A hierarchy of eigencomputations for polynomial optimization on the sphere

Nathaniel Johnston [*,†], Benjamin Lovitz[*,‡], and Aravindan Vijayaraghavan[*,§]

[†]Department of Mathematics and Computer Science, Mount Allison University, Sackville, New Brunswick, Canada
[‡]Department of Mathematics, Northeastern University, Boston, Massachusetts, USA
[§]Department of Computer Science, Northwestern University, Evanston, Illinois, USA

October 30, 2023

Abstract

We introduce a convergent hierarchy of lower bounds on the minimum value of a real homogeneous polynomial over the sphere. The main practical advantage of our hierarchy over the sum-of-squares (SOS) hierarchy is that the lower bound at each level of our hierarchy is obtained by a minimum eigenvalue computation, as opposed to the full semidefinite program (SDP) required at each level of SOS. In practice, this allows us to go to much higher levels than are computationally feasible for the SOS hierarchy. For both hierarchies, the underlying space at the k-th level is the set of homogeneous polynomials of degree $2k$. We prove that our hierarchy converges as $O(1/k)$ in the level k, matching the best-known convergence of the SOS hierarchy when the number of variables n is less than the half-degree d (the best-known convergence of SOS when $n \geq d$ is $O(1/k^2)$). More generally, we introduce a convergent hierarchy of minimum eigenvalue computations for minimizing the inner product between a real tensor and an element of the spherical Segre-Veronese variety, with similar convergence guarantees. As examples, we obtain hierarchies for computing the (real) tensor spectral norm, and for minimizing biquadratic forms over the sphere. Hierarchies of eigencomputations for more general constrained polynomial optimization problems are discussed.

[*]emails: njohnston@mta.ca, b.lovitz@northeastern.edu, aravindv@northwestern.edu

Contents

1 Introduction

We consider the fundamental task of minimizing a homogeneous degree-D polynomial $p(x) \in \mathbb{R}[x]_D$ in n variables $x = (x_1, \ldots, x_n)$ over the unit sphere

$$p_{\min} = \min_{x \in \mathcal{S}(\mathbb{R}^n)} p(x), \tag{1}$$

where $\mathcal{S}(\mathbb{R}^n) = \{x \in \mathbb{R}^n : \|x\| = 1\}$ and $\|\cdot\|$ denotes the Euclidean norm. Optimization problems of this form have applications in several areas [FF21]. For example, for a special class of degree-three polynomials this corresponds to computing the largest stable set of a graph [N+03, DK08]. As another example, computing the $2 \to 4$ norm (i.e., hypercontractivity) of a matrix A is equivalent to maximizing the degree-four polynomial $p(x) = \|Ax\|_4^4$ on the unit sphere and has many connections to problems in computational complexity, quantum information and designing matrices for compressive sensing [BBH+12]. For $D = 2$ this problem is equivalent to computing the minimum eigenvalue of a symmetric matrix, which can be solved efficiently. However, already for $D = 3$ this problem is NP-hard as it contains the stable set problem as a special case [N+03].

The sum-of-squares (SOS) hierarchy is a hierarchy of semidefinite programs (SDPs) of increasing size whose optimum values approach $p_{\min}$ from below. We propose a hierarchy of minimum eigenvalue computations that also approach $p_{\min}$ from below. At the k-th level of both our hierarchy and the SOS hierarchy, the underlying space is the set of homogeneous polynomials of degree $2k$. However, our hierarchy has the advantage of merely requiring an eigencomputation at each level, as opposed to the full SDP required at each level of SOS.[1] In addition, we prove that the convergence of our hierarchy in the level k is similar to SOS: The difference between the lower bound computed by our hierarchy and the true minimum $p_{\min}$ goes as $O(1/k)$ in the level k. This is quadratically slower than the best-known convergence $O(1/k^2)$ of SOS when $D \leq 2n$, and matches the best-known convergence of SOS when $D > 2n$ [Rez95, DW12, FF21].[2] In particular, our convergence result gives an alternate proof of the $O(1/k)$ convergence results of [Rez95, DW12] for the SOS hierarchy. The main ingredient in our proof is the quantum de Finetti theorem, which is a convergence result for a certain complex optimization problem known as the *quantum separability problem* (see [DPS04, Wat18]). We bypass the need to develop a specialized *real* quantum de Finetti theorem (as was done in [DW12]) by proving that the real optimization problem (1) is related to a certain complex optimization problem up to a constant (Proposition 6). These results are presented in Section 3.

In Section 4 we describe an implementation of our hierarchy in MATLAB, and present several examples to demonstrate its performance in comparison to the SOS and Diagonally Dominant Sum-of-Squares (DSOS) hierarchies [AM19]. The latter hierarchy optimizes over polynomials that admit a diagonally dominant Gram matrix, which is stronger than being a sum-of-squares, in hopes of computational savings.[3] We find that we can compute much higher levels of our hierarchy than either of these, and can often outperform them in terms of time budget. For example, for the Motzkin polynomial we can compute our hierarchy up to level 2000, as opposed to level ~ 10 for SOS or DSOS (Table 1). As another example, for random homogeneous polynomials our

[1] We note that other works have also designed alternatives to the SOS hierarchy that are based on spectral or eigenvector computation [HSSS16, SS17]. However these algorithms are specialized for average-case analysis (i.e., random instances) of tensor decompositions and finding planted sparse vectors.

[2] See also [BGG+17] which gives multiplicative approximation guarantees for polynomial optimization over the sphere using the sum of squares hierarchy.

[3] In contrast to our hierarchy and the SOS hierarchy, the DSOS hierarchy does not always converge [AM19, Proposition 3.15].

hierarchy outperforms DSOS in numerical tests even at a fixed level (Table 2). Here, both hierarchies are performing computations in the space of homogeneous polynomials of degree $2k$, but our hierarchy is using less computational resources and obtaining better bounds. Our hierarchy can also compute non-trivial bounds for random dense polynomials in as many as 90 variables, which surpasses the 25 variables possible with SOS and 70 variables possible with DSOS (Table 3).

Let $S^D(\mathbb{R}^n) \subseteq (\mathbb{R}^n)^{\otimes D}$ be the *symmetric subspace* (see Section 2). Under the standard isomorphism between $\mathbb{R}[x]_D$ and $S^D(\mathbb{R}^n)$, one can view the optimization problem (1) as minimizing the inner product between a symmetric tensor and a unit symmetric product tensor (i.e. an element of the *spherical Veronese variety*). More generally, we develop a hierarchy of eigencomputations for minimizing the inner product between a real tensor and a unit partially-symmetric product tensor (i.e. an element of the *spherical Segre-Veronese variety*). We prove that this more general hierarchy also converges as $O(1/k)$. In particular, this gives hierarchies for computing the real spectral norm of a real tensor, and for minimizing a biquadratic form over the unit sphere. Computing the tensor spectral norm is a well-studied problem with connections to planted clique, tensor PCA and tensor decompositions [FK08, BV09, RM14]. These results are presented in Section 5.

In Section 6 we consider minimizing a real homogeneous polynomial under more general constraints. We develop a similar hierarchy of eigencomputations to lower bound the constrained optimum, which converge to a certain analogous *complex* constrained optimization problem. As mentioned above, in the case of real polynomial optimization over the sphere (and more generally tensor optimization over the spherical Segre-Veronese), we obtained our convergent hierarchy by showing that this complex optimum is related to the real optimum up to a constant (Proposition 6). It remains an interesting problem for future work to determine how the complex optimum is related to the real optimum under more general constraints.

In the remainder of this introduction, we describe the sum-of-squares hierarchy for the polynomial optimization problem (1), introduce our hierarchy for this problem, and describe the generalization of our hierarchy to tensor optimization problems. For the sake of readability, we will defer some definitions until Section 2.

1.1 The sum-of-squares hierarchy

Let us begin by considering the optimization problem (1). By [DW12, Lemma B.2], we can (and will) assume D is even.[4] We let $d = D/2$. The sum-of-squares (SOS) hierarchy is a hierarchy of lower bounds on $p_{\min}$ which can be computed by semidefinite programming (see e.g. [Par03, Section 3] and the references therein for background on semidefinite programming). Let $\mathbb{R}[x]$ be the real polynomial ring in n variables $x = (x_1, \dots, x_n)$, let $\mathbb{R}[x]_k$ be the set of homogeneous polynomials of degree k, let $\Sigma_{n,k} \subseteq \mathbb{R}[x]_{2k}$ (or simply Σ_k when n is understood) be the set of n-variate polynomials which are sums of squares of homogeneous polynomials of degree k, let

$$s(x) = \sum_{i=1}^{n} x_i^2 \in \Sigma_1$$

be the Euclidean norm squared, and let $s_{n,d}(x) = s(x)^d$ (see Section 2 for more details). For an integer $k \geq d$, the k-th level of the SOS hierarchy computes the following SDP:

[4] If D is odd, then

$$\min_{x \in \mathcal{S}(\mathbb{R}^n)} p(x) = \frac{(D+1)^{(D+1)/2}}{D^{D/2}} \min_{(x, x_{n+1}) \in \mathcal{S}(\mathbb{R}^{n+1})} (x_{n+1} \cdot p)(x, x_{n+1}),$$

where $(x_{n+1} \cdot p)(x, x_{n+1})$ is an (even) degree $D+1$ homogeneous polynomial.

$$\begin{aligned} \text{maximize:} \quad & \gamma \\ \text{subject to:} \quad & p(x) \cdot s(x)^{k-d} - \gamma \cdot s(x)^k \in \Sigma_k \end{aligned} \tag{2}$$

This optimization can indeed be computed by semidefinite programming by maximizing over γ for which $p(x) \cdot s(x)^{k-d} - \gamma \cdot s(x)^k$ admits a positive semidefinite *Gram matrix* (see Section 2). To see equivalences between the formulation (2) and other, perhaps more familiar forms of the SOS hierarchy, see e.g. [dKLP05, Proposition 2] or [Lau19, Lemma 1.3]. Let γ_k be the optimum value of this SDP. Then $\gamma_1 \leq \gamma_2 \leq \dots$ and $\lim_{k\to\infty} \gamma_k = p_{\min}$ [Rez95] (see also [DW12]). Moreover, it is known that $p_{\min} - \gamma_k = O(1/k)$ [Rez95, DW12], which can be improved to $O(1/k^2)$ when $d \leq n$ [FF21].[5]

The dual SDP to (2) is given as follows:

$$\begin{aligned} \text{minimize:} \quad & \tilde{\mathbb{E}}(p(x) \cdot s(x)^{k-d}) \\ \text{subject to:} \quad & \tilde{\mathbb{E}} : \mathbb{R}[x]_{2k} \to \mathbb{R} \quad \text{linear} \\ & \tilde{\mathbb{E}}\left(s(x)^k\right) = 1 \\ & \tilde{\mathbb{E}}(q^2) \geq 0 \quad \text{for all} \quad q \in \mathbb{R}[x]_k. \end{aligned} \tag{3}$$

The linear forms $\tilde{\mathbb{E}}$ are called *pseudoexpectations*. As both the primal and dual problems are strictly feasible, there is no duality gap and the optimum values of the SDPs (2) and (3) are equal.

1.2 Our polynomial optimization hierarchy

Like the SOS hierarchy, ours is a hierarchy of computations of increasing size which converge to $p_{\min}$. The k-th level of both hierarchies perform computations in the space $\mathbb{R}[x]_{2k}$. Our hierarchy has the computational advantage of merely performing a minimum eigenvalue computation at each level, as opposed to the full SDP required at each level of SOS. As a result, we can compute hundreds (and in some cases, thousands) of levels of our hierarchy in practice.

Let $S^d(\mathbb{R}^n) \subseteq (\mathbb{R}^n)^{\otimes d}$ be the *symmetric subspace*, and let $M(p) \in \mathrm{Hom}_{\mathbb{R}}(S^d(\mathbb{R}^n))$ be the polynomial p after the sequence of maps

$$\mathbb{R}[x]_{2d} \cong S^{2d}(\mathbb{R}^n) \subseteq S^d(\mathbb{R}^n) \otimes S^d(\mathbb{R}^n) \cong \mathrm{Hom}_{\mathbb{R}}(S^d(\mathbb{R}^n))$$

(see Section 2 for details). We often view $M(p)$ as a linear map on $(\mathbb{R}^n)^{\otimes d}$ by setting it equal to zero on the orthogonal complement to $S^d(\mathbb{R}^n)$. It is not difficult to check that $M(p)$ is a Gram matrix for p, and we call it the *canonical* Gram matrix for p. In Section 2 we prove that $M(s_{n,d}) \in \mathrm{Hom}_{\mathbb{R}}(S^d(\mathbb{R}^n))$ is positive definite.

At the k-th level, our hierarchy computes $\nu_k := \lambda_{\min}(M_k(p) M_k(s_{n,d})^{-1})$, where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue and

$$M_k(p) := \Pi_{n,k}(M(p) \otimes \mathbb{1}_n^{\otimes k-d})\Pi_{n,k}.$$

Here, $\Pi_{n,k} : (\mathbb{R}^n)^{\otimes k} \to (\mathbb{R}^n)^{\otimes k}$ is the orthogonal projection onto the symmetric subspace $S^k(\mathbb{R}^n)$. An explicit (and efficient) formula for the coordinates of $M_k(p)$ is derived in Appendix A. We prove that $\nu_1 \leq \nu_2 \leq \dots$ and $\lim_{k\to\infty} \nu_k = p_{\min}$. Moreover, we prove that $p_{\min} - \nu_k = O(1/k)$.

[5] In particular, our analysis provides an alternate proof of the $O(1/k)$ convergence result of [Rez95, DW12] for the SOS hierarchy.

Note that ν_k is equal to the optimum value of the SDP

$$\begin{aligned} \text{minimize:} \quad & \tilde{\Omega}(M_k(p)M_k(s_{n,d})^{-1}) \\ \text{subject to:} \quad & \tilde{\Omega} : \mathrm{Hom}(S^k(\mathbb{R}^n)) \to \mathbb{R} \quad \text{linear} \\ & \tilde{\Omega}(\mathbb{1}_n^{\otimes k}) = 1 \\ & \tilde{\Omega}(qq^\intercal) \geq 0 \quad \text{for all} \quad q \in S^k(\mathbb{R}^n). \end{aligned} \tag{4}$$

Indeed, the first, second and third conditions on $\tilde{\Omega}$ are equivalent to i) $\Omega(M) = \mathrm{Tr}(\rho M)$ where $\rho : S^k(\mathbb{R}^n) \to S^k(\mathbb{R}^n)$ is linear, ii) $\mathrm{Tr}(\rho) = 1$ and iii) ρ is positive semidefinite. This expression for ν_k reveals similarities between our hierarchy and the SOS hierarchy: These conditions are analogous to the first, second, and third conditions on the pseudoexpectation $\tilde{\mathbb{E}}$ in the SDP (3). Note that $\nu_k = \max\{\nu : M_k(p - \nu \cdot s(x)^d) \succeq 0\}$ where $\succeq$ is the Loewner order; i.e. ν_k is the largest ν for which the *particular* Gram matrix $M_k(p - \nu \cdot s(x)^d)$ for the polynomial $p(x) \cdot s(x)^{k-d} - \nu \cdot s(x)^k$ is positive semidefinite.[6] This reveals that $\nu_k \leq \gamma_k$, i.e. the k-th level of our hierarchy is weaker than the k-th level of the SOS hierarchy.

The dual SDP to (4) also has a satisfying comparison to the SOS SDP (2):

$$\begin{aligned} \text{maximize:} \quad & \nu \\ \text{subject to:} \quad & M_k(p - \nu \cdot s_{n,d}) \succeq 0. \end{aligned}$$

For example, suppose we wish to determine whether $p_{\min} \geq 0$.[7] We would use the SOS hierarchy to do this by checking whether $\gamma_k \geq 0$ at each level, or equivalently whether there exists a positive semidefinite Gram matrix for $p(x) \cdot s(x)^{k-d}$. In contrast, we would use our hierarchy to do this by checking whether $\nu_k \geq 0$, or equivalently whether the *single* Gram matrix $M_k(p)$ for $p(x) \cdot s(x)^{k-d}$ is positive semidefinite. Surprisingly, our hierarchy still converges at a rate of $O(1/k)$. This is quadratically slower than the best-known convergence $O(1/k^2)$ of SOS when $d \leq n$, and matches the best-known convergence of SOS when $d > n$ [Rez95, DW12, FF21].

1.3 Our tensor optimization hierarchy

More generally, our techniques can be used to minimize the inner product between a real tensor and an element of the spherical Segre-Veronese variety. Let $p(x) \in \mathbb{R}[x]_D$ be a homogeneous degree-D polynomial in n variables. Note that

$$\min_{x \in \mathcal{S}(\mathbb{R}^n)} p(x) = \min_{v \in \mathcal{S}(\mathbb{R}^n)} \langle \vec{p}, v^{\otimes D} \rangle,$$

where $\vec{p} \in S^D(\mathbb{R}^n)$ is the polynomial p after the isomorphism $\mathbb{R}[x]_D \cong S^D(\mathbb{R}^n)$ (see Section 2). So minimizing $p(x)$ over the unit sphere is equivalent to minimizing the inner product of the

[6] To see this, note that

$$\begin{aligned} \nu_k &= \lambda_{\min}(M_k(p)M_k(s_{n,d})^{-1}) \\ &= \lambda_{\min}(M_k(s_{n,d})^{-1/2}M_k(p)M_k(s_{n,d})^{-1/2}) \\ &= \max\{\nu : M_k(s_{n,d})^{-1/2}M_k(p)M_k(s_{n,d})^{-1/2} - \nu\, \mathbb{1}_{S^k(\mathbb{R}^n)} \succeq 0\} \\ &= \max\{\nu : M_k(p - \nu \cdot s(x)^d) \succeq 0\}. \end{aligned}$$

[7] Both our hierarchy and the SOS hierarchy can more generally be used to determine whether an *arbitrary* polynomial (not necessarily homogeneous) is globally non-negative: Given a polynomial $q(x)$ of degree D, let $p(x, y) = y^D \cdot q(x_1/y, \ldots, x_n/y)$ be its homogenization. Then q is globally non-negative if and only if $p_{\min} \geq 0$ [Mar08].

symmetric tensor $\vec{p}$ with a unit symmetric product tensor $v^{\otimes D}$ (i.e. an element of the *(real) spherical Veronese variety*). More generally, one can choose positive integers $n_1, \ldots, n_m$ and $D_1, \ldots, D_m$, and a tensor $p \in (\mathbb{R}^{n_1})^{\otimes D_1} \otimes \cdots \otimes (\mathbb{R}^{n_m})^{\otimes D_m}$, and consider the minimization problem

$$\min_{v_j \in \mathcal{S}(\mathbb{R}^{n_j})} \langle p, v_1^{\otimes D_1} \otimes \cdots \otimes v_m^{\otimes D_m} \rangle. \tag{5}$$

Tensors of the form $v_1^{\otimes D_1} \otimes \cdots \otimes v_m^{\otimes D_m}$ are by definition elements of the *(real) Segre-Veronese variety*. Our hierarchy extends naturally to this setting, with similar convergence guarantees. As was the case for real polynomial optimization, we can assume that each $D_j = 2d_j$ is even (see Proposition 11), and that $p \in S^{2d_1}(\mathbb{R}^{n_1}) \otimes \cdots \otimes S^{2d_m}(\mathbb{R}^{n_m})$. For $k \geq \max_j d_j$, the k-th level of our hierarchy computes the minimum eigenvalue of the matrix $M_k(p) M_k(s_{n_1,d_1} \otimes \cdots \otimes s_{n_m,d_m})^{-1}$, where

$$M_k(p) = (\Pi_{n_1,k} \otimes \cdots \otimes \Pi_{n_m,k})(M(p) \otimes \mathbb{1}_{n_1}^{\otimes k-d_1} \otimes \cdots \otimes \mathbb{1}_{n_m}^{\otimes k-d_m})(\Pi_{n_1,k} \otimes \cdots \otimes \Pi_{n_m,k}).$$

Here, $\Pi_{n_j,k}$ is the orthogonal projection onto $S^k(\mathbb{R}^{n_j})$, and $M(p)$ is p viewed as an element of $\mathrm{Hom}_{\mathbb{R}}(S^{d_1}(\mathbb{R}^{n_1}) \otimes \cdots \otimes S^{d_m}(\mathbb{R}^{n_m}))$. Just like our hierarchy for polynomial minimization, we prove that this gives rise to a sequence of lower bounds on the true minimum (5) converging as $O(1/k)$ in additive error. As applications, we use this hierarchy to compute the tensor spectral norm and to minimize biquadratic forms over the unit sphere.

2 Background

Let $\mathbb{F}$ be either $\mathbb{R}$ or $\mathbb{C}$, and let n, d be positive integers. Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{F}^n$. Let $\mathbb{F}[x_1, \ldots, x_n]_d$ be the vector space of homogeneous degree-d polynomials in n variables over $\mathbb{F}$. The permutation group S_d on d letters acts on $(\mathbb{F}^n)^{\otimes d}$ by permuting factors, i.e. for $\sigma \in S_d$ we have

$$\sigma \cdot (v_1 \otimes \cdots \otimes v_d) = v_{\sigma^{-1}(1)} \otimes \cdots \otimes v_{\sigma^{-1}(d)}.$$

Define the *symmetric subspace* $S^d(\mathbb{F}^n) \subseteq (\mathbb{F}^n)^{\otimes d}$ to be the linear subspace of tensors invariant under S_d. The symmetric subspace $S^d(\mathbb{F}^n)$ is isomorphic as an $\mathbb{F}$-vector space to $\mathbb{F}[x_1, \ldots, x_n]_d$ via the map which sends $x_{i_1} \cdots x_{i_d}$ to $\frac{1}{d!} f_{(i_1,\ldots,i_d)}$, where

$$f_{(i_1,\ldots,i_d)} := \sum_{\sigma \in S_d} e_{i_{\sigma(1)}} \otimes \cdots \otimes e_{i_{\sigma(d)}}.$$

For a polynomial $p \in \mathbb{F}[x_1, \ldots, x_n]_d$ we let $\vec{p} \in S^d(\mathbb{F}^n)$ be the polynomial p after this isomorphism.

For an $\mathbb{F}$-vector space $\mathcal{U}$, let $\mathrm{Hom}_{\mathbb{F}}(\mathcal{U})$ denote the space of $\mathbb{F}$-linear maps from $\mathcal{U}$ to $\mathcal{U}$. We will also write $\mathrm{Hom}(\mathcal{U})$ when the field is clear from the context, and identify $\mathrm{Hom}(\mathbb{F}^n)$ with $n \times n$ matrices. For positive integers $d \leq k$, let $[d] = \{1, 2, \ldots, d\}$ and $[d+1 \cdot\cdot k] = \{d+1, d+2, \ldots, k\}$.

Let $\langle \cdot, \cdot \rangle$ be the standard bilinear form on $\mathbb{F}^n$, extended to a bilinear form on $(\mathbb{F}^n)^{\otimes d}$ (and hence $S^d(\mathbb{F}^n)$) by setting $\langle u_1 \otimes \cdots \otimes u_d, v_1 \otimes \cdots \otimes v_d \rangle = \langle u_1, v_1 \rangle \cdots \langle u_d, v_d \rangle$ and extending linearly. For an $\mathbb{R}$-vector space $\mathcal{U}$ with a bilinear form $\langle \cdot, \cdot \rangle$ we say that a map $M \in \mathrm{Hom}_{\mathbb{R}}(\mathcal{U})$ is *positive semidefinite* ($M \succeq 0$) if $\langle u, Mu \rangle \geq 0$ for all $u \in \mathcal{S}(\mathcal{U})$, and *positive definite* ($M \succ 0$) if $\langle u, Mu \rangle > 0$ for all $u \in \mathcal{S}(\mathcal{U})$. Let $\langle \cdot | \cdot \rangle$ be the standard sesquilinear form on $\mathbb{C}^n$, extended to a sesquilinear form on $(\mathbb{C}^n)^{\otimes d}$. For a $\mathbb{C}$-vector space $\mathcal{U}$ with a sesquilinear form $\langle \cdot | \cdot \rangle$ we say that a map $M \in \mathrm{Hom}_{\mathbb{C}}(\mathcal{U})$ is *positive semidefinite* ($M \succeq 0$) if $\langle u | Mu \rangle \geq 0$ for all $u \in \mathcal{S}(\mathcal{U})$, and *positive definite* ($M \succ 0$) if $\langle u | Mu \rangle > 0$ for all $u \in \mathcal{S}(\mathcal{U})$. For a vector $v \in \mathbb{C}^n$, let v^* be the conjugate-transpose of v. For a matrix $M \in \mathrm{Hom}(\mathbb{F}^n)$ and a real number $p \in [1,\infty) \cup \{\infty\}$ we let $\|M\|_p$ be the *Schatten p-norm* of M (see e.g. [Wat18]). We define the *condition number* of a positive definite matrix M to be $\kappa(M) := \lambda_{\max}(M)/\lambda_{\min}(M)$.

An orthonormal basis for $S^d(\mathbb{F}^n)$ is given by

$$\left\{ e_{(i_1,\ldots,i_d)} := f(i_1,\ldots,i_d) f_{(i_1,\ldots,i_d)} : 1 \le i_1 \le \cdots \le i_d \le n \right\},$$

where

$$f(i_1,\ldots,i_d) := (d! \cdot r_1^{(i_1,\ldots,i_d)}! \cdots r_n^{(i_1,\ldots,i_d)}!)^{-1/2}$$

and $r_j^{(i_1,\ldots,i_d)}$ is the number of times j appears in $(i_1,\ldots,i_d)$ (see Appendix A).

The *transpose map* $T \in \mathrm{Hom}(\mathrm{Hom}(\mathbb{F}^n))$ is the map which sends a matrix M to its transpose $T(M) = M^\intercal$. The *partial transpose map* on the first factor of $(\mathbb{F}^n)^{\otimes d}$ is defined as $T \otimes \mathbb{1}_{\mathrm{Hom}((\mathbb{F}^n)^{\otimes d-1})}$, where $\mathbb{1}_{\mathrm{Hom}((\mathbb{F}^n)^{\otimes d-1})} \in \mathrm{Hom}(\mathrm{Hom}((\mathbb{F}^n)^{\otimes d-1}))$ is the identity map. For a matrix $M \in \mathrm{Hom}((\mathbb{F}^n)^{\otimes d})$, we let

$$M^{\Gamma_1} := (T \otimes \mathbb{1}_{\mathrm{Hom}((\mathbb{F}^n)^{\otimes d-1})})(M)$$

be the partial transpose of M in the first factor. We define the partial transpose map in the j-th factor for $j \in [d]$ similarly, and let M^{Γ_j} be the partial transpose of M in the j-th factor. We extend these definitions to linear maps on $\mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_d}$ for potentially distinct $n_1,\ldots,n_d$ in the obvious way. See also [Wat18, Section 6.2.3].

The *trace map* $\mathrm{Tr} \in \mathrm{Hom}(\mathrm{Hom}(\mathbb{F}^n))$ is the map which sends a matrix M to its trace. The *partial trace map* on the first factor of $(\mathbb{F}^n)^{\otimes d}$ is defined as $\mathrm{Tr} \otimes \mathbb{1}_{\mathrm{Hom}((\mathbb{F}^n)^{\otimes d-1})}$. For a matrix $M \in \mathrm{Hom}((\mathbb{F}^n)^{\otimes d})$, we let

$$\mathrm{Tr}_1(M) := (\mathrm{Tr} \otimes \mathbb{1}_{\mathrm{Hom}((\mathbb{F}^n)^{\otimes d-1})})(M) \in \mathrm{Hom}((\mathbb{F}^n)^{\otimes d-1})$$

be the partial trace of M in the first factor. We define the partial trace map on the j-th factor for $j \in [d]$ similarly, and let $\mathrm{Tr}_j(M)$ be the partial trace of M in the j-th factor. For a subset $S \subseteq [d]$, we let $\mathrm{Tr}_S(M) \in \mathrm{Hom}((\mathbb{F}^n)^{\otimes d-|S|})$ be the composition of the partial traces of M with respect to the factors indexed by S. See also [DW12] or [Wat18]. We extend these definitions to linear maps on $\mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_d}$ for potentially distinct $n_1,\ldots,n_d$ in the obvious way.

For a homogeneous polynomial $p \in \mathbb{R}[x_1,\ldots,x_n]_{2d}$, a *Gram matrix* for p is a matrix $M \in \mathrm{Hom}_{\mathbb{R}}(S^d(\mathbb{R}^n))$ for which $p(x) = \langle x^{\otimes d}, Mx^{\otimes d}\rangle$ for all $x \in \mathbb{R}^n$. We say that p is a *sum of squares* if there exist homogeneous polynomials $q_1,\ldots,q_\ell \in \mathbb{R}[x_1,\ldots,x_n]_d$ for which $p = q_1^2 + \cdots + q_\ell^2$. We denote the set of homogeneous polynomials of degree $2d$ which are sums of squares by $\Sigma_{n,d}$, or simply Σ_d when the number of variables n is clear from the context. It is well-known that $p \in \Sigma_d$ if and only if there exists a positive semi-definite Gram matrix for p (see e.g. [Net16, Lemma 12]).

As mentioned in the introduction, we let $M(p)$ be the polynomial p after the sequence of maps

$$\mathbb{R}[x]_{2d} \cong S^{2d}(\mathbb{R}^n) \subseteq S^d(\mathbb{R}^n) \otimes S^d(\mathbb{R}^n) \cong \mathrm{Hom}_{\mathbb{R}}(S^d(\mathbb{R}^n)).$$

The first isomorphism is described above. The inclusion is clear: A $(2d)$-factor tensor that is invariant under permutations of all $2d$ factors is in particular invariant under permutations of the first d factors and permutations of the second d factors (one could alternatively choose any bipartition of the $2d$ factors into equally numbered parts). The last isomorphism is standard and invokes the bilinear pairing to identity $S^d(\mathbb{R}^n)$ with $S^d(\mathbb{R}^n)^*$ (here, $(\cdot)^*$ denotes the dual vector space). It is easy to check that $M(p)$ forms a Gram matrix for p. We will call $M(p)$ the *canonical* Gram matrix for p. Note that $M(p)^{\intercal} = M(p)$ and $M(p)^{\Gamma_j} = M(p)$ for all $j \in [d]$.

A recipe for obtaining $M(p)$ is also given in [Lau19, Section 1.1.3], and we write down an explicit formula for the coordinates of $M(p)$ in Appendix A. We often view $M(p)$ as a linear map on $(\mathbb{R}^n)^{\otimes d}$ by setting it equal to zero on the orthogonal complement to $S^d(\mathbb{R}^n)$. For an integer $k \geq d$ we define

$$M_k(p) := \Pi_{n,k}(M(p) \otimes \mathbb{1}_n^{\otimes k-d})\Pi_{n,k} \in \mathrm{Hom}(S^k(\mathbb{R}^n)),$$

where $\Pi_{n,k}$ is the orthogonal projection onto $S^k(\mathbb{R}^n)$. Note that $M_k(p)$ is a Gram matrix for the polynomial $p(x) \cdot s(x)^{k-d}$, where $s(x) := x_1^2 + \cdots + x_n^2$. Recall the definition $s_{n,d}(x) = s(x)^d$.

We close this section by providing a few examples of $M(p)$, and proving that $M(s_{n,d})$ is positive definite.

Example 1. Let $p(x,y) = x^2 + y^2 + xy$. Then

$$M(p) = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix} \in \mathrm{Hom}(\mathbb{R}^2).$$

Example 2. Let $p(x,y) := s(x,y)^2 = (x^2+y^2)^2$. Viewing $M(p)$ as a linear map on $(\mathbb{R}^2)^{\otimes 2}$, we obtain

$$M(p) = \begin{bmatrix} 1 & 0 & 0 & 1/3 \\ 0 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1 \end{bmatrix} \in \mathrm{Hom}((\mathbb{R}^2)^{\otimes 2}).$$

This can be verified by noting that $p(x,y) = \langle (x,y)^{\otimes 2}, M(p)(x,y)^{\otimes 2}\rangle$, and if we reshape $M(p)$ into an element of $(\mathbb{R}^2)^{\otimes 4}$ (a 16-dimensional vector), then it is invariant under S_4. With respect to the orthonormal basis $e_{(1,1)}, e_{(1,2)}, e_{(2,2)}$ of $S^2(\mathbb{R}^2)$, $M(p)$ is given by

$$M(p) = \begin{bmatrix} 1 & 0 & 1/3 \\ 0 & 2/3 & 0 \\ 1/3 & 0 & 1 \end{bmatrix} \in \mathrm{Hom}(S^2(\mathbb{R}^2)).$$

Example 3. Let $p(x,y) := s(x,y)^3 = (x^2+y^2)^3$. With respect to the orthonormal basis $e_{(1,1,1)}, e_{(1,1,2)}, e_{(1,2,2)}, e_{(2,2,2)}$ of $S^3(\mathbb{R}^2)$, $M(p)$ is given by

$$M(p) = \begin{bmatrix} 1 & 0 & \sqrt{3}/5 & 0 \\ 0 & 3/5 & 0 & \sqrt{3}/5 \\ \sqrt{3}/5 & 0 & 3/5 & 0 \\ 0 & \sqrt{3}/5 & 0 & 1 \end{bmatrix} \in \mathrm{Hom}(S^3(\mathbb{R}^2)).$$

This can be verified directly or using the formula derived in Appendix A.

Proposition 4. *Let n, d be positive integers, and let $s_{n,d} = (x_1^2 + \cdots + x_n^2)^d \in \mathbb{R}[x_1, \ldots, x_n]_{2d}$. It holds that $M(s_{n,d}) \succ 0$.*

Proof. We need to prove that $\langle v, M(s_{n,d})v\rangle > 0$ for all $v \in \mathcal{S}(S^d(\mathbb{R}^n))$. This follows from

$$\begin{aligned}\langle v, M(s_{n,d})v\rangle &= \langle \Pi_{n,2d}(v \otimes v), \vec{s}_{n,d}\rangle \\ &= \langle \vec{q^2}, \vec{s}_{n,d}\rangle \\ &= \int_{x\in\mathcal{S}(\mathbb{R}^n)} q^2(x)d\mu(x) \\ &> 0,\end{aligned}$$

where μ is the normalized Lebesgue measure on $\mathcal{S}(\mathbb{R}^n)$ and $q(x) \in \mathbb{R}[x]_d$ is the symmetric tensor v after the isomorphism $S^d(\mathbb{R}^n) \cong \mathbb{R}[x]_d$. The third line is [Rez95, Proposition 6.6]. This completes the proof. □

3 Proof of convergence

In this section we prove that our polynomial optimization hierarchy converges as $O(1/k)$. We fix positive integers n, d, and $k \geq d$, and a homogeneous n-variate polynomial $p \in \mathbb{R}[x]_{2d}$. We also recall the definitions $s_{n,d} = (x_1^2 + \cdots + x_n^2)^d$, $p_{\min} = \min_{x\in\mathcal{S}(\mathbb{R}^n)} p(x)$, $M_k(p) = \Pi_{n,k}(M(p) \otimes \mathbb{1}_n^{\otimes k-d})\Pi_{n,k}$ and $\nu_k = \lambda_{\min}(M_k(p)M_k(s_{n,d})^{-1})$. Recall also the *condition number* of a positive definite matrix M is defined as $\kappa(M) = \lambda_{\max}(M)/\lambda_{\min}(M)$. Note that $M_k(s_{n,d})$ is invertible and moreover positive definite by Proposition 4. Furthermore, let

$$c_d = \sqrt{\frac{1}{2^d}\sum_{j=0}^{d}\frac{\binom{d}{j}^2}{\binom{2d}{2j}}}.$$

Theorem 5. *For each natural number $k \geq d$, it holds that $\nu_k \leq p_{\min}$ and*

$$p_{\min} - \nu_k \leq \|M(p)\|_\infty \left(1 + \kappa(M(s_{n,d}))\right) \frac{4d(n-1)}{c_d(k+1)} = O\left(\frac{1}{k}\right).$$

In particular, $\lim_{k\to\infty} \nu_k = p_{\min}$.

Numerics suggest that

$$\kappa(M(s_{n,d})) = \binom{n/2 + d - 1}{\lfloor d/2 \rfloor},$$

where the binomial function is extended to non-integer inputs via the Gamma function. While the quantity $\|M(p)\|_\infty$ is perhaps unnatural in the setting of polynomial optimization, it is equal to $C_{n,d}\|p\|_{S(1)}$, where $\|p\|_{S(1)}$ is the somewhat more natural quantity $\max_{x\in\mathcal{S}(\mathbb{R}^n)}|p(x)|$, for some constant $C_{n,d}$ which depends only on n and d. This follows from the equivalence of norms on finite-dimensional spaces and the fact that $\|\cdot\|_{S(1)}$ defines a norm on $\mathbb{R}[x]_{2d}$. Unsatisfying though this may be, it is not unprecedented: a similar phenomenon occurs in the convergence results presented in [DW12] (and clarified in [Lau19]) for the SOS hierarchy.

To prove Theorem 5 we require the following proposition, which relates $p_{\min}$ to the optimum value of a certain complex optimization problem.

Proposition 6. *Let*

$$p_{\min} := \min_{x \in \mathcal{S}(\mathbb{R}^n)} p(x),$$
$$p_{\min}^{\mathbb{C}} := \min_{v \in \mathcal{S}(\mathbb{C}^n)} \operatorname{Tr}((vv^*)^{\otimes d} M(p)).$$

Then

$$p_{\min}^{\mathbb{C}} \leq p_{\min} \leq \frac{p_{\min}^{\mathbb{C}}}{c_d}.$$

This proposition will allow us to apply the quantum de Finetti theorem (Theorem 7) directly in our analysis. Our $O(1/k)$ convergence result in Theorem 5 in particular reproduces the known $O(1/k)$ convergence result for the sum of squares hierarchy [DW12] (and also strengthens this to a convergence result for our weaker hierarchy of eigencomputations).

Proof of Proposition 6. Let $v \in \mathcal{S}(\mathbb{C}^n)$ be such that $p_{\min}^{\mathbb{C}} = \operatorname{Tr}((vv^*)^{\otimes d} M(p))$. Let $u = \operatorname{Re}(v)$ and $w = \operatorname{Im}(v)$. Recall that $M(p)$ is invariant under partial transposition along any of the d factors of $\mathbb{C}^n$. In particular, $M(p) = \frac{1}{2}(M(p) + M(p)^{\Gamma_1})$, where $(\cdot)^{\Gamma_1} = (T \otimes \mathbb{1}_{\operatorname{Hom}(\mathbb{R}^n)}^{\otimes d-1})(\cdot)$ denotes the partial transpose on the first factor (see Section 2). Thus,

$$\begin{aligned}
\operatorname{Tr}((vv^*)^{\otimes d} M(p)) &= \frac{1}{2}\operatorname{Tr}((vv^*)^{\otimes d} M(p)) + \frac{1}{2}\operatorname{Tr}((vv^*)^{\otimes d} M(p)^{\Gamma_1}) \\
&= \frac{1}{2}\operatorname{Tr}((vv^*)^{\otimes d} M(p)) + \frac{1}{2}\operatorname{Tr}((\overline{v}v^{\mathsf{T}}) \otimes (vv^*)^{\otimes d-1} M(p)) \\
&= \operatorname{Tr}\left((uu^{\mathsf{T}} + ww^{\mathsf{T}}) \otimes (vv^*)^{\otimes d-1} M(p)\right).
\end{aligned}$$

Continuing in this way for the other factors, we obtain

$$\begin{aligned}
\operatorname{Tr}((vv^*)^{\otimes d} M(p)) &= \operatorname{Tr}\left((uu^{\mathsf{T}} + ww^{\mathsf{T}})^{\otimes d} M(p)\right) \\
&= \langle \vec{p}, (u \otimes u + w \otimes w)^{\otimes d} \rangle \\
&= \langle \vec{p}, \Pi_{n,2d}((u \otimes u + w \otimes w)^{\otimes d}) \rangle,
\end{aligned}$$

where $\vec{p}$ is the polynomial p after the isomorphism $\mathbb{R}[x]_{2d} \cong S^{2d}(\mathbb{R}^n)$ (see Section 2), and $\Pi_{n,2d}$ is the orthogonal projection onto $S^{2d}(\mathbb{R}^n)$. Let

$$c = \|\Pi_{n,2d}((u \otimes u + w \otimes w)^{\otimes d})\|.$$

By [Rez13, Corollary 5.6] there exist real unit vectors $v_1, \ldots, v_{d+1} \in \mathcal{S}(\mathbb{R}^n)$ for which

$$\frac{1}{c}\Pi_{n,2d}((u \otimes u + w \otimes w)^{\otimes d}) \in \operatorname{conv}\{v_1^{\otimes 2d}, \ldots, v_{d+1}^{\otimes 2d}\},$$

where conv denotes the convex hull. Viewing these as elements of $\operatorname{Hom}(S^d(\mathbb{R}^n))$ and taking the trace of both sides, we obtain

$$\frac{p_{\min}^{\mathbb{C}}}{c} \in \operatorname{conv}\{\operatorname{Tr}((v_i {v_i}^{\mathsf{T}})^{\otimes d} M(p)) : i \in [d+1]\}.$$

Thus, there exists $i \in [d+1]$ for which $\operatorname{Tr}((v_i {v_i}^{\mathsf{T}})^{\otimes d} M(p)) \leq \frac{p_{\min}^{\mathbb{C}}}{c}$.

To complete the proof, it suffices to show that $c \geq c_d$. Let $U \in \mathrm{Hom}(\mathbb{R}^n)$ be an orthogonal matrix for which $U(uu^\intercal + ww^\intercal)U^\intercal = te_1e_1^\intercal + (1-t)e_2e_2^\intercal$ for some $t \in [0,1]$. Let $p(x,y) = (tx^2 + (1-t)y^2)^d \in \mathbb{R}[x,y]_{2d}$. Note that

$$\begin{aligned}\vec{p} &= \Pi_{n,2d}((te_1 \otimes e_1 + (1-t)e_2 \otimes e_2)^{\otimes d}) \\ &= \frac{1}{(2d)!}\sum_{j=0}^{d}\binom{d}{j}t^j(1-t)^{d-j}f_{I_j}e_{I_j},\end{aligned}$$

where $I_j = (1^{(2j)}, 2^{(2(d-j))})$, and $i^{(j)}$ represents the number i repeated j times. It follows that

$$\begin{aligned}c^2 &= \|\Pi_{n,2d}((te_1 \otimes e_1 + (1-t)e_2 \otimes e_2)^{\otimes d})\|^2 \\ &= \|\vec{p}\|^2 \\ &= \frac{1}{(2d)!}\sum_{j=0}^{d}\binom{d}{j}^2(2d)!(2j)!(2(d-j))! \;\; t^j(1-t)^{d-j} \\ &= \sum_{j=0}^{d}\frac{\binom{d}{j}^2}{\binom{2d}{2j}} \;\; t^j(1-t)^{d-j}.\end{aligned}$$

To complete the proof, we will show that this quantity is minimized when $t = 1/2$. If we make the change of variables $x = t - 1/2$ then we see that

$$\begin{aligned}c^2 &= \sum_{j=0}^{d}\frac{\binom{d}{j}^2}{\binom{2d}{2j}}\;\;\left(\frac{1}{2}+x\right)^j\left(\frac{1}{2}-x\right)^{d-j} \\ &= \sum_{j=0}^{d}\frac{\binom{d}{j}^2}{\binom{2d}{2j}}\;\;\left(\sum_{k=0}^{j}\binom{j}{k}\frac{x^k}{2^{j-k}}\right)\left(\sum_{\ell=0}^{d-j}\binom{d-j}{\ell}\frac{(-x)^\ell}{2^{d-j-\ell}}\right) \\ &= \sum_{j=0}^{d}\frac{\binom{d}{j}^2}{\binom{2d}{2j}}\;\;\sum_{k=0}^{j}\sum_{\ell=0}^{d-j}(-1)^\ell\binom{j}{k}\binom{d-j}{\ell}\frac{x^{k+\ell}}{2^{d-k-\ell}}.\end{aligned}$$

If we define $s = k + \ell$ then the above expression can be rewritten as

$$c^2 = \frac{1}{2^d}\sum_{j,s=0}^{d}\frac{\binom{d}{j}^2}{\binom{2d}{2j}}(-2)^s\sum_{k=0}^{s}(-1)^k\binom{j}{k}\binom{d-j}{s-k}x^s.$$

If s is odd then the coefficient of x^s above equals

$$-2^{s-d}\sum_{j=0}^{d}\sum_{k=0}^{s}(-1)^k\frac{\binom{d}{j}^2\binom{j}{k}\binom{d-j}{s-k}}{\binom{2d}{2j}},$$

which equals 0 since replacing j by $d-j$ and k by $s-k$ in the term

$$(-1)^k\frac{\binom{d}{j}^2\binom{j}{k}\binom{d-j}{s-k}}{\binom{2d}{2j}}$$

results in another term with the same absolute value but its sign flipped. It follows that c^2 is a sum of even powers of x, so it is minimized when $x = 0$ (i.e., when $t = 1/2$), completing the proof. $\square$

Our proof of convergence also relies on the quantum de Finetti theorem [CKMR07]. We use the form of this theorem stated and proven in [Wat18, Theorem 7.26].

Theorem 7 (Quantum de Finetti theorem). *For any symmetric unit vector $v \in \mathcal{S}(S^k(\mathbb{C}^n))$, there exists a matrix*

$$\tau \in \operatorname{conv}\{(uu^*)^{\otimes d} : u \in \mathcal{S}(\mathbb{C}^n)\} \subseteq \operatorname{Hom}(S^d(\mathbb{C}^n))$$

for which

$$\|\operatorname{Tr}_{[d+1..k]}(vv^*) - \tau\|_1 \leq \frac{4d(n-1)}{k+1}.$$

In the theorem statement, conv denotes the convex hull, and $\operatorname{Tr}_{[d+1..k]}(vv^*) \in \operatorname{Hom}((\mathbb{C}^n)^{\otimes d-k})$ denotes the partial trace of vv^* over the subsystems $d+1, \ldots, k$ (see Section 2). Now we can prove Theorem 5.

Proof of Theorem 5. Note that $\nu_k = \max\{\nu : M_k(p - \nu \cdot s_{n,d}) \succeq 0\}$ (see Footnote 6). For the inequality $\nu_k \leq p_{\min}$, note that since $M_k(p(x) - \nu_k s_{n,d}) \succeq 0$, we have

$$p(x) - \nu_k = \langle x^{\otimes k}, M_k(p(x) - \nu_k\, s_{n,d}) x^{\otimes k} \rangle \geq 0$$

for all $x \in \mathcal{S}(\mathbb{R}^n)$, so $p_{\min} - \nu_k \geq 0$.

For the bound, let $q_k(x) = p(x) - \nu_k \cdot s_{n,d} \in \mathbb{R}[x]_{2d}$, and let $v \in \mathcal{S}(S^k(\mathbb{R}^n))$ be a minimum (zero) eigenvector of $M_k(q_k)$. By the quantum de Finetti theorem, there exists a matrix

$$\tau \in \operatorname{conv}\{(uu^*)^{\otimes d} : u \in \mathcal{S}(\mathbb{C}^n)\} \subseteq \operatorname{Hom}(S^d(\mathbb{C}^n))$$

for which

$$\|\operatorname{Tr}_{[d+1..k]}(vv^\intercal) - \tau\|_1 \leq \frac{4d(n-1)}{k+1}.$$

Let

$$\begin{aligned} q_{k,\min} &:= \min_{x \in \mathcal{S}(\mathbb{R}^n)} q_k(x) = p_{\min} - \nu_k \\ q^{\mathbb{C}}_{k,\min} &:= \min_{v \in \mathcal{S}(\mathbb{C}^n)} \operatorname{Tr}((vv^*)^{\otimes d} M(q_k)). \end{aligned}$$

Then

$$\begin{aligned} q^{\mathbb{C}}_{k,\min} &\leq \operatorname{Tr}(M(q_k)\tau) \\ &= \operatorname{Tr}(M(q_k)(\tau - \operatorname{Tr}_{[d+1..k]}(vv^\intercal))) \\ &\leq \|M(q_k)\|_\infty \frac{4d(n-1)}{k+1} \\ &\leq \|M(p)\|_\infty\, (1 + \kappa(M(s_{n,d})))\, \frac{4d(n-1)}{k+1} \end{aligned}$$

The first line follows from convexity, the second line follows from the chain of equalities

$$\operatorname{Tr}(M(q_k) \operatorname{Tr}_{[d+1..k]}(vv^\intercal)) = \operatorname{Tr}((M(q_k) \otimes \mathbb{1}_n^{\otimes k-d}) vv^\intercal) = \operatorname{Tr}(M_k(q_k) vv^\intercal) = 0,$$

the third line follows from the quantum de Finetti theorem and the matrix norm inequality $\mathrm{Tr}(AB) \leq \|A\|_\infty \|B\|_1$, and the fourth line follows from

$$\begin{aligned}\|M(q_k)\|_\infty &\leq \|M(p)\|_\infty + |\nu_k| \|M(s_{n,d})\|_\infty \\ &\leq \|M(p)\|_\infty \left(1 + \frac{\lambda_{\max}(M(s_{n,d}))}{\lambda_{\min}(M(s_{n,d}))}\right)\end{aligned}$$

Here, the first line is the triangle inequality. The second line follows from $\|M(s_{n,d})\|_\infty = \lambda_{\max}(M(s_{n,d}))$ by Proposition 4, and $|\nu_k| \leq \|M(p)\|_\infty \, \lambda_{\min}(M(s_{n,d}))^{-1}$ since choosing ν equal to minus the righthand side would guarantee $M(p - \nu \, s_{n,d}) \succeq 0$. It follows from Proposition 6 that

$$\begin{aligned}p_{\min} - \nu_k &= q_{k,\min} \\ &\leq \frac{q^{\mathbb{C}}_{k,\min}}{c_d} \\ &= \|M(p)\|_\infty \, (1 + \kappa(M(s_{n,d}))) \, \frac{4d(n-1)}{c_d(k+1)}.\end{aligned}$$

This completes the proof. □

4 Numerical implementation and examples

While the quantity $\nu_k = \lambda_{\min}(M_k(p) M_k(s_{n,d})^{-1})$ that defines the k-th level of our hierarchy is the minimum eigenvalue of a matrix, it should not be computed as such, since doing so would require computation of the inverse of the matrix $M_k(s_{n,d})$.[8] An alternate way of computing ν_k is to instead solve the generalized eigenvalue/eigenvector problem

$$M_k(p)\mathbf{v} = \lambda M_k(s_{n,d})\mathbf{v}. \tag{6}$$

It is straightforward to show that the minimum generalized eigenvalue λ (i.e., the minimal λ for which there is a vector $\mathbf{v}$ solving Equation (6)) is equal to $\nu_k = \lambda_{\min}(M_k(p) M_k(s_{n,d})^{-1})$. However, this generalized eigenvalue can be found without inverting or multiplying any matrices. Furthermore, there are extremely fast numerical algorithms for solving this problem that can exploit the extreme sparsity of $M_k(p)$ and $M_k(s_{n,d})$ [Ste02]; $M_k(s_{n,d})$ is sparse because $s_{n,d}$ itself is sparse (i.e., most of its coefficients are equal to 0), and $M_k(p)$ is sparse when k is large even if p is dense (since

$$M_k(p) = \Pi_{n,k}(M(p) \otimes \mathbb{1}_n^{\otimes k-d})\Pi_{n,k},$$

and $\mathbb{1}_n^{\otimes k-d}$ is sparse). These generalized eigenvalue algorithms have been implemented in ARPACK [LSY98], making them available out-of-the-box in SciPy, Mathematica, MATLAB, and many other popular computational software packages. We have implemented the computation of ν_k in the QETLAB package for MATLAB [Joh16].

Example 8 (Homogeneous Motzkin polynomial)**.** Let

$$p(x) = x_1^2 x_2^2 (x_1^2 + x_2^2 - 3x_3^2) + x_3^6$$

[8] One could also store $M_k(s_{n,d})^{-1}$ in a lookup table, or possibly write down an explicit formula similar to Appendix A. We have chosen to rephrase the problem as a generalized eigenvalue/eigenvector problem instead.

be the homogeneous Motzkin polynomial [Lau09, Section 3.2] of degree $2d = 6$. This polynomial is non-negative but not a sum of squares; in fact, its minimum value on the unit sphere is exactly 0. Because this polynomial has so few non-zero coefficients, the matrix $M_k(p)$ is extremely sparse, so our hierarchy can be run at extremely high levels—on standard desktop hardware we have been able to go up to level $k - d = 2000$. Lower bounds on the minimum value of this polynomial at various levels of the hierarchy, and the time required to compute those bounds, are provided in Table 1.

$k-d$	lower bound	time
0	-0.500000	< 0.01 s
1	-0.200649	< 0.01 s
2	-0.127006	< 0.01 s
3	-0.084855	< 0.01 s
4	-0.053542	0.01 s
5	-0.045059	0.02 s
10	-0.018898	0.06 s
15	-0.011980	0.12 s
20	-0.008835	0.21 s
25	-0.007004	0.33 s
30	-0.005804	0.77 s
50	-0.003445	1.85 s

$k-d$	lower bound	time
75	-0.002285	4.96 s
100	-0.001710	10.9 s
200	-0.000852	1.27 min
300	-0.000567	4.02 min
400	-0.000425	10.4 min
500	-0.000340	19.4 min
750	-0.000227	1.36 h
1000	-0.000170	3.99 h
1250	-0.000136	5.86 h
1500	-0.000113	10.5 h
1750	-0.000097	19.2 h
2000	-0.000085	30.4 h

Table 1: Lower bounds on the minimum value of the Motzkin polynomial, as computed by the $(k-d)$-th level of our hierarchy, as well as the time required to compute those bounds.

Another lightweight alternative to the sum-of-squares hierarchy was introduced in [AM19]. Their diagonally-dominant sum-of-squares (DSOS) hierarchy can compute bounds on homogeneous polynomials via linear programming, which is less memory-intensive than the semidefinite programs required by the sum-of-squares hierarchy. However, our hierarchy based on (generalized) eigenvalues is even less memory-intensive and can thus be used to bound polynomials of even higher degree and even more variables. Furthermore, it seems to produce better bounds on randomly-generated polynomials than the DSOS hierarchy does.

Example 9 (Random dense quartic polynomials)**.** Our hierarchy seems to perform quite well on randomly-generated polynomials. To illustrate this fact, we generated a 10-variable degree-4 homogeneous polynomial with all $\binom{10+4-1}{4} = 715$ coefficients independently drawn from a standard normal distribution. Upper bounds on the maximum value of this polynomial on the unit sphere, as computed by the SOS hierarchy, DSOS hierarchy, and our hierarchy, are given in Table 2.[9]

Of particular note is the fact the bound arising from the first (i.e., $k - d = 0$) level of our hierarchy is better than the bound arising from the first level of the DSOS hierarchy (or even the 2nd level of that hierarchy), despite requiring fewer computational resources. We saw similar behaviour with every randomly-generated example that we tried.

Our hierarchy is also less memory intensive; it was noted in [AM19] that while the SOS hierarchy can only be used for quartic polynomials with up to 25 variables or so, the DSOS hierarchy can

[9] In both Table 2 and Table 3, we use the timings reported in [AM19] for the DSOS and SOS hierarchies, since their implementations of those hierarchies seem to be a bit quicker than our implementations of them.

hierarchy	$k-d$	upper bound	time
DSOS	0	5.9578	0.30 s
	1	4.7756	0.92 s
Ours	0	3.5674	0.16 s
	1	2.6235	0.87 s
	2	2.3612	5.94 s
	3	2.2320	46.5 s
	4	2.1547	5.03 min
	5	2.1025	31.8 min
	6	2.0650	3.48 h
	7	2.0365	16.4 h
SOS	0	1.8290	0.24 s

Table 2: Upper bounds on the maximum value of a particular (randomly-generated) dense 10-variable quartic polynomial, and the times required to compute these bounds by various hierarchies.

be used for much larger quartic polynomials with as many as 70 variables. Our hierarchy can go even farther, producing bounds on dense quartic polynomials with up to 90 variables, as detailed in Table 3.

	SOS	DSOS		Ours		
n	$k-d=0$	$k-d=0$	$k-d=1$	$k-d=0$	$k-d=1$	$k-d=2$
10	0.24 s	0.30 s	0.92 s	0.16 s	0.86 s	6.61 s
15	5.60 s	0.38 s	6.26 s	0.99 s	8.80 s	4.24 min
20	1.37 min	0.74 s	38.0 s	4.42 s	1.47 min	1.90 h
25	17.8 min	15.51 s	6.15 min	12.7 s	10.5 min	27.7 h
30	∞	7.88 s	∞	31.6 s	1.07 h	∞
40	∞	10.7 s	∞	2.56 min	21.1 h	∞
50	∞	26.0 s	∞	9.39 min	∞	∞
60	∞	58.1 s	∞	31.2 min	∞	∞
70	∞	5.71 min	∞	1.55 h	∞	∞
80	∞	∞	∞	4.56 h	∞	∞
90	∞	∞	∞	11.4 h	∞	∞

Table 3: The time it takes for various levels of various hierarchies to produce an upper bound on the maximum value of an n-variable quartic polynomial. Values of ∞ indicate that memory limitations were exceeded, so no bound could be computed on a standard desktop computer running MATLAB with 16 GB of RAM.

5 Tensor optimization hierarchy

Let $n_1, \ldots, n_m$ and $D_1, \ldots, D_m$ be positive integers, and let $p \in (\mathbb{R}^{n_1})^{\otimes D_1} \otimes \cdots \otimes (\mathbb{R}^{n_m})^{\otimes D_m}$ be a tensor. In this section we generalize our hierarchy to solve the following minimization problem:

$$\min_{v_j \in \mathcal{S}(\mathbb{R}^{n_j})} \langle p, v_1^{\otimes D_1} \otimes \cdots \otimes v_m^{\otimes D_m} \rangle.$$

We let $\mathbf{n} = (n_1, \ldots, n_m)$ and $\mathbf{D} = (D_1, \ldots, D_m)$, and denote this minimum value by $p_{\min}^{(\mathbf{n},\mathbf{D}/2)}$ (soon we will restrict without loss of generality to the case when $D_1, \ldots, D_m$ are even, so dividing by 2 in the index now will ease notation later).[10]

Remark 10. Note that if any of $D_1, \ldots, D_m$ are odd, then

$$p_{\min}^{(\mathbf{n},\mathbf{D}/2)} = - \max_{v_j \in \mathcal{S}(\mathbb{R}^{n_j})} |\langle p, v_1^{\otimes D_1} \otimes \cdots \otimes v_m^{\otimes D_m} \rangle|.$$

However these quantities may differ if $D_1, \ldots, D_m$ are all even. For example, if $p(x_1, x_2) = x_1^2$, then $p_{\min} = 0$, but

$$\max_{v \in \mathcal{S}(\mathbb{R}^2)} |\langle \vec{p}, v^{\otimes 2} \rangle| = 1.$$

Note also that

$$p_{\min}^{(2,2),(1,1)/2} = \min_{u,v \in \mathcal{S}(\mathbb{R}^2)} \langle e_1^{\otimes 2}, u \otimes v \rangle = -1 \neq p_{\min},$$

so the minimum depends on the choice of $\mathbf{D}$.

We can assume $p \in S^{D_1}(\mathbb{R}^{n_1}) \otimes \cdots \otimes S^{D_m}(\mathbb{R}^{n_m})$ without loss of generality by replacing $p \to (\Pi_{n_1,D_1} \otimes \cdots \otimes \Pi_{n_m,D_m})p$, where Π_{n_j,D_j} is the orthogonal projection onto $S^{D_j}(\mathbb{R}^{n_j})$. The following proposition establishes that it suffices to assume $D_1, \ldots, D_m$ are all even.

Proposition 11. *If D_1 is odd, then*

$$\begin{aligned} p_{\min}^{(\mathbf{n},\mathbf{D}/2)} &= \frac{(D_1+1)^{(D_1+1)/2}}{D_1^{D_1/2}} (e_{n_1+1} \otimes p)_{\min}^{(\tilde{\mathbf{n}},\tilde{\mathbf{D}}/2)} \\ &= \frac{(D_1+1)^{(D_1+1)/2}}{D_1^{D_1/2}} \min_{\substack{v_j \in \mathcal{S}(\mathbb{R}^{n_j}) \\ 1 \neq j \in [m]}} \min_{v_1 \in \mathcal{S}(\mathbb{R}^{n_1+1})} \langle e_{n_1+1} \otimes p, v_1^{\otimes D_1+1} \otimes v_2^{\otimes D_2} \otimes \cdots \otimes v_m^{\otimes D_m} \rangle, \end{aligned}$$

where $\tilde{\mathbf{D}} = (D_1+1, D_2, \ldots, D_m)$ and $\tilde{\mathbf{n}} = (n_1+1, n_2, \ldots, n_m)$, and we have embedded $\mathbb{R}^{n_1}$ into the first n_1 coordinates of $\mathbb{R}^{n_1+1}$.

---

[10] Note that $p_{\min}^{\mathbf{n},\mathbf{D}/2}$ can also be computed using standard polynomial optimization techniques (such as SOS) by setting

$$\tilde{p}^{\mathbf{n},\mathbf{D}/2}(v_1, \ldots, v_m) = \langle p, v_1^{\otimes D_1} \otimes \cdots \otimes v_m^{\otimes D_m} \rangle,$$

which is a homogeneous degree-($\sum_i D_i$) polynomial in the $m(\sum_i n_i)$ variables $v_1, \ldots, v_m$, and minimizing over $\mathcal{S}(\mathbb{R}^{n_1}) \times \cdots \times \mathcal{S}(\mathbb{R}^{n_m})$. However it is not obvious that $p_{\min}^{\mathbf{n},\mathbf{D}/2}$ can be computed by a hierarchy of minimum eigenvalue computations, which is the main result of this section.

Note that $\tilde{D}_1 = D_1 + 1$ is even and $\tilde{D}_j = D_j$ for all $1 \neq j \in [m]$. By symmetry, Proposition 11 implies analogous statements for other $j \in [m]$. So one can assume $D_1, \ldots, D_m$ are all even, at the expense of adding up to m new variables.

Proof of Proposition 11. For $v := (v_2, \ldots, v_m)$ fixed, define a polynomial $p_v \in \mathbb{R}[x_1, \ldots, x_{n_1}]_{D_1}$ as

$$p_v(v_1) = \langle p, v_1^{\otimes D_1} \otimes \cdots \otimes v_m^{\otimes D_m} \rangle.$$

By [DW12, Lemma B.2], we have

$$\begin{aligned} p_{\min}^{(\mathbf{n},\mathbf{D}/2)} &= \min_{\substack{v_j \in \mathcal{S}(\mathbb{R}^{n_j}) \\ 1 \neq j \in [m]}} \min_{v_1 \in \mathcal{S}(\mathbb{R}^{n_1})} p_v(v_1) \\ &= \frac{(D_1+1)^{(D_1+1)/2}}{D_1^{D_1/2}} \min_{\substack{v_j \in \mathcal{S}(\mathbb{R}^{n_j}) \\ 1 \neq j \in [m]}} \min_{v_1 \in \mathcal{S}(\mathbb{R}^{n_1+1})} (x_{n+1} \cdot p_v)(v_1) \\ &= \frac{(D_1+1)^{(D_1+1)/2}}{D_1^{D_1/2}} \min_{\substack{v_j \in \mathcal{S}(\mathbb{R}^{n_j}) \\ 1 \neq j \in [m]}} \min_{v_1 \in \mathcal{S}(\mathbb{R}^{n_1+1})} \langle e_{n_1+1} \otimes p, v_1^{\otimes D_1+1} \otimes v_2^{\otimes D_2} \otimes \cdots \otimes v_m^{\otimes D_m} \rangle. \end{aligned}$$

This completes the proof. □

In the remainder of this section, we assume $D_1, \ldots, D_m$ are even without loss of generality. Let $d_j = D_j/2$ for all $j \in [m]$, and let $\mathbf{d} = \mathbf{D}/2 = (d_1, \ldots, d_m)$. For positive integers n, d let $\vec{s}_{n,d} \in S^{2d}(\mathbb{R}^n)$ be the polynomial $s_{n,d} = (x_1^2 + \cdots + x_n^2)^d \in \mathbb{R}[x_1, \ldots, x_n]_{2d}$ after the isomorphism $\mathbb{R}[x_1, \ldots, x_n]_{2d} \cong S^{2d}(\mathbb{R}^n)$. For a tensor $q \in S^{2d_1}(\mathbb{R}^{n_1}) \otimes \cdots \otimes S^{2d_m}(\mathbb{R}^{n_m})$ and an integer $k \geq \max_j d_j$, let

$$M_k^{(\mathbf{n},\mathbf{d})}(q) := (\Pi_{n_1,k} \otimes \cdots \otimes \Pi_{n_m,k})(M(q) \otimes \mathbb{1}_{n_1}^{\otimes k-d_1} \otimes \cdots \otimes \mathbb{1}_{n_m}^{\otimes k-d_m})(\Pi_{n_1,k} \otimes \cdots \otimes \Pi_{n_m,k}),$$

where $\Pi_{n_j,k}$ is the orthogonal projection onto $S^k(\mathbb{R}^{n_j})$, and $M(q)$ is q viewed as an element of $\mathrm{Hom}_{\mathbb{R}}(S^{d_1}(\mathbb{R}^{n_1}) \otimes \cdots \otimes S^{d_m}(\mathbb{R}^{n_m}))$. The k-th level of our hierarchy computes

$$\nu_k^{(\mathbf{n},\mathbf{d})} := \lambda_{\min}(M_k^{(\mathbf{n},\mathbf{d})}(p)\ M_k^{(\mathbf{n},\mathbf{d})}(\vec{s}_{\mathbf{n},\mathbf{d}})^{-1}),$$

where $\vec{s}_{\mathbf{n},\mathbf{d}} := \vec{s}_{n_1,d_1} \otimes \cdots \otimes \vec{s}_{n_m,d_m}$. Note that

$$M_k^{(\mathbf{n},\mathbf{d})}(\vec{s}_{\mathbf{n},\mathbf{d}}) = M_k(s_{n_1,d_1}) \otimes \cdots \otimes M_k(s_{n_m,d_m})$$

is positive definite (hence invertible) by Proposition 4. In Section 5.1 we prove that the $\nu_k^{(\mathbf{n},\mathbf{d})}$ converge to $p_{\min}^{(\mathbf{n},\mathbf{d})}$ from below at rate $O(1/k)$, and in Section 5.2 we use this hierarchy to minimize biquadratic forms and compute the real spectral norm of a tensor.

5.1 Proof of convergence

The following theorem proves that our hierarchy of minimum eigenvalue computations $\nu_k^{(\mathbf{n},\mathbf{d})}$ converges to $p_{\min}^{(\mathbf{n},\mathbf{d})}$ at a rate of $O(1/k)$. Let $c_{\mathbf{d}} = c_{d_1} \cdots c_{d_m}$, where for a positive integer d we define

$$c_d = \sqrt{\frac{1}{2^d} \sum_{j=0}^{d} \frac{\binom{d}{j}^2}{\binom{2d}{2j}}}.$$

Theorem 12. *It holds that* $\nu_k^{(\mathbf{n},\mathbf{d})} \leq p_{\min}^{(\mathbf{n},\mathbf{d})}$*, and*

$$p_{\min}^{(\mathbf{n},\mathbf{d})} - \nu_k^{(\mathbf{n},\mathbf{d})} \leq \|M(p)\|_\infty \left(1 + \kappa(M(s_{\mathbf{n},\mathbf{d}}))\right) \frac{4d(\max_j n_j - 1)}{c_{\mathbf{d}}(k+1)} = O\left(\frac{1}{k}\right),$$

where $d = d_1 + \cdots + d_m$*. In particular,* $\lim_{k\to\infty} \nu_k^{(\mathbf{n},\mathbf{d})} = p_{\min}^{(\mathbf{n},\mathbf{d})}$.

Numerics indicate that

$$\kappa(M(s_{\mathbf{n},\mathbf{d}})) = \binom{n_1/2 + d_1 - 1}{\lfloor d_1/2 \rfloor} \cdots \binom{n_m/2 + d_m - 1}{\lfloor d_m/2 \rfloor},$$

where the binomial function is extended to non-integer inputs via the Gamma function. We can make a similar remark as we did after Theorem 5: While the quantity $\|M(p)\|_\infty$ is perhaps unnatural, it is equal to $C_{(\mathbf{n},\mathbf{d})}\|p\|_{S_{(\mathbf{n},\mathbf{d})}(1)}$, where $\|p\|_{S_{(\mathbf{n},\mathbf{d})}(1)}$ is the somewhat more natural quantity

$$\max_{v_j \in \mathcal{S}(\mathbb{R}^{n_j})} |\langle p, v_1^{\otimes 2d_1} \otimes \cdots \otimes v_m^{\otimes 2d_m} \rangle|,$$

for some constant $C_{(\mathbf{n},\mathbf{d})}$ which depends only on $\mathbf{n}$ and $\mathbf{d}$. This follows from the equivalence of norms on finite-dimensional spaces and the fact that $\|\cdot\|_{S_{(\mathbf{n},\mathbf{d})}(1)}$ defines a norm on $S^{2d_1}(\mathbb{R}^{n_1}) \otimes \cdots \otimes S^{2d_m}(\mathbb{R}^{n_m})$. Alternatively, one can use the bound $\|M(p)\|_\infty \leq \|p\|_2$.

To prove the theorem we require the following proposition. In the proposition we view $M(p)$ as an element of $\mathrm{Hom}_{\mathbb{C}}(S^{d_1}(\mathbb{C}^{n_1}) \otimes \cdots \otimes S^{d_m}(\mathbb{C}^{n_m}))$ under the standard inclusion $\mathbb{R} \subseteq \mathbb{C}$, and recall that v^* denotes the conjugate-transpose of v.

Proposition 13. *Let* $p_{\min} = p_{\min}^{(\mathbf{n},\mathbf{d})}$ *and*

$$p_{\min}^{\mathbb{C}} = \min_{v_j \in \mathcal{S}(\mathbb{C}^{n_j})} \mathrm{Tr}((v_1 v_1^*)^{\otimes d_1} \otimes \cdots \otimes (v_m v_m^*)^{\otimes d_m} M(p)).$$

It holds that

$$p_{\min}^{\mathbb{C}} \leq p_{\min} \leq \frac{p_{\min}^{\mathbb{C}}}{c_{\mathbf{d}}}.$$

Proof. It will be convenient to view $M(p)$ as a map on $(\mathbb{C}^n)^{\otimes d}$ by setting it equal to zero on the orthogonal complement to $S^d(\mathbb{C}^n)$. Let $v_1 \in \mathcal{S}(\mathbb{C}^{n_1}), \ldots, v_m \in \mathcal{S}(\mathbb{C}^{n_m})$ be such that

$$p_{\min}^{\mathbb{C}} = \mathrm{Tr}((v_1 v_1^*)^{\otimes d_1} \otimes \cdots \otimes (v_m v_m^*)^{\otimes d_m} M(p)).$$

Let $u_j = \mathrm{Re}(v_j)$ and $w_j = \mathrm{Im}(v_j)$ for each $j \in [m]$. Note that $M(p)$ is invariant under partial transposition along any of the d_1 factors of $\mathbb{C}^{n_1}$. In particular, $M(p) = \frac{1}{2}(M(p) + M(p)^{\Gamma_1})$, where $(\cdot)^{\Gamma_1}$ denotes the partial transpose on the first factor of $\mathbb{C}^{n_1}$. Thus,

$$\begin{aligned}
&\mathrm{Tr}((v_1 v_1^*)^{\otimes d_1} \otimes \cdots \otimes (v_m v_m^*)^{\otimes d_m} M(p)) \\
&= \frac{1}{2}\mathrm{Tr}((v_1 v_1^*)^{\otimes d_1} \otimes \cdots \otimes (v_m v_m^*)^{\otimes d_m} M(p)) + \frac{1}{2}\mathrm{Tr}((v_1 v_1^*)^{\otimes d_1} \otimes \cdots \otimes (v_m v_m^*)^{\otimes d_m} M(p)^{\Gamma_1}) \\
&= \frac{1}{2}\mathrm{Tr}((v_1 v_1^*)^{\otimes d_1} \otimes \cdots \otimes (v_m v_m^*)^{\otimes d_m} M(p)) + \frac{1}{2}\mathrm{Tr}((\overline{v}_1 v_1^\intercal) \otimes (v_1 v_1^*)^{\otimes d_1 - 1} \otimes \cdots \otimes (v_m v_m^*)^{\otimes d_m} M(p)) \\
&= \mathrm{Tr}\left((u_1 u_1^\intercal + w_1 w_1^\intercal) \otimes (v_1 v_1^*)^{\otimes d_1 - 1} \otimes \cdots \otimes (v_m v_m^*)^{\otimes d_m} M(p)\right).
\end{aligned}$$

Continuing in this way for the other factors, we obtain

$$\begin{aligned}
&\mathrm{Tr}((v_1 v_1^*)^{\otimes d_1} \otimes \cdots \otimes (v_m v_m^*)^{\otimes d_m} M(p)) \\
&\qquad = \mathrm{Tr}((u_1 u_1^\intercal + w_1 w_1^\intercal)^{\otimes d_1} \otimes \cdots \otimes (u_m u_m^\intercal + w_m w_m^\intercal)^{\otimes d_m} M(p)) \\
&\qquad = \langle p, (u_1 \otimes u_1 + w_1 \otimes w_1)^{\otimes d_1} \otimes \cdots \otimes (u_m \otimes u_m + w_m \otimes w_m)^{\otimes d_m} \rangle \\
&\qquad = \langle p, \Pi_{n_1, 2d_1}((u_1 \otimes u_1 + w_1 \otimes w_1)^{\otimes d_1}) \otimes \cdots \otimes \Pi_{n_m, 2d_m}((u_m \otimes u_m + w_m \otimes w_m)^{\otimes d_m}) \rangle,
\end{aligned}$$

where $\Pi_{n_j, 2d_j}$ is the orthogonal projection onto $S^{2d_j}(\mathbb{R}^{n_j})$. For each $j \in [m]$, let

$$c^{(j)} = \| \Pi_{n_j, 2d_j}((u_j \otimes u_j + w_j \otimes w_j)^{\otimes d_j}) \|$$

By [Rez13, Corollary 5.6] for each $j \in [m]$ there exist real unit vectors $z_{j,1}, \ldots, z_{j,d_j+1} \in \mathcal{S}(\mathbb{R}^{n_j})$ for which

$$\frac{1}{c^{(j)}} \Pi_{n_j, 2d_j}((u_j \otimes u_j + w_j \otimes w_j)^{\otimes d_j}) \in \mathrm{conv}\{z_{j,1}^{\otimes 2d_j}, \ldots, z_{j,d_j+1}^{\otimes 2d_j}\}.$$

Thus,

$$\frac{p_{\min}^{\mathbb{C}}}{c^{(1)} \cdots c^{(m)}} \in \mathrm{conv}\{\mathrm{Tr}((z_{1,i_1} z_{1,i_1}^\intercal)^{\otimes d_1} \otimes \cdots \otimes (z_{m,i_m} z_{m,i_m}^\intercal)^{\otimes d_m} M(p)) : i_j \in [d_j + 1]\}.$$

Thus, there exist $i_1 \in [d_1 + 1], \ldots, i_m \in [d_m + 1]$ for which

$$\mathrm{Tr}((z_{1,i_1} z_{1,i_1}^\intercal)^{\otimes d_1} \otimes \cdots \otimes (z_{m,i_m} z_{m,i_m}^\intercal)^{\otimes d_m} M(p)) \leq \frac{p_{\min}^{\mathbb{C}}}{c^{(1)} \cdots c^{(m)}}.$$

Note that $c^{(j)} \geq c_{d_j}$ for each $j \in [m]$, by the proof of Proposition 6. This completes the proof. □

We also require a slightly more general quantum de Finetti theorem.

Theorem 14 (More general quantum de Finetti theorem). *Let $k \geq \max_j d_j$ be an integer, and let $v \in \mathcal{S}(S^k(\mathbb{C}^{n_1}) \otimes \cdots \otimes S^k(\mathbb{C}^{n_m}))$. Then there exists a matrix*

$$\tau \in \mathrm{conv}\{(u_1 u_1^*)^{\otimes d_1} \otimes \cdots \otimes (u_m u_m^*)^{\otimes d_m} : u_j \in \mathcal{S}(\mathbb{C}^{n_j})\}$$

for which

$$\| \mathrm{Tr}_{[d_1+1..k], \ldots, [d_m+1..k]}(vv^*) - \tau \|_1 \leq \frac{4d(\max_j n_j - 1)}{k+1},$$

where $d = d_1 + \cdots + d_m$, and $\mathrm{Tr}_{[d_1+1..k], \ldots, [d_m+1..k]}(vv^*)$ denotes the partial trace over $k - d_j$ copies of $\mathbb{C}^{n_j}$ for each $j \in [m]$.

This theorem can be seen as a special case of [KM09, Theorem III.3, Remark III.4], or a slight generalization of [CKMR07, Theorem II.2′] and [Wat18, Theorem 7.26]. We will prove the theorem as a corollary to [KM09, Theorem III.3, Remark III.4].

Proof of Theorem 14. Let

$$\begin{aligned}
\mathcal{A} &= S^{d_1}(\mathbb{C}^{n_1}) \otimes \cdots \otimes S^{d_m}(\mathbb{C}^{n_m}) \\
\mathcal{B} &= S^{k-d_1}(\mathbb{C}^{n_1}) \otimes \cdots \otimes S^{k-d_m}(\mathbb{C}^{n_m}) \\
\mathcal{C} &= S^{k}(\mathbb{C}^{n_1}) \otimes \cdots \otimes S^{k}(\mathbb{C}^{n_m}) \\
\mathcal{X} &= u_1^{\otimes d_1} \otimes \cdots \otimes u_m^{\otimes d_m},
\end{aligned}$$

where the $u_j \in \mathcal{S}(\mathbb{C}^{n_j})$ are arbitrary but fixed. Then $\mathcal{C} \subseteq \mathcal{A} \otimes \mathcal{B}$ is an irreducible subrepresentation (with multiplicity one) of the product unitary group $U(\mathbb{C}^{n_1}) \times \cdots \times U(\mathbb{C}^{n_m})$. It is also straightforward to check that the quantity $\delta(\mathcal{X})$ defined in [KM09, Definition III.2] is equal to

$$\begin{aligned}\frac{\dim(\mathcal{B})}{\dim(\mathcal{C})} &= \prod_{j=1}^{m} \frac{\binom{n_j+k-d_j-1}{k-d_j}}{\binom{n_j+k-1}{k}} \\ &\geq \prod_{j=1}^{m} \left(1 - \frac{d_j(n_j-1)}{k+1}\right) \\ &\geq 1 - \frac{d(\max_j n_j - 1)}{k+1},\end{aligned}$$

where the second line is a standard inequality that can be found e.g. in [Wat18, Eq. (7.196)] The desired bound then follows directly from the bound given in [KM09, Theorem III.3, Remark III.4] in terms of $\delta(\mathcal{X})$. □

Now we can prove Theorem 12.

Proof of Theorem 12. Note that

$$\nu_k^{(\mathbf{n},\mathbf{d})} = \max\{\nu : M_k^{(\mathbf{n},\mathbf{d})}(p - \nu\ s_{\mathbf{n},\mathbf{d}}) \succeq 0\}.$$

For the inequality $\nu_k^{(\mathbf{n},\mathbf{d})} \leq p_{\min}^{(\mathbf{n},\mathbf{d})}$, note that since $M_k^{(\mathbf{n},\mathbf{d})}(p - \nu_k^{(\mathbf{n},\mathbf{d})} s_{\mathbf{n},\mathbf{d}}) \succeq 0$, we have

$$\langle p, v_1^{\otimes 2d_1} \otimes \cdots \otimes v_m^{\otimes 2d_m}\rangle - \nu_k^{(\mathbf{n},\mathbf{d})} = \mathrm{Tr}((v_1 v_1^{\intercal})^{\otimes k} \otimes \cdots \otimes (v_m v_m^{\intercal})^{\otimes k} M_k^{(\mathbf{n},\mathbf{d})}(p - \nu_k^{(\mathbf{n},\mathbf{d})}\ \ s_{\mathbf{n},\mathbf{d}})) \geq 0$$

for any $v_1 \in \mathcal{S}(\mathbb{R}^{n_1}), \ldots, v_m \in \mathcal{S}(\mathbb{R}^{n_m})$. For the bound, let

$$q_k = p - \nu_k^{(\mathbf{n},\mathbf{d})} s_{\mathbf{n},\mathbf{d}} \in S^{2d_1}(\mathbb{R}^{n_1}) \otimes \cdots \otimes S^{2d_m}(\mathbb{R}^{n_m}),$$

and let $v \in \mathcal{S}(S^{2k}(\mathbb{R}^{n_1}) \otimes \cdots \otimes S^{2k}(\mathbb{R}^{n_m})$ be a minimum (zero) eigenvector of $M_k^{(\mathbf{n},\mathbf{d})}(q_k)$. By Theorem 14, there exists a matrix

$$\tau \in \mathrm{conv}\{(u_1 u_1^*)^{\otimes d_1} \otimes \cdots \otimes (u_m u_m^*)^{\otimes d_m} : u_j \in \mathcal{S}(\mathbb{C}^{n_j})\}$$

for which

$$\| \mathrm{Tr}_{[d_1+1..k],\ldots,[d_m+1..k]}(vv^{\intercal}) - \tau \|_1 \leq \frac{4d(\max_j n_j - 1)}{k+1}.$$

Let

$$\begin{aligned} q_{k,\min} = q_{k,\min}^{(\mathbf{n},\mathbf{d})} &= \min_{v_j \in \mathcal{S}(\mathbb{R}^{n_j})} \langle q_k, v_1^{\otimes 2d_1} \otimes \cdots \otimes v_m^{\otimes 2d_m}\rangle \\ q_{k,\min}^{\mathbb{C}} &= \min_{v_j \in \mathcal{S}(\mathbb{C}^{n_j})} \mathrm{Tr}((v_1 v_1^*)^{\otimes d_1} \otimes \cdots \otimes (v_m v_m^*)^{\otimes d_m} M(q_k)).\end{aligned}$$

Then

$$\begin{aligned} q_{k,\min}^{\mathbb{C}} &\leq \mathrm{Tr}(M(q_k)\tau) \\ &= \mathrm{Tr}(M(q_k)(\tau - \mathrm{Tr}_{[d_1+1..k],\ldots,[d_m+1..k]}(vv^{\intercal}))) \\ &\leq \|M(q_k)\|_\infty \frac{4d(\max_j n_j - 1)}{k+1} \\ &\leq \|M(p)\|_\infty (1 + \kappa(M(s_{\mathbf{n},\mathbf{d}}))) \frac{4d(\max_j n_j - 1)}{k+1}.\end{aligned}$$

The first line follows from convexity. The second line follows from the chain of equalities

$$\begin{aligned}\mathrm{Tr}(M(q_k)\,\mathrm{Tr}_{[d_1+1..k],\dots,[d_m+1..k]}(vv^\intercal)) &= \mathrm{Tr}((M(q_k)\otimes \mathbb{1}_{n_1}^{\otimes k-d_1}\otimes\cdots\otimes \mathbb{1}_{n_m}^{\otimes k-d_m})vv^\intercal)\\ &= \mathrm{Tr}(M_k(q_k)vv^\intercal)\\ &= 0.\end{aligned}$$

The third line follows from Theorem 14 and the matrix norm inequality $\mathrm{Tr}(AB) \le \|A\|_\infty \|B\|_1$. The fourth line follows from

$$\begin{aligned}\|M(q_k)\|_\infty &\le \|M(p)\|_\infty + |\nu_k^{(\mathbf{n},\mathbf{d})}|\|M(s_{\mathbf{n},\mathbf{d}})\|_\infty\\ &\le \|M(p)\|_\infty\,(1+\kappa(M(s_{\mathbf{n},\mathbf{d}}))) .\end{aligned}$$

Here, the first line is the triangle inequality, and the second line follows from $\|M(s_{\mathbf{n},\mathbf{d}})\|_\infty = \lambda_{\max}(M(s_{\mathbf{n},\mathbf{d}}))$ by Proposition 4, and $|\nu_k^{(\mathbf{n},\mathbf{d})}| \le \|M(p)\|_\infty \lambda_{\min}(M(s_{\mathbf{n},\mathbf{d}}))^{-1}$ since choosing ν equal to minus the righthand side would guarantee $M(p) - \nu M(s_{\mathbf{n},\mathbf{d}}) \succeq 0$. It follows from Proposition 13 that

$$\begin{aligned}p_{\min}^{(\mathbf{n},\mathbf{d})} - \nu_k &= q_{k,\min}\\ &\le \frac{q_{k,\min}^{\mathbb{C}}}{c_{\mathbf{d}}}\\ &\le \|M(p)\|_\infty\,(1+\kappa(M(s_{\mathbf{n},\mathbf{d}})))\,\frac{4d(\max_j n_j - 1)}{c_{\mathbf{d}}(k+1)}.\end{aligned}$$

This completes the proof. □

5.2 Examples: Biquadratic forms and the real spectral norm of a tensor

In this section we use our tensor optimization hierarchy and Theorem 12 to give hierarchies of eigencomputations for two tasks: minimizing a biquadratic form over the unit sphere, and computing the real spectral norm of a real tensor.

Example 15 (Biquadratic forms and bihomogeneous forms)**.** A *bihomogeneous form* of degree d is a polynomial $r(x,y) \in \mathbb{R}[x,y]_{2d}$ in two sets of variables $x = (x_1,\dots,x_{n_1})$ and $y = (y_1,\dots,y_{n_2})$ for which there exists a tensor $p \in S^d(\mathbb{R}^{n_1}) \otimes S^d(\mathbb{R}^{n_2})$ that satisfies

$$r(x,y) = \langle p, x^{\otimes d}\otimes y^{\otimes d}\rangle.$$

A *biquadratic form* is a bihomogeneous form of degree 2. We will focus on biquadratic forms, but analogous results are easily shown for bihomogeneous forms of higher degree. Let $q(x,y) \in \mathbb{R}[x,y]_4$ be a biquadratic form, let

$$M_k = (\Pi_{n_1,k}\otimes\Pi_{n_2,k})(M(q)\otimes \mathbb{1}_{n_1}^{\otimes k-1}\otimes \mathbb{1}_{n_2}^{\otimes k-1})(\Pi_{n_1,k}\otimes\Pi_{n_2,k})$$

and let $\nu_k = \lambda_{\min}(M_k)$. By Theorem 12, the ν_k converge to $\min_{x\in\mathcal{S}(\mathbb{R}^{n_1}),y\in\mathcal{S}(\mathbb{R}^{n_2})} q(x,y)$ from below at a rate of $O(1/k)$.[11] In particular, suppose we wish to determine whether q is *strictly positive* i.e. $q(x,y) > 0$ for all $x, y \ne 0$. Our hierarchy establishes that this holds if and only if M_k is positive definite for some k. Note that

$$\langle x^{\otimes k}\otimes y^{\otimes k}, M_k\ \ x^{\otimes k}\otimes y^{\otimes k}\rangle = q(x,y)\cdot s(x)^k\cdot s(y)^k,$$

[11] To see this, note that $M_k^{(\mathbf{n},(1,1))}(q) = M_k$ and $M_k^{(\mathbf{n},(1,1))}(s_{\mathbf{n},(1,1)}) = \mathbb{1}_{S^k(\mathbb{R}^{n_1})}\otimes \mathbb{1}_{S^k(\mathbb{R}^{n_2})}$.

so M_k is a *Gram matrix* for q.[12] In particular, a biquadratic form is strictly positive on non-zero inputs if and only if $q(x,y)\cdot s(x)^k \cdot s(y)^k$ admits a positive definite Gram matrix for some k.

Example 16 (The real spectral norm of a tensor)**.** Let $p \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_m}$ be a tensor. Assume without loss of generality that $\|p\|_2 = 1$, where $\|\cdot\|_2$ denotes the Euclidean norm. The *(real) spectral norm* of p is defined as

$$\|p\|_{\sigma,\mathbb{R}} := \max_{v_j \in \mathcal{S}(\mathbb{R}^{n_j})} |\langle p, v_1 \otimes \cdots \otimes v_m \rangle|.$$

We can use our hierarchy to compute the spectral norm of p as follows. First note that

$$-\|p\|_{\sigma,\mathbb{R}} = \min_{v_j \in \mathcal{S}(\mathbb{R}^{n_j})} \langle p, v_1 \otimes \cdots \otimes v_m \rangle,$$

so it suffices to compute the righthand side of this expression. Let

$$q = p \otimes e_{n_1+1} \otimes \cdots \otimes e_{n_m+1} \in \left(\bigotimes_{j=1}^{m} \mathbb{R}^{n_j+1} \right)^{\otimes 2},$$

where we embed each $\mathbb{R}^{n_j}$ into the first n_j coordinates of $\mathbb{R}^{n_j+1}$. Let

$$r = (\Pi_{n_1+1,k} \otimes \cdots \otimes \Pi_{n_m+1,k})(\sigma \cdot q) \in S^2(\mathbb{R}^{n_1+1}) \otimes \cdots \otimes S^2(\mathbb{R}^{n_m+1}),$$

where $\sigma \in S_{2m}$ is the permutation that sends $(1,2,\ldots,2m)$ to $(1, m+1, 2, m+2, \ldots, m, 2m)$. Direct calculation shows that $\|r\|_2 = 2^{-m/2}$. By Proposition 11 it holds that

$$2^{-m} \cdot \min_{v_j \in \mathcal{S}(\mathbb{R}^{n_j})} \langle p, v_1 \otimes \cdots \otimes v_m \rangle = \min_{v_j \in \mathcal{S}(\mathbb{R}^{n_j+1})} \langle r, v_1^{\otimes 2} \otimes \cdots \otimes v_m^{\otimes 2} \rangle. \tag{7}$$

Our hierarchy can be used to compute the righthand side of (7) as follows. Let $M(r) \in \mathrm{Hom}(\otimes_j \mathbb{R}^{n_j+1})$ be the tensor r after the isomorphism $\otimes_j S^2(\mathbb{R}^{n_j+1}) \cong \mathrm{Hom}(\otimes_j \mathbb{R}^{n_j+1})$. For each $k \in \mathbb{N}$, let

$$M_k = (\Pi_{n_1+1,k} \otimes \cdots \otimes \Pi_{n_m+1,k})(M(r) \otimes \mathbb{1}_{n_1+1}^{\otimes k-1} \otimes \cdots \otimes \mathbb{1}_{n_m+1}^{\otimes k-1})(\Pi_{n_1+1,k} \otimes \cdots \otimes \Pi_{n_m+1,k}).$$

Then by our Theorem 12, the minimum eigenvalues $\nu_k := \lambda_{\min}(M_k)$ converge to the righthand side of (7) from below at a rate of $O(1/k)$ (for similar reasons as in Footnote 11). More precisely, we have the following convergence guarantee:

Theorem 17 (Convergence guarantee for spectral norm)**.** *For each $k \in \mathbb{N}$ let $\mu_k = -2^m \lambda_{\min}(M_k)$. Then $\mu_k \geq \|p\|_{\sigma,\mathbb{R}}$ for all k, and*

$$\begin{aligned} \mu_k - \|p\|_{\sigma,\mathbb{R}} &\leq 2^m \cdot \|M(r)\|_\infty \left(1 + \kappa(M(s_{\mathbf{n},1^{(m)}}))\right) \frac{4m(\max_j n_j - 1)}{c_{(1^{(m)})}(k+1)} \\ &\leq \frac{2^{m/2+3} m(\max_j n_j - 1)}{k+1} \\ &= O\left(\frac{1}{k}\right), \end{aligned} \tag{8}$$

where $1^{(m)} = (1,\ldots,1)$ (m times). In particular, $\lim_{k\to\infty} \kappa_k = \|p\|_{\sigma,\mathbb{R}}$.

[12]Here we define the *Gram matrix* of a bihomogeneous form $r(x,y) \in \mathbb{R}[x,y]_{2k}$ as a matrix $M \in \mathrm{Hom}(S^k(\mathbb{R}^{n_1}) \otimes S^k(\mathbb{R}^{n_2}))$ for which $r(x,y) = \langle x^{\otimes k} \otimes y^{\otimes k}, M\, x^{\otimes k} \otimes y^{\otimes k} \rangle$.

Proof. For the first line in (8), note that

$$\begin{aligned}\mu_k - \|p\|_{\sigma,\mathbb{R}} &= -2^m \nu_k + \min_{v_j \in \mathcal{S}(\mathbb{R}^{n_j})} \langle p, v_1 \otimes \cdots \otimes v_m \rangle \\ &= -2^m \nu_k + 2^m \min_{v_j \in \mathcal{S}(\mathbb{R}^{n_j+1})} \langle r, v_1^{\otimes 2} \otimes \cdots \otimes v_m^{\otimes 2} \rangle \\ &= 2^m (r_{\min}^{(\mathbf{n},(1^{(m)}))} - \nu_k).\end{aligned}$$

So the first line follows from Theorem 12. The second line follows from $c_{(1^{(m)})} = 1$, $\kappa(M(s_{\mathbf{n},1^{(m)}})) = 1$, and $\|M(r)\|_\infty \leq \|r\|_2 = 2^{-m/2}$. This completes the proof. □

6 Constrained polynomial optimization

To obtain our hierarchy for computing the minimum value $p_{\min}$ of a real polynomial p over the real sphere, we showed that it was related to the minimum value $p^{\mathbb{C}}_{\min}$ of a certain minimization problem over the complex sphere (Proposition 6), and invoked the quantum de Finetti theorem for this complex problem (Theorem 7). In more details, we proved that $p^{\mathbb{C}}_{\min} \leq p_{\min} \leq \frac{1}{c_d} p^{\mathbb{C}}_{\min}$ for a constant c_d; in particular,

$$p_{\min} = 0 \iff p^{\mathbb{C}}_{\min} = 0. \tag{9}$$

In Section 5 we generalized these results to obtain hierarchies of eigencomputations for a certain class of constrained polynomial optimization problems: minimizing the inner product between a real tensor and an element of the real spherical Segre-Veronese variety.

In this section, we consider whether our hierarchy can be generalized to handle other polynomial constraints. We focus here on hierarchies of *eigencomputations* (see the end of this section for two natural choices of hierarchies which are not eigencomputations). We first note that the complex minimization hierarchy can indeed be generalized, but find an example where the real minimum is zero but the complex minimum is non-zero (i.e. the corresponding generalization of (9) does not hold). This indicates to us that a corresponding generalization of Proposition 6 seems non-trivial. We will require a bit of algebraic geometry in this section; we refer the reader to [Har13] for background.

Let us consider constrained optimization problems of the form

$$p_{q,\min} = \min_{\substack{x \in \mathcal{S}(\mathbb{R}^n) \\ q_1(x) = \cdots = q_\ell(x) = 0}} p(x), \tag{10}$$

where $p \in \mathbb{R}[x]_{2d}$ and $q_1, \ldots, q_\ell \in \mathbb{R}[x]$ are homogeneous polynomials.[13] Identifying the polynomials q_i with symmetric tensors, let $I = \langle q_1, \ldots, q_\ell \rangle \subseteq S(\mathbb{R}^n)$ be the ideal generated by $q_1, \ldots, q_\ell$ in the symmetric algebra $S(\mathbb{R}^n) := \bigoplus_{d \geq 0} S^d(\mathbb{R}^n)$. For each positive integer k, let $I_k \subseteq S^k(\mathbb{R}^n)$ be the degree-k part of I, and let $\Pi_{n,k,I} \in \mathrm{Hom}((\mathbb{R}^n)^{\otimes k})$ be the orthogonal projection onto the orthogonal complement of I_k. Note that $\Pi_{n,k,0} = \Pi_{n,k}$ is the orthogonal projection onto $S^k(\mathbb{R}^n)$.

For a field $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and subset $J \subseteq \mathbb{F}[x]$, we let

$$V_{\mathbb{F}}(J) = \{x \in \mathbb{F}^n : q(x) = 0 \quad \text{for all} \quad q \in J\}.$$

[13] One can replace the multiple equality constraints $q_1(x) = \cdots = q_\ell(x) = 0$ with the single constraint $q(x) := q_1(x)^2 + \cdots + q_\ell(x)^2 = 0$. However we will soon consider a complex analogue to this problem for which this cannot necessarily be done. We therefore stick with multiple equality constraints throughout to keep the notation uniform.

The sets $V_{\mathbb{F}}(J) \subseteq \mathbb{F}^n$ form the closed sets in the *Zariski topology* on $\mathbb{F}^n$. We denote the Zariski closure of a set $S \subseteq \mathbb{F}^n$ by $\mathrm{Cl}_{\mathbb{F}}(S)$. An $\mathbb{F}$-*variety* is a set of the form $V_{\mathbb{F}}(J)$ for some $J \subseteq \mathbb{F}[x]$.

The following result gives a hierarchy of minimum eigenvalue computations for solving a certain complex analogue to the constrained minimization problem (10).

Theorem 18. *Let $M \in \mathrm{Hom}(S^d(\mathbb{R}^n))$ be any Gram matrix of p for which $M = M^{\mathsf{T}}$. For each integer $k \geq d$, let $\nu_k = \lambda_{\min}(M_k)$, where*

$$M_k = \Pi_{n,k,I}(M \otimes \mathbb{1}_n^{\otimes k-d})\Pi_{n,k,I}.$$

Then $\nu_d \leq \nu_{d+1} \leq \nu_{d+2} \leq \dots$, and

$$\lim_{k\to\infty} \nu_k = \min_{\substack{v\in\mathcal{S}(\mathbb{C}^n)\\ q_1(v)=\dots=q_\ell(v)=0}} \mathrm{Tr}((vv^*)^{\otimes d} M).$$

In the unconstrained case $I = 0$, Theorem 18 follows from the quantum de Finetti theorem (Theorem 7). A proof of Theorem 18 in full generality will be presented in the forthcoming work [DJLV23].

Proposition 19. *Let*

$$\begin{aligned} p(x,y,z) &= xz \in \mathbb{R}[x,y,z]_2, \\ q(x,y,z) &= x^2 + y^2 - \frac{1}{\sqrt{3}}xz \in \mathbb{R}[x,y,z]_2. \end{aligned} \tag{11}$$

Then

$$p_{q,\min} = 0 > -\frac{1}{2\sqrt{3}} \geq \min_{\substack{v\in\mathcal{S}(\mathbb{C}^3)\\ q(v)=0}} \mathrm{Tr}(vv^* M(p)).$$

This proposition indicates to us that there is no obvious generalization of Proposition 6, because in particular no natural generalization of (9) seems to hold. The example provided by this proposition is non-trivial in two ways:

1. For the general constrained polynomial optimization problem (10) it always holds that

$$p_{q,\min} \geq \min_{\substack{v\in\mathcal{S}(\mathbb{C}^n)\\ q_1(v)=\dots=q_\ell(v)=0}} \mathrm{Tr}((vv^*)^{\otimes d} M)$$

   for any symmetric ($M = M^{\mathsf{T}}$) Gram matrix M, since $V_{\mathbb{C}}(q) \cap \mathbb{R}^n = V_{\mathbb{R}}(q)$. It is natural to ask if one could get equality in the above proposition by instead minimizing the righthand side over a smaller variety in $\mathbb{C}^n$ whose intersection with $\mathbb{R}^n$ still equals $V_{\mathbb{R}}(q)$ (optimizing over larger varieties can only make the inequality worse). This is not the case, since for q defined as in (11) it holds that $\mathrm{Cl}_{\mathbb{C}}(V_{\mathbb{R}}(q)) = V_{\mathbb{C}}(q)$ i.e. $V_{\mathbb{C}}(q)$ is the smallest $\mathbb{C}$-variety containing $V_{\mathbb{R}}(q)$. Indeed, it is easily checked that q is irreducible over $\mathbb{C}$ and that $V_{\mathbb{C}}(q)$ contains at least one real smooth point (the only singular point of $V_{\mathbb{C}}(q)$ is $(0,0,0)$). Hence $\mathrm{Cl}_{\mathbb{C}}(V_{\mathbb{R}}(q)) = V_{\mathbb{C}}(q)$ by [Man20, Theorem 2.2.9(2)].

2. It is also natural to ask if equality holds under a different choice of Gram matrix for p that satisfies $M = M^{\mathsf{T}}$ (so that Theorem 18 applies). This is not the case because when p is quadratic, $M(p)$ is the unique Gram matrix for p satisfying this property.

In short, there does not seem to be an obvious way to compute $p_{q,\min}$ by some natural adaptation of the complex problem.

Proof of Proposition 19. Let us first verify that $p_{q,\min} = 0$. It is clear that $p_{q,\min} \geq 0$ because any solution must satisfy $xz = \sqrt{3}(x^2 + y^2) \geq 0$. For the equality to zero, note that $q(0,0,1) = p(0,0,1) = 0$. For the last inequality, consider the point $v = (\frac{1}{\sqrt{6}}, \frac{i}{\sqrt{3}}, -\frac{1}{\sqrt{2}})^\intercal \in \mathcal{S}(\mathbb{C}^3)$, which is a zero of q. Note that

$$M(p) = \begin{bmatrix} 0 & 0 & 1/2 \\ 0 & 0 & 0 \\ 1/2 & 0 & 0 \end{bmatrix},$$

and $\mathrm{Tr}(vv^* M(p)) = -\frac{1}{2\sqrt{3}}$. This completes the proof. □

We conclude this section by noting that there are two other natural ways to adapt our hierarchy to handle constraints, neither of which produce hierarchies of eigencomputations: First, one can use the method described in [AH19] for adapting certificates of global positivity to perform constrained polynomial optimization, but each step of the resulting hierarchy would require the use of bisection. Second, one could simply replace the sum-of-squares polynomials that appear in standard positivstellensatze with polynomials that our hierarchy certifies to be globally non-negative at some level. For example, a special case of Putinar's positivestellensatz says that a polynomial $p(x)$ is strictly positive on $V_{\mathbb{R}}(q)$ if and only if there exists a sum-of-squares polynomial $\sigma(x)$ and a polynomial $h(x)$ for which $p(x) = \sigma(x) + h(x)q(x)$ [Put93]. Instead of searching for a sum-of-squares polynomial σ, one could impose that some fixed level of our hierarchy certifies global non-negativity of σ. This would give a sequence of semidefinite programs, indexed by the degrees of σ and h, and by the level of our hierarchy that we consider, which successfully finds such a pair (σ, h) in some high enough degrees and level if and only if $p(x)$ is strictly positive on $V_{\mathbb{R}}(q)$.

7 Acknowledgments

We thank Hamza Fawzi, Monique Laurent, Tim Netzer, Bruce Reznick, Jurij Volcic and Timo de Wolff for enlightening discussions. N.J. was supported by NSERC Discovery Grant RGPIN2022-04098. B.L. acknowledges that this material is based upon work supported by the National Science Foundation under Award No. DMS-2202782. A.V. was supported by the National Science Foundation under Grant Nos. CCF-1652491, CCF-1934931, ECCS-2216970. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[AH19] Amir Ali Ahmadi and Georgina Hall. On the construction of converging hierarchies for polynomial optimization based on certificates of global positivity. *Mathematics of Operations Research*, 44(4):1192–1207, 2019.

[AM19] Amir Ali Ahmadi and Anirudha Majumdar. DSOS and SDSOS optimization: More tractable alternatives to sum of squares and semidefinite optimization. *SIAM Journal on Applied Algebra and Geometry*, 3(2):193–230, 2019.

[BBH$^+$12] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326, 2012.

[BGG$^+$17] V. Bhattiprolu, M. Ghosh, V. Guruswami, E. Lee, and M. Tulsiani. Weak decoupling, polynomial folds and approximate optimization over the sphere. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1008–1019, 2017.

[BV09] S. Charles Brubaker and Santosh S. Vempala. Random tensors and planted cliques. In *Proceedings of the 12th International Workshop and 13th International Workshop on Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, APPROX '09 / RANDOM '09, pages 406–419, Berlin, Heidelberg, 2009. Springer-Verlag.

[CKMR07] Matthias Christandl, Robert König, Graeme Mitchison, and Renato Renner. One-and-a-half quantum de finetti theorems. *Communications in mathematical physics*, 273(2):473–498, 2007.

[DJLV23] Harm Derksen, Nathaniel Johnston, Benjamin Lovitz, and Aravindan Vijayaraghavan. Beyond separability: a generalization of the symmetric extensions hierarchy. *In preparation*, 2023+.

[DK08] Etienne De Klerk. The complexity of optimizing over a simplex, hypercube or sphere: a short survey. *Central European Journal of Operations Research*, 16:111–125, 2008.

[dKLP05] Etienne de Klerk, Monique Laurent, and Pablo Parrilo. On the equivalence of algebraic approaches to the minimization of forms on the simplex. *Positive Polynomials in Control*, pages 121–132, 2005.

[DPS04] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri. A complete family of separability criteria. *Physical Review A*, 69:022308, 2004.

[DW12] Andrew C Doherty and Stephanie Wehner. Convergence of sdp hierarchies for polynomial optimization on the hypersphere. *arXiv preprint arXiv:1210.5048*, 2012.

[FF21] Kun Fang and Hamza Fawzi. The sum-of-squares hierarchy on the sphere and applications in quantum information theory. *Mathematical Programming*, 190:331–360, 2021.

[FK08] Alan M. Frieze and Ravi Kannan. A new approach to the planted clique problem. In *Foundations of Software Technology and Theoretical Computer Science*, 2008.

[Har13] Joe Harris. *Algebraic Geometry: A First Course*. Graduate Texts in Mathematics. Springer New York, 2013.

[HSSS16] Samuel B. Hopkins, Tselil Schramm, Jonathan Shi, and David Steurer. Fast spectral algorithms from sum-of-squares proofs: Tensor decomposition and planted sparse vectors. In *ACM Symposium on Theory of Computing (STOC)*, 2016.

[Joh16] Nathaniel Johnston. QETLAB: A MATLAB toolbox for quantum entanglement. https://qetlab.com, 2016. See the "PolynomialOptimize" function in particular: https://qetlab.com/PolynomialOptimize.

[KM09] Robert Koenig and Graeme Mitchison. A most compendious and facile quantum de finetti theorem. *Journal of Mathematical Physics*, 50(1):012105, 2009.

[Lau09] Monique Laurent. *Sums of Squares, Moment Matrices and Optimization Over Polynomials*, pages 157–270. Springer New York, 2009.

[Lau19] Monique Laurent. Notes on the paper: "Convergence of SDP hierarchies for polynomial optimization on the hypersphere", by A.C Doherty and S. Wehner. https://homepages.cwi.nl/ monique/Doherty-Wehner-notesv2.pdf, 2019.

[LSY98] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK Users' Guide*. Society for Industrial and Applied Mathematics, 1998.

[Man20] Frédéric Mangolte. *Real algebraic varieties*. Springer, 2020.

[Mar08] Murray Marshall. *Positive polynomials and sums of squares*. Number 146. American Mathematical Society, 2008.

[N^+03] Jurij Evgenevič Nesterov et al. Random walk in a simplex and quadratic optimization over convex polytopes. Technical report, Center for Operations Research and Econometrics, 2003.

[Net16] Tim Netzer. Real algebraic geometry and its applications. *arXiv preprint arXiv:1606.07284*, 2016.

[Par03] Pablo A Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical programming*, 96:293–320, 2003.

[Put93] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993.

[Rez95] Bruce Reznick. Uniform denominators in Hilbert's seventeenth problem. *Mathematische Zeitschrift*, 220(1):75–97, 1995.

[Rez13] Bruce Reznick. On the length of binary forms. In *Quadratic and higher degree forms*, pages 207–232. Springer, 2013.

[RM14] Emile Richard and Andrea Montanari. A statistical model for tensor PCA. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[SS17] Tselil Schramm and David Steurer. Fast and robust tensor decomposition with applications to dictionary learning. In *Conference on Learning Theory*, pages 1760–1793, 2017.

[Ste02] G. W. Stewart. A Krylov–Schur algorithm for large eigenproblems. *SIAM Journal on Matrix Analysis and Applications*, 23(3):601–614, 2002.

[Wat18] J. Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.

A Formula for $M_k(p)$

In this section we derive an explicit formula for the coordinates of $M_k(p) = \Pi_{n,k}(M(p) \otimes \mathbb{1}_n^{\otimes k-d})\Pi_{n,k}$. We make the following definitions.

1. For positive integers n and d, let $[n]^{\vee d} = \{(i_1, \ldots, i_d) \in [n]^{\times d} : 1 \leq i_1 \leq \cdots \leq i_d \leq n\}$ be the non-decreasing d-tuples of integers from 1 to n.
2. For $I \in [n]^{\vee d}$, $j \in [n]$, let r_j^I be the number of times j appears in I.
3. For $I \in [n]^{\vee d}$, let $f(I) = (d! r_1^I! \cdots r_n^I!)^{1/2}$.
4. For $I \in [n]^{\vee d}$, let $m(I) = \frac{d!}{r_1^I! \cdots r_n^I!}$ be the multinomial coefficient.
5. For non-negative integers a and b, let $a^{\underline{b}} = \prod_{i=0}^{b-1}(a-i)$ be the falling factorial.
6. For a positive integer $k \geq d$ and tuples $I \in [n]^{\vee k}$, $H \in [n]^{\vee(k-d)}$, let $g(I,H) = (r_1^I)^{\underline{r_1^H}} \cdots (r_n^I)^{\underline{r_n^H}}$.
7. For $I, J \in [n]^{\times k}$, we say $I \sim J$ if $I = J$ up to reordering indices. We extend all of the above definitions to tuples in $[n]^{\times k}$ by reordering indices. Let $\delta_{I,J}$ equal 1 if $I \sim J$ and zero otherwise.

We will use the orthonormal basis $\{e_I : I \in [n]^{\vee d}\}$ for $S^d(\mathbb{R}^n)$, where

$$e_I := \frac{1}{f(I)} \sum_{\sigma \in S_d} e_{i_{\sigma(1)}} \otimes \cdots \otimes e_{i_{\sigma(d)}}.$$

For each $I \in [n]^{\vee d}$, let $f_I = f(I) e_I$.

Proposition 20. *For $I, L \in [n]^{\vee k}$, let $M_k(p)_{I,L}$ be the (I,L)-coordinate of $M_k(p)$ in the basis $\{e_I : I \in [n]^{\vee k}\}$. It holds that*

$$M_k(p)_{I,L} = (d!)^2 f(I)^{-1} f(L)^{-1} \sum_{J,K \in [n]^{\vee d}} \alpha_{(J,K)} m((J,K))^{-1} \sum_{H \in [n]^{\vee k-d}} \delta_{I,(J,H)} \delta_{L,(K,H)} m(H) g(I,H) g(L,H).$$

In particular, for $J, K \in [n]^{\vee d}$ we have

$$M(p)_{J,K} = \alpha_{(J,K)}\ (d!)^2\ m((J,K))^{-1}\ f(J)^{-1}\ f(K)^{-1}.$$

Proof. First let's write down the coordinates of $M(p) = M_d(p)$. If

$$p = \sum_{I \in [n]^{\vee 2d}} \alpha_I x_{i_1} \cdots x_{i_{2d}},$$

then after the isomorphism $\mathbb{R}[x]_{2d} \cong S^{2d}(\mathbb{R}^n)$ we obtain

$$\vec{p} = \frac{1}{(2d)!} \sum_{I \in [n]^{\vee 2d}} \alpha_I f_I.$$

Note that

$$\begin{aligned}
f_I &= (\Pi_{n,d} \otimes \Pi_{n,d}) f_I \\
&= (\Pi_{n,d} \otimes \Pi_{n,d}) \sum_{\sigma \in S_{2d}} e_{i_{\sigma(1)}} \otimes \cdots \otimes e_{i_{\sigma(2d)}} \\
&= \frac{1}{(d!)^2} \sum_{\sigma \in S_{2d}} f_{\sigma(I)_{[d]}} \otimes f_{\sigma(I)_{[d+1..2d]}} \\
&= \frac{1}{(d!)^2} \sum_{\substack{J,K \in [n]^{\vee d} \\ (J,K) \sim I}} \mu_{I,J} f_J \otimes f_K \\
&= \sum_{\substack{J,K \in [n]^{\vee d} \\ (J,K) \sim I}} \binom{r_1^I}{r_1^J} \cdots \binom{r_n^I}{r_n^J} f_J \otimes f_K
\end{aligned}$$

where $\sigma(I)_{[d]}$ is the first d elements of $\sigma(I)$; $\sigma(I)_{[d+1..2d]}$ is the last d elements of $\sigma(I)$; and $\mu_{I,J}$ is the number of permutations $\sigma \in S_{2d}$ for which $\sigma(I)_{[d]} \sim J$. The last line follows from the fact that

$$\mu_{I,J} = (d!)^2 \binom{r_1^I}{r_1^J} \cdots \binom{r_n^I}{r_n^J}.$$

This formula holds because, for each j, σ can take r_j^J elements from r_j^I to $[d]$, and permute the resulting d elements arbitrarily, and also permute the leftover d elements arbitrarily. This gives

$$\begin{aligned}
\vec{p} &= \frac{1}{(2d)!} \sum_{I \in [n]^{\vee 2d}} \alpha_I \sum_{\substack{J,K \in [n]^{\vee d} \\ (J,K) \sim I}} \binom{r_1^I}{r_1^J} \cdots \binom{r_n^I}{r_n^J} f_J \otimes f_K \\
&= \frac{1}{(2d)!} \sum_{I \in [n]^{\vee 2d}} \alpha_I \sum_{\substack{J,K \in [n]^{\vee d} \\ (J,K) \sim I}} f(J) f(K) \binom{r_1^I}{r_1^J} \cdots \binom{r_n^I}{r_n^J} f_J \otimes f_K
\end{aligned}$$

Thus,

$$\begin{aligned}
M(p)_{J,K} &= \alpha_{(J,K)} \frac{1}{(2d)!} f(J) f(K) \binom{r_1^I}{r_1^J} \cdots \binom{r_n^I}{r_n^J} \\
&= \alpha_{(J,K)}\ (d!)^2\ m((J,K))^{-1}\ f(J)^{-1}\ f(K)^{-1}.
\end{aligned}$$

Now let's compute the coordinates of $M_k(p)$. First note that for $I, L \in [n]^{\vee k}$, $J, K \in [n]^{\vee d}$, and $H \in [n]^{\vee k-d}$, we have

$$e_I^*(e_J \otimes e_{h_1} \otimes \cdots \otimes e_{h_{k-d}}) = \delta_{I,(J,H)} f(I)^{-1} f(J) g(I,H),$$

and

$$\begin{aligned}
e_I^*(e_J e_K^* \otimes \mathbb{1}_n^{\otimes k-d}) e_L &= \sum_{H \in [n]^{\vee k-d}} \delta_{I,(J,H)} \delta_{L,(K,H)}\ m(H)\ e_I^*(e_J \otimes e_{h_1} \otimes \cdots \otimes e_{h_{k-d}})(e_K^* \otimes e_{h_1}^* \otimes \cdots \otimes e_{h_{k-d}}^*) e_L \\
&= f(I)^{-1} f(J) f(K) f(L)^{-1} \sum_{H \in [n]^{\vee k-d}} \delta_{I,(J,H)} \delta_{L,(K,H)} m(H) g(I,H) g(L,H).
\end{aligned}$$

Thus,

$$\begin{aligned} M_k(p)_{I,L} &= e_I^* \left(\sum_{J,K \in [n]^{\vee d}} M(p)_{(J,K)} e_J e_K^* \otimes \mathbb{1}_n^{\otimes k-d} \right) e_L \\ &= (d!)^2 f(I)^{-1} f(L)^{-1} \sum_{J,K \in [n]^{\vee d}} \alpha_{(J,K)} m((J,K))^{-1} \sum_{H \in [n]^{\vee k-d}} \delta_{I,(J,H)} \delta_{L,(K,H)} m(H) g(I,H) g(L,H). \end{aligned}$$

This completes the proof. □