Mathematical Programming
https://doi.org/10.1007/s10107-020-01537-7

FULL LENGTH PAPER

Series A

The sum-of-squares hierarchy on the sphere and applications in quantum information theory

Kun Fang[1] · **Hamza Fawzi**[1]

Received: 27 August 2019 / Accepted: 22 June 2020

Abstract

We consider the problem of maximizing a homogeneous polynomial on the unit sphere and its hierarchy of sum-of-squares relaxations. Exploiting the *polynomial kernel technique*, we obtain a quadratic improvement of the known convergence rate by Reznick and Doherty and Wehner. Specifically, we show that the rate of convergence is no worse than $O(d^2/\ell^2)$ in the regime $\ell = \Omega(d)$ where ℓ is the level of the hierarchy and d the dimension, solving a problem left open in the recent paper by de Klerk and Laurent (arXiv:1904.08828). Importantly, our analysis also works for matrix-valued polynomials on the sphere which has applications in quantum information for the Best Separable State problem. By exploiting the duality relation between sums of squares and the Doherty–Parrilo–Spedalieri hierarchy in quantum information theory, we show that our result generalizes to nonquadratic polynomials the convergence rates of Navascués, Owari and Plenio.

Mathematics Subject Classification 90C22 · 90C23 · 81P42

1 Introduction

We consider in this paper a fundamental computational task, that of maximizing a multivariate polynomial $p \in \mathbb{R}[x]$ in d variables $x = (x_1, \ldots, x_d)$ on the unit sphere:

$$p_{\max} = \max_{x \in S^{d-1}} p(x) \tag{1}$$

where $S^{d-1} = \{x \in \mathbb{R}^d : x_1^2 + \cdots + x_d^2 = 1\}$. Optimization problems of the above form have applications in many areas. For example, computing the largest stable set of

✉ Kun Fang
kf383@cam.ac.uk

Hamza Fawzi
h.fawzi@damtp.cam.ac.uk

[1] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK

Published online: 06 July 2020

a graph is a special case of (1) for a suitable polynomial p of degree three, see [13,27]. Computing the $2 \to 4$ norm of a matrix A corresponds to the maximization of the degree-four polynomial $p(x) = \|Ax\|_4^4$ on the sphere, see e.g., [4] for more on this. In quantum information, the so-called *Best Separable State problem* very naturally relates to polynomial optimization on the sphere, as we explain later.

When $p(x)$ is quadratic, problem (1) reduces to an eigenvalue problem which can be solved efficiently. However for general polynomials of degree greater than two, the problem is NP-hard as it contains as a special case the stable set problem [27]. The sum-of-squares hierarchy is a hierarchy of semidefinite relaxations that approximate the value $p_{\max}$ by a sequence of semidefinite programs of increasing size [25,31]. In this paper we study the approximation quality of this sequence of semidefinite relaxations.

1.1 Sum-of-squares hierarchy

The sum-of-squares hierarchy for (1) that we study in this paper is defined by

$$p_\ell = \min \left\{ \gamma \in \mathbb{R} \text{ s.t. } \gamma - p \text{ is sum-of-squares of degree } \ell \text{ on } S^{d-1} \right\}. \tag{2}$$

The sequence $(p_\ell)_{\ell \in \mathbb{N}}$ consists of monotone upper bounds on $p_{\max}$, i.e., for any ℓ we have $p_{\max} \leq p_\ell$ and $p_\ell \leq p_{\ell-1}$. For each ℓ, the value p_ℓ can be computed by a semidefinite program of size $d^{O(\ell)}$, see e.g., [25,31]. As explained in "Appendix A", (2) coincides with the usual sum-of-squares hierarchy defined in terms of Putinar and Schmüdgen-type Positivstellensatz.

A result of Reznick [34] (see also [19]) shows that $p_\ell \to p_{\max}$ as $\ell \to \infty$. In fact Reznick shows, assuming $p_{\min} = \min_{x \in S^{d-1}} p(x) = 0$, that $p_\ell / p_{\max}$ converges to 1 at the rate d/ℓ, for ℓ large enough. In this paper we show that the sum-of-squares hierarchy actually converges at the faster rate of $(d/\ell)^2$. More precisely, we prove the following

Theorem 1 *Assume $p(x_1, \ldots, x_d)$ is a homogeneous polynomial of degree $2n$ in d variables with $n \leq d$, and let $p_{\min}$ denote the minimum of p on S^{d-1}. Then for any $\ell \geq C_n d$*

$$1 \leq \frac{p_\ell - p_{\min}}{p_{\max} - p_{\min}} \leq 1 + (C_n d/\ell)^2 \tag{3}$$

for some constant C_n that depends only on n.

In a recent paper, de Klerk and Laurent [14] proved that a semidefinite hierarchy of *lower bounds* on $p_{\max}$ converges at a rate of $O(1/\ell^2)$ and left open the question of whether the same is true for the hierarchy (p_ℓ) of upper bounds. Our Theorem 1 answers this question positively.

1.2 Matrix-valued polynomials

The proof technique we use in this paper actually allows us to get a significant generalization of Theorem 1, related to matrix-valued polynomials. Let $\mathbf{S}^k$ be the space of real

symmetric matrices of size $k \times k$, and let $\mathbf{S}^k[x]$ be the space of $\mathbf{S}^k$-valued polynomials in $x = (x_1, \ldots, x_d)$. We will often use the lighter notation $F \in \mathbf{S}[x]$ when the size k is unimportant for the discussion. A polynomial $F(x_1, \ldots, x_d) \in \mathbf{S}[x]$ is *positive* if $F(x) \geq 0$ for all $x \in \mathbb{R}^d$ where the inequality is interpreted in the positive semidefinite sense[1]. We say that $F(x) \in \mathbf{S}^k[x]$ is a *sum of squares* if there exist polynomials $U_j(x) \in \mathbb{R}^{k \times k}[x]$ such that $F(x) = \sum_j U_j(x) U_j(x)^\mathsf{T}$ for all $x \in \mathbb{R}^d$. We say that $F(x)$ is *ℓ-sos on S^{d-1}* if it agrees with a sum-of-squares polynomial on the sphere with $\deg U_j \leq \ell$. We are now ready to state our main theorem on sum of squares representations for matrix-valued polynomials.

Theorem 2 *Assume $F(x_1, \ldots, x_d) \in \mathbf{S}[x]$ is a homogeneous matrix-valued polynomial of degree $2n$ in d variables with $n \leq d$, such that $F(x)$ is symmetric for all x. Assume furthermore that $0 \leq F(x) \leq I$ for all $x \in S^{d-1}$, where I is the identity matrix. There are constants C_n and C'_n that depend only on n such that for any $\ell \geq C_n d$, $F + C'_n \left(\frac{d}{\ell}\right)^2 I$ is ℓ-sos on S^{d-1}.*

Some remarks concerning the statement are in order:

- Theorem 1 is a direct corollary of Theorem 2 where $F(x)$ is the scalar polynomial given by $F(x) = (p_{\max} - p)/(p_{\max} - p_{\min})$.
- A remarkable fact of Theorem 2 is that the result is totally independent on the size of the matrix $F(x)$.
- Theorem 2 can be applied to get sum-of-squares certificates for scalar *bihomogeneous* polynomials on products of two spheres $S^{k-1} \times S^{d-1}$. Indeed, one way to think about a matrix-valued polynomial $F(x_1, \ldots, x_d) \in \mathbf{S}^k[x]$ is to consider the real-valued polynomial $p(x, y) = y^\mathsf{T} F(x) y$ where $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^k$. This polynomial is bihomogeneous of degree $(2n, 2)$ in the variables (x, y). One important application of this setting is in quantum information theory for the best separable state problem which we explain later in Sect. 1.3 and more details can be found in Sect. 4.
- As stated, Theorem 2 is concerned only with levels $\ell \geq \Omega(d)$ of the sum-of-squares hierarchy. The main technical result we prove in this paper (Theorem 6 below) actually allows us to get a bound on the performance of the sum-of-squares hierarchy for *all* values of level ℓ, and not just the regime $\ell \geq \Omega(d)$. The bounds we get however do not have closed-form expressions in general, and they depend on the eigenvalues of some generalized Toeplitz matrices. For small values of n (namely $2n = 2$ and $2n = 4$) our bounds can be computed efficiently though, as we explain later in Eqs. (17) and (18).
- For more details about the regime $\ell = o(d)$ of the sum-of-squares hierarchy, we refer the reader to the recent works [6,8] and references therein.

1.3 The Best Separable State problem in quantum information theory

The notion of entanglement plays a fundamental role in quantum mechanics. The set of separable states (i.e., non-entangled states) on the Hilbert space $\mathbb{C}^d \otimes \mathbb{C}^d$ is defined

[1] In the following discussion, the relation $X \geq Y$ for any two matrices is always interpreted as the positive semidefinite order, i.e., $X - Y$ is a positive semidefinite matrix.

as the convex hull of all pure product states

$$\mathrm{Sep}(d) = \mathrm{conv}\left\{xx^\dagger \otimes yy^\dagger : (x, y) \in \mathbb{C}^d \times \mathbb{C}^d \text{ and } \|x\| = \|y\| = 1\right\}. \quad (4)$$

Here $x^\dagger = \bar{x}^\mathsf{T}$ is the conjugate transpose and $\|x\|^2 = x^\dagger x = \sum_{i=1}^d |x_i|^2$. $\mathrm{Sep}(d)$ is a convex subset of the set $\mathrm{Herm}(d^2)$ of Hermitian matrices of size $d^2 \times d^2$. A key computational task in quantum information theory is the so-called *Best Separable State* (BSS) problem: given $M \in \mathrm{Herm}(d^2)$, compute

$$h_{\mathrm{Sep}}(M) = \max_{\rho \in \mathrm{Sep}(d)} \mathrm{Tr}[M\rho] = \max_{\substack{x,y \in \mathbb{C}^d \\ \|x\|=\|y\|=1}} \sum_{1 \le i,j,k,l \le d} M_{ij,kl} x_i \bar{x}_k y_j \bar{y}_l. \quad (5)$$

In words, $h_{\mathrm{Sep}}(M)$ is the *support function* of the convex set $\mathrm{Sep}(d)$ evaluated at M. Note that $h_{\mathrm{Sep}}(M)$ is simply the maximum of the *Hermitian polynomial*[2]

$$p_M(x, \bar{x}, y, \bar{y}) := \sum_{1 \le i,j,k,l \le d} M_{ij,kl} x_i \bar{x}_k y_j \bar{y}_l \quad (6)$$

over the product of spheres $S_{\mathbb{C}^d} \times S_{\mathbb{C}^d} = \{(x, y) \in \mathbb{C}^d \times \mathbb{C}^d : \|x\| = \|y\| = 1\}$. In that sense the BSS problem is very related to the polynomial optimization problem (1).

The Doherty–Parrilo–Spedalieri (DPS) hierarchy [18] is a hierarchy of semidefinite relaxations to the set of separable states, which is defined in terms of so-called *state extensions* (we recall the precise definitions later in the paper). It satisfies

$$\mathrm{Sep}(d) \subseteq \cdots \subseteq \mathrm{DPS}_\ell(d) \subseteq \cdots \subseteq \mathrm{DPS}_2(d) \subseteq \mathrm{DPS}_1(d)$$

where $\mathrm{DPS}_\ell(d)$ is the ℓ'th level of the DPS hierarchy. It turns out that the DPS hierarchy can be interpreted, from the dual point of view, as a sum of squares hierarchy. This duality relation has been mentioned multiple times in the literature, however we could not find any formal and complete proof of this equivalence. In this paper we give a proof of this duality relation. To do this, we first need to specify the definition of *sum of squares* for Hermitian polynomials. We say that a Hermitian polynomial is a *real sum of squares (rsos)* if it can be written as a sum of squares of Hermitian polynomials.[3] To state the result it is more convenient to work in the conic setting and we denote the convex cones associated to Sep and DPS_k by $\mathcal{SEP}$ and $\mathcal{DPS}_k$ respectively (these convex cones simply correspond to dropping a trace normalization condition).

[2] A Hermitian polynomial is a polynomial of complex variables and their conjugates that takes only real values. See Sect. 4.2 for more details.

[3] Another common definition is to require that the polynomial is a sum of squares of modulus squares of (holomorphic) complex polynomials. This is a different condition, and it corresponds from the dual point of view to the DPS hierarchy without the Positive Partial Transpose conditions. See Sect. 4.2 for more details on this.

Theorem 3 (Duality DPS/sum-of-squares) *Let $\mathcal{SEP}(d)$ be the convex cone of separable states on $\mathbb{C}^d \otimes \mathbb{C}^d$, and let $\mathcal{DPS}_\ell(d)$ be the convex cone of quantum states corresponding to the ℓ'th level of the DPS hierarchy. Then we have:*

(i) $\mathcal{SEP}(d)^* = \left\{M \in Herm(d^2) : p_M \text{ is nonnegative}\right\}$
(ii) $\mathcal{DPS}_\ell(d)^* = \left\{M \in Herm(d^2) : \|y\|^{2(\ell-1)} p_M \text{ is a real sum-of-squares}\right\}$,

where K^* denotes the dual cone to K and p_M is the Hermitian polynomial of Eq. (6).

Using this connection, our results on the convergence of the sum-of-squares hierarchy can be easily translated to bound the convergence rate of the DPS hierarchy. More precisely, since the polynomial p_M of Eq. (6) is bihomogeneous of degree (2, 2) (i.e., it is quadratic in x and y independently) we can get a bound on the rate of convergence of the DPS hierarchy from Theorem 2 where $\deg F = 2$. The rate of convergence we get in this way actually coincides with the rate of convergence obtained by Navascues, Owari and Plenio [28], who use a completely different (quantum-motivated) argument based on the primal definition of the DPS hierarchy using state extensions. From the sum-of-squares point of view, the theorem of Navascues et al. can thus be seen as a special case of Theorem 2 when $\deg F = 2$. We conclude by stating the result on the convergence rate of the DPS hierarchy.

Theorem 4 (Convergence rate of DPS hierarchy, see also [28]) *Let $M \in Herm(d^2)$ and assume that $(x \otimes y)^\dagger M (x \otimes y) \geq 0$ for all $(x, y) \in \mathbb{C}^d \times \mathbb{C}^d$. Then*

$$h_{Sep}(M) \leq h_{DPS_\ell}(M) \leq (1 + Cd^2/\ell^2) h_{Sep}(M) \tag{7}$$

for any $\ell \geq C'd$, where $C, C' > 0$ are absolute constants.

Remark 1 (Multiplicative vs. additive approximations) The guarantee of Eq. (7) for $h_{\mathrm{Sep}}(M)$ is *multiplicative*. In the quantum literature however, most results produce guarantees on $h_{\mathrm{Sep}}(M)$ that are *additive*, assuming the spectral norm of M is at most 1 [7,22], i.e., those take the form

$$h_{\mathrm{Sep}}(M) \leq h_{\mathrm{DPS}_\ell}(M) \leq h_{\mathrm{Sep}}(M) + \epsilon \|M\| \tag{8}$$

where $\|M\|$ is the spectral norm and ϵ depends on ℓ and d. We note that multiplicative guarantees are stronger because $h_{\mathrm{Sep}}(M) \leq \|M\|$.

One important tool used to get additive guarantees like (8) are *de Finetti* results [7,11]. One notable result on additive approximations of $h_{\mathrm{Sep}}(M)$ is that if M has the *1-LOCC structure*, i.e., $M = \sum_{i=1}^k X_i \otimes Y_i$ where $0 \leq X_i, \sum_{i=1}^k X_i \leq I$ and $0 \leq Y_i \leq I$, then after at most $\ell = O(\log d/\epsilon^2)$ we get $h_{\mathrm{DPS}_\ell}(M) \leq h_{\mathrm{Sep}}(M) + \epsilon$ [5].

1.4 Overview of proof

We give a brief overview of the proof of Theorem 2. We will focus on the case where $F(x)$ is a scalar-valued polynomial for simplicity of exposition.

Given a univariate polynomial $q(t)$ of degree ℓ consider the *kernel* $K(x, y) = q(\langle x, y\rangle)^2$ for $(x, y) \in S^{d-1} \times S^{d-1}$. Define the integral transform, for $h : S^{d-1} \to \mathbb{R}$

$$(Kh)(x) = \int_{y \in S^{d-1}} K(x, y)h(y)d\sigma(y) \qquad \forall x \in S^{d-1} \tag{9}$$

where $d\sigma$ is the rotation-invariant probability measure on S^{d-1}. As such, we think of K as a linear map acting on functions on S^{d-1}. A crucial property of K is that if $h \geq 0$ then the function Kh is ℓ-sos[4] on S^{d-1}, by construction of $K(x, y)$.

Let $F(x)$ be a scalar-valued polynomial such that $0 \leq F(x) \leq 1$ on S^{d-1}. Our goal is to find $\delta > 0$ such that $\tilde{F} = F + \delta = F + \delta\|x\|_2^{2n}$ is ℓ-sos on S^{d-1}. It can be shown (details later) that the linear map K defined by (9) is invertible on the space of homogeneous polynomials on S^{d-1}. As such we can always write $\tilde{F} = Kh$ with $h = K^{-1}\tilde{F}$ on S^{d-1}. If K is close to the identity (i.e., the kernel $K(x, y)$ is close to a Dirac kernel $\delta(x, y)$) then we expect that $h \approx \tilde{F}$, i.e., that $\|h - \tilde{F}\|_\infty$ is small. Since $\tilde{F} \geq \delta$, if we can guarantee that $\|h - \tilde{F}\|_\infty \leq \delta$ it would follow that $h \geq 0$, in which case the equation $\tilde{F} = Kh = K(K^{-1}\tilde{F})$ gives a degree-ℓ sum-of-squares representation of $\tilde{F}$ on the sphere.

To make the argument above precise we need to measure how close the kernel K is to the identity. This is best done in the Fourier domain, where we analyze how close the Fourier coefficients of the the kernel $K(x, y)$ are to 1. The Fourier coefficients of $K(x, y)$ depend in a quadratic way on the coefficients in the expansion of $q(t)$ in the basis of Gegenbauer polynomials. We show that there is a choice of $q(t)$ such that the Fourier coefficients of $K(x, y)$ converge to 1 at the rate $\frac{d^2}{\ell^2}$, as $\ell \to \infty$. The kernel we construct is the solution of an eigenvalue maximization for a generalized Toeplitz matrix, associated to the family of Gegenbauer polynomials. We use known results on the roots of such polynomials to obtain the desired rate of convergence.

The idea of proof here is similar to the approaches in Reznick [34], and Doherty and Wehner [19], and Parrilo [32]. The work of Reznick and Doherty and Wehner use the kernel $K(x, y) = \langle x, y\rangle^{2\ell}/c$ for some normalizing constant c for which the Fourier coefficients can be computed explicitly.[5] The Fourier coefficients of this kernel happen to converge to 1 at a rate of $\frac{d}{\ell}$, which is slower than the kernels we construct.

Organization

In Sect. 2 we review some background material concerning Fourier decompositions on the sphere. The proof of Theorem 2 is in Sect. 3. Section 4 is devoted to the Best Separable State problem in quantum information theory.

[4] Because of the integral, Kh is an "infinite" sum of squares. Standard convexity results can be used however to turn this into a finite sum of squares. We provide more details later.

[5] The fact that Reznick's proof is based on this choice of kernel was observed by Blekherman in [9, Remark 7.3].

2 Background

Spherical harmonics We review the basics of Fourier analysis on the sphere S^{d-1}. Any polynomial p of degree n on the sphere has a unique decomposition

$$p = p_0 + p_1 + \cdots + p_n, \qquad p_i \in \mathcal{H}_i^d \tag{10}$$

where each p_i is a *spherical harmonic of degree* i. The decomposition (10) is known as the Fourier-Laplace decomposition of p. The space $\mathcal{H}_i^d$ is defined as the restriction on S^{d-1} of the set of homogeneous harmonic polynomials of degree i, i.e.,

$$\mathcal{H}_i^d = \left\{ f|_{S^{d-1}} : f \in \mathbb{R}[x_1, \ldots, x_d], \text{ homogeneous of degree } i \text{ and } \Delta f = \sum_{k=1}^{d} \frac{\partial^2 f}{\partial x_k^2} = 0 \right\}.$$

Equivalently, the spaces $\mathcal{H}_i^d$ are also the irreducible subspaces of $L^2(S^{d-1})$ under the action of $SO(d)$. For example $\mathcal{H}_0^d$ is the set of constant functions, $\mathcal{H}_1^d$ is the set of linear functions, and $\mathcal{H}_2^d$ is the set of traceless quadratic forms. The spaces $\mathcal{H}_i^d$ are mutually orthogonal with respect to the L^2 inner product $\langle f, g \rangle = \int f g d\sigma$ where $d\sigma$ is the rotation-invariant probability measure on the sphere. Note that if p is an *even* polynomial (i.e., $p(x) = p(-x)$) then the only nonzero harmonic components of p are the ones of even order.

Integral transforms Consider a general $SO(n)$-invariant kernel $K(x, y) = \phi(\langle x, y \rangle)$ where ϕ is some univariate polynomial of degree L. The kernel K acts on functions $f : S^{d-1} \to \mathbb{R}$ as follows

$$(Kf)(x) = \int_{S^{d-1}} K(x, y) f(y) d\sigma(y) \qquad \forall x \in S^{d-1}.$$

To understand the action of K on arbitrary polynomials f, it is very convenient to decompose ϕ into the basis of *Gegenbauer polynomials* (also known as *ultraspherical polynomials*) $(C_k(t))_{k \in \mathbb{N}}$ which are orthogonal polynomials on $[-1, 1]$ with respect to the weight $(1 - t^2)^{\frac{d-3}{2}} dt$. Using appropriate normalization (which we adopt here) these polynomials satisfy the following important property:

$$\int_{S^{d-1}} C_k(\langle x, y \rangle) p_i(y) d\sigma(y) = \delta_{ik} p_i(x) \qquad \forall x \in S^{d-1}$$

for any $p_i \in \mathcal{H}_i^d$. In other words, the kernel $(x, y) \mapsto C_k(\langle x, y \rangle)$ is a *reproducing kernel* for $\mathcal{H}_k^d$. Now going back to the kernel $K(x, y) = \phi(\langle x, y \rangle)$, if we expand $\phi = \lambda_0 C_0 + \lambda_1 C_1 + \cdots + \lambda_L C_L$, then it follows that for any polynomial p with Fourier expansion (10) we have

$$(Kp)(x) = \int_{y \in S^{d-1}} K(x, y)p(y)d\sigma(y) = \lambda_0 p_0(x) + \lambda_1 p_1(x) + \cdots + \lambda_L p_L(x) \quad \forall x \in S^{d-1}. \tag{11}$$

The equation above tells us that the harmonic decomposition $\mathcal{H}_0 \oplus \mathcal{H}_1 \oplus \ldots$ diagonalizes K, with the Gegenbauer coefficients $(\lambda_i)_{i=0,\ldots,L}$ being the eigenvalues. Equation (11) is also known as the *Funk-Hecke formula*. The coefficients $(\lambda_i)_{i=1,\ldots,L}$ in the expansion of ϕ in the basis of Gegenbauer polynomials are given by the following integral

$$\lambda_i = \frac{\omega_{d-1}}{\omega_d} \int_{-1}^{1} \phi(t) \frac{C_i(t)}{C_i(1)} (1 - t^2)^{\frac{d-3}{2}} dt \tag{12}$$

where ω_d is the surface area of S^{d-1}. If we let $w(t) = (1 - t^2)^{\frac{d-3}{2}}$, one can check that $\int_{-1}^{1} C_i(t)^2 w(t)dt = \frac{\omega_d}{\omega_{d-1}} C_i(1)$; in other words, $\sqrt{\frac{\omega_{d-1}}{\omega_d}} \frac{C_i(t)}{\sqrt{C_i(1)}}$ has unit norm with respect to $w(t)dt$.

Remark 2 Note that if the univariate polynomial $\phi(t)$ is nonnegative on $[-1, 1]$, then the coefficients $\lambda_0, \ldots, \lambda_L$ in (12) satisfy $\lambda_i \leq \lambda_0$ for all $i = 0, \ldots, L$ since $C_i(t) \leq C_i(1)$ for all $t \in [-1, 1]$. We will use this simple property of the coefficients later in the proof.

A technical lemma The following lemma will be important for our proof later. It shows that the sup-norm of the harmonic components of a polynomial f can be bounded by a constant independent of the dimension d, times the sup-norm of f.

Proposition 5 *For any integer n there exists a constant B_{2n} such that the following is true. For any homogeneous polynomial f with degree $2n$ and with decomposition into spherical harmonics $f = \sum_{k=0}^{n} f_{2k}$ on S^{d-1} with $f_j \in \mathcal{H}_j^d$ it holds $\|f_{2k}\|_\infty \leq \mathsf{B}_{2n}\|f\|_\infty$. Also $\mathsf{B}_2 \leq 2$ and $\mathsf{B}_4 \leq 10$.*

Proof The proof is in "Appendix B".

The remarkable property in the previous proposition is that the constant B_{2n} is independent of the dimension d.

Remark 3 When f is a homogeneous polynomial of degree $2n$ such that $0 \leq m \leq f \leq M$ on S^{d-1}, Proposition 5 gives us that $\|f_{2k}\|_\infty \leq \mathsf{B}_{2n} M$. However one can get a better bound by applying Proposition 5 instead to $f - \|x\|_2^{2n}(m + M)/2$; this gives $\|f_{2k}\|_\infty \leq B_{2n}(M - m)/2$ for all $k = 1, \ldots, n$.

3 Proof of main approximation result

In this section we prove our main theorem, Theorem 2. We will actually prove a more general result giving bounds on the performance of the sum-of-squares hierarchy for *all* values of the level ℓ. (In Theorem 2 stated in the introduction, only the regime $\ell \geq \Omega(d)$ was presented.)

For the statement of our theorem we need to introduce two quantities that play an important role in our analysis.

- The first quantity, which we denote $\rho_{2n}(d, \ell)$, is defined as (where n, d, ℓ are integers)

$$\rho_{2n}(d, \ell) = \min_{\substack{q \in \mathbb{R}[t], \deg(q) = \ell \\ \lambda_0 = 1}} \sum_{k=1}^{2n} |\lambda_{2k}^{-1} - 1|. \tag{13}$$

  Here, the minimization is over polynomials $q(t)$ of degree ℓ, and λ_{2k} is the $2k$'th coefficient of $\phi(t) = (q(t))^2$ in its Gegenbauer expansion, see Equation (12). In words, $\rho_{2n}(d, \ell)$ quantifies how close we can get the Gegenbauer coefficients of $\phi(t) = (q(t))^2$ to 1 (note however that the distance to 1 is measured by $|\lambda_{2k}^{-1} - 1|$ and not linearly).
- The second quantity is the constant B_{2n} introduced in Proposition 5. It is the smallest constant such that for any homogeneous polynomial f of degree $2n$, we have $\|f_{2k}\|_\infty \leq \mathsf{B}_{2n} \|f\|_\infty$ for all $k = 0, \ldots, n$, where f_{2k} are the $2k$'th harmonic components of f. In other words, B_{2n} is an upper bound on the $\infty \to \infty$ operator norm of the linear map that projects a homogeneous polynomial of degree $2n$ onto its $2k$'th harmonic component. Proposition 5 says that such an upper bound that only depends on n (i.e., independent of d) does exist. One can get explicit upper bounds on B_{2n} for small values of n. For example one can show that $\mathsf{B}_2 \leq 2$ and $\mathsf{B}_4 \leq 10$.

We are now ready to state our main theorem:

Theorem 6 *Assume $F(x_1, \ldots, x_d)$ is a homogeneous matrix-valued polynomial of degree $2n$ in d variables, such that $F(x)$ is symmetric for all x, and $0 \leq F \leq I$ on S^{d-1}. Then $F + (\mathsf{B}_{2n}/2)\rho_{2n}(d, \ell)I$ is ℓ-sos on S^{d-1}.*

Furthermore, the quantity $\rho_{2n}(d, \ell)$ satisfies the following: for any $n \leq d$, there are constants $\mathsf{C}_n, \mathsf{C}'_n$ such that for $\ell \geq \mathsf{C}'_n d$, $\rho_{2n}(d, \ell) \leq \mathsf{C}_n (d/\ell)^2$.

Proof (Proof of first part of Theorem 6) We will start by proving the first part of the theorem. For clarity of exposition, we will assume that F is a scalar-valued polynomial, and we explain later why the argument also works for matrices. Let thus F be a homogeneous polynomial of degree $2n$ such that $0 \leq F \leq 1$ on S^{d-1}. Let

$$F = F_0 + F_2 + \cdots + F_{2n} \qquad (F_{2k} \in \mathcal{H}_{2k}^d)$$

be the decomposition of F into spherical harmonics on S^{d-1} (since F is even, only harmonics of even order are nonzero). Given $\delta > 0$ to be specified later, we will exhibit a sum-of-squares decomposition of $\tilde{F} = F + \delta$ by writing $\tilde{F} = K K^{-1} \tilde{F}$ where K is an integral transform defined as

$$(Kh)(x) := \int_{S^{d-1}} \phi(\langle x, y \rangle) h(y) d\sigma(y), \quad \forall x \in S^{d-1} \tag{14}$$

where $\phi(t) = (q(t))^2$ is a univariate polynomial of degree 2ℓ. Recall that if $h \geq 0$ then the function Kh is ℓ-sos on S^{d-1} by construction, where the "infinite" sum of squares in the integral can be turned into a finite sum of squares by standard convexity

results.[6] In order for $\tilde{F} = KK^{-1}\tilde{F}$ to be a valid sum-of-squares decomposition of $\tilde{F}$ on the sphere, we need that $K^{-1}\tilde{F} \geq 0$. The polynomial $q(t)$ will be chosen so that K is close to a Dirac kernel; when combined with $\tilde{F} \geq \delta > 0$ we will be able to conclude that $K^{-1}\tilde{F} \geq 0$ from the fact that $\|\tilde{F} - K^{-1}(\tilde{F})\|_\infty \leq \delta$.

Let $(\lambda_i)_{0\leq i\leq 2\ell}$ be the coefficients in the Gegenbauer expansion of ϕ, i.e., $\phi = \lambda_0 C_0 + \lambda_1 C_1 + \cdots + \lambda_{2\ell} C_{2\ell}$. By the Funk-Hecke formula we have $K^{-1}(\tilde{F}) = \lambda_0^{-1}(F_0 + \delta) + \lambda_2^{-1}F_2 + \cdots + \lambda_{2n}^{-1}F_{2n}$. Our analysis does not depend on the scaling of K so we will assume $\lambda_0 = 1$. Thus we get

$$\begin{aligned}\|K^{-1}(\tilde{F}) - \tilde{F}\|_\infty &= \left\|\sum_{k=1}^{n}\left(\frac{1}{\lambda_{2k}} - 1\right)F_{2k}\right\|_\infty \leq \sum_{k=1}^{n}\left|\frac{1}{\lambda_{2k}} - 1\right| \|F_{2k}\|_\infty \\ &\leq (\mathsf{B}_{2n}/2)\sum_{k=1}^{n}\left|\frac{1}{\lambda_{2k}} - 1\right|\end{aligned}$$

where in the last inequality we used Proposition 5 (see also Remark 3) together with the fact that $0 \leq F \leq 1$. It thus follows that if

$$(\mathsf{B}_{2n}/2)\sum_{k=1}^{n}|\lambda_{2k}^{-1} - 1| \leq \delta \tag{15}$$

then $K^{-1}(\tilde{F}) \geq 0$ and the equation $\tilde{F} = KK^{-1}(\tilde{F})$ gives a valid sum-of-squares decomposition of $\tilde{F} = F + \delta$ on the sphere. We have thus proved the first part of Theorem 1. □

It now remains to prove the second part of the theorem, which leads us to the analysis of the quantity $\rho_{2n}(d, \ell)$. Before doing so, we explain how the proof above applies in the case where F is a matrix-valued polynomial.

Matrix-valued polynomials Assume $F \in \mathbf{S}[x]$ homogeneous of degree $2n$. We can decompose each entry of F into spherical harmonics to get $F = F_0 + F_2 + \cdots + F_{2n}$. Define $\tilde{F} = F(x) + \delta I$ for a $\delta > 0$ to be specified later. The steps in the argument above are identical, where $\|\cdot\|_\infty$ is defined as the maximum of $\|F(x)\|$ over $x \in S^{d-1}$, where $\|F(x)\|$ is the spectral norm of $F(x)$, and the bound on $\|F_{2k}\|_\infty$ follows from Proposition 17. If $\|K^{-1}\tilde{F} - \tilde{F}\|_\infty \leq \delta$ then $K^{-1}\tilde{F} \geq 0$ in the positive semidefinite sense. Letting $H = K^{-1}\tilde{F} \geq 0$, we get $\tilde{F}(x) = (KH)(x) = \int_{S^{d-1}} q(\langle x, y\rangle)^2 H(y)d\sigma(y) = \int_{S^{d-1}} U_y(x)U_y(x)^\mathsf{T} d\sigma(y)$ where $U_y(x) = q(\langle x, y\rangle)H(y)^{1/2}$ is a polynomial of degree ℓ in x. This is what we wanted.

We now proceed to the analysis of $\rho_{2n}(d, \ell)$.

[6] More precisely, note that one can find a finite number of points $y_1, \ldots, y_N \in S^{d-1}$ and coefficients $\alpha_1, \ldots, \alpha_N \geq 0$ such that the following is true: for any polynomial P of degree 2ℓ we have $\int_{S^{d-1}} P(y)d\sigma(y) = \sum_{i=1}^{N}\alpha_i P(y_i)$. (This essentially follows by Carathéodory's theorem, see e.g., [2, Chapter 1].) Then it follows that $Kh(x) = \sum_{i=1}^{N}\alpha_i\phi(\langle x, y_i\rangle)h(y_i)$ which implies that Kh is a finite sum of squares since $\phi(\langle x, y_i\rangle) = q(\langle x, y_i\rangle)^2$.

Reformulating $\rho_{2n}(d,\ell)$ using generalized Toeplitz matrices It will be convenient to reformulate the optimization problem (13) in terms of certain suitable (generalized) Toeplitz matrices. We parametrize the degree-ℓ polynomial $q(t)$ as

$$q(t) = \sum_{i=0}^{\ell} e_i \frac{C_i(t)}{\sqrt{C_i(1)}}$$

where $e_0, \ldots, e_\ell \in \mathbb{R}$. The presence of the term $\sqrt{C_i(1)}$ is for convenience later. The Gegenbauer coefficients of $\phi(t) = (q(t))^2$ are then equal to (cf. Equation (12))

$$\begin{aligned}\lambda_k &= \frac{\omega_{d-1}}{\omega_d}\int_{-1}^{1} \phi(t)\frac{C_k(t)}{C_k(1)}(1-t^2)^{\frac{d-3}{2}}dt \\ &= \sum_{i,j=0}^{\ell} e_i e_j \left(\frac{\omega_{d-1}}{\omega_d}\int_{-1}^{1}\frac{C_i(t)}{\sqrt{C_i(1)}}\frac{C_j(t)}{\sqrt{C_j(1)}}\frac{C_k(t)}{C_k(1)}(1-t^2)^{\frac{d-3}{2}}dt\right) \\ &= e^{\mathsf{T}}\mathcal{T}\left[C_k/C_k(1)\right]e,\end{aligned}$$

where for $h : [-1,1] \to \mathbb{R}$, $\mathcal{T}[h]$ is the $(\ell+1)\times(\ell+1)$ symmetric matrix

$$\mathcal{T}[h]_{i,j} = \frac{\omega_{d-1}}{\omega_d}\int_{-1}^{1}\frac{C_i(t)}{\sqrt{C_i(1)}}\frac{C_j(t)}{\sqrt{C_j(1)}}h(t)(1-t^2)^{\frac{d-3}{2}}dt.$$

It can be easily checked that $\mathcal{T}[1] = I$ is the identity matrix (this follows from the fact that the polynomials $\sqrt{\frac{\omega_{d-1}}{\omega_d}}\frac{C_i}{\sqrt{C_i(1)}}$ have unit norm with respect to the weight function $(1-t^2)^{(d-3)/2}$), and so $\lambda_0 = e^{\mathsf{T}}e = \sum_k e_k^2$. It thus follows that $\rho_{2n}(d,\ell)$ can be formulated as:

$$\rho_{2n}(d,\ell) = \min_{\substack{e\in\mathbb{R}^{\ell+1}\\ \sum_k e_k^2=1}} \sum_{k=1}^{n}\left|\left(e^{\mathsf{T}}\mathcal{T}\left[\frac{C_{2k}}{C_{2k}(1)}\right]e\right)^{-1} - 1\right|. \tag{16}$$

Case $2n = 2$ Let us first analyze the case $2n = 2$ which corresponds to quadratic polynomials. In this case the sum in (16) has simply one term. It is then not difficult to see that $\rho_2(d,\ell)$ is given by

$$\rho_2(d,\ell) = \left\|\mathcal{T}\left[\frac{C_2}{C_2(1)}\right]\right\|^{-1} - 1 \tag{17}$$

where $\|\cdot\|$ denotes the spectral norm. Thus we see that $\rho_2(d,\ell)$ can be computed efficiently by simply evaluating the spectral norm of $\mathcal{T}[C_2/C_2(1)]$. The latter matrix can be formed explicitly using known formulas for the integrals of Gegenbauer polynomials (see e.g., [23]). Note that $\mathcal{T}[C_2/C_2(1)]$ is a banded matrix with bandwidth 3.

Case $2n = 4$ We now turn to quartic polynomials. In this case $\rho_4(d, \ell)$ takes the form

$$\rho_4(d,\ell) = \min_{\substack{e\in\mathbb{R}^{\ell+1} \\ \sum_k e_k^2=1}} \left|\left(e^\mathsf{T}\mathcal{T}\left[\frac{C_2}{C_2(1)}\right]e\right)^{-1} - 1\right| + \left|\left(e^\mathsf{T}\mathcal{T}\left[\frac{C_4}{C_4(1)}\right]e\right)^{-1} - 1\right|. \tag{18}$$

Let $\mathcal{R}$ be the *joint numerical range* (also known as the field of values) of the matrices $\mathcal{T}\left[\frac{C_2}{C_2(1)}\right]$ and $\mathcal{T}\left[\frac{C_4}{C_4(1)}\right]$, i.e.,

$$\mathcal{R} = \left\{\left(e^\mathsf{T}\mathcal{T}\left[\frac{C_2}{C_2(1)}\right]e \;,\; e^\mathsf{T}\mathcal{T}\left[\frac{C_4}{C_4(1)}\right]e\right) : e \in \mathbb{R}^{\ell+1}, \sum_{k=0}^{\ell} e_k^2 = 1\right\}.$$

From results about joint numerical ranges, it is known that $\mathcal{R} \subset \mathbb{R}^2$ is convex, see [10] and also [33, Theorem 5.6]. It is not difficult to see then that $\mathcal{R}$ has a semidefinite representation, and that $\rho_4(d, \ell)$ can be computed using semidefinite programming.

General degree $2n$ We now analyze the case of general degree $2n$. To do this we formulate a proxy for the optimization problem that defines $\rho_{2n}(d, \ell)$ that is easier to analyze. Instead of minimizing $\sum_{k=1}^{2n} |\lambda_{2k}^{-1} - 1|$ we will seek instead to minimize $\sum_{k=1}^{2n}(1 - \lambda_{2k})$. Since $\lambda_{2k} \leq \lambda_0 = 1$, both problems seek to bring the λ_{2k} close to 1, but the latter problem is easier to analyze because it is linear in the λ_{2k}. Define

$$\tilde{\rho}_{2n}(d,\ell) = \min_{\substack{e\in\mathbb{R}^{\ell+1} \\ \sum_i e_i^2=1}} \sum_{k=1}^{n}\big(1 - e^\mathsf{T}\mathcal{T}[C_{2k}/C_{2k}(1)]e\big). \tag{19}$$

Since $\mathcal{T}$ is linear, i.e., $\mathcal{T}[h_1 + h_2] = \mathcal{T}[h_1] + \mathcal{T}[h_2]$ we get that $\tilde{\rho}_{2n}(d, \ell) = n - n\lambda_{\max}(\mathcal{T}[h])$ where $h = \frac{1}{n}\sum_{k=1}^{n} C_{2k}/C_{2k}(1)$. It thus remains to analyze $\lambda_{\max}(\mathcal{T}[h])$. This is what we do next.

Proposition 7 *Let* $h = \frac{1}{n}\sum_{k=1}^{n} \frac{C_{2k}}{C_{2k}(1)}$. *Then* $\lambda_{\max}(\mathcal{T}[h]) \geq 1 - \frac{7n}{12}\frac{d^2}{\ell^2}$.

Proof We use the following standard result on orthogonal polynomials which gives the eigenvalues of $\mathcal{T}[f]$ for any linear polynomial f. (The result below is stated in full generality for clarity, in our case the (p_k) is the family of normalized Gegenbauer polynomials.)

Proposition 8 (Standard result on orthogonal polynomials) *Let* $(p_k)_{k\in\mathbb{N}}$ *be a family of orthogonal polynomials with respect to a weight function* $w(x) > 0$. *We assume the* (p_k) *are normalized, i.e.,* $\int p_k^2 w = 1$. *Given a linear polynomial* f, *define the* $(\ell + 1) \times (\ell + 1)$ *matrix*

$$\mathcal{T}[f]_{ij} = \int_a^b p_i(t)p_j(t)f(t)w(t)dt \qquad \forall 0 \leq i, j \leq \ell. \tag{20}$$

Then the eigenvalues of $\mathcal{T}[f]$ *are precisely the* $f(x_{\ell+1,i})$ *where the* $(x_{\ell+1,i})_{i=1,\ldots,\ell+1}$ *are the roots of* $p_{\ell+1}$.

Proof This follows from standard results on orthogonal polynomials. When $f = 1$ then $\mathcal{T}[f]$ is the identity matrix. When $f(t) = t$, the matrix $\mathcal{T}[f]$ is the tridiagonal matrix that encodes the three-term recurrence formula for the (p_k). It is well-known that the eigenvalues of this tridiagonal matrix are the roots of $p_{\ell+1}$. See e.g., [30, Lemma 3.9]. □

Our function $h(t) = \frac{1}{n}\sum_{k=1}^{n} \frac{C_{2k}(t)}{C_{2k}(1)}$ is not linear. However one can verify (see Proposition 18) that it is lower bounded by its linear approximation at $t = 1$, i.e., we have

$$h(t) \geq h'(1)(t-1) + h(1).$$

It is easy to check that if h_1, h_2 are two functions such that $h_1(t) \geq h_2(t)$ for all $t \in [-1, 1]$, then $\mathcal{T}[h_1] \geq \mathcal{T}[h_2]$ (positive semidefinite order) and thus the largest eigenvalue of $\mathcal{T}[h_1]$ is at least the largest eigenvalue of $\mathcal{T}[h_2]$. Let $\bar{h}(t) = h'(1)(t-1) + h(1)$. The largest eigenvalue of $\mathcal{T}[\bar{h}]$ is equal to $\bar{h}(x_{\ell+1,\ell+1})$ where $x_{\ell+1,\ell+1}$ is the largest root of $C_{\ell+1}$. It is known [12, Section 2.3 (last displayed equation)] that $x_{\ell+1,\ell+1}$ satisfies

$$x_{\ell+1,\ell+1} \geq 1 - \frac{1}{4}\frac{d^2}{\ell^2}.$$

It thus follows, using the fact that $h(1) = 1$ and $h'(1) > 0$, that

$$\lambda_{\max}(\mathcal{T}[h]) \geq \lambda_{\max}(\mathcal{T}[\bar{h}]) = \bar{h}(x_{\ell+1,\ell+1}) \geq -h'(1)\frac{d^2}{4\ell^2} + 1 \geq 1 - \frac{7n}{12}\cdot\frac{d^2}{\ell^2},$$

where in the last inequality we used the exact value of $h'(1)$ given by $h'(1) = (n+1)(3d+4n-4)/(3(d-1))$ and the fact that $n \leq d$. □

Our proof of Theorem 6 (i) is now almost complete. We just need to relate $\tilde{\rho}$ back to ρ. We use the following easy proposition.

Proposition 9 *If $\tilde{\rho} < 1$ then $\rho \leq \tilde{\rho}/(1-\tilde{\rho})$.*

Proof Let (λ_{2k}) be the optimal choice in the solution to $\tilde{\rho}$ (Eq. (19)). Then $\lambda_{2k} = 1-(1-\lambda_{2k}) \geq 1-\tilde{\rho} > 0$. Thus $\sum_{k=1}^{n} |\lambda_{2k}^{-1} - 1| = \sum_{k=1}^{n}(1-\lambda_{2k})/\lambda_{2k} \leq \tilde{\rho}/(1-\tilde{\rho})$.

Proposition 7 tells us that $\tilde{\rho}_{2n}(d,\ell) \leq (7n^2/12)(d/\ell)^2$. For $\ell \geq 2nd$, we will have $\tilde{\rho}_{2n}(d,\ell) \leq 1/2$ and so $\rho_{2n}(d,\ell) \leq 2\tilde{\rho}_{2n}(d,\ell) \leq 2n^2(d/\ell)^2$. This completes the proof of Theorem 6.

Tightness Our analysis of $\rho_{2n}(d,\ell)$ in the regime $\ell \geq \Omega(d)$ can be shown to be tight. We show this in the case $2n = 2$ below.

Theorem 10 (Tightness of convergence rate) *There is an absolute constant $C > 0$ such that for $\ell \geq \Omega(d)$, $\rho_2(d,\ell) \geq C(d/\ell)^2$.*

Proof Given the expression for $\rho_2(d, \ell)$ in (17), we need to produce an upper bound on $\|\mathcal{T}[C_2/C_2(1)]\|$. Note that $C_2(t)/C_2(1) = \frac{d}{d-1}t^2 - \frac{1}{d-1}$. It thus follows that $\mathcal{T}[C_2/C_2(1)] = \frac{d}{d-1}\mathcal{T}[t^2] - \frac{I}{d-1}$. We now use the following property of generalized Toeplitz matrices constructed from sequences of orthogonal polynomials: If $\mathcal{T}_\infty$ denotes the semi-infinite version of (20), then $\mathcal{T}_\infty[f]\mathcal{T}_\infty[g] = \mathcal{T}_\infty[fg]$ for any polynomials f, g (this property follows immediately from the fact that the sequence of orthogonal polynomials $(p_k)_{k=0}^\infty$ is an orthonormal basis of the space of polynomials, see e.g., [3, Lemma 2.4]). In particular we have $\mathcal{T}_\infty[t^2] = \mathcal{T}_\infty[t]^2$. Now noting that $\mathcal{T}_\infty[t]$ is tridiagonal, we see that $\mathcal{T}[t^2]$ is a submatrix of $(\mathcal{T}_{\ell+2}[t])^2$, where the subscript indicates the truncation level (so $\mathcal{T}_{\ell+2}[t]$ is $(\ell+3)\times(\ell+3)$). Thus it follows that $\mathcal{T}[C_2/C_2(1)]$ is a submatrix of $\frac{d}{d-1}(\mathcal{T}_{\ell+2}[t])^2 - \frac{1}{d-1}I$. Since $(\mathcal{T}_{\ell+2}[t])^2$ is positive semidefinite it then follows that

$$\|\mathcal{T}[C_2/C_2(1)]\| \leq \frac{d}{d-1}\lambda_{\max}(\mathcal{T}_{\ell+2}[t])^2 - \frac{1}{d-1}$$

Recall that $\lambda_{\max}(\mathcal{T}_{\ell+2}[t])$ is the largest root of $C_{\ell+3}$. From [1, Corollary 2.3] we get, for $\ell \geq \Omega(d)$, $\lambda_{\max}(\mathcal{T}_{\ell+2}[t])^2 \leq 1 - \mathsf{C}(d/\ell)^2$ for some constant C. Thus we get $\rho_2(d, \ell) = \|\mathcal{T}[C_2/C_2(1)]\|^{-1} - 1 \geq \mathsf{C}(d/\ell)^2$ as desired. □

4 Relation to quantum state extendibility

Quantum entanglement is one of the key ingredients in quantum information processing. Certifying whether a given state is entangled or not is a hard computational task [20] and considerable effort has been dedicated to this problem, e.g., [21,26]. Of particular interest is the hierarchy of tests known as the *DPS hierarchy* [17,18], applying semidefinite programs to verify quantum entanglement.

In this section, we explore the duality relation between the DPS hierarchy and sums of squares, and explain how our results from the previous section can be used to bound the convergence rate of the DPS hierarchy. We show that the result of Navascues et al. [28] can be seen as the special case of our Theorem 6 when the polynomial F is quadratic.

4.1 Quantum extendible states

A quantum state is usually represented by a positive semidefinite operator normalized with unit trace. In this work, we mainly work with *unnormalized quantum states* and consider its convex cone. Given Hilbert spaces $\mathcal{H}_A \simeq \mathbb{C}^{d_A}$ and $\mathcal{H}_B \simeq \mathbb{C}^{d_B}$, denote the cone of bipartite quantum states as $\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, i.e., the cone of positive semidefinite matrices of size $d_A d_B$. A bipartite quantum state $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is *separable* if and only if it can be written as a conic combination of tensor product states, i.e.,

$$\rho_{AB} = \sum_i p_i (x_i x_i^\dagger) \otimes (y_i y_i^\dagger) \quad \text{with} \quad p_i \geq 0, x_i \in \mathcal{H}_A, y_i \in \mathcal{H}_B. \tag{21}$$

The convex cone of quantum separable states is denoted as $\mathrm{SEP}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and it is strictly included in $\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$.

Positive partial transpose A well-known necessary condition for a state ρ_{AB} to be in SEP is that it has a *positive partial transpose* (PPT). If we let T denote the transpose operation on Hermitian matrices of size $d_B \times d_B$, then for ρ_{AB} of the form (21) we have

$$(I \otimes \mathsf{T})(\rho_{AB}) = \sum_i p_i (x_i x_i^\dagger) \otimes (y_i y_i^\dagger)^\mathsf{T} = \sum_i p_i (x_i x_i^\dagger) \otimes (\bar{y_i} \bar{y_i}^\dagger) \geq 0.$$

If we let $\mathcal{PPT}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be the set of states with a positive partial transpose then we have the inclusions

$$\mathcal{SEP}(\mathcal{H}_A \otimes \mathcal{H}_B) \subset \mathcal{PPT}(\mathcal{H}_A \otimes \mathcal{H}_B) \subset \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B).$$

A well-known result due to Woronowicz [35] asserts we have equality $\mathrm{SEP}(\mathcal{H}_A \otimes \mathcal{H}_B) = \mathcal{PPT}(\mathcal{H}_A \otimes \mathcal{H}_B)$ if and only if dimensions of $\mathcal{H}_A$ and $\mathcal{H}_B$ satisfy $d_A + d_B \leq 5$.

Extendibility When the inclusion $\mathcal{SEP} \neq \mathcal{PPT}$ is strict, one can find more accurate relaxations of SEP based on the notion of *state extendibility*. For simplicity of the following discussion, we introduce the notation $[s_1 : s_2] := \{s_1, s_1 + 1, \cdots, s_2\}$ and $[s] := [1 : s]$ for short. Given a separable state expressed as Eq. (21) with[7] $\|x_i\| = \|y_i\| = 1$ we can consider its *extension* (on the B subsystem) as:

$$\rho_{AB_{[\ell]}} = \sum_i p_i x_i x_i^\dagger \otimes \left(y_i y_i^\dagger\right)^{\otimes \ell}. \tag{22}$$

The new system $\rho_{AB_{[\ell]}}$ lies in $\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{B_1} \otimes \cdots \otimes \mathcal{H}_{B_\ell})$ where each $\mathcal{H}_{B_i} \simeq \mathbb{C}^{d_B}$; i.e., it is a Hermitian matrix of size $d_A (d_B)^\ell \times d_A (d_B)^\ell$. The system $\rho_{AB_{[\ell]}}$ satisfies a number of properties, as follows:

(a) *Positivity*: $\rho_{AB_{[\ell]}}$ is positive semidefinite

(b) *Reduction under partial traces*: If we *trace out*[8] the systems $B_2, \ldots, B_\ell$ from $\rho_{AB_{[\ell]}}$ we get back the original system ρ_{AB}. Indeed we have:

$$\mathrm{Tr}_{B_{[2:\ell]}} \rho_{AB_{[\ell]}} = \sum_i p_i x_i x_i^\dagger \otimes y_i y_i^\dagger \cdot \mathrm{Tr}\left[(y_i y_i^\dagger)^{\otimes \ell - 1}\right] \overset{(*)}{=} \sum_i p_i x_i x_i^\dagger \otimes y_i y_i^\dagger = \rho_{AB}. \tag{23}$$

In $(*)$ we used the fact that $\|y_i\| = 1$.

(c) *Symmetry*: define the *symmetric subspace* of $\mathcal{H}^{\otimes \ell}$ as

$$\mathrm{Sym}(\mathcal{H}^{\otimes \ell}) = \left\{ Y \in \mathcal{H}^{\otimes \ell} : P \cdot Y = Y \quad \forall P \in \mathfrak{S}_\ell \right\}$$

[7] We can always impose such condition without losing generality by changing the coefficients p_i accordingly.

[8] The partial trace operator is the unique linear map $\mathrm{Tr}_B : \mathrm{Herm}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathrm{Herm}(\mathcal{H}_A)$ such that $\mathrm{Tr}_B(\rho_A \otimes \sigma_B) = \mathrm{Tr}(\sigma_B)\rho_A$ for all $\rho_A \in \mathrm{Herm}(\mathcal{H}_A)$ and $\sigma_B \in \mathrm{Herm}(\mathcal{H}_B)$.

where $\mathfrak{S}_\ell$ is the symmetric group on ℓ elements which naturally acts on $\mathcal{H}^{\otimes \ell}$ by permutation of the components. The dimension of $\mathrm{Sym}(\mathcal{H}^{\otimes \ell})$ is equal to $\binom{\ell+d-1}{\ell}$ where $d = \dim \mathcal{H}$. If we let $\Pi = \Pi^\dagger$ be the projector on the symmetric subspace of $\mathcal{H}_B^{\otimes \ell}$ then one can easily verify that $\Pi \left(yy^\dagger\right)^{\otimes \ell} \Pi = \left(yy^\dagger\right)^{\otimes \ell}$. It thus follows that the extension $\rho_{AB_{[\ell]}}$ of Eq. (22) satisfies

$$(I \otimes \Pi)\rho_{AB_{[\ell]}}(I \otimes \Pi) = \rho_{AB_{[\ell]}}. \tag{24}$$

(d) *Positive Partial Transpose*: If we let T be the transpose map on Hermitian matrices of size $d_B \times d_B$ then $\rho_{AB_{[\ell]}}$ satisfies

$$(I_A \otimes \underbrace{\mathsf{T}_{B_1} \otimes \cdots \mathsf{T}_{B_s}}_{s} \otimes I_{B_{i+1}} \otimes \cdots \otimes I_{B_\ell})(\rho_{AB_{[\ell]}}) \geq 0 \tag{25}$$

for any $s = 1, \ldots, \ell$. For convenience later the state on the left of (25) will be denoted $\rho_{AB_{[\ell]}}^{\mathsf{T}_{B_{[s]}}}$.

The DPS hierarchy Define now the set $\mathcal{DPS}_\ell(\mathcal{H}_A \otimes \mathcal{H}_B)$ as

$$\mathcal{DPS}_\ell(\mathcal{H}_A \otimes \mathcal{H}_B) = \Big\{\omega \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B) \text{ s.t. } \exists\, \omega_{AB_{[\ell]}} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{B_1} \otimes \cdots \mathcal{H}_{B_\ell}) \text{ s.t. conditions (23), (24), (25) are satisfied}\Big\}.$$

By the previous reasoning, each set $\mathcal{DPS}_\ell(\mathcal{H}_A \otimes \mathcal{H}_B)$ is a convex cone containing $\mathcal{SEP}(\mathcal{H}_A \otimes \mathcal{H}_B)$, i.e., we have

$$\mathcal{SEP} \subseteq \cdots \subseteq \mathcal{DPS}_\ell \subseteq \cdots \subseteq \mathcal{DPS}_2 \subseteq \mathcal{DPS}_1 \subseteq \mathcal{S}.$$

Note that $\mathcal{DPS}_1 = \mathcal{PPT}$. Also it is known that the hierarchy is complete in the sense that if $\rho \notin \mathcal{SEP}$ then there exists a $\ell \in \mathbb{N}$ such that $\rho \notin \mathcal{DPS}_\ell$ [17,18].

Remark 4 (Extendibility without PPT conditions) One can also consider the weaker hierarchy where the Positive Partial Transpose constraints are dropped:

$$\mathcal{EXT}_\ell(\mathcal{H}_A \otimes \mathcal{H}_B) = \Big\{\omega \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B) \text{ s.t. } \exists\, \omega_{AB_{[\ell]}} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{B_1} \otimes \cdots \mathcal{H}_{B_\ell}) \text{ s.t. conditions (23) and (24) are satisfied}\Big\}. \tag{26}$$

It turns out that this weaker hierarchy $\mathcal{EXT}_\ell$ is already complete in the sense stated above. This is usually proven using de Finetti theorems [11,24].

4.2 Hermitian polynomials and sums of squares

In this section we leave the quantum world and introduce some terminology pertaining to Hermitian polynomials. A Hermitian polynomial $p(z, \bar{z})$ is a polynomial with

complex coefficients in the variables $z = (z_1, \ldots, z_n)$ and $\bar{z} = (\bar{z}_1, \ldots, \bar{z}_n)$ such that $p(z, \bar{z}) \in \mathbb{R}$ for all $z \in \mathbb{C}^n$. If $u \in \mathbb{N}^n$ we define the monomial $z^u := z_1^{u_1} \cdots z_n^{u_n}$. The general form of a Hermitian polynomial is

$$p(z, \bar{z}) = \sum_{(u,v)\in\mathcal{A}} p_{uv} z^u \bar{z}^v \qquad (p_{uv} \in \mathbb{C} \text{ and } \mathcal{A} \subset \mathbb{N}^n \times \mathbb{N}^n)$$

where the coefficients p_{uv} satisfy $p_{uv} = \overline{p_{vu}}$. We say that $p(z)$ is nonnegative if $p(z) \geq 0$ for all $z \in \mathbb{C}^n$.

Definition 11 (Hermitian polynomials and sums of squares) Let $p(z, \bar{z})$ be a nonnegative Hermitian polynomial. We say that $p(z, \bar{z})$ is a *real sum-of-squares (rsos)* if we can write $p(z, \bar{z}) = \sum_i g_i(z, \bar{z})^2$ where $g_i(z, \bar{z})$ are Hermitian polynomials. We say that $p(z, \bar{z})$ is a *complex sum-of-squares (csos)* if we can write $p(z, \bar{z}) = \sum_i |q_i(z)|^2$ where $q_i(z)$ are (holomorphic) polynomial maps in z (i.e., q_i are functions of z alone and not $\bar{z}$).

Clearly if $p(z, \bar{z})$ is csos then it is also rsos since $|q(z)|^2 = \mathrm{Re}[q(z)]^2 + \mathrm{Im}[q(z)]^2$ and $\mathrm{Re}[q(z)]$ and $\mathrm{Im}[q(z)]$ are both Hermitian polynomials. The converse however is not true. For example $p(z, \bar{z}) = (z + \bar{z})^2$ is evidently rsos, however it is not csos [16, Example (a)]. Indeed the zero-set of a csos polynomial must be a complex variety, and the zero set of $p(z, \bar{z})$ here is the imaginary axis. Note that a Hermitian polynomial $p(z, \bar{z})$ is rsos iff the *real* polynomial $P(a, b) = p(a + ib, a - ib)$ is a sum-of-squares (in the usual sense for real polynomials).

4.3 The duality relation

An element $M \in \mathrm{Herm}(d_A d_B)$ is in the dual of $\mathcal{SEP}(\mathcal{H}_A \otimes \mathcal{H}_B)$ if, and only if the Hermitian polynomial p_M defined by

$$p_M(x, \bar{x}, y, \bar{y}) := \sum_{ijkl} M_{ij,kl}\, x_i\, \bar{x}_k\, y_j\, \bar{y}_l, \quad \forall x \in \mathbb{C}^{d_A}, y \in \mathbb{C}^{d_B} \tag{27}$$

is nonnegative for all $(x, y) \in \mathbb{C}^{d_A} \times \mathbb{C}^{d_B}$. We prove our first main result on the duality between the DPS hierarchy and sums of squares.

Theorem 12 (Duality between extendibility hierarchy and sums of squares) *For $M \in Herm(d_A d_B)$, we let p_M be the associated Hermitian polynomial in (27). Then we have:*

(i) $\mathcal{SEP}^* = \{M \in Herm(d_A d_B) : p_M \text{ is nonnegative}\}$.
(ii) $\mathcal{DPS}_\ell^* = \left\{M \in Herm(d_A d_B) : \|y\|^{2(\ell-1)} p_M \text{ is rsos}\right\}$.
(iii) $\mathcal{EXT}_\ell^* = \left\{M \in Herm(d_A d_B) : \|y\|^{2(\ell-1)} p_M \text{ is csos}\right\}$.

Proof Point (i) is immediate and follows from the definition of duality. Points (ii) and (iii) are proved in "Appendix C". □

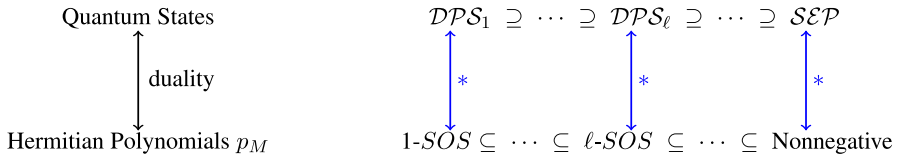

Fig. 1 A summary of the duality relations between the DPS hierarchy and sums of squares. The notation ℓ-SOS is a shorthand for $\|y\|^{2(\ell-1)} p_M$ is a real sum-of-squares

The above diagram summarizes the situation (Fig. 1).

In terms of the support functions The support function of the set DPS_ℓ is defined as

$$h_{\mathrm{DPS}_\ell}(M) = \max_{\rho \in \mathrm{DPS}_\ell} \quad \mathrm{Tr}[M\rho]$$

where $M \in \mathrm{Herm}(d_A d_B)$. The duality relation of Theorem 12 allows us to express $h_{\mathrm{DPS}_\ell}(M)$ in the following way:

$$h_{\mathrm{DPS}_\ell}(M) = \min \ \gamma \ \text{ s.t. } \ \|y\|^{2(\ell-1)}(\gamma \|x\|^2 \|y\|^2 - p_M) \text{ is rsos.}$$

A somewhat more convenient formulation using matrix polynomials is as follows. For $x \in \mathbb{C}^d$, we let $\tilde{x} \in \mathbb{R}^{2d}$ be the vector of real and imaginary components of x. Given $M \in \mathrm{Herm}(d_A d_B)$, let also $\tilde{\mathrm{P}}_M(\tilde{y}) \in \mathbf{S}[\tilde{y}]$ such that, for any $(x, y) \in \mathbb{C}^{d_A} \times \mathbb{C}^{d_B}$ we have

$$(x \otimes y)^\dagger M (x \otimes y) = \tilde{x}^T \tilde{\mathrm{P}}_M(\tilde{y}) \tilde{x}.$$

Then one can show the following equivalent formulation of $h_{\mathrm{DPS}_\ell}(M)$:

$$h_{\mathrm{DPS}_\ell}(M) = \min \ \gamma \ \text{ s.t. } \ \gamma I - \tilde{\mathrm{P}}_M(\tilde{y}) \text{ is } \ell - \text{sos on } S^{2d_B - 1}. \tag{28}$$

This can be proved using the following lemma, which is a straightforward generalization of [15, Proposition 2] to the matrix case.

Lemma 13 *Let $G(y_1, \ldots, y_d)$ be a homogeneous matrix-valued polynomial of even degree $2n$, such that $G(y)$ is symmetric for all y. Then G is ℓ-sos on S^{d-1}, if and only if, $\|y\|^{2(\ell-n)} G(y)$ is a sum of squares.*

4.4 Convergence rate of the DPS hierarchy

In [28, Theorem 3], Navascues, Owari and Plenio's proved the following result on the convergence of the sequence of relaxations (DPS_ℓ) to Sep.

Theorem 14 (NOP09) *For any quantum state $\rho_{AB} \in DPS_\ell$ with reduced state $\rho_A := \mathrm{Tr}_B[\rho_{AB}]$, we have*

$$(1-t)\rho_{AB} + t\rho_A \otimes \frac{I_B}{d_B} \quad \in \quad Sep(d) \tag{29}$$

where $t = O\left(\frac{d_B^2}{\ell^2}\right)$, $d_B = \dim(\mathcal{H}_B)$, and I_B is the identity matrix of dimension d_B.

Note that the state $\rho_A \otimes I_B/d_B$ is clearly separable. In words, the result above says that if ρ_{AB} in DPS_ℓ, then by moving ρ_{AB} in the direction $\rho_A \otimes I_B/d_B$ by $t = O(d_B^2/\ell^2)$ results in a separable state. In terms of the *Best Separable State* problem, the result of [28] has the following immediate implication. We show below how we can recover this result using our Theorem 6 from the previous section.

Theorem 15 *Let $M \in Herm(d_A d_B)$ and assume that $(x \otimes y)^\dagger M (x \otimes y) \geq 0$ for all $(x, y) \in \mathbb{C}^{d_A} \times \mathbb{C}^{d_B}$. Then*

$$h_{Sep}(M) \leq h_{DPS_\ell}(M) \leq (1 + Cd_B^2/\ell^2) h_{Sep}(M)$$

for any $\ell \geq C' d_B$, where $C, C' > 0$ is some absolute constant.

Proof We know from (28) that

$$h_{\mathrm{DPS}_\ell}(M) = \min \gamma \text{ s.t. } \gamma I - \tilde{\mathrm{P}}_M(\tilde{y}) \text{ is } \ell - \text{sos on } \tilde{y} \in S^{2d_B - 1}.$$

By assumption we have $0 \leq \tilde{\mathrm{P}}_M(\tilde{y}) \leq h_{\mathrm{Sep}}(M) I$ for all $\tilde{y} \in S^{2d_B-1}$. Our Theorem 6 from previous section tells us that for $\ell \geq C d_B$, $C' d_B^2/\ell^2 + \frac{h_{\mathrm{Sep}}(M) I - \tilde{\mathrm{P}}_M}{h_{\mathrm{Sep}}(M)}$ is ℓ-sos on S^{2d_B-1}. This implies that $h_{\mathrm{DPS}_\ell}(M) \leq h_{\mathrm{Sep}}(M)(1 + C' d_B^2/\ell^2)$ which is what we wanted. □

5 Conclusions

We have shown a quadratic improvement on the convergence rate of the SOS hierarchy on the sphere compared to the previous analysis of Reznick [34] and Doherty and Wehner [19]. The proof technique also works for matrix-valued polynomials on the sphere and surprisingly, the bounds we get are independent of the dimension of the matrix polynomial. In the special case of quadratic matrix polynomials, we recover the same rate obtained by Navascues, Owari and Plenio [28] using arguments from quantum information theory.

Acknowledgements Funding was provided by Isaac Newton Trust (Grant No. RG74916).

A The sum-of-squares hierarchy on the sphere

In this appendix we clarify the sum-of-squares hierarchy that we study, defined in Eq. (2). Let $p(x) = p(x_1, \ldots, x_d)$ be a homogeneous polynomial of even degree $2n$. The ℓ'th level of the hierarchy for the maximization of p on the sphere S^{n-1} is defined by:

$$p_\ell = \min \left\{ \gamma \in \mathbb{R} \text{ s.t. } \gamma - p \text{ is a sum of squares of degree } \ell \text{ on } S^{d-1} \right\}. \tag{30}$$

The constraint in the equation above means that there exist a sum-of-squares polynomial $s(x)$ such that $\gamma - p = s(x)$ for all $x \in S^{n-1}$ (i.e., $\gamma - p(x)$ is equal to $s(x)$ in the quotient ring $\mathbb{R}[x]/I$ where I is the ideal generated by $\sum_{i=1}^d x_i^2 - 1$). Since two polynomials are equal on S^{d-1} if, and only if, their difference is a (polynomial) multiple of $\sum_{i=1}^d x_i^2 - 1$, our hierarchy can be equivalently written as:

$$\begin{aligned} p_\ell = \min\ & \gamma \\ \text{s.t. }\ & \gamma - p(x) = s(x) + \left(1 - \textstyle\sum_{i=1}^d x_i^2\right) h(x) \\ & s \in \mathbb{R}[x] \text{ is a sum of squares, } \deg s \leq 2\ell \\ & h \in \mathbb{R}[x], \deg h \leq 2\ell - 2 \end{aligned} \tag{31}$$

The equality $\gamma - p(x) = s(x) + (1 - \sum_{i=1}^d x_i^2)h(x)$ in the constraint is an equality of polynomials in the polynomial ring $\mathbb{R}[x]$. Formulation (31) coincides precisely with the usual way the sum-of-squares hierarchy is formulated on general semialgebraic sets (the hierarchies arising from Putinar's and Schmüdgen-type Positivstellensatz coincide here because the semialgebraic set S^{d-1} is defined using a single polynomial equation). Also note that since $\deg p = 2n$ the first meaningful level of the hierarchy is $\ell = n$; otherwise (31) is not feasible. We finally note that another equivalent way of writing p_ℓ is the following:

$$p_\ell = \min \left\{ \gamma \in \mathbb{R} \text{ s.t. } \|x\|_2^{2(\ell-n)} (\gamma \|x\|_2^{2n} - p(x)) \text{ is a sum of squares in } \mathbb{R}[x] \right\}.$$

This is proved in [15, Theorem 3].

B Some technical results on polynomials on the sphere

We use the following lemma which appears in [34]. Recall that the Laplacian of a twice differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is $\Delta f = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2}$.

Lemma 16 [34] *If f is homogeneous polynomial of degree n on $\mathbb{R}^d$ and $\|f\|_\infty \leq M$, then $\|\Delta^k f\|_\infty \leq d^k (n)_{2k} M$, where Δ is the Laplace operator and $(n)_m := n(n-1)\cdots(n-(m-1))$ is the falling factorial.*

Proposition 5 (restatement) *For any homogeneous polynomial f with degree $2n$, denote its spherical harmonics decomposition as $f = \sum_{k=0}^{n} f_{2k}$ on S^{d-1} with $f_j \in \mathcal{H}_j^d$. Then for any k, it holds $\|f_{2k}\|_\infty \leq \mathsf{B}_{2n}\|f\|_\infty$, where B_{2n} is a constant that depends only on n (and independent of d). Also $\mathsf{B}_2 \leq 2$ and $\mathsf{B}_4 \leq 10$.*

Proof For simplicity of exposition we prove first the cases $2n = 2$ and $2n = 4$, before considering the general case. The result is immediate when $2n = 2$ with $\mathsf{B}_2 = 2$ since the harmonic decomposition of a quadratic polynomial is $f = f_0 + f_2$ with $f_0 = \int_{S^{d-1}} f d\sigma$. Then $|f_0| \leq \|f\|_\infty$ and $\|f_2\|_\infty = \|f - f_0\|_\infty \leq 2\|f\|_\infty$.

The first nontrivial case is $2n = 4$. The decomposition of a quartic polynomial on the sphere is $f = f_0 + f_2 + f_4$. Clearly $\|f_0\|_\infty \leq \|f\|_\infty$. We thus focus on bounding $\|f_2\|_\infty$. Since f is homogeneous note that $f(x)$ can be written as $f(x) = \|x\|^4 f_0 + \|x\|^2 f_2(x) + f_4(x)$ for all $x \in \mathbb{R}^d$. Using well-known identities concerning Laplacian one can check that

$$\begin{aligned}
\Delta(f(x)) &= \Delta(\|x\|^4 f_0 + \|x\|^2 f_2(x) + f_4(x)) \\
&= \Delta(\|x\|^4) f_0 + \Delta(\|x\|^2 f_2(x)) \\
&= 4(d+2)\|x\|^2 f_0 + \Delta(\|x\|^2) f_2(x) + 2\langle \nabla \|x\|^2, \nabla f_2(x)\rangle + \Delta(f_2(x)) \\
&= 4(d+2)\|x\|^2 f_0 + 2(d+2) f_2(x).
\end{aligned}$$

In the second and the last lines we used that $\Delta f_{2k}(x) = 0$ since the f_{2k} are harmonic. In the third line we used $\Delta(\|x\|^{2k}) = 2k(2k+d-2)\|x\|^{2k-2}$ and $\Delta(h \circ g) = (\Delta h)g + 2\langle \nabla h, \nabla g\rangle + h(\Delta g)$. In the last line we used the identity $\langle x, \nabla g(x)\rangle = 2kg(x)$ for any homogeneous polynomial g of degree $2k$ and $x \in S^{d-1}$. It thus follows that $f_2(x) = \frac{1}{2d+4}\Delta f(x) - 2\|x\|^2 f_0$. By Lemma 16 we know that $\|\Delta f\|_\infty \leq 12d\|f\|_\infty$. It thus follows that $\|f_2\|_\infty \leq \frac{12d}{2d+4}\|f\|_\infty + 2\|f\|_\infty \leq 8\|f\|_\infty$. Finally $\|f_4\|_\infty = \|f - (f_0 + f_2)\|_\infty \leq 10\|f\|_\infty$ by the triangle inequality. Thus $\mathsf{B}_4 \leq 10$.

We now proceed to prove the general case. Let $f(x)$ be a homogeneous polynomial of degree $2n$ and let $f = f_0 + f_2 + \cdots + f_{2n}$ be its spherical harmonic decomposition. Note that $f(x)$ has the following expression $f(x) = \sum_{k=0}^{n} \|x\|^{2(n-k)} f_{2k}(x)$ for all $x \in \mathbb{R}^d$. Since f_{2k} is a spherical harmonic, direct calculations give us

$$\Delta^m\left(\|x\|^{2(n-k)} f_{2k}(x)\right) = \begin{cases} r_{n,d,m,k}\|x\|^{2(n-k-m)} f_{2k}(x) & \text{if } m \leq n-k \\ 0 & \text{otherwise,} \end{cases}$$

where,

$$r_{n,d,m,k} = 4^m (n-k)_m \, (n+k+d/2-1)_m = O(d^m).$$

Thus by linearity, we have

$$\Delta^m f(x) = \sum_{k=0}^{n-m} \Delta^m\left(\|x\|^{2(n-k)} f_{2k}(x)\right) = \sum_{k=0}^{n-m} r_{n,d,m,k}\|x\|^{2(n-k-m)} f_{2k}(x) \qquad \forall x \in \mathbb{R}^d.$$

When restricted on the unit sphere, we have the inequalities

$$\begin{aligned}\|f_{2(n-m)}\|_\infty &\le \frac{\|\Delta^m f\|_\infty + \sum_{k=0}^{n-m-1} r_{n,d,m,k}\|f_{2k}\|_\infty}{r_{n,d,m,n-m}} \\ &\le \frac{d^m(2n)_{2m}\|f\|_\infty + \sum_{k=0}^{n-m-1} r_{n,d,m,k}\|f_{2k}\|_\infty}{r_{n,d,m,n-m}},\end{aligned}$$

where the second inequality follows from Lemma 16. For any $0 \le k \le n-m-1$, we have

$$\frac{r_{n,d,m,k}}{r_{n,d,m,n-m}} = \frac{4^m(n-k)_m\,(n+k+d/2-1)_m}{4^m(m)_m\,(n+(n-m)+d/2-1)_m} \le \frac{(n-k)_m}{(m)_m} \le n! \le (2n)!.$$

Moreover, for any $m \le n$ we also have

$$\frac{d^m(2n)_{2m}}{r_{n,d,m,n-m}} = \frac{(d/2)^m(2n)_{2m}}{2^m(m)_m\,(n+(n-m)+d/2-1)_m} \le \frac{(2n)_{2m}}{2^m(m)_m} \le (2n)!,$$

where the first inequality holds since each term in the falling factorial $(n+(n-m)+d/2-1)_m$ is no smaller than $d/2$. Thus we have

$$\|f_{2(n-m)}\|_\infty \le (2n)!\left[\|f\|_\infty + \sum_{k=0}^{n-m-1}\|f_{2k}\|_\infty\right].$$

By induction, we have the estimation

$$\|f_{2k}\|_\infty \le \|f\|_\infty(2n)!\,[1+(2n)!]^k \le \|f\|_\infty(2n)!\,[1+(2n)!]^n, \quad \forall k.$$

Thus we have $\mathsf{B}_{2n} \le (2n)!\,[1+(2n)!]^n$ independent of d. □

We can extend Proposition 5 to matrix-valued polynomials on the sphere.

Proposition 17 *Let $F(x) \in \mathbf{S}^k[x]$ be a $k \times k$ symmetric matrix-valued polynomial of degree $2n$. Let $\|F\|_\infty = \max_{x\in S^{d-1}}\|F(x)\|$ where $\|\cdot\|$ denotes the spectral norm. If $F = F_0 + F_2 + \cdots + F_{2n}$ is the harmonic decomposition of F, then $\|F_{2k}\|_\infty \le \mathsf{B}_{2n}\|F\|_\infty$ where B_{2n} is the constant from Proposition 5.*

Proof We can assume without loss of generality that $\|F\| = 1$. Note that $\|F\| = \max_{x\in S^{d-1}}\max_{y\in S^{k-1}}|y^{\mathsf{T}}F(x)y|$. For any fixed $y \in S^{k-1}$ define the real-valued polynomial $f_y(x) = y^{\mathsf{T}}F(x)y$. By assumption on F we know that

$\|f_y\|_\infty = \max_{x\in S^{d-1}} |y^\mathsf{T} F(x) y| \leq 1$. The spherical harmonic decomposition of f_y is given by $f_y(x) = \sum_{k=0}^{2n} y^\mathsf{T} F_{2k}(x) y$ since, for fixed y, $y^\mathsf{T} F_{2k}(x) y$ is a linear combination of the entries of F_{2k} which are all in $\mathcal{H}^d_{2k}$. It thus follows from Proposition 5 that $\|y^\mathsf{T} F_{2k}(\cdot) y\|_\infty \leq \mathsf{B}_{2n}$. This is true for all $y \in S^{k-1}$ thus we get $\max_{y\in S^{k-1}} \max_{x\in S^{d-1}} |y^\mathsf{T} F_{2k}(x) y| \leq \mathsf{B}_{2n}$, i.e., $\|F_{2k}\| \leq \mathsf{B}_{2n}$ as desired. □

We will also need the following technical result about Gegenbauer polynomials.

Proposition 18 *Let $C_i(t)$ be the Gegenbauer polynomial of degree i. Then for the curve of C_i lies above its tangent at $t = 1$, i.e., $C_i(t) \geq C_i'(1)(t-1) + C_i(1)$ for all $t \in [-1, 1]$.*

Proof Let $l(t) = C_i'(1)(t-1) + C_i(1)$. Let $\alpha = \max\left\{x \in (0,1) : C_i'(x) = 0\right\}$. It is known that $C_i(\alpha) < 0$ and that $|C_i(t)| \leq |C_i(\alpha)|$ for all $t \in [0, \alpha]$ (see [29, 18.14.16]). By standard arguments on orthogonal polynomials, we know that $C_i'' \geq 0$ on $[\alpha, \infty)$. Thus the inequality $C_i(t) \geq l(t)$ is true on $t \in [\alpha, 1]$. For $t \in [0, \alpha]$ it also has to be true since

$$C_i(t) \geq -|C_i(\alpha)| = C_i(\alpha) \geq l(\alpha) \geq l(t).$$

Using the fact that $C_i'(1)/C_i(1) = i(i+d-2)/(d-1)$, one can easily check that $l(0) \leq -C_i(1)$, and so $C_i(t) \geq -C_i(1) \geq l(t)$ for all $t \in [-1, 0]$.

C Duality relations DPS and SOS (Theorem 12)

In this section we prove that for any integer $\ell \geq 1$, we have the duality relation

$$\mathcal{DPS}^*_\ell = \left\{M : \|y\|^{2(\ell-1)} p_M \text{ is rsos}\right\}. \tag{32}$$

The key is the following lemma which gives a semidefinite programming characterization of the right-hand side of (32).

Lemma 19 *For any $M_{AB} \in Herm(\mathcal{H}_A \otimes \mathcal{H}_B)$ and integer $\ell \geq 1$, then $\|y\|^{2(\ell-1)} p_M$ is a rsos if and only if there exist positive semidefinite operators $W_{s,AB_{[\ell]}} \geq 0$, $s = 0, 1, \cdots, \ell$ such that*

$$\|y\|^{2(\ell-1)} p_M = \sum_{s=0}^{\ell} \left(x \otimes \bar{y}^{\otimes s} \otimes y^{\otimes \ell - s}\right)^\dagger W_{s,AB_{[\ell]}} \left(x \otimes \bar{y}^{\otimes s} \otimes y^{\otimes \ell - s}\right) \tag{33}$$

$$= \sum_{s=0}^{\ell} \left(x \otimes y^{\otimes \ell}\right)^\dagger W^{T_{B_{[s]}}}_{s,AB_{[\ell]}} \left(x \otimes y^{\otimes \ell}\right) \qquad \forall x \in \mathcal{H}_A, y \in \mathcal{H}_B. \tag{34}$$

The proof of the previous lemma is based on analyzing the biquadratic structure of p_M to see which monomials can appear in a sum-of-squares decomposition of $\|y\|^{2(\ell-1)} p_M$. The proof is deferred to the end of this section.

Using Lemma 19, the proof of (32) follows from standard duality arguments which we now explain.

First, we can dualize the semidefinite programming definition of DPS_ℓ to get

$$\mathcal{DPS}_\ell^* = \Big\{ M_{AB_1} : \; M_{AB_1} \otimes I_{B_{[2:\ell]}} = \big(Y_{AB_{[\ell]}} - \Pi_\ell Y_{AB_{[\ell]}} \Pi_\ell\big) + \sum_{s=0}^{\ell} W_{s,AB_{[\ell]}}^{\mathsf{T}_{B_{[s]}}} \\ \text{where } Y_{AB_{[\ell]}} \in \mathrm{Herm}, \; W_{s,AB_{[\ell]}} \geq 0, \forall s \in [0:\ell] \Big\}.$$

The variable $W_{s,AB_{[\ell]}}$ for $s = 0$ (resp. $s = 1, \ldots, \ell$) is the dual variable for the positivity constraint on $\rho_{AB_{[\ell]}}$ (resp. PPT constraint (25)).

Proof of Theorem 12 The proof consists of two directions. Assume $M \in \mathcal{DPS}_\ell^*$. Then there exists a Hermitian operator $Y_{AB_{[\ell]}}$, and positive semidefinite operators $W_{s,AB_{[\ell]}} \geq 0$, $s = 0, 1, \cdots, \ell$ such that

$$M_{AB_1} \otimes I_{B_{[2:\ell]}} = \big[Y_{AB_{[\ell]}} - (I \otimes \Pi_\ell) Y_{AB_{[\ell]}} (I \otimes \Pi_\ell)\big] + \sum_{s=0}^{\ell} W_{s,AB_{[\ell]}}^{\mathsf{T}_{B_{[s]}}}. \tag{35}$$

Recalling that Π_ℓ is the projector onto the symmetric subspace, we have $\Pi y^{\otimes \ell} = y^{\otimes \ell}$ for any vector y. Thus

$$\left(x \otimes y^{\otimes \ell}\right)^\dagger \left(Y_{AB_{[\ell]}} - (I \otimes \Pi_\ell) Y_{AB_{[\ell]}} (I \otimes \Pi_\ell)\right) \left(x \otimes y^{\otimes \ell}\right) = 0, \\ \forall\, x \in \mathcal{H}_A, y \in \mathcal{H}_B.$$

Evaluating Eq. (35) on both sides at the state $x \otimes y^{\otimes \ell}$, we have

$$\begin{aligned} \|y\|^{2(\ell-1)} p_M &= \left(x \otimes y^{\otimes \ell}\right)^\dagger M_{AB_1} \otimes I_{B_{[2:\ell]}} \left(x \otimes y^{\otimes \ell}\right) \\ &= \sum_{s=0}^{\ell} \left(x \otimes y^{\otimes \ell}\right)^\dagger W_{s,AB_{[\ell]}}^{\mathsf{T}_{B_{[s]}}} \left(x \otimes y^{\otimes \ell}\right). \end{aligned} \tag{36}$$

According to Proposition 19, we have $\|y\|^{2(\ell-1)} p_M$ is a rsos.

On the other hand, suppose $\|y\|^{2(\ell-1)} p_M$ is a rsos. From Proposition 19, there exists positive semidefinite operators $W_{s,AB_{[\ell]}} \geq 0$, $s = 0, 1, \cdots, \ell$ such that Eq. (36) holds. Since $y^{\otimes \ell}$ forms a basis on the symmetric subspace of $\mathcal{H}_{B_1} \otimes \mathcal{H}_{B_2} \otimes \cdots \otimes \mathcal{H}_{B_\ell}$, it implies that the operators

$$M_{AB_1} \otimes I_{B_{[2:\ell]}} \quad \text{and} \quad \sum_{s=0}^{\ell} W_{s,AB_{[\ell]}}^{\mathsf{T}_{B_{[s]}}}$$

coincide when restricted on the symmetric subspace $\mathrm{Sym}(\mathcal{H}^{\otimes \ell})$. That is,

$$(I \otimes \Pi_\ell)(M_{AB_1} \otimes I_{B_{[2:\ell]}})(I \otimes \Pi_\ell) = (I \otimes \Pi_\ell)\left(\sum_{s=0}^{\ell} W_{s,AB_{[\ell]}}^{\mathsf{T}_{B_{[s]}}}\right)(I \otimes \Pi_\ell).$$

Take the Hermitian operator

$$Y_{AB_{[\ell]}} := M_{AB_1} \otimes I_{B_{[2:\ell]}} - \sum_{s=0}^{\ell} W_{s,AB_{[\ell]}}^{\mathsf{T}_{B_{[s]}}}.$$

Then by the definition of $Y_{AB_{[\ell]}}$, we have $\Pi_\ell Y_{AB_{[\ell]}} \Pi_\ell = 0$ and

$$M_{AB_1} \otimes I_{B_{[2:\ell]}} = \left[Y_{AB_{[\ell]}} - (I \otimes \Pi_\ell) Y_{AB_{[\ell]}} (I \otimes \Pi_\ell)\right] + \sum_{s=0}^{\ell} W_{s,AB_{[\ell]}}^{\mathsf{T}_{B_{[s]}}},$$

which implies $M_{AB_1} \in \mathcal{DPS}_\ell^*$. □

It remains to prove Lemma 19. To have an easier understanding of the result in Lemma 19, let us first have a look at the special case on the second level of the hierarchy, i.e, $\ell = 2$. This will give us the key idea without loss of generality, and the higher level case is just a straightforward generalization.

Lemma 19 (special case $\ell = 2$) *For any Hermitian operator M_{AB_1}, we have that $\|y\|^2 p_M$ is rsos if and only if there exist positive semidefinite operators $W_{0,AB_1B_2}, W_{1,AB_1B_2}, W_{2,AB_1B_2} \geq 0$, such that*

$$\begin{aligned} \|y\|^2 p_M(x, \bar{x}, y, \bar{y}) = {} & (x \otimes y \otimes y)^\dagger W_{0,AB_1B_2}(x \otimes y \otimes y) \\ & + (x \otimes \bar{y} \otimes y)^\dagger W_{1,AB_1B_2}(x \otimes \bar{y} \otimes y) \\ & + (x \otimes \bar{y} \otimes \bar{y})^\dagger W_{2,AB_1B_2}(x \otimes \bar{y} \otimes \bar{y}) \qquad \forall x \in \mathcal{H}_A, y \in \mathcal{H}_B. \end{aligned} \tag{37}$$

Proof If there exist operators $W_0, W_1, W_2 \geq 0$ such that Eq. (37) holds, then $\|y\|^2 p_M$ can be shown to be rsos by using the spectral decompostion of W_i. For the converse suppose $\|y\|^2 p_M$ is rsos. Then there exist polynomials $f_m(x, \bar{x}, y, \bar{y})$ such that $\|y\|^2 p_M(x, \bar{x}, y, \bar{y}) = \sum_m f_m(x, \bar{x}, y, \bar{y})^2$. Since the monomials of $\|y\|^2 p_M(x, \bar{x}, y, \bar{y})$ are all of the forms $\bar{x}_i x_m \bar{y}_j y_k y_r \bar{y}_r$ (they are degree 2 in $(x, \bar{x})$ and degree 4 in $(y, \bar{y})$, then the possible monomials of $f_m(x, \bar{x}, y, \bar{y})$ can only be given by

$$\left\{x_i y_j y_k, x_i \bar{y}_j y_k, x_i \bar{y}_j \bar{y}_k, \bar{x}_i y_j y_k, \bar{x}_i y_j \bar{y}_k, \bar{x}_i \bar{y}_j \bar{y}_k\right\}.$$

The existence of any other monomials in $f_m(x, \bar{x}, y, \bar{y})$, such as $x_i x_j y_k$, will not be compatible with the monomials in $\|y\|^2 p_M(x, \bar{x}, y, \bar{y})$. Thus the most general form

of $f_m(x, \bar{x}, y, \bar{y})$ can be written as the linear combinations,

$$f_m(x, \bar{x}, y, \bar{y}) = \sum_{i,j,k} a^{m,0}_{i,j,k} x_i y_j y_k + \sum_{i,j,k} a^{m,1}_{i,j,k} x_i \bar{y}_j y_k + \sum_{i,j,k} a^{m,2}_{i,j,k} x_i \bar{y}_j \bar{y}_k + \sum_{i,j,k} b^{m,0}_{i,j,k} \bar{x}_i \bar{y}_j \bar{y}_k + \sum_{i,j,k} b^{m,1}_{i,j,k} \bar{x}_i y_j \bar{y}_k + \sum_{i,j,k} b^{m,2}_{i,j,k} \bar{x}_i y_j y_k.$$

Since $f_m(x, \bar{x}, y, \bar{y}) \in \mathbb{R}$, we have

$$\bar{a}^{m,0}_{i,j,k} = b^{m,0}_{i,j,k},\ \bar{a}^{m,1}_{i,j,k} = b^{m,1}_{i,j,k},\ \bar{a}^{m,2}_{i,j,k} = b^{m,2}_{i,j,k}, \quad \forall\, i, j, k, m.$$

Comparing the monomials of $\sum_m f_m(x, \bar{x}, y, \bar{y})^2$ and $\bar{x}_i x_m \bar{y}_j y_k y_r \bar{y}_r$, the terms, such as

$$\sum_m \left(\sum_{i,j,k} a^{m,0}_{i,j,k} x_i y_j y_k \right) \left(\sum_{i,j,k} a^{m,1}_{i,j,k} x_i \bar{y}_j y_k \right)$$

have to vanish, since the resulting monomial $x_i y_j y_k x_{i'} \bar{y}_{j'} y_{k'}$ is not compatible with $\bar{x}_i x_m \bar{y}_j y_k y_r \bar{y}_r$. After we get rid of those incompatible monomials, we have

$$\sum_m f_m(x, \bar{x}, y, \bar{y})^2 = 2 \sum_m \left(\left| \sum_{i,j,k} a^{m,0}_{i,j,k} x_i y_j y_k \right|^2 + \left| \sum_{i,j,k} a^{m,1}_{i,j,k} x_i \bar{y}_j y_k \right|^2 + \left| \sum_{i,j,k} a^{m,2}_{i,j,k} x_i \bar{y}_j \bar{y}_k \right|^2 \right).$$

Then we can construct matrices W_0, W_1, W_2 whose elements are respectively given by

$$(W_0)_{i,j,k;r,s,t} = 2 \sum_m \bar{a}^{m,0}_{i,j,k} a^{m,0}_{r,s,t},\ (W_1)_{i,j,k;r,s,t} = 2 \sum_s \bar{a}^{m,1}_{i,j,k} a^{m,1}_{r,s,t},\ (W_2)_{i,j,k;r,s,t} = 2 \sum_s \bar{a}^{m,2}_{i,j,k} a^{m,2}_{r,s,t}.$$

By construction, we know that $W_0, W_1, W_2 \geq 0$ and

$$\|y\|^2 p_M(x, \bar{x}, y, \bar{y}) = \sum_m f_m(x, \bar{x}, y, \bar{y})^2 = (x \otimes y \otimes y)^\dagger W_{0,AB_1B_2} (x \otimes y \otimes y) + (x \otimes \bar{y} \otimes y)^\dagger W_{1,AB_1B_2} (x \otimes \bar{y} \otimes y) + (x \otimes \bar{y} \otimes \bar{y})^\dagger W_{2,AB_1B_2} (x \otimes \bar{y} \otimes \bar{y}),$$

which completes the proof. □

Lemma 19 (general result, restatement) *For any Hermitian operator M_{AB_1} and integer $\ell \geq 1$, we have that $\|y\|^{2(\ell-1)} p_M$ is rsos if and only if there exist positive semidefinite operators $W_{s,AB_{[\ell]}} \geq 0$, $s \in [0:\ell]$ such that*

$$\|y\|^{2(\ell-1)} p_M = \sum_{s=0}^{\ell} \left(x \otimes \bar{y}^{\otimes s} \otimes y^{\otimes \ell - s}\right)^{\dagger} W_{s,AB_{[\ell]}} \left(x \otimes \bar{y}^{\otimes s} \otimes y^{\otimes \ell - s}\right) \tag{38}$$

$$= \sum_{s=0}^{\ell} \left(x \otimes y^{\otimes \ell}\right)^{\dagger} W_{s,AB_{[\ell]}}^{T_{B_{[s]}}} \left(x \otimes y^{\otimes \ell}\right) \qquad \forall x \in \mathcal{H}_A, y \in \mathcal{H}_B. \tag{39}$$

Proof Note that the second equality trivially holds due to the equation $x^{\dagger} Z x = (\bar{x})^{\dagger} Z^{\mathsf{T}} (\bar{x})$. We will prove the first equality. If Eq. (38) holds for positive semidefinite operators $W_{s,AB_{[\ell]}}$, then it is easy to check that $\|y\|^{2(\ell-1)} p_M(x, \bar{x}, y, \bar{y})$ is a rsos by using the spectral decomposition of $W_{s,AB_{[\ell]}}$.

On the other hand, if $\|y\|^{2(\ell-1)} p_M(x, \bar{x}, y, \bar{y})$ is a rsos, by definition there exist Hermitian polynomials $f_m(x, \bar{x}, y, \bar{y})$ such that $\|y\|^{2(\ell-1)} p_M(x, \bar{x}, y, \bar{y}) = \sum_m f_m(x, \bar{x}, y, \bar{y})^2$. In the following, we will compare the monomials on both sides of this equation and explicitly construct $W_{s,AB_{[\ell]}}$ from the coefficients of $\sum_m f_m(x, \bar{x}, y, \bar{y})^2$. We first note that the monomials of $\|y\|^{2(\ell-1)} p_M(x, \bar{x}, y, \bar{y})$ are all of the form

$$x_t \bar{x}_{t'} \prod_{i=1}^{\ell} y_{r_i} \bar{y}_{r_i'}, \tag{40}$$

which is of degree 2 and 2ℓ with respect to x and y, respectively. Then the possible monomials of $f_m(x, \bar{x}, y, \bar{y})$ can only be of degree 1 and ℓ with respect to x and y, respectively. That is, the possible monomials are given by

$$\left\{ x_t \prod_{i=1}^{s} \bar{y}_{r_i} \prod_{i=s+1}^{\ell} y_{r_i} \right\}_{s=0}^{\ell} \quad \text{and} \quad \left\{ \bar{x}_t \prod_{i=1}^{s} y_{r_i} \prod_{i=s+1}^{\ell} \bar{y}_{r_i} \right\}_{s=0}^{\ell},$$

where we denote the term $\prod_{i=s_1}^{s_2}(\cdot) = 1$ if $s_2 < s_1$. These monomials are basically formed by the ones with different number of complex conjugation over the symbol y. Therefore, the most general form of $f_m(x, \bar{x}, y, \bar{y})$ can be written as a linear combination of these monomials:

$$f_m(x, \bar{x}, y, \bar{y}) = \sum_{s=0}^{\ell} \sum_{t, r_{[\ell]}} a_{t,r_{[\ell]}}^{m,s} \left(x_t \prod_{i=1}^{s} \bar{y}_{r_i} \prod_{i=s+1}^{\ell} y_{r_i} \right) + \sum_{s=0}^{\ell} \sum_{t, r_{[\ell]}} b_{t,r_{[\ell]}}^{m,s} \left(\bar{x}_t \prod_{i=1}^{s} y_{r_i} \prod_{i=s+1}^{\ell} \bar{y}_{r_i} \right).$$

Since $f_m(x, \bar{x}, y, \bar{y}) \in \mathbb{R}$ for all x, y, we know that the coefficients between the conjugate monomials have to be conjugate with each other. That is, $\bar{a}^{m,s}_{t,r_{[\ell]}} = b^{m,s}_{t,r_{[\ell]}}$ holds for all $m, s, t, r_{[\ell]}$. Comparing the monomials of $\sum_m f_m(x, \bar{x}, y, \bar{y})^2$ and the monomials in Eq. (40), we have

$$\begin{aligned}\|y\|^{2(\ell-1)} p_M(x, \bar{x}, y, \bar{y}) &= \sum_m f_m(x, \bar{x}, y, \bar{y})^2 \\ &= 2 \sum_m \sum_{s=0}^{\ell} \left| \sum_{t, r_{[\ell]}} a^{m,s}_{t,r_{[\ell]}} \left(x_t \prod_{i=1}^{s} \bar{y}_{r_i} \prod_{i=s+1}^{\ell} y_{r_i} \right) \right|^2 .\end{aligned}$$

For any $s \in [0 : \ell]$, we construct the matrix W_s whose elements are given by

$$(W_s)_{t', r'_{[\ell]}; t, r_{[\ell]}} := 2 \sum_m \bar{a}^{m,s}_{t', r'_{[\ell]}} a^{m,s}_{t, r_{[\ell]}}.$$

Then we have that $W_s \geq 0, \forall s \in [0 : \ell]$ and

$$\begin{aligned}\|y\|^{2(\ell-1)} p_M(x, \bar{x}, y, \bar{y}) &= \sum_m f_m(x, \bar{x}, y, \bar{y})^2 \\ &= \sum_{s=0}^{\ell} \left(x \otimes \bar{y}^{\otimes s} \otimes y^{\otimes \ell - s} \right)^{\dagger} W_s \left(x \otimes \bar{y}^{\otimes s} \otimes y^{\otimes \ell - s} \right),\end{aligned}$$

which completes the proof. □

The above argument also works for csos polynomials with slight modifications.

Lemma 20 *For any Hermitian operator M_{AB_1} and integer $\ell \geq 1$, we have that $\|y\|^{2(\ell-1)} p_M$ is csos if and only if there exists a positive semidefinite operator $W_{AB_{[\ell]}} \geq 0$, such that*

$$\|y\|^{2(\ell-1)} p_M(x, \bar{x}, y, \bar{y}) = \left(x \otimes y^{\otimes \ell} \right)^{\dagger} W_{AB_{[\ell]}} \left(x \otimes y^{\otimes \ell} \right) \qquad \forall x \in \mathcal{H}_A, y \in \mathcal{H}_B. \tag{41}$$

Proof If there exists $W_{AB_{[\ell]}} \geq 0$ such that Eq. (41) holds, we can check that $\|y\|^{2(\ell-1)} p_M(x, \bar{x}, y, \bar{y})$ is a csos by using the spectral decomposition of W. On the other hand, if $\|y\|^{2(\ell-1)} p_M(x, \bar{x}, y, \bar{y})$ is a csos, by definition there exist polynomials $f_m(x, y)$ such that $\|y\|^{2(\ell-1)} p_M(x, \bar{x}, y, \bar{y}) = \sum_m |f_m(x, y)|^2$. In the following, we will compare the monomials on both sides of this equation and explicitly construct $W_{AB_{[\ell]}}$ from the coefficients of $\sum_m |f_m(x, y)|^2$. We first note that the monomials of $\|y\|^{2(\ell-1)} p_M(x, \bar{x}, y, \bar{y})$ are all of the form $x_t \bar{x}_{t'} \prod_{i=1}^{\ell} y_{r_i} \bar{y}_{r'_i}$, which is of degree 2 and 2ℓ with respect to x and y, respectively. Then the possible monomials of $f_m(x, y)$ can only be of degree 1 and ℓ with respect to x and y, respectively. Furthermore, by definition $f_m(x, y)$ are polynomials with respect to x, y alone, thus the only possible

monomial of $f_m(x, y)$ is $x_t \prod_{i=1}^{\ell} y_{r_i}$ and we have the general form of $f_m(x, y)$ as $f_m(x, y) = \sum_{t, r_{[\ell]}} a^m_{t, r_{[\ell]}} \left(x_t \prod_{i=1}^{\ell} y_{r_i}\right)$ with coefficients $a^m_{t, r_{[\ell]}}$. Define the matrix W with elements

$$W_{t', r'_{[\ell]}; t, r_{[\ell]}} := \sum_m \bar{a}^m_{t', r'_{[\ell]}} a^m_{t, r_{[\ell]}}.$$

Then we have $W \geq 0$, and

$$\|y\|^{2(\ell-1)} p_M(x, \bar{x}, y, \bar{y}) = \sum_m |f_m(x, y)|^2 = \left(x \otimes y^{\otimes \ell}\right)^{\dagger} W_{AB_{[\ell]}} \left(x \otimes y^{\otimes \ell}\right),$$

which completes the proof. □

Finally the result of $\mathcal{EXT}^*_{\ell} = \left\{M : \|y\|^{2(\ell-1)} p_M \text{ is csos}\right\}$ can be proved in a similar way by using Lemma 20.

References

1. Area, I., Dimitrov, D., Godoy, E., Ronveaux, A.: Zeros of Gegenbauer and Hermite polynomials and connection coefficients. Math. Comput. **73**(248), 1937–1951 (2004)
2. Barvinok, A.: A course in convexity, vol. 54. American Mathematical Society, Providence (2002)
3. Baxley, J.V.: Extreme eigenvalues of Toeplitz matrices associated with certain orthogonal polynomials. SIAM J. Math. Anal. **2**(3), 470–482 (1971)
4. Barak, B., Brandao, F.G., Harrow, A.W., Kelner, J., Steurer, D., Zhou, Y.: Hypercontractivity, sum-of-squares proofs, and their applications. In: Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing, pp. 307–326. ACM (2012)
5. Brandão, F.G., Christandl, M., Yard, J.: A quasipolynomial-time algorithm for the quantum separability problem. In: Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing, pp. 343–352 (2011)
6. Bhattiprolu, V., Ghosh, M., Guruswami, V., Lee, E., Tulsiani, M.: Weak decoupling, polynomial folds and approximate optimization over the sphere. In: 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), pp. 1008–1019. IEEE (2017)
7. Brandao, F.G., Harrow, A.W.: Quantum de finetti theorems under local measurements with applications. Commun. Math. Phys. **353**(2), 469–506 (2017)
8. Barak, B., Kothari, P.K., Steurer, D.: Quantum entanglement, sum of squares, and the log rank conjecture. In: Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, pp. 975–988. ACM (2017)
9. Blekherman, G.: Convexity properties of the cone of nonnegative polynomials. Discrete Comput. Geom. **32**(3), 345–371 (2004)
10. Brickman, L.: On the field of values of a matrix. Proc. Am. Math. Soc. **12**(1), 61–66 (1961)
11. Christandl, M., König, R., Mitchison, G., Renner, R.: One-and-a-half quantum de Finetti theorems. Commun. Math. Phys. **273**(2), 473–498 (2007)
12. Driver, K., Jordaan, K.: Bounds for extreme zeros of some classical orthogonal polynomials. J. Approx. Theory **164**(9), 1200–1204 (2012)
13. De Klerk, E.: The complexity of optimizing over a simplex, hypercube or sphere: a short survey. Central Eur. J. Oper. Res. **16**(2), 111–125 (2008)
14. de Klerk, E., Laurent, M.: Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere. arXiv:1904.08828 (2019)
15. de Klerk, E., Laurent, M., Parrilo, P.: On the equivalence of algebraic approaches to the minimization of forms on the simplex. In: Henrion, D., Garulli, A. (eds.) Positive Polynomials in Control, pp. 121–132. Springer, Berlin (2005)

16. D'Angelo, J.P., Putinar, M.: Polynomial optimization on odd-dimensional spheres. In: Putinar, M., Sullivant, S. (eds.) Emerging Applications of Algebraic Geometry, pp. 1–15. Springer, Berlin (2009)
17. Doherty, A.C., Parrilo, P.A., Spedalieri, F.M.: Distinguishing separable and entangled states. Phys. Rev. Lett. **88**(18), 187904 (2002)
18. Doherty, A.C., Parrilo, P.A., Spedalieri, F.M.: Complete family of separability criteria. Phys. Rev. A **69**(2), 022308 (2004)
19. Doherty, A.C., Wehner, S.: Convergence of SDP hierarchies for polynomial optimization on the hypersphere. arXiv:1210.5048, (2012)
20. Gurvits, L.: Classical deterministic complexity of Edmonds' problem and quantum entanglement. In: Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing, pp. 10–19. ACM (2003)
21. Horodecki, M., Horodecki, P., Horodecki, R.: Mixed-state entanglement and quantum communication. In: Alber, G., Beth, T., Horodecki, M., Horodecki, P., Horodecki, R., Rötteler, M., Weinfurter, H., Werner, R., Zeilinger, A. (eds.) Quantum Information, pp. 151–195. Springer, Berlin (2001)
22. Harrow, A.W., Montanaro, A.: Testing product states, quantum merlin-arthur games and tensor optimization. J. ACM (JACM) **60**(1), 1–43 (2013)
23. Hsu, H.-Y.: Certain integrals and infinite series involving ultra-spherical polynomials and Bessel functions. Duke Math. J. **4**(2), 374–383 (1938)
24. Koenig, R., Mitchison, G.: A most compendious and facile quantum de Finetti theorem. J. Math. Phys. **50**(1), 012105 (2009)
25. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**(3), 796–817 (2001)
26. Lewenstein, M., Bruß, D., Cirac, J.I., Kraus, B., Kuś, M., Samsonowicz, J., Sanpera, A., Tarrach, R.: Separability and distillability in composite quantum systems-a primer. J. Modern Opt. **47**(14), 2481–2499 (2000)
27. Nesterov, Y.: Random walk in a simplex and quadratic optimization over convex polytopes. Technical report, CORE (2003)
28. Navascues, M., Owari, M., Plenio, M.B.: The power of symmetric extensions for entanglement detection. Phys. Rev. A Atomic Mol. Opt. Phys. **80**(5), 1–16 (2009)
29. NIST Digital Library of Mathematical Functions. In: Olver, F.W.J., Olde Daalhuis, A.B., Lozier, D.W., Schneider, B.I., Boisvert, R.F., Clark, C.W., Miller, B.R., Saunders, B.V. (eds) http://dlmf.nist.gov/, Release1.0.22of2019-03-15
30. Parter, S.V.: Remarks on the extreme eigenvalues of Toeplitz forms associated with orthogonal polynomials. J. Math. Anal. Appl. **12**(3), 456–470 (1965)
31. Parrilo, P.A.: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D. thesis, California Institute of Technology (2000)
32. Parrilo, P.A.: Approximation quality of SOS relaxations (2013). Talk at ICCOPT 2013
33. Pólik, I., Terlaky, T.: A survey of the s-lemma. SIAM Rev. **49**(3), 371–418 (2007)
34. Reznick, B.: Uniform denominators in Hilbert's seventeenth problem. Mathematische Zeitschrift **220**(1), 75–97 (1995)
35. Woronowicz, S.L.: Positive maps of low dimensional matrix algebras. Rep. Math. Phys. **10**(2), 165–183 (1976)