomial $p(x) = x^4$. Using the Legendre basis $l_0(x) = 1$, $l_1(x) = x$, $l_2(x) = \frac{1}{2}(3x^2 - 1)$, with inner products $\langle l_n, l_m \rangle = \int_{-1}^1 l_n l_m \, dx = \frac{2}{2n+1} \delta_{nm}$, we compute the projection coefficients*

$$p_n = \frac{\langle p, l_n \rangle}{\langle l_n, l_n \rangle}, \qquad n = 0, 1, 2.$$

Since p is even, $\langle p, l_1 \rangle = p_1 = 0$. Then

$$\langle p, l_0 \rangle = \int_{-1}^1 x^4 \, dx = \frac{2}{5}, \qquad \langle p, l_2 \rangle = \frac{1}{2} \Big(3 \int_{-1}^1 x^6 dx - \int_{-1}^1 x^4 dx \Big) = \frac{1}{2} \Big(3 \cdot \frac{2}{7} - \frac{2}{5} \Big) = \frac{8}{35}.$$

Hence, using $\langle l_0, l_0 \rangle = 2$ *and* $\langle l_2, l_2 \rangle = \frac{2}{5}$,

$$p_0 = \frac{\langle p, l_0 \rangle}{\langle l_0, l_0 \rangle} = \frac{1}{5}, \qquad p_2 = \frac{\langle p, l_2 \rangle}{\langle l_2, l_2 \rangle} = \frac{8/35}{2/5} = \frac{4}{7}.$$

Therefore

$$\Pi_2 p = p_0 l_0 + p_2 l_2 = \frac{1}{5} + \frac{4}{7} \cdot \frac{3x^2 - 1}{2} = \frac{6}{7} x^2 - \frac{3}{35}$$

which is negative at $x = 0$.

Remark 2. *The cone* $\Sigma \cap \mathbb{R}[x]_{2d}$ *is a closed cone in a finite-dimensional space. Since there exist positive polynomials whose neighborhoods are not SOS,* $\Sigma \cap \mathbb{R}[x]_{2d}$ *cannot be dense in* $P \cap \mathbb{R}[x]_{2d}$. *Theorem 4 states that if we allow an SOS of possibly higher degree, then its degree 2d truncation approximates any polynomial in* $P \cap \mathbb{R}[x]_{2d}$ *arbitrarily well.*

Example 1. *The Motzkin polynomial*

$$p_M(x) := 1 - 3x_1^2 x_2^2 + x_1^4 x_2^2 + x_1^2 x_2^4$$

is a well-known example of positive but not SOS polynomial, i.e. $p_M \in P \setminus \Sigma$, *see e.g.* [13, *Section 2.4.2]. Consider*

$$s(x) := \tfrac{1}{2} (x_1^2 x_2 + x_1 x_2^2)^2 + \tfrac{1}{2} (x_1^2 x_2 - x_1 x_2^2)^2 + (1 - \tfrac{3}{2} x_1^2 x_2^2)^2.$$

It holds $\frac{1}{2} (x_1^2 x_2 + x_1 x_2^2)^2 + \frac{1}{2} (x_1^2 x_2 - x_1 x_2^2)^2 = x_1^4 x_2^2 + x_1^2 x_2^4$ *and* $(1 - \frac{3}{2} x_1^2 x_2^2)^2 = 1 - 3x_1^2 x_2^2 + \frac{9}{4} x_1^4 x_2^4$ *and hence* $s(x) = p_M(x) + \frac{9}{4} x_1^4 x_2^4$. *Therefore* $s \in \Sigma \cap \mathbb{R}[x]_8$ *matches* $p_M \in P \cap \mathbb{R}[x]_6$ *in all degrees* ≤ 6, *i.e.*

$$\|p_M - \Pi_6 s\|_{L^2(\mu)} = 0,$$

where μ *is any measure on* $\mathbb{R}^2$ *satisfying Carleman.*

4 Quadratic Module Density

Let $g = (g_1, \ldots, g_m) \subset \mathbb{R}[x]$ and let $g_0 := 1$. Define the closed basic semialgebraic set

$$K := \{x \in \mathbb{R}^n : g_j(x) \geq 0 \text{ for } j = 1, \ldots, m\}, \tag{7}$$

the quadratic module

$$Q(g) := \left\{ \sum_{j=0}^m s_j g_j : s_j \in \Sigma \right\} \subset \mathbb{R}[x],$$

and the cone of polynomials positive on K

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

Theorem 5. *If* μ *satisfies Carleman, then the cones* $Q(g)$ *and* $P(K)$ *have the same closures w.r.t.* $L^2(\mu)$. *Equivalently, for every* $p \in P(K)$ *and every* $\varepsilon > 0$ *there exists* $s \in Q(g)$ *such that* $\|p - s\|_{L^2(\mu)} < \varepsilon$.

Proof. All the closures in this proof are w.r.t. $L^2(\mu)$. The inclusion of the closure of $Q(g)$ into the closure of $P(K)$ is immediate since $Q(g) \subset P(K)$ by construction.

For the reverse inclusion, choose p in the closure of $P(K)$ and suppose by contradiction that p does not belong to C, defined as the closure of $Q(g)$. By the Separating Hyperplane Theorem in $L^2(\mu)$ – see e.g. [2, Thm. 3.38] – there exists $h \in L^2(\mu) \setminus \{0\}$ such that

$$\langle h, p \rangle_{L^2(\mu)} < 0 \leq \langle h, s \rangle_{L^2(\mu)} \quad \forall s \in C. \tag{8}$$

Since $\Sigma \subset Q(g) \subset C$, taking $s = q^2$ yields $\int q(x)^2 h(x) \, d\mu(x) \geq 0$ for all $q \in \mathbb{R}[x]$. By Lemma 7, $h \geq 0$ μ-a.e.

Next, for each j we also have $q^2 g_j \in Q(g) \subset C$, hence $\int q(x)^2 g_j(x) h(x) \, d\mu(x) \geq 0$ for all $q \in \mathbb{R}[x]$. By Lemma 7 we get $g_j h \geq 0$ μ-a.e. for every j. In particular, on $\{x : g_j(x) < 0\}$ one must have $h(x) = 0$. Therefore $h = 0$ μ-a.e. on $\{x : \min_j g_j(x) < 0\} = \mathbb{R}^n \setminus K$, i.e., $\operatorname{spt} h \subseteq K$.

Combining $h \geq 0$ a.e. with $\operatorname{spt} h \subseteq K$ we now use that p lies in the closure of $P(K)$: consider a sequence $p_k \in P(K)$ with $p_k \to p$ in $L^2(\mu)$. Then $\langle h, p_k \rangle_{L^2(\mu)} \geq 0$ for every k, and by Cauchy-Schwarz, $\langle h, p_k \rangle \to \langle h, p \rangle$ as $k \to \infty$. Hence $\langle h, p \rangle \geq 0$, contradicting (8). Therefore p belongs to the closure of $Q(g)$. This proves that the closures of $P(K)$ and $Q(g)$ coincide. $\qquad \square$

Remark 3. *No assumption on compactness of K or Archimedean property of $Q(g)$ is required.*

Theorem 6. *Let μ satisfy Carleman. Given $d \in \mathbb{N}$, the cone $\Pi_d(Q(g))$ is dense in $P(K) \cap \mathbb{R}[x]_d$ w.r.t. $L^2(\mu)$. Equivalently, for every $p \in P(K) \cap \mathbb{R}[x]_d$ and every $\varepsilon > 0$ there exists $s \in Q(g)$ such that*

$$\left\| p - \Pi_d s \right\|_{L^2(\mu)} < \varepsilon.$$

Proof. By Theorem 5, the $L^2(\mu)$ closures of $Q(g)$ and $P(K)$ coincide. Fix $p \in P(K) \cap \mathbb{R}[x]_d$ and $\varepsilon > 0$. Since p belongs to the closure of $Q(g)$, there exists $s \in Q(g)$ satisfying $\|p - s\|_{L^2(\mu)} < \varepsilon$. Because $p \in \mathbb{R}[x]_d$ and Π_d is an orthogonal projection (hence a contraction),

$$\left\| p - \Pi_d s \right\|_{L^2(\mu)} = \left\| \Pi_d(p - s) \right\|_{L^2(\mu)} \leq \left\| p - s \right\|_{L^2(\mu)} < \varepsilon.$$

Since $\Pi_d s \in \Pi_d(Q(g))$, this shows that $\Pi_d(Q(g))$ is dense in $P(K) \cap \mathbb{R}[x]_d$ with respect to $L^2(\mu)$. $\qquad \square$

Example 2. *Let $g(x) = -x^2$ and $p(x) = x$. The quadratic module $Q(g)$ is Archimedean since it contains $1 - x^2$. For any $\varepsilon > 0$, define the quadratic*

$$q_\varepsilon(x) := \varepsilon + \frac{x}{2\varepsilon} - \frac{x^2}{8\,\varepsilon^3}, \qquad s_\varepsilon(x) := q_\varepsilon(x)^2 \in Q(g).$$

A direct expansion yields

$$s_\varepsilon(x) = \varepsilon^2 + x - \frac{1}{8\varepsilon^4}\, x^3 + \frac{1}{64\varepsilon^6}\, x^4,$$

so its degree-2 projection is

$$\Pi_2\big(s_\varepsilon\big) = \varepsilon^2 + x.$$

With μ the Gaussian measure, we obtain the exact error

$$\|x - \Pi_2(s_\varepsilon)\|_{L^2(\mu)}^2 = \varepsilon^4 \xrightarrow[\varepsilon \to 0]{} 0.$$

Moreover, inspection reveals that an exact match $\Pi_2(s) = x$ is impossible in $Q(g)$.

Example 3. *Let $g(x) := (1 - x^2)^3$ and $p(x) := 1 - x^2$. The quadratic module $Q(g)$ is Archimedean since $4 - x^2 = (2x^3 - 3x)^2 + 2x^2 + 4(1 - x^2)^3 \in Q(g)$. However, it can be checked that $p \in P(K) \setminus Q(g)$, see [18] and the developments in [1, Section 2.4]. Yet, with the choice*

$$s(x) := \tfrac{2}{3} + \tfrac{1}{3} g(x) = 1 - x^2 + x^4 - \tfrac{1}{3} x^6$$

it holds

$$\Pi_2(s) = p.$$

Example 4. Let $g(x) := (x_1 - \frac{1}{2}, x_2 - \frac{1}{2}, 1 - x_1 x_2)$ and $p(x) := \frac{17}{4} - x_1^2 - x_2^2$. The set K is compact, but it can be checked that $Q(g)$ is not Archimedean – see [15, Ex. 6.3.1] and [12, Ex 6.5] – and hence $p \in P(K) \setminus Q(g)$. Yet, with the choice

$$s(x) := \tfrac{17}{4} + 2x_1^2 \, g_1(x) + 2x_2^2 \, g_2(x) = \tfrac{17}{4} - x_1^2 - x_2^2 + 2x_1^3 + 2x_2^3 \in Q(g)$$

it holds

$$\Pi_2(s) = p.$$

Example 5. Let $g_1(x) := x^3$, $g_2(x) := 1 - x^2$ and $p(x) := x$. Therefore $K = [0, 1]$ and the quadratic module $Q(g)$ is Archimedean since it trivially contains $1 - x^2 = g_2$. However, $p \in P(K) \setminus Q(g)$. To prove this by contradiction, assume that there exist SOS polynomials s_0, s_1, s_2 such that $x = s_0(x) + s_1(x) \, x^3 + s_2(x) \, (1 - x^2)$. At $x = 0$ it yields $0 = s_0(0) + s_2(0)$, and since s_0 and s_2 are SOS, this implies $s_0(0) = s_2(0) = 0$. It follows that $s_0(x) = x^2 a(x)$ and $s_2(x) = x^2 b(x)$ with $a, b \in \Sigma[x]$. The SOS decomposition becomes $x = x^2 a(x) + x^3 s_1(x) + (1 - x^2) \, x^2 b(x) = x^2 (a(x) + x s_1(x) + (1 - x^2) b(x))$. Thus the right-hand side is divisible by x^2, whereas the left-hand side x is not divisible by x^2, a contradiction. Therefore no SOS representation exists, and $x \notin Q(g)$.

However, for any $\varepsilon \geq \frac{1}{2}$, letting $\delta := \sqrt{4\varepsilon^2 - 1}$, we have the simple SOS decomposition

$$x + \varepsilon = \underbrace{\frac{x^2 (x + \delta)^2}{4\varepsilon}}_{s_0(x) \in \Sigma[x]} + \underbrace{\frac{2(2\varepsilon - \delta)}{4\varepsilon}}_{s_1(x) \in \Sigma[x]} \, x^3 + \underbrace{\frac{(x + 2\varepsilon)^2}{4\varepsilon}}_{s_2(x) \in \Sigma[x]} \, (1 - x^2).$$

Putinar's Theorem guarantees that a SOS decomposition exists for every value of $\varepsilon > 0$, but it becomes more complicated (higher degree SOS and higher algebraic degree in ε) when $\varepsilon \to 0$.

Although $p \notin Q(g)$, we can approximate p arbitrarily well by first degree truncations of elements of $Q(g)$. For any $\varepsilon > 0$, let

$$s_\varepsilon(x) := (\varepsilon + \frac{x}{2\varepsilon})^2 = \varepsilon^2 + x + \frac{1}{4\varepsilon^2} x^2 \in \Sigma \subset Q(g)$$

and hence

$$\Pi_1(s_\varepsilon) = \varepsilon^2 + x.$$

Therefore, for any Carleman measure μ with $1 \in L^2(\mu)$,

$$\left\| p - \Pi_1(s_\varepsilon) \right\|_{L^2(\mu)} = \| \varepsilon^2 \|_{L^2(\mu)} = \varepsilon^2 \, \| 1 \|_{L^2(\mu)} \xrightarrow[\varepsilon \downarrow 0]{} 0.$$

Note however that an exact identity $\Pi_1(s) = x$ with $s \in Q(g)$ is impossible. Indeed, write $s = s_0 + x^3 s_1 + (1 - x^2) s_2$ with $s_i \in \Sigma$. Only s_0 and $(1 - x^2) s_2$ contribute to degrees zero and one. Let a_0 (resp. b_0) be the constant coefficient of the polynomial whose square is s_0 (resp. s_2). Then the constant term of $\Pi_1(s)$ equals $a_0^2 + b_0^2 \geq 0$. If $\Pi_1(s)$ had zero constant term (as x does), we would need $a_0 = b_0 = 0$, and for any polynomial r, $r(0) = 0$ implies $(r^2)'(0) = 2 \, r(0) \, r'(0) = 0$, so both linear contributions from s_0 and $(1 - x^2) s_2$ would vanish, forcing the linear coefficient of $\Pi_1(s)$ to be 0, a contradiction. Thus $\Pi_1(s) = x$ cannot hold, although $\Pi_1(s_\varepsilon) \to x$ in $L^2(\mu)$.

5 A regularized moment-SOS hierarchy

Given a polynomial $p \in \mathbb{R}[x]$ and a basic semialgebraic set K as in (7), we are consider the polynomial optimization problem (POP)

$$p^* := \inf_{x \in K} p(x).$$

11

Assumption 1 (Full-dimensional reference measure). *The set K has non-empty interior and the probability measure μ satisfies Carleman and it is absolutely continuous w.r.t. the Lebesgue measure with a continuous density w such that there exists an open neighborhood $U \supset K$ and a constant w_m such that $w(x) \geq w_m > 0$ for all $x \in U$.*

Assumption 1 implies that (i) no nonzero polynomial vanishes μ-a.e. on K; (ii) every nonempty relatively open subset $V \subset K$ satisfies $\mu(V) \geq w_m |V| > 0$; (iii) $Q(g)$ is dense in $P(K)$ in $L^2(\mu)$, see Theorem 5.

5.1 Dual SOS problem

Let $d_j := \lceil \deg g_j/2 \rceil$ for $j = 0, 1, \ldots, m$. Given a relaxation degree d and a parameter $\varepsilon_d > 0$, consider the SOS optimization problem

$$v_d^*(\varepsilon_d) := \sup \ v$$

$$\text{s.t.} \quad p - v = r + \sum_{j=0}^{m} s_j g_j, \tag{9}$$

$$\|r\|_{L^2(\mu)} \leq \varepsilon_d$$

where the maximization is w.r.t. scalar v, residual polynomial $r \in \mathbb{R}[x]_{2d}$ and SOS multipliers $s_j \in \Sigma[x]_{2(d-d_j)}$. Note that $\|r\|_{L^2(\mu)} = \|r\|_2$ can be readily implemented by a second-order cone constraint on the coefficient vector of r in an orthonormal basis w.r.t. $L^2(\mu)$, recall identity (6).

In SOS problem (9), polynomial $p - v - r$ belongs to the truncated quadratic module

$$Q(g)_d := \left\{ \sum_{j=0}^{m} s_j g_j : \ s_j \in \Sigma[x], \ \deg(s_j g_j) \leq 2d \right\} \subset P(K).$$

5.2 Primal pseudo-moment problem

Introduce the Riesz functional $\ell_y : \mathbb{R}[x]_{2d} \to \mathbb{R}$, $\ell_y(q) := \sum_{|\alpha| \leq 2d} q_\alpha y_\alpha$. In the primal problem (9) write each SOS in Gram form $s_j(x) = b_{d-d_j}^T(x) Q_j b_{d-d_j}(x)$ with $Q_j \succeq 0$. The Lagrangian with dual variable y acting on the identity $p - v_d = r + \sum_j s_j g_j$ reads

$$L(v_d, r, (s_j), y) := v_d + \ell_y(p - v_d - r - \sum_{j=0}^{m} s_j g_j),$$

together with conic indicators for $Q_j \succeq 0$ and $\|r\|_2 \leq \varepsilon_d$. Maximizing L over v_d forces $\ell_y(1) = 1$. Maximizing over the Gram matrices gives the standard moment and localizing matrix constraints $M_{d-d_j}(g_j y) \succeq 0$. Maximizing over r_d against the Euclidean ball $\|r\|_2 \leq \varepsilon_d$ produces the support function $\varepsilon_d \|y\|_2$. Thus the primal problem reads:

$$u_d^*(\varepsilon_d) := \inf \quad \ell_y(p) + \varepsilon_d \|y\|_2$$
$$\text{s.t.} \quad \ell_y(1) = 1, \tag{10}$$
$$M_{d-d_j}(g_j y) \succeq 0, \quad j = 0, 1, \ldots, m$$

where the minimization is w.r.t. the vector of pseudo-moments $(y_\alpha)_{|\alpha| \leq 2d}$. Alternatively, it can be formulated with a second-order cone constraint

$$\inf \quad \ell_y(p) + \varepsilon_d t$$
$$\text{s.t.} \quad \ell_y(1) = 1,$$
$$M_{d-d_j}(g_j y) \succeq 0, \quad j = 0, 1, \ldots, m$$
$$\|y\|_2 \leq t.$$

Weak duality ($u_d^* \geq v_d^*$) follows readily between dual SOS problem (9) and primal moment problem (10). Strong duality ($u_d^* = v_d^*$) follows from Slater conditions:

- there exists a pseudo-moment vector y such that $\ell_y(1) = 1$, $M_{d-d_j}(g_j\, y) \succ 0$, $j = 0, 1, \ldots, m$, $\|y\|_2 < t$. A concrete sufficient condition is the existence of a strict feasible point in K. Then, for any small ball B around an interior point of K, the moments of the normalized Lebesgue measure on B will be strictly feasible.

- there exist $(v, r, (s_j))$ such that $\|r\|_{L^2(\mu)} < \varepsilon_d$, $s_j \in \Sigma[x]_{2(d-d_j)}$ with Gram matrices $Q_j \succ 0$. A practical way to enforce strict primal feasibility (for large d) is: pick any $\delta > 0$, use the $L^2(\mu)$-density of $Q(g)$ in $P(K)$ to find $q \in Q(g)$ with $\|p - (p^* - \delta) - q\|_{L^2(\mu)} < \varepsilon_d$, let $r := p - (p^* - \delta) - q$, and then perturb the SOS Gram matrices by a tiny ϵI (absorbing the change into r) so that $Q_j \succ 0$ while keeping $\|r\|_{L^2(\mu)} < \varepsilon_d$.

5.3 Convergence

Theorem 7. *Let Assumption 1 hold. If $\varepsilon_d \downarrow 0$, then $v_d^\star(\varepsilon_d) \to p^\star$ as $d \to \infty$.*

Proof. Upper limit. Fix $v > p^\star$. Then $r := p - v$ is strictly negative at some point of K, hence by continuity there exist $\gamma > 0$ and a relatively open nonempty $V \subset K$ such that $r(x) \leq -\gamma$ on V. For any $q \in Q(g)$ we have $q \geq 0$ on K, so $r - q \leq -\gamma$ on V. Using $w \geq w_m$ on K,

$$\|r - q\|_{L^2(\mu)}^2 = \int_K |r-q|^2\, w\, dx \;\geq\; \int_V \gamma^2\, w\, dx \;\geq\; \gamma^2 w_m\, |V| \;=:\; \delta(v)^2 \;>\; 0.$$

In particular $\mathrm{dist}_{L^2(\mu)}(p - v, Q(g)_d) \geq \delta(v)$ for every d. If $\varepsilon_d < \delta(v)$ then v is infeasible in (9). Since $\varepsilon_d \to 0$, we obtain $\limsup_{d\to\infty} v_d^\star \leq p^\star$.

Lower limit. Fix $\eta > 0$ and set $p_\eta := p - (p^\star - \eta) \in P(K)$. By the $L^2(\mu)$-density of $Q(g)$ in $P(K)$, for large enough d, there exists $q \in Q(g)_d$ with $\|p_\eta - q\|_{L^2(\mu)} \leq \varepsilon_d$. Let $v := p^\star - \eta$ and $r := p_\eta - q$; then $p - v = r + q$ and $\|r\|_{L^2(\mu)} \leq \varepsilon_d$, so v is feasible in (9) and $v_d^\star \geq p^\star - \eta$. Let $d \to \infty$ and then $\eta \downarrow 0$ to get $\liminf_{d\to\infty} v_d^\star \geq p^\star$. $\qquad\square$

Assumption 2. *Assume that (K, μ) satisfies a Bernstein-Markov (BM) inequality at degree $2d$, i.e., there exist a constant $c_{2d} \geq 1$ such that for all $r \in \mathbb{R}[x]_{2d}$,*

$$\|r\|_{L^\infty(K)} \;\leq\; c_{2d}\, \|r\|_{L^2(\mu)}, \tag{11}$$

see e.g. [10, Section 4.3.3].

Corollary 1. *Under Assumptions 1 and 2, every optimal solution of (9) yields a certified lower bound*

$$p_d^*(\varepsilon_d) \;:=\; v_d^\star - c_{2d}\, \varepsilon_d \;\leq\; p^\star.$$

Moreover, if $c_{2d}\, \varepsilon_d \to 0$, then $p_d^*(\varepsilon_d) \to p^\star$.

Proof. For any feasible $(v, r, (s_j))$ in (9) and all $x \in K$,

$$p(x) = v + r(x) + \sum_j s_j(x) g_j(x) \;\geq\; v + r(x) \;\geq\; v - \|r\|_{L^\infty(K)} \;\geq\; v - c_{2d}\|r\|_{L^2(\mu)} \;\geq\; v - c_{2d}\varepsilon_d.$$

Taking the infimum over $x \in K$ and then the supremum over feasible v gives the bound. If $c_{2d}\varepsilon_d \to 0$, then $p_d^* \uparrow p^\star$ by Theorem 7. $\qquad\square$

Remark 4. *Let* $\{b_\alpha\}_{|\alpha|\leq 2d}$ *be an* $L^2(\mu)$ *orthonormal polynomial basis and define the degree* $2d$ *Christoffel-Darboux polynomial*

$$p^\mu_{2d}(x) := \sum_{|\alpha|\leq 2d} b_\alpha(x)^2.$$

Then the tighest BM constant in (11) *is*

$$c_{2d} = \sup_{x\in K} \sqrt{p^\mu_{2d}(x)},$$

see e.g. [10, Section 4.3.3].

5.4   Lower bounds with penalized formulation

Let us now describe an equivalent reformulation of regularized problems (9) and (10) that does not involve the parameter ϵ_d. Consider the ℓ_2 penalized formulation:

$$\hat{v}^*_d := \sup_{} v - c_{2d}\|r\|_{L^2(\mu)}$$
$$\text{s.t.}\quad p - v = r + \sum_{j=0}^{m} s_j g_j \tag{12}$$

where the maximization is w.r.t. scalar v, residual polynomial $r\in\mathbb{R}[x]_{2d}$ and SOS multipliers $s_j\in\Sigma[x]_{2(d-d_j)}$. As in the proof of Corollary 1, any feasible tuple certifies on K the pointwise lower bound $p(x)\geq v - c_{2d}\|r\|_2$ for $x\in K$, so $\hat{v}^*_d\leq\min_{x\in K} p(x)$ is a valid lower bound. SOS problem (12) is the dual to the primal moment problem

$$\hat{u}^*_d := \inf_{} \ell_y(p)$$
$$\text{s.t.}\quad \ell_y(1) = 1,$$
$$M_{d-d_j}(g_j\, y) \succeq 0,\quad j = 0, 1, \ldots, m \tag{13}$$
$$\|y\|_2 \leq c_{2d}.$$

As above, strong duality $(\hat{u}^*_d = \hat{v}^*_d)$ follows from Slater conditions. The next result shows that no tuning of ε is needed: any optimal solution v^*, r^* of (12) both selects the optimal $\varepsilon_d = c_{2d}\|r^*\|_2$ and returns the tightest BM certified bound $p^*_d = v^* - c_{2d}\|r^*\|_2$.

Lemma 8. *It holds*

$$\hat{v}^*_d = \sup_{\varepsilon\geq 0} (v^*_d(\varepsilon) - c_{2d}\,\varepsilon). \tag{14}$$

Proof. Let $\mathcal{F}_d$ be the degree-$2d$ SOS–feasible set of triples $(v, r, (s_j))$ with $p - v = r + \sum_{j=0}^{m} s_j g_j$, $r\in\mathbb{R}[x]_{2d}$, $s_j\in\Sigma[x]_{2(d-d_j)}$. By definition,

$$\hat{v}^*_d = \sup_{(v,r,(s_j))\in\mathcal{F}_d} \left\{ v - c_{2d}\|r\|_{L^2(\mu)} \right\}.$$

For each fixed r,

$$v - c_{2d}\|r\|_{L^2(\mu)} = \sup_{\varepsilon\geq 0} \left\{ v - c_{2d}\varepsilon \ : \ \|r\|_{L^2(\mu)} \leq \varepsilon \right\},$$

with the supremum attained at $\varepsilon = \|r\|_{L^2(\mu)}$. Hence

$$\hat{v}^*_d = \sup_{(v,r,(s_j))\in\mathcal{F}_d} \sup_{\varepsilon\geq 0} \left\{ v - c_{2d}\varepsilon \ : \ \|r\|_{L^2(\mu)} \leq \varepsilon \right\}.$$

Interchanging the two suprema (over a product set) yields

$$\hat{v}_d^* = \sup_{\varepsilon \geq 0} \ \sup \left\{ v - c_{2d}\varepsilon \ : \ (v, r, (s_j)) \in \mathcal{F}_d, \ \|r\|_{L^2(\mu)} \leq \varepsilon \right\}.$$

The inner supremum is precisely $v_d^*(\varepsilon) - c_{2d}\varepsilon$ by the definition of $v_d^*(\varepsilon)$, whence

$$\hat{v}_d^* = \sup_{\varepsilon \geq 0}\big(v_d^*(\varepsilon) - c_{2d}\varepsilon\big).$$

$\square$

Theorem 8. *Under Assumptions 1 and 2, any optimal solution v^*, r^* of (12) generates a sequence $p_d^* := v^* - c_{2d}\|r^*\|_2$ of lower bounds which is monotone nondecreasing and converging to p^*:*

$$p_d^* \leq p_{d+1}^* \leq \cdots \leq \lim_{d \to \infty} p_d^* = p^*.$$

Proof. Set $p_d^* := \hat{v}_d^*$ (the optimal value of (12)); at any optimizer $(v^*, r^*, (s_j^*))$ we have $p_d^* = v^* - c_{2d}\|r^*\|_{L^2(\mu)}$.

(i) Certified lower bound. By Assumption 2, for any feasible $(v, r, (s_j))$ and all $x \in K$,

$$p(x) = v + r(x) + \sum_j s_j(x)g_j(x) \ \geq \ v + r(x) \ \geq \ v - \|r\|_{L^\infty(K)} \ \geq \ v - c_{2d}\|r\|_{L^2(\mu)}.$$

Taking $\inf_{x \in K}$ and then sup over feasible tuples gives $p^* \geq \hat{v}_d^* = p_d^*$. Thus each p_d^* is a valid lower bound.

(ii) Monotonicity. If $d' < d$, then $\hat{v}_d^* \geq \hat{v}_{d'}^*$. Indeed, since $\mathbb{R}[x]_{2d'} \subset \mathbb{R}[x]_{2d}$ and $Q_{d'}(g) \subset Q(g)_d$, the feasible set of (12) at level d contains that at level d', so the supremum cannot decrease. Therefore $p_d^* = \hat{v}_d^*$ is nondecreasing in d.

(iii) Convergence. Fix $\eta > 0$. Choose a sequence $\{\varepsilon_d\}_{d \geq 1}$ with $\varepsilon_d \downarrow 0$ and $c_{2d}\varepsilon_d \downarrow 0$ (e.g. $\varepsilon_d := \min\{1/d, \ 1/(d\,c_{2d})\}$). By Theorem 7 under Assumption 1, there exists d_0 such that $v_d^*(\varepsilon_d) \geq p^* - \eta/2$ for all $d \geq d_0$. Using Lemma 8 and $c_{2d}\varepsilon_d \leq \eta/2$ for d large,

$$p_d^* = \hat{v}_d^* = \sup_{\varepsilon \geq 0}\big(v_d^*(\varepsilon) - c_{2d}\varepsilon\big) \ \geq \ v_d^*(\varepsilon_d) - c_{2d}\varepsilon_d \ \geq \ p^* - \eta.$$

Combining with (i) gives $p^* \geq p_d^* \geq p^* - \eta$ for all $d \geq d_0$. Since $\eta > 0$ was arbitrary and $(p_d^*)_d$ is nondecreasing, we conclude $p_d^* \uparrow p^*$. $\square$

5.5 Recovering the standard moment-SOS hierarchy

At level d, the standard moment-SOS hierarchy is

$$\inf \left\{ \ell_y(p) : \ \ell_y(1) = 1, \ M_{d-d_j}(g_j y) \succeq 0 \ \forall j \ \right\} = \sup \left\{ v : \ p - v \in Q(g)_d \ \right\}$$

Setting $\varepsilon_d = 0$ in the constrained form (9) enforces $\|r\|_{L^2(\mu)} \leq 0$, hence $r = 0$ and $p - v = r + \sum_j s_j g_j = \sum_j s_j g_j \in Q(g)_d$. Dually, the objective function in (10) becomes $\ell_y(p) + r\|y\|_2 = \ell_y(p)$.

In the penalized form (12), we introduce a penalty weight $\lambda > 0$:

$$\sup \left\{ v - \lambda\|r\|_{L^2(\mu)} : \ p - v = r + \sum_j s_j g_j \ \right\}$$

and our choice $\lambda = c_{2d}$ is the BM certified instance. As $\lambda \uparrow \infty$, any optimal sequence must satisfy $\|r\| \to 0$ (otherwise the objective would go to $-\infty$), hence $r \to 0$. On the moment side, we have

$$\inf \ \ell_y(p) \quad \text{s.t.} \quad \ell_y(1) = 1, \ M_{d-d_j}(g_j y) \succeq 0, \ \ \|y\|_2 \leq \lambda,$$

and letting $\lambda \uparrow \infty$ removes the norm constraint.

6 Examples

In the following examples we used the MOSEK semidefinite solver and we modeled the moment-SOS hierarchy in Python on a standard PC running under Ubuntu 24. Numerical values are reported to 5 significant digits.

6.1 Motzkin

Let us apply the regularized hierarchy to the Motzkin polynomial $p_M(x) = 1 - 3x_1^2x_2^2 + x_1^4x_2^2 + x_1^2x_2^4$ of Example 1. Since the problem is unconstrained, we can choose the normalized Gaussian (zero mean unit variance). To express the moment matrices and SOS decompositions we use orthonormal Hermite polynomials b_k so that $\|q\|_{L^2(\mu)}$ is the Euclidean norm of the coefficient vector of q in the $\{b_k\}$ basis, recall (6). The Hermite polynomials satisfy $b_0(x) = 1$, $b_1(x) = x$ and the three-term recurrence $xb_k(x) = \sqrt{k+1}b_{k+1}(x) + \sqrt{k}b_{k-1}(x)$, $k \geq 1$.

At relaxation order $d = 3$ (degree $2d = 6$), after solving the moment-SOS problem (9)-(10) we obtain the following numerical bounds:

ε_3	1	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
v_3^*	1.2472	-0.68628	-3.0515	-10.104	-32.325	-102.57	-324.69

We observe that the bounds follow a square-root law $v_3^*(\varepsilon_3) \sim -\kappa_3\varepsilon_3^{-1/2}$ with $\kappa_3 \approx 0.32452$. Because $p_M \notin \Sigma_6$, the best degree-6 SOS approximation sits on a rank-2 face generated by the cubic forms $f_\pm = x_1^2x_2 \pm x_1x_2^2$, with $s_0 = \frac{1}{2}f_+^2 + \frac{1}{2}f_-^2$. The feasible SOS variations issuing from s_0 at first order are the tangent directions generated by the two active squares, i.e., linear combinations of f_+u and f_-v (with u, v cubic). Writing $p_M = s_0 + w$, the residual $w = 1 - 3x_1^2x_2^2$ decomposes into a part that can be created by these first-order SOS variations (tangent image) and a part that cannot (transverse to that image). The parameter κ_3 quantifies, in the Gaussian L^2 metric used by the hierarchy, the size of this transverse component of w relative to the constant shift direction. Geometrically, it measures how stiff the cone Σ_6 is at the face through s_0 against moving along the affine line $p_M - v$, and it is precisely this second-order stiffness that produces the square-root law. In short, κ_3 is the intrinsic curvature-controlled obstruction, at s_0, to absorbing the degree-6 residual of p_M by first-order SOS deformations.

Raising the relaxation order to $d = 4$ enriches the active face of the SOS cone by adding a *quartic* square direction. Concretely, besides the rank-2 face through s_0 the degree-8 cone Σ_8 also contains the square of the quartic $g(x) = 1 - \frac{3}{2}x_1^2x_2^2$, i.e. $g^2 \in \Sigma_8$. This extra square direction lets one cancel, at s_0, not only the tangent (first-order) component of the residual $w = 1 - 3x_1^2x_2^2$, but also its second-order effect along the enlarged face spanned by the two cubic squares and the quartic square. The first non-removable piece appears at third order in a local SOS expansion. Geometrically, this means the affine line $p_M - v$ now meets Σ_8 with third-order contact at the augmented face, and the remaining transverse obstruction is governed by this third-order curvature. As $\varepsilon \downarrow 0$, this yields the cubic-root law $v_4^*(\varepsilon) \sim -\kappa_4\varepsilon^{-1/3}$ with a prefactor $\kappa_4 \approx 0.22902$ determined by that third-order stiffness in the Gaussian L^2 metric. Numerically, the bounds behave in line with this picture:

ε_4	1	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
v_4^*	1.2722	-0.081974	-0.69080	-1.9431	-4.6598	-10.518	-23.142

At $d = 5$, the additional degree-5 square directions do not remove the remaining third-order transverse component at that face: the affine line $p_M - v$ continues to have third-order contact with the (now larger) cone, so the same cubic-root law persists: $v_5(\varepsilon) \sim -\kappa_5\varepsilon^{-1/3}$ but with a

smaller prefactor $\kappa_5 \approx 0.053771$ reflecting the reduced stiffness of Σ_{10} near the face containing s_0:

ε_5	1	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
v_5^*	1.4503	0.18558	-0.073781	-0.45841	-1.2176	-2.6123	-5.1260

At higher relaxation orders d the trend is similar with decreasing prefactors κ_d.

6.2 Origin

Consider Example 2 with $p(x) = x$ and $g(x) = -x^2$, so that $K = \{0\}$ reduces to the origin, with the obvious solution $x^* = p^* = 0$. Let μ be the normalized Gaussian as in Section 6.1.

Fix a relaxation order $d \in \mathbb{N}$ and a regularization parameter $\varepsilon_d > 0$. The (dual) SOS problem is

$$\max_{v,\, r,\, s_0,\, s_1} \left\{ v : \ x - v = r + s_0 - s_1\, x^2, \ \ s_0 \in \Sigma[x]_{2d}, \ s_1 \in \Sigma[x]_{2(d-1)}, \ \ \|r\|_{L^2(\mu)} \leq \varepsilon_d \right\},$$

and the (primal) pseudo-moment problem reads

$$\min_y \ \ell_y(p) + \varepsilon_d \|y\|_2 \quad \text{s.t.} \quad y_0 = 1, \ \ M_d(y) \succeq 0, \ \ M_{d-1}(-x^2\, y) \succeq 0, \tag{15}$$

where $y = (y_0, \ldots, y_{2d})$ acts by $\ell_y(q) = \sum_{k=0}^{2d} q_k\, y_k$ on Hermite coefficients q_k, and $[M_r(y)]_{ij} = \ell_y(b_i b_j)$, $[M_{r-1}(-x^2\, y)]_{ij} = \ell_y(-x^2 b_i b_j)$. Let v_d^* denote the value of the primal-dual problems.

Now observe that if y satisfies the constraints in (15), then $\ell_y(x^2 q^2) = 0$, $\forall q \in \mathbb{R}[x]_{d-1}$. In particular, in the Hermite basis one has $y_k = b_k(0)$, $k = 0, 1, \ldots, 2d$. That is, y coincides with the truncated moment sequence of the Dirac mass δ_0. One has $p(x) = x = b_1(x)$, hence $\ell_y(p) = y_1 = b_1(0) = 0$. For the Hermite polynomials, $b_{2m+1}(0) = 0$ and $b_{2m}(0) = (-1)^m \sqrt{(2m)!}/(2^m m!)$. It follows that the optimal value is

$$v_d^\star \ = \ \varepsilon_d \left\| y \right\|_2 \ = \ \varepsilon_d \sqrt{\sum_{k=0}^{2d} b_k(0)^2} \ = \ \varepsilon_d \sqrt{\sum_{m=0}^{d} \frac{\binom{2m}{m}}{4^m}}.$$

Solving the corresponding regularized moment-SOS hierarchy gives the following values matching with the analytic values:

d	2	3	4	5	6	7	8	9	10
v_d^*/ε_d	1.3693	1.4790	1.5687	1.6453	1.7125	1.7726	1.8271	1.8772	1.9236

As $d \to \infty$, since $\binom{2m}{m}/4^m \sim (\pi m)^{-1/2}$, $v_d^* \sim \frac{\sqrt{2}}{\pi^{1/4}} \varepsilon_d\, d^{1/4}$ and hence if $\varepsilon_d = o(d^{-1/4})$ then $v_d^\star \to p^* = 0$. Note that if the decrease of ε_d is too slow, e.g. $\varepsilon_d = d^{-1/4}$, then v_d^* does not converge to p^*.

In the present case $K = \{0\}$, even if μ is Gaussian (so $w > 0$ everywhere), every relatively open subset $V \subset K$ is of the form $V = \{0\}$ and thus has Lebesgue measure $|V| = 0$ and $\mu(V) = 0$. Consequently, the key estimate used in the proof of Theorem 7, $\|r - q\|_{L^2(\mu)}^2 \geq \int_V |r - q|^2\, w\, dx \geq \gamma^2 \mu(V) > 0$ fails because $\mu(V) = 0$ for every nonempty relatively open $V \subset K$. Thus the upper bound argument (which needs a set of positive μ-measure where $p - v$ is strictly negative) cannot be carried out on $K = \{0\}$.

On the other hand, the sharp BM constant (11) at degree $2d$ is

$$c_{2d} \ = \ \sqrt{p_{2d}^\mu(0)} \ = \ \sqrt{\sum_{k=0}^{2d} b_k(0)^2} \ = \ \sqrt{\sum_{m=0}^{d} \frac{\binom{2m}{m}}{4^m}}.$$
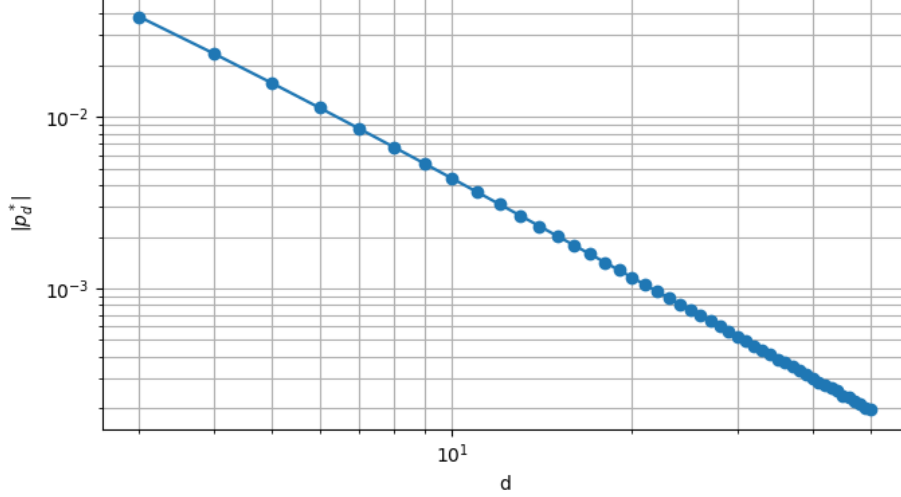
Figure 1: Log-log plot of the gaps $|p_d^* - p^*|$ versus the relaxation order d for the Stengle POP. The near-linearity indicates an algebraic decay $|p_d^* - p^*| \approx C\, d^{-2}$.

Therefore the BM certified bound of Corollary 1 becomes

$$p_d^* \; := \; v_d^* - c_{2d}\, \varepsilon_d \; = \; \varepsilon_d \sqrt{\sum_{m=0}^{d} \frac{\binom{2m}{m}}{4^m}} \; - \; \varepsilon_d \sqrt{\sum_{m=0}^{d} \frac{\binom{2m}{m}}{4^m}} \; = \; 0.$$

Since the true optimum is $p^* = 0$, we have $p_d^* = p^*$ for all $d \in \mathbb{N}$, $\varepsilon_d > 0$. For this degenerate POP, the BM certificate is exact (tight) at every relaxation order and every regularization level.

6.3 Stengle

As in Example 3, let $p(x) = 1 - x^2$, and $g(x) = (1 - x^2)^3$ so that $K = [-1, 1]$. If μ is the arcsine measure, then its orthonormal basis are Chebyshev polynomials $b_0 = T_0$, $b_k = \sqrt{2}\, T_k$ for $k \geq 1$. The corresponding BM constant is

$$c_{2d} \; = \; \sup_{x \in K} \sqrt{\sum_{k=0}^{2d} b_k(x)^2} \; = \; \sqrt{4d + 1}.$$

Solving the penalized mom-SOS hierarchy (12)-(13), we obtain the lower bounds $p_3^* \approx -3.8252 \cdot 10^{-2} \leq p_{50}^* \approx -1.9605 \cdot 10^{-4} \leq p^* = \min_{x \in [-1,1]}(1 - x^2) = 0$. Figure 1 shows neatly the algebraic decay of the gaps $|p_d^* - p^*| \approx C\, d^{-2}$.

6.4 Prestel and Delzell

Consider the POP of Example 4 with $p = \frac{17}{4} - x_1^2 - x_2^2$, $g_1(x) = x_1 - \frac{1}{2}$, $g_2(x) = x_2 - \frac{1}{2}$, $g_3(x) = 1 - x_1 x_2$ and where $Q(g)$ is a non-Archimedean quadratic module. The minimum is $p^* = 0$ attained at $x \in \{(1/2, 2), (2, 1/2)\}$.

We use the affine map $x_i \mapsto \frac{3}{4} x_i + \frac{5}{4}$ so that $[-1, 1]^2$ is sent to $[\frac{1}{2}, 2]^2 \supset K$. On $[-1, 1]$ we use the arcsine reference measure

$$d\mu(y) = \frac{1}{\pi} \frac{dy}{\sqrt{1 - y^2}},$$

18

for which the Chebyshev polynomials of the first kind are orthogonal. Using the orthonormalized 1D basis $b_0 = T_0$, $b_k = \sqrt{2}\, T_k$ for $k \geq 1$, in 2D we use the tensor basis $b_{i,j} = b_i b_j$. Since $\sup_{x \in [-1,1]} |b_0(x)| = 1$ and $\sup_{x \in [-1,1]} |b_k(x)| = \sqrt{2}$ for $k \geq 1$, for an expansion

$$r(x) = \sum_{i+j \leq 2d} r_{ij} b_{i,j}(x)$$

the Cauchy-Scharwz inequality gives

$$\sup_{x \in [-1,1]^2} |r(x)| \leq \sqrt{\sum_{i+j \leq 2d} \alpha_i \alpha_j} \sqrt{\sum_{i+j \leq 2d} r_{ij}^2}$$

where $\alpha_0 = 1$ and $\alpha_k = 2$ for $k \geq 1$. Therefore the BM constant is $c_{2d}^2 = \sum_{i+j \leq 2d} \alpha_i \alpha_j$. Counting pairs: exactly one of i, j positive gives $4d$ pairs with weight 2; both positive gives $(2d+1)(d+1) - (4d+1) = 2d^2 - d$ pairs with weight 4; and $(0,0)$ contributes 1. Hence $c_{2d}^2 = 1 + 8d + (8d^2 - 4d) = 8d^2 + 4d + 1$, i.e.

$$c_{2d} = \sqrt{8d^2 + 4d + 1}.$$

The table below reports the bounds and internal quantities for $d = 1, 2, 3$ obtained by solving the penalized moment-SOS hierarchy (12)-(13):

d	$p_d^* = v^* - c_{2d} t^*$	v^*	t^*
1	-0.40399	1.4382	0.51093
2	-2.1928×10^{-3}	5.9888×10^{-3}	1.2777×10^{-3}
3	-1.7033×10^{-9}	5.1568×10^{-8}	5.7781×10^{-9}

The certified lower bounds are negative and monotonically increasing to $p^* = 0$, demonstrating rapid convergence of the regularized hierarchy: from $-4.0 \cdot 10^{-1}$ at $d = 1$ to about $-2.2 \cdot 10^{-3}$ at $d = 2$, and essentially machine precision at $d = 3$. Concurrently, the optimizer chooses a small v^* and drives the residual norm t^* to zero, so that $v^* - c_{2d} t^* \uparrow 0$. The Chebyshev-arcsine tensor basis together with exact product rules is numerically stable, and the explicit BM constant $c_{2d} = \sqrt{8d^2 + 4d + 1}$ provides tight control on the sup-norm via the ℓ_2-penalty on coefficients.

Acknowledgment

References

[1] L. Baldi, L. Slot. Degree bounds for Putinar's positivstellensatz on the hypercube. SIAM J. Appl. Alg. Geom. 8(1):1-25, 2024.

[2] H. H. Bauschke, P. L. Combettes. Convex analysis and monotone operator theory in Hilbert spaces. Springer, 2011.

[3] C. Dunkl, Y. Xu. Orthogonal polynomials of several variables, 2nd edition. Cambridge Univ. Press, 2014.

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

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

[6] J. B. Lasserre. A sum of squares approximation of nonnegative polynomials. SIAM J. Optim 16:751–765, 2006.

[7] J. B. Lasserre. Moments, positive polynomials and their applications. Imperial College Press, 2010.

[8] J. B. Lasserre. The K-moment problem for continuous linear functionals. Trans. Amer. Math. Soc. 365(5):2489-2504, 2013.

[9] J. B. Lasserre, T. Netzer. SOS approximations of nonnegative polynomials via simple high degree perturbations. Math. Zeitschrift 256:99-112, 2007.

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

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

[12] T. Netzer. Positive polynomials, sums of squares and the moment problem. PhD Thesis, Univ. Konstanz, 2008.

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

[14] T. Theobald. Real algebraic geometry and optimization. AMS, 2024.

[15] A. Prestel, C. N. Delzell. Positive polynomials. Springer, 2001.

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

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

[18] G. Stengle. Complexity estimates for the Schmüdgen Positivstellensatz, J. Complex. 12:167-174, 1996.