

# MOMENT-SOS AND SPECTRAL HIERARCHIES FOR POLYNOMIAL OPTIMIZATION ON THE SPHERE AND QUANTUM DE FINETTI THEOREMS

ALEXANDER TAVEIRA BLOMENHOFER AND MONIQUE LAURENT

**ABSTRACT.** We revisit the convergence analysis of two approximation hierarchies for polynomial optimization on the unit sphere. The first one is based on the moment-sos approach and gives semidefinite bounds for which Fang and Fawzi (2021) showed an analysis in  $O(1/r^2)$  for the  $r$ th level bound, using the polynomial kernel method. The second hierarchy was recently proposed by Lovitz and Johnston (2023) and gives spectral bounds for which they show a convergence rate in  $O(1/r)$ , using a quantum de Finetti theorem of Christandl et al. (2007) that applies to complex Hermitian matrices with a “double” symmetry. We investigate links between these approaches, in particular, via duality of moments and sums of squares.

Our main results include showing that the spectral bounds cannot have a convergence rate better than  $O(1/r^2)$  and that they do not enjoy generic finite convergence. In addition, we propose alternative performance analyses that involve explicit constants depending on intrinsic parameters of the optimization problem. For this we develop a novel “banded” real de Finetti theorem that applies to real matrices with “double” symmetry. We also show how to use the polynomial kernel method to obtain a de Finetti type result in  $O(1/r^2)$  for real maximally symmetric matrices, improving an earlier result in  $O(1/r)$  of Doherty and Wehner (2012).

## 1. INTRODUCTION

Throughout we consider the polynomial optimization problem, asking to minimize a polynomial  $p$  over the unit sphere  $\mathbb{S}^{n-1}$ , i.e., to compute

$$(1) \quad p_{\min} = \min_{x \in \mathbb{S}^{n-1}} p(x) = \min\{p(x) : x \in \mathbb{R}^n, \|x\|_2 = 1\}.$$

Therein,  $p \in \mathbb{R}[x] = \mathbb{R}[x_1, \dots, x_n]$  is an  $n$ -variate polynomial of degree  $2d$ . In other words,  $p_{\min}$  is the largest scalar  $\lambda$  for which the polynomial  $p - \lambda$  is nonnegative on the unit sphere  $\mathbb{S}^{n-1}$ . Replacing this nonnegativity condition by existence of a sum-of-squares representation leads to the well-known sum-of-squares (sos, for short) bounds: for an integer  $r \geq d$ ,

$$(2) \quad \text{sos}_r(p) = \max_{\lambda \in \mathbb{R}} \{\lambda : p - \lambda = \sigma + u(1 - \|x\|_2^2), \sigma \in \mathbb{R}[x]_{\leq 2r} \text{ sos}, u \in \mathbb{R}[x]_{2r-2}\}.$$

The parameters  $\text{sos}_r(p)$  provide a hierarchy of lower bounds that converge asymptotically to  $p_{\min}$  (Lasserre [26]). They are also often referred to as the moment-sos bounds (see Lemma 5.4 for their ‘moment’ reformulation).

An alternative hierarchy of lower bounds  $\text{sp}_r(p)$  on  $p_{\min}$  has been recently introduced by Lovitz and Johnston [29] for homogeneous polynomials  $p$ , that are

---

*Date:* September 22, 2025.

*2020 Mathematics Subject Classification.* 90C23, 90C22, 81P45 .

*Key words and phrases.* polynomial optimization, sums of squares, quantum de Finetti, semidefinite programming, spectral bounds, moment problem.

University of Copenhagen.

Centrum Wiskunde & Informatica (CWI), and Tilburg University.

based on a (generalized) eigenvalue computation. Roughly speaking, these spectral bounds rely on very specific sums of squares representations of  $p - \lambda$  on the sphere, based on *doubly symmetric* representing matrices. We refer to Section 5 for the precise definition of the spectral bounds. We have  $\text{sp}_r(p) \leq \text{sos}_r(p) \leq p_{\min}$ . So, the sos bounds are stronger, but they are also more costly to compute than  $\text{sp}_r(p)$ ; indeed the latter boil down to an eigenvalue computation while the former rely on semidefinite programming.

Convergence analysis has been carried out for both hierarchies. Interestingly, there are common ingredients to both cases, like the use of the polynomial kernel method and the relevance to Quantum de Finetti type results. We discuss the polynomial kernel method in Section 3 and introduce the relevant Quantum de Finetti theorems in Section 2.4. However, there are also some important differences, as we now briefly sketch.

**Convergence analysis.** Lovitz and Johnston [29] show the following convergence rate  $p_{\min} - \text{sp}_r(p) = O(1/r)$  for their spectral bounds (see Theorem 5.1). A crucial ingredient in their proof is the Quantum de Finetti (abbreviated as QdF) theorem of Christandl et al. [9] for *doubly symmetric* complex Hermitian matrices (see Theorem 2.17). Since this result is stated for complex matrices, additional work is needed to be able to reach a conclusion for the spectral bounds, which deal with *real* polynomials. The strategy adopted in [29] is to establish a relation between the real optimization problem (1) and its complex analog (asking to minimize a polynomial on the complex unit sphere) to which one can then apply the above mentioned QdF theorem. The  $O(1/r)$  rate in the QdF theorem results in the  $O(1/r)$  convergence rate for the spectral bounds  $\text{sp}_r(p)$ .

The convergence rate of the sos bounds in (2) has been analyzed by Fang and Fawzi [18] who show  $p_{\min} - \text{sos}_r(p) = O(1/r^2)$  when  $p$  is homogeneous (see Theorem 3.2). This improves the earlier convergence analyses in  $O(1/r)$  by Reznick [38] and Doherty and Wehner [17].

These results follow roughly a similar approach, but there are some subtle differences. A common ingredient in both proofs is the use of the *polynomial kernel method*, a well-known method for constructing good approximations of a given function via convolution by a given polynomial kernel. The difference lies in the selection of the polynomial kernel, which relies on a kernel of the form  $(x^T y)^{2r}$  in [38, 17] and on a suitably optimized kernel in [18]. Fang and Fawzi [18] use their polynomial kernel to construct a sos polynomial representation that yields a feasible solution for the program (2), leading to their convergence analysis.

On the other hand, Doherty and Wehner [17] work in the symmetric tensor setting. Additionally, they use their kernel to derive a result on the dual moment side that leads to a ‘real Quantum de Finetti’ type result (see Theorem 2.19), which they then exploit to show their convergence analysis for the sos bounds. However, their real QdF theorem applies to real *maximally symmetric* matrices, a symmetry property that is much stronger than being doubly symmetric (see Definition 2.2).

**Our contributions.** As sketched above, there are similarities and differences between the sos and spectral bounds and their analyses. This raises several questions that we address in this paper.

A first question is whether the QdF theorem of [9] (Theorem 2.17) admits an analog for *real* doubly symmetric matrices. A positive answer would immediately imply a simple proof of the Doherty-Wehner real QdF (Theorem 2.19). As a first minor contribution, we give a negative answer: it is not possible to approximate a sequence of real  $r$ -doubly symmetric matrices (in fact, their partial traces) by real

separable ones (see Proposition 4.1 for the precise statement). We show this using a family of counterexamples inspired by Caves et al. [6].

A second question that we address in Section 3 is whether one can improve the analysis in the real QdF theorem (Theorem 2.19) of [17]. We give a positive answer (see Theorem 3.1). For this, we make the simple but important observation that the polynomial kernel method can also be used to show approximation results on the *dual moment side*, by considering *adjoint kernel operators*: we indicate how the kernel operator of Fang and Fawzi [18] can be used to strengthen the analysis of Doherty and Wehner [17] from  $O(1/r)$  to  $O(1/r^2)$ .

As a third contribution, we further investigate the spectral bounds. We show in Section 5.4 that, unlike the moment-sos bounds, they do not enjoy generic finite convergence (Theorem 5.14). Another natural question is whether the analysis of [29] in  $O(1/r)$  for the spectral parameters can possibly be improved and what is a *lower bound* on their exact convergence rate. As our fourth and main technical contribution, we show in Section 5.3 that  $p_{\min} - \text{sp}_r(p) = \Omega(1/r^2)$  for a suitably crafted 5-variate form of degree 4 (that we construct from the Choi-Lam form, well-known to be nonnegative but not sos). So, there is still a gap between the lower bound  $\Omega(1/r^2)$  and the upper bound  $O(1/r)$  and it would be very interesting to close this gap. In addition, at this moment, we are not aware of any lower bound for the convergence rate of the sos bounds; showing such a lower bound would also be very interesting.

In addition, we revisit in Section 4 the convergence analysis in  $O(1/r)$  for the spectral bounds from [29]. Rather than going via optimization on the complex unit sphere, we propose to directly use the QdF theorem to derive a ‘banded’ real QdF theorem (see Theorem 4.2). Roughly, this result approximates the symmetrized partial trace of a real doubly symmetric positive semidefinite matrix by a *scaled* real separable matrix. So, in a nutshell, the banded real QdF theorem interpolates between the QdF theorem (Theorem 2.17, that applies to complex Hermitian, doubly symmetric matrices) and the real QdF (Theorem 2.19, that applies to real maximally symmetric matrices). It thus falls within the realm of *real* quantum theory, where physical phenomena are modelled using real numbers that to many physicists appear more natural than complex ones. Whether real numbers suffice to model quantum theory is debated. In some settings they do, but Renou et al. [37] show that different predictions may be obtained when restricting to real numbers in certain (so-called entanglement-swapping) network scenarios.

The new banded real QdF theorem (Theorem 4.2) can then be directly used to recover the convergence rate in  $O(1/r)$  of [29] (recalled in Theorem 5.1) for the spectral bounds  $\text{sp}_r(p)$ , albeit with other constants than in [29]. First, in Theorem 5.5, our error analysis involves an explicit constant depending on  $n$  and  $d$ , instead of the unknown parameter  $\kappa(M)$  (the condition number of the maximally symmetric matrix representing  $s^d$ ) used in [29]. Second, in Theorem 5.7, our analysis involves the range  $p_{\max} - p_{\min}$  of values that the polynomial  $p$  takes on the sphere, instead of other constants of  $p$ . This dependency on the range of values is used, e.g., in the error analysis of the moment-sos bounds (see Theorem 3.2), and it is commonly used in continuous optimization (see, e.g., [45, 14]). Note that it is well-behaved under translating  $p$  by a scalar multiple of  $s^d$ : The value range does not change while the constant  $\|Q(p)\|_{\infty}$  (used in [29, Theorem 4.1], see Theorem 5.1) or  $\gamma(Q(p))$  (from Theorem 5.5) grows with the scaling factor (see Appendix D).

**Some background on existing literature.** Problem (1) is computationally hard already for homogeneous polynomials of degree 4. Indeed, it contains some known hard optimization problems, such as the maximum stable set graph problem (due

to a reformulation of Motzkin and Straus [30]), or deciding convexity of a quartic homogeneous polynomial, shown to be NP-hard in [1]. See, e.g., [11] for more about complexity issues. Huang [21] shows that the sos hierarchy has finite convergence for generic polynomials. Polynomial optimization on the unit sphere is relevant to a broad range of applications, e.g., for modelling matrix norms  $\|\cdot\|_{2 \rightarrow p}$ , for computing eigenvalues of tensors (see, e.g., [28]), for capturing entanglement versus separability of states in quantum information theory (see, e.g., [17], [18], [19], and further references therein).

Prior to the spectral lower bounds  $\text{sp}_r(p) \leq p_{\min}$  of [29] for optimization on the sphere, another spectral approach had been introduced earlier for minimization over a general compact set  $S$  by Lasserre [25]. He defines *upper bounds*  $\text{ub}_r(p) \geq p_{\min}$ , obtained by minimizing the expected value  $\int_S pq d\mu$  of  $p$  on  $S$ , taken over all sos polynomials  $q$  of degree at most  $2r$  such that  $\int_S q d\mu = 1$ , where  $\mu$  is a given measure supported on  $S$ . Computing  $\text{ub}_r(g)$  boils down to an eigenvalue computation [25]. These spectral upper bounds have been analyzed: a convergence rate in  $O(1/r^2)$  for  $\text{ub}_r(p) - p_{\min}$  is shown for the interval  $S = [-1, 1]$  in [12] (by establishing links to extremal roots of orthogonal polynomials), for the sphere in [13], and for many more sets  $S$  in [42]. Interestingly, there is an intimate connection between the spectral upper bounds and the sos lower bounds, which is (implicitly) exploited in the analysis of the sos lower bounds for the sphere in [18] and, more explicitly, for the boolean cube in [43] and the ball and the simplex in [41].

Finally, let us briefly mention some background on (quantum) de Finetti type results. De Finetti's theorem is a classical result about infinite sequences of random variables. In its original form [10], it roughly says the following: Assume the joint probability distribution of an infinite sequence of random variables  $X_1, X_2, \dots$  is invariant under permutations, in the sense that  $X_{\sigma(1)}, X_{\sigma(2)}, \dots$  has the same joint distribution for any bijective map  $\sigma: \mathbb{N} \rightarrow \mathbb{N}$  (also known as being *exchangeable*). Then, the joint distribution of  $X_1, X_2, X_3, \dots$  is a convex sum of product distributions.

Diaconis and Freedman [16] showed a finite version of this theorem: If  $X_1, \dots, X_r$  is an exchangeable sequence of  $r$  random variables with joint density  $p_r(x_1, \dots, x_r)$ , then, for any  $d \leq r$ , the joint density  $p_d(x_1, \dots, x_d)$  of the first  $d$  random variables may be approximated in a suitable norm by a convex combination of product densities  $q(x_1) \dots q(x_d)$ , where the  $q$ 's are univariate density functions. The error in the approximation is of the form  $O(\frac{1}{r})$  for constant  $d$ . See, e.g., [17, Section 6].

There exist "quantum" versions of de Finetti's theorem, both for the infinite and finite settings. In Theorem 2.17, we will present a finite quantum de Finetti theorem due to Christandl et al. [9]. It allows to approximate the partial trace of a complex density matrix  $M: (\mathbb{C}^n)^{\otimes r} \rightarrow (\mathbb{C}^n)^{\otimes r}$  with row and column permutation invariance (i.e., double symmetry) by a convex combination of product states. The approximation error is  $O(\frac{1}{r})$ , just as for the Diaconis-Freedman result [16], and  $O(\frac{1}{r})$  is known to be optimal for both results. An explanation of the philosophical aspects behind the classical and the quantum de Finetti theorem and its significance for quantum information is given by Caves, Fuchs and Schack [6]. The paper by Caves, Fuchs and Schack also serves as a friendly introduction to de Finetti theorems in quantum information.

Finally, let us mention the work of Navascués, Owari and Plenio [32] who show a finite quantum de Finetti result with an improved rate in  $O(1/r^2)$  for quantum bipartite states with a positive semidefinite partial transpose (aka the PPT condition).

We will mention in Section 6 some close connections between de Finetti type results and the classical moment problem.

**Organization of the paper.** In Section 2 we gather definitions and preliminaries needed for the paper, about polynomials, symmetric tensors, linear functionals, quantum de Finetti theorems, and harmonic polynomials. In Section 3 we show an improved real quantum de Finetti theorem for maximally symmetric matrices. We present a new ‘banded’ analog for real doubly symmetric matrices in Section 4, which we use to provide alternative analyses of the spectral bounds in Section 5. In the same section, we also prove a lower bound on the convergence rate of the spectral hierarchy and we show that the spectral hierarchy does not enjoy generic finite convergence. We close the paper with some discussion in Section 6.

## 2. PRELIMINARIES

We begin with some notation that is used throughout the paper.  $\mathbb{N} = \{0, 1, 2, \dots\}$  is the set of nonnegative integers. For an integer  $d \in \mathbb{N}$ ,  $\mathbb{N}_d^n$  denotes the set of sequences  $\alpha \in \mathbb{N}^n$  with  $|\alpha| := \alpha_1 + \dots + \alpha_n = d$ . We let  $\mathfrak{S}_d$  denote the set of permutations of the  $d$ -element set  $[d] = \{1, \dots, d\}$ . We let  $e_1, \dots, e_n$  denote the standard unit vectors in  $\mathbb{C}^n$ , where all entries of  $e_i$  are 0 except 1 at the  $i$ th coordinate. For a vector  $u \in \mathbb{C}^n$ ,  $\|u\|_2 = \sqrt{u^* u}$  denote its Euclidean norm. The unit sphere in  $\mathbb{R}^n$  is denoted  $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2} = 1\}$ . We let  $\text{conv}(S)$  denote the convex hull of a set  $S$  and  $\text{cone}(S)$  is its conic hull.

For  $\mathbb{K} = \mathbb{R}$  or  $\mathbb{C}$ , we use the trace inner product on the space  $\mathbb{K}^{n \times n}$  of matrices:  $\langle A, B \rangle = \text{Tr}(A^* B)$  for  $A, B \in \mathbb{K}^{n \times n}$ . Here,  $A^* = \bar{A}^T$  denotes the complex conjugate transpose of  $A$ ;  $A$  is Hermitian if  $A^* = A$ . For a Hermitian matrix  $A$ ,  $\lambda_{\max}(A)$  and  $\lambda_{\min}(A)$  denote, respectively, its largest and smallest eigenvalues. The matrix  $A$  is positive semidefinite, written as  $A \succeq 0$ , if  $\lambda_{\min}(A) \geq 0$ . For Hermitian matrices  $A, B$ , we write  $A \preceq B$  if  $B - A$  is positive semidefinite. Throughout,  $I_n$  denotes the  $n \times n$  identity matrix.

Let  $A$  be an  $n \times n$  Hermitian matrix with eigenvalues  $\lambda_1, \dots, \lambda_n$ . Its Schatten- $p$  norm is  $\|A\|_p = \sqrt[p]{\sum_{i=1}^n |\lambda_i|^p}$ . Hence,  $\|A\|_1 = \text{Tr}(A)$  if  $A \succeq 0$ . The dual norm of  $\|A\|_1$  is the Schatten  $\infty$ -norm  $\|A\|_\infty$ , which is the largest absolute value of an eigenvalue, i.e.,  $\|A\|_\infty = \max\{\lambda_{\max}(A), -\lambda_{\min}(A)\}$ . Note also that the Schatten 2-norm coincides with the Frobenius norm, so  $\|A\|_2 = \sqrt{\langle A, A \rangle}$ . We record the following relation for later use: if  $A$  has  $r$  nonzero eigenvalues, then

$$(3) \quad \|A\|_1 \leq \sqrt{r} \|A\|_2 \leq \sqrt{r} \|A\|_\infty.$$

In what follows we group definitions and preliminary facts needed for this paper about tensors, (homogeneous) polynomials and linear functionals on polynomials, Bose symmetric and maximally symmetric matrices, quantum de Finetti type results, and harmonic polynomials. These are somewhat scattered notions, so we give pointers to the literature throughout the text. In addition, for general background, we refer, e.g., to the survey about tensors by Kolda and Bader [22] and to the monographs by Lasserre [27] or Nie [34] about polynomials and moment theory, by Nielsen and Chuang [35] or Watrous [46] about quantum information theory, and by Müller [31] about harmonic polynomials.

**2.1. Symmetric tensors, doubly symmetric matrices and maximally symmetric matrices.** Let  $\mathbb{K} = \mathbb{R}$  or  $\mathbb{C}$  be the real or complex field. For integers  $r, n \geq 1$ ,  $(\mathbb{K}^n)^{\otimes r}$  denotes the space of  $r$ -tensors on  $\mathbb{K}^n$ , with elements  $v = (v_i)_{i \in [n]^r}$ . Here, we use the notation  $\underline{i} = (i_1, \dots, i_r) \in [n]^r$  (or simply  $i_1 \dots i_r$ ) for an  $r$ -sequence in  $[n]^r$ . Given two integers  $1 \leq d \leq r$ , it may also be convenient to denote  $r$ -sequences as the concatenation  $\underline{j}\underline{k}$  of a  $d$ -sequence  $\underline{j} \in [n]^d$  and an  $(r-d)$ -sequence  $\underline{k} \in [n]^{r-d}$ . For vectors  $u_i \in \mathbb{K}^{n_i}$ ,  $i = 1, 2$ , their tensor product is the vector  $u_1 \otimes u_2 \in \mathbb{K}^{n_1} \otimes \mathbb{K}^{n_2}$  with entries  $(u_1 \otimes u_2)_{i_1 i_2} = (u_1)_{i_1} (u_2)_{i_2}$  for  $(i_1, i_2) \in [n_1] \times [n_2]$ . For  $u \in \mathbb{K}^n$ ,  $u^{\otimes r} = u \otimes \dots \otimes u$  (taking the tensor product of  $r$  copies).

We now recall several notions of symmetry for tensors. In Appendix A we illustrate some of these notions on examples.

Permutations  $\sigma \in \mathfrak{S}_r$  act on  $[n]^r$  and thus on  $r$ -tensors, in the following way. For  $\underline{i} \in [n]^r$ , set  $\underline{i}^\sigma = (i_{\sigma(1)}, \dots, i_{\sigma(r)})$  and, for  $v \in (\mathbb{K}^n)^{\otimes r}$ , define  $v^\sigma$  as the  $r$ -tensor with entries  $(v^\sigma)_{\underline{i}} = v_{\underline{i}^\sigma}$  for  $\underline{i} \in [n]^r$ . Then,  $v$  is said to be a *symmetric tensor* if  $v^\sigma = v$  for all  $\sigma \in \mathfrak{S}_r$ , and  $S^r(\mathbb{K}^n)$  denotes the space of symmetric  $r$ -tensors on  $\mathbb{K}^n$ . We let  $\Pi_r$  denote the projection from the space  $(\mathbb{K}^n)^{\otimes r}$  to the symmetric subspace  $S^r(\mathbb{K}^n)$ . Concretely, we have

$$(4) \quad \Pi_r = \frac{1}{r!} \sum_{\sigma \in \mathfrak{S}_r} P_\sigma,$$

where  $P_\sigma$  is the square matrix indexed by  $[n]^r \times [n]^r$ , with entries  $(P_\sigma)_{\underline{i}, \underline{j}} = 1$  if  $\underline{j} = \underline{i}^\sigma$  and 0 otherwise, for  $\underline{i}, \underline{j} \in [n]^r$ .

Consider a matrix  $M$  indexed by the set  $[n]^r \times [n]^r$ , so that  $M$  represents an endomorphism from  $(\mathbb{K}^n)^{\otimes r}$  to  $(\mathbb{K}^n)^{\otimes r}$ . Then, any permutation  $\sigma \in \mathfrak{S}_r$  induces naturally a permutation of the columns (and rows) of  $M$ .

**Definition 2.1** (Doubly symmetric matrix). *A matrix  $M \in \mathbb{K}^{[n]^r \times [n]^r}$  is called  $r$ -column symmetric (resp.,  $r$ -row symmetric) if  $M\Pi_r = M$  (resp.,  $\Pi_r M = M$ ). If  $M$  is complex Hermitian or real symmetric, then  $M$  is  $r$ -column symmetric if and only if it is  $r$ -row symmetric, or equivalently if  $\Pi_r M \Pi_r = M$ , in which case  $M$  is called  $r$ -doubly symmetric (or just doubly symmetric if  $r$  is clear from the context).*

Hermitian doubly symmetric matrices are also known as *Bose symmetric* matrices in the quantum information literature. The following stronger symmetry property is sometimes needed (it is considered, e.g., in [17]).

**Definition 2.2** (Maximally symmetric matrix). *A matrix  $M \in \mathbb{K}^{[n]^r \times [n]^r}$  can be viewed as a  $2r$ -tensor  $\text{vec}(M) \in (\mathbb{K}^n)^{\otimes 2r}$ , with entries  $\text{vec}(M)_{\underline{i}\underline{j}} = M_{\underline{i}, \underline{j}}$  for  $\underline{i}, \underline{j} \in [n]^r$ . Then,  $M$  is said to be maximally symmetric when  $\text{vec}(M)$  is a symmetric  $2r$ -tensor.*

Note that a Hermitian matrix that is maximally symmetric is in fact real-valued (see also Lemma 2.5 below). Let  $\Pi_{\text{MaxSym}}$  denote the projection from the space  $\mathbb{K}^{[n]^r \times [n]^r}$  onto  $\text{MaxSym}_r(\mathbb{K}^n)$ , the space of maximally symmetric matrices. Concretely, the projection  $\Pi_{\text{MaxSym}}(M)$  is obtained by first constructing the symmetric  $2r$ -tensor  $\Pi_{2r}(\text{vec}(M))$  and then reshaping it as a square matrix indexed by  $[n]^r$  (by splitting the  $2r$  indices into two classes of size  $r$ , which can be done arbitrarily as the tensor is  $2r$ -symmetric).

**Example 2.3.** *As an illustration, for  $n = r = 2$ , we have*

$$(5) \quad \Pi_{\text{MaxSym}}(I_2^{\otimes 2}) = \begin{pmatrix} 1 & 0 & 0 & 1/3 \\ 0 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1 \end{pmatrix},$$

ordering the index set  $[2]^2$  as 11, 12, 21, 22 (corresponding to the standard basis elements ordered as  $e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2$ ). See Appendix A for details.

**Example 2.4.** *Given two vectors  $a, b \in \mathbb{R}^n$ , one can check that*

$$\begin{aligned} \Pi_{\text{MaxSym}}((aa^T + bb^T)^{\otimes 2}) &= aa^T \otimes aa^T + bb^T \otimes bb^T + \\ &\frac{1}{3}(aa^T \otimes bb^T + ab^T \otimes ab^T + ab^T \otimes ba^T + ba^T \otimes ab^T + ba^T \otimes ba^T + bb^T \otimes aa^T) \end{aligned}$$

We now examine what happens if we project an element of the form  $(uu^*)^{\otimes d}$  to the maximally symmetric subspace. The next result is shown in [29], but we give a proof in order to provide some insight.

**Lemma 2.5.** [29] *Let  $u \in \mathbb{C}^n$ . Write  $u = a + ib$ , with  $a, b \in \mathbb{R}^n$ . Then,*

$$\Pi_{\text{MaxSym}}((uu^*)^{\otimes d}) = \Pi_{\text{MaxSym}}((aa^T + bb^T)^{\otimes d}).$$

*In particular, the maximal symmetrization of  $(uu^*)^{\otimes d}$  is real valued for complex  $u$ .*

The expansion of  $\Pi_{\text{MaxSym}}((aa^T + bb^T)^{\otimes d})$  for  $d = 2$  is given in Example 2.4. As a tool to show Lemma 2.5, we use the notion of *partial transpose*. For a matrix  $M$  indexed by  $[n]^d \times [n]^d$ , taking the partial transpose w.r.t. the first register yields the matrix  $\tilde{M}$  with entries  $\tilde{M}_{i_1 i_2 \dots i_d, j_1 j_2 \dots j_d} = M_{j_1 i_2 \dots i_d, i_1 j_2 \dots j_d}$  for  $i_1, j_1, \dots, i_d, j_d \in [n]$ . This corresponds to applying to  $\text{vec}(M)$  the permutation  $\sigma \in \mathfrak{S}_{2d}$  that flips 1 and  $d+1$ . Taking the partial transpose w.r.t. to any subset of the  $d$  registers is defined analogously. Note  $\Pi_{\text{MaxSym}}(M) = \Pi_{\text{MaxSym}}(\tilde{M})$  for any partial transpose  $\tilde{M}$  of  $M$ .

*Proof of Lemma 2.5.* Note that  $uu^* + (uu^*)^T = uu^* + \bar{u}\bar{u}^* = 2aa^T + 2bb^T$ . Hence, taking the average of  $(uu^*)^{\otimes d}$  and its first partial transpose gives the matrix

$$\begin{aligned} \frac{1}{2}(uu^*)^{\otimes d} + \frac{1}{2}(\bar{u}\bar{u}^*) \otimes (uu^*)^{\otimes d-1} &= \frac{1}{2}(uu^* + \bar{u}\bar{u}^*) \otimes (uu^*)^{\otimes d-1} \\ &= (aa^T + bb^T) \otimes (uu^*)^{\otimes d-1}. \end{aligned}$$

Taking iteratively the average of the partial transposes at all other registers yields the matrix  $(aa^T + bb^T)^{\otimes d}$ . Hence,  $\Pi_{\text{MaxSym}}((uu^*)^{\otimes d}) = \Pi_{\text{MaxSym}}((aa^T + bb^T)^{\otimes d})$  by the above observation.  $\square$

We will also use the notion of *partial trace*.

**Definition 2.6** (Partial trace). *Consider a matrix  $M$  indexed by  $[n]^r \times [n]^r$  and an integer  $1 \leq d \leq r$ . Then, the partial trace  $\text{Tr}_{r-d}(M)$  is the matrix indexed by  $[n]^d \times [n]^d$  obtained by ‘tracing out the last  $r-d$  registers’ of  $(\mathbb{K}^n)^{\otimes r}$ , i.e., with entries  $(\text{Tr}_{r-d}(M))_{\underline{i}, \underline{j}} = \sum_{\underline{k} \in [n]^{r-d}} M_{\underline{i}\underline{k}, \underline{j}\underline{k}}$  for  $\underline{i}, \underline{j} \in [n]^d$ . In other words,*

$$\langle M, A \otimes I_n^{\otimes (r-d)} \rangle = \langle \text{Tr}_{r-d}(M), A \rangle \quad \text{for any } A \in \mathbb{K}^{[n]^d \times [n]^d}.$$

**2.2. Polynomials and linear functionals on polynomials.** Throughout, let  $R := \mathbb{R}[x] = \mathbb{R}[x_1, \dots, x_n]$  denote the polynomial ring in variables  $x = (x_1, \dots, x_n)$ . For an integer  $d \in \mathbb{N}$ ,  $R_d$  denotes its  $d$ -th graded component, which consists of the homogeneous polynomials (aka forms) with degree  $d$ , and  $R_{\leq d} = R_0 \oplus R_1 \oplus \dots \oplus R_d$  consists of the polynomials with degree at most  $d$ . We let  $[x]_d = (x^\alpha)_{\alpha \in \mathbb{N}_d^n}$  denote the vector of all monomials with degree  $d$ . Throughout we use the notation

$$s = x_1^2 + \dots + x_n^2.$$

**Definition 2.7** (Representing matrix of a polynomial). *A symmetric matrix  $G$  indexed by  $\mathbb{N}_d^n$  is said to be a representing matrix of a polynomial  $f \in R_{2d}$  if the polynomial identity  $f = [x]_d^T G [x]_d$  holds.*

A polynomial  $f \in R$  is a *sum of squares* (abbreviated as sos) if  $f = p_1^2 + \dots + p_N^2$  for some  $N \in \mathbb{N}$  and  $p_1, \dots, p_N \in R$ . Then, all  $p_i$  lie in  $R_d$  (resp.,  $R_{\leq d}$ ) if  $f \in R_{2d}$  (resp.,  $f \in R_{\leq 2d}$ ). The following characterization is well-known (see [8, 36]).

**Lemma 2.8.** *A polynomial  $f$  is a sum of squares if and only if it admits a positive semidefinite representing matrix (then, often called a Gram matrix of  $f$ ).*

Let  $\Sigma$  denote the cone of sos polynomials and set  $\Sigma_d = \Sigma \cap R_{2d}$ . Let  $\mathcal{P}$  denote the cone of globally nonnegative polynomials and set  $\mathcal{P}_d = \mathcal{P} \cap R_{2d}$ . So,  $\Sigma_d \subseteq \mathcal{P}_d$  are convex cones in  $R_{2d}$ .

Let  $\mathcal{P}(\mathbb{S}^{n-1})$  denote the cone of polynomials that are nonnegative on the sphere  $\mathbb{S}^{n-1}$  and set  $\mathcal{P}_d(\mathbb{S}^{n-1}) = \mathcal{P}(\mathbb{S}^{n-1}) \cap R_{2d}$ . Note that  $\mathcal{P}_d(\mathbb{S}^{n-1}) = \mathcal{P}_d$  (since nonnegativity on the sphere implies global nonnegativity for a homogeneous polynomial).

The dual cones  $\Sigma_d^*$  and  $\mathcal{P}_d^*$  live in the dual vector space  $R_{2d}^*$ , which consists of the linear functionals  $L : R_{2d} \rightarrow \mathbb{R}$ . The following notion of moment matrix is useful to characterize membership in the dual cones.

**Definition 2.9** (Moment matrix). *The moment matrix of  $L \in R_{2d}^*$  is the matrix  $M_d(L) = L([x]_d[x]_d^T)$ , where  $L$  is applied entry-wise.*

We refer to Lemma 2.13 and Proposition 2.14 below for a characterization of the dual cones  $\Sigma_d^*$  and  $\mathcal{P}_d^*$  in terms of properties of moment matrices.

**2.3. Homogeneous polynomials and symmetric tensors.** Let  $f \in R_d$  be homogeneous of degree  $d$  in  $n$  variables  $x_1, \dots, x_n$ . It is most common to write

$$(6) \quad f = \sum_{\alpha \in \mathbb{N}_d^n} f_\alpha x^\alpha = \langle (f_\alpha)_{\alpha \in \mathbb{N}_d^n}, [x]_d \rangle$$

in the (usual) monomial basis, for some scalars  $f_\alpha$  ( $\alpha \in \mathbb{N}_d^n$ ); we call this *multiplicity indexing* or *multi-indexing* of  $f$ . For a sequence  $\underline{i} = (i_1, \dots, i_d) \in [n]^d$ , set  $x^{\underline{i}} = x_{i_1} \cdots x_{i_d}$ . Then, one can alternatively write

$$(7) \quad f = \sum_{\underline{i} \in [n]^d} v_{\underline{i}} x^{\underline{i}} = \langle v, x^{\otimes d} \rangle,$$

for some (uniquely determined) symmetric  $d$ -tensor  $v \in S^d(\mathbb{R}^n)$ ; we call this *tensor-indexing* of  $f$ . To see this, for  $\underline{i} = (i_1, \dots, i_d) \in [n]^d$ , define  $\alpha(\underline{i}) \in \mathbb{N}_d^n$  with  $\alpha(\underline{i})_k$  being the number of occurrences of the symbol  $k \in [n]$  in the multiset  $\{i_1, \dots, i_d\}$ . Then,  $x^{\underline{i}} = x^{\alpha(\underline{i})}$  and  $v_{\underline{i}} = v_{\underline{j}}$  if  $\alpha(\underline{i}) = \alpha(\underline{j})$  (since  $v$  is a symmetric tensor). For a given  $\beta \in \mathbb{N}^n$ , the number of sequences  $\underline{i} \in [n]^d$  with  $\alpha(\underline{i}) = \beta$  equals the multinomial coefficient  $\binom{d}{\beta} = \frac{d!}{\beta_1! \cdots \beta_n!}$ . Comparing (6) and (7) gives  $f_\beta = \binom{d}{\beta} v_{\underline{i}}$  if  $\alpha(\underline{i}) = \beta$ .

As a consequence, moment matrices and representing matrices of polynomials may also be defined in tensor-indexing, as analogs of Definitions 2.7 and 2.9.

**Definition 2.10.** *Let  $f \in R_{2d}$ . A symmetric matrix  $G$  indexed by  $[n]^d$  is called a tensor-indexed representing matrix of  $f$  if the following polynomial identity holds:*

$$f = (x^{\otimes d})^T G x^{\otimes d} = \langle G, (xx^T)^{\otimes d} \rangle.$$

Then, any polynomial  $f \in R_{2d}$  admits a (unique) tensor-indexed representing matrix that is maximally symmetric, denoted by  $\text{MaxSym}(f)$ . Namely, we have  $\text{MaxSym}(f) = \Pi_{\text{MaxSym}}(G)$ , where  $G$  is any tensor-indexed representing matrix of  $f$ . Moreover, in analogy with Lemma 2.8, a polynomial  $f$  is a sum of squares if and only if it has a tensor-indexed positive semidefinite representing matrix.

**Definition 2.11.** *Let  $L \in R_{2d}^*$ . Then, the tensor-indexed moment matrix of  $L$  is*

$$L(x^{\otimes d}(x^{\otimes d})^T) = (L(x^{\underline{i}}x^{\underline{j}}))_{\underline{i}, \underline{j} \in [n]^d}.$$

Note that the tensor-indexed moment matrix  $L(x^{\otimes d}(x^{\otimes d})^T)$  of  $L \in R_{2d}^*$  arises from its multi-indexed moment matrix  $M_d(L) = L([x]_d([x]_d)^T)$  by repetitions of its rows/columns (with two rows/columns indexed by  $\underline{i}, \underline{j} \in [n]^d$  being equal when  $\alpha(\underline{i}) = \alpha(\underline{j})$ ). It is easy but important to realize that maximally symmetric matrices and tensor-indexed moment matrices are in fact equivalent notions (see [17]).

**Lemma 2.12.** *Let  $M$  be a real matrix indexed by  $[n]^d$ . The following are equivalent:*

- (i)  *$M$  is the tensor-indexed moment matrix of some linear functional  $L \in R_{2d}^*$ .*
- (ii)  *$M$  is maximally symmetric.*

*Proof.* (i)  $\implies$  (ii) follows since  $x^{\otimes d}(x^{\otimes d})^T$  is maximally symmetric, which implies  $M = L(x^{\otimes d}(x^{\otimes d})^T)$  is maximally symmetric. Conversely, assume  $M$  is maximally symmetric. Then, one can define  $L \in R_{2d}^*$  as follows. For  $\gamma \in \mathbb{N}_{2d}^n$ , pick  $\underline{k} \in [n]^{2d}$  such that  $\alpha(\underline{k}) = \gamma$ , i.e.,  $x^\gamma = x^{\underline{k}}$ . Then, write  $\underline{k} = \underline{i}\underline{j}$  as the concatenation of two sequences  $\underline{i}, \underline{j} \in [n]^d$  and set  $L(x^\gamma) = M_{\underline{i}, \underline{j}}$ . Note that this definition does not depend on the choice of  $\underline{k}$  since  $M$  is maximally symmetric. By construction,  $M$  is the tensor-indexed moment matrix of  $L$ , as desired.  $\square$

We now characterize the dual cones  $\Sigma_d^*$  and  $\mathcal{P}_d^* = \mathcal{P}_d(\mathbb{S}^{n-1})^*$ , both in terms of the (usual) moment matrices  $M_d(L)$  and their tensor analogs  $L(x^{\otimes d}(x^{\otimes d})^T)$ . The following is well-known and easy to see.

**Lemma 2.13.** *For  $L \in R_{2d}^*$ ,  $L \in \Sigma_d^* \iff M_d(L) \succeq 0 \iff L(x^{\otimes d}(x^{\otimes d})^T) \succeq 0$ .*

Given  $L \in R_{2d}^*$ , recall that a (Borel positive) measure  $\mu$  is called a *representing measure* of  $L$  if  $L(x^\alpha) = \int x^\alpha d\mu(x)$  holds for all  $\alpha \in \mathbb{N}_{2d}^n$ . Clearly,  $L \in \mathcal{P}_d^*$  if it has a representing measure. Classical results of Haviland [20] and Tchakaloff [44] show that linear functionals that are nonnegative on nonnegative polynomials of a given degree have a representing measure that can be chosen to be finite atomic. We need the following variant of these results for linear functionals on *homogeneous* polynomials, which may be found, e.g., in [40, Lemma 4.2].

**Proposition 2.14.** *Let  $L \in R_{2d}^*$ . The following assertions are equivalent.*

- (i)  *$L \in \mathcal{P}_d^*$  (equivalently,  $L \in \mathcal{P}(\mathbb{S}^{n-1})^*$ ).*
- (ii)  *$M_d(L) \in \text{cone}\{[u]_d[u]_d^T : u \in \mathbb{R}^n, \|u\| = 1\}$ .*
- (iii)  *$L(x^{\otimes d}(x^{\otimes d})^T) \in \text{cone}\{u^{\otimes d}(u^{\otimes d})^T : u \in \mathbb{R}^n\}$ .*

Throughout, we will use the norm  $\|\cdot\|_\infty$  on the polynomial space  $R_d$ , its dual norm  $\|\cdot\|_1$  on the dual space  $R_d^*$ , and (following [17]) the associated norm  $\|\cdot\|_{F1}$  for maximally symmetric matrices, defined as follows.

**Definition 2.15.** *For  $f \in R_d$ ,  $\|f\|_\infty = \max\{|f(x)| : x \in \mathbb{S}^{n-1}\}$  and, for  $L \in R_d^*$ ,  $\|L\|_1 = \max\{L(f) : f \in R_d, \|f\|_\infty = 1\}$ . For a maximally symmetric matrix  $M$  indexed by  $[n]^d$ , we set  $\|M\|_{F1} = \|L\|_1$ , where  $L \in R_{2d}^*$  is the linear functional whose moment matrix is  $M$ , i.e., such that  $L(x^{\otimes d}(x^{\otimes d})^T) = M$ .*

**Lemma 2.16.** *If  $f \in R_{2d}$  and  $G$  is a (tensor-indexed) matrix representing  $f$ , then*

$$(8) \quad \|f\|_\infty \leq \|G\|_\infty.$$

*If  $M$  is a real maximally symmetric matrix indexed by  $[n]^d$ , then*

$$(9) \quad \|M\|_1 \leq \|M\|_{F1}.$$

*Here,  $\|G\|_\infty$  and  $\|M\|_1$  are the Schatten norms.*

*Proof.* We first show (8). By definition, we have  $f(x) = \langle G, (xx^T)^{\otimes d} \rangle$ . Hence, for any unit vector  $x \in \mathbb{R}^n$ ,  $|f(x)| \leq \max\{\lambda_{\max}(G), -\lambda_{\min}(G)\} = \|G\|_\infty$ , showing (8). Now, we show (9). By duality of the Schatten 1- and  $\infty$ -norms,  $\|M\|_1 = \langle M, G \rangle$ , for some matrix  $G$  with  $\|G\|_\infty = 1$ . Let  $L \in R_{2d}^*$  be the linear functional with tensor-indexed moment matrix  $M$ , i.e.,  $M = L((xx^T)^{\otimes d})$ , and consider the polynomial  $f(x) = \langle G, (xx^T)^{\otimes d} \rangle$ , with representing matrix  $G$ . Then, we have

$$\langle M, G \rangle = \langle L((xx^T)^{\otimes d}), G \rangle = L(\langle (xx^T)^{\otimes d}, G \rangle) = L(f).$$

Hence,  $\|M\|_1 = \langle M, G \rangle = L(f) \leq \|L\|_1 \|f\|_\infty \leq \|M\|_{F1} \|G\|_\infty = \|M\|_{F1}$ , where in the last inequality we use  $\|L\|_1 = \|M\|_{F1}$  and (8). So, (9) holds.  $\square$

**2.4. Quantum de Finetti theorems.** We now have all tools in hand to introduce the Quantum de Finetti type theorems that are relevant to our treatment.

We first present the Quantum de Finetti theorem, due to Christandl, König, Mitchison and Renner [9]; we also refer to the exposition by Watrous [46, Theorem 7.26] and König and Renner [23]. We first state it for matrices  $M = vv^*$  with  $v \in S^r(\mathbb{C}^n)$ , as is customary in the literature. The result easily extends to the case when  $M$  is a positive semidefinite  $r$ -doubly symmetric Hermitian matrix.

**Theorem 2.17** (Christandl et al. [9], Quantum de Finetti, abbreviated as QdF). *Let  $v \in S^r(\mathbb{C}^n)$  be a unit vector. Then, for any integer  $1 \leq d \leq r$ , there exists*

$$(10) \quad \tau \in \mathcal{C}_d(\mathbb{C}^n) = \text{conv}(u^{\otimes d}(u^{\otimes d})^* : u \in \mathbb{C}^n, \|u\|_2 = 1)$$

such that

$$(11) \quad \|\text{Tr}_{r-d}(vv^*) - \tau\|_1 \leq \frac{4d(n-1)}{r+1}.$$

**Corollary 2.18** (Quantum de Finetti, for doubly symmetric Hermitian matrices). *Let  $M$  be a Hermitian matrix indexed by  $[n]^r$ . Assume  $M$  is doubly symmetric, positive semidefinite and  $\text{Tr}(M) = 1$ . Then, for any integer  $1 \leq d \leq r$ , there exists  $\tau \in \mathcal{C}_d(\mathbb{C}^n)$  (as defined in (10)) such that*

$$(12) \quad \|\text{Tr}_{r-d}(M) - \tau\|_1 \leq \frac{4d(n-1)}{r+1}.$$

*Proof.* Consider the spectral decomposition  $M = \sum_l \lambda_l v_l v_l^*$ , where the eigenvalues  $\lambda_l$  are nonnegative and the eigenvectors  $v_l$  are unit vectors in  $S^r(\mathbb{C}^n)$  (since  $M$  is doubly symmetric). By Theorem 2.17, for each  $l$  there exists  $\tau_l \in \mathcal{C}_d(\mathbb{C}^n)$  that satisfies (11) w.r.t.  $v_l$ . For  $\tau := \sum_l \lambda_l \tau_l$  one can easily see that (12) holds, since  $\|\text{Tr}_{r-d}(M) - \tau\|_1 \leq (\sum_l \lambda_l) \max_l \|\text{Tr}_{r-d}(v_l v_l^*) - \tau_l\|_1$  and  $\sum_l \lambda_l = \text{Tr}(M) = 1$ .  $\square$

The above results are stated for complex Hermitian matrices. One may wonder whether they extend to real doubly symmetric matrices, now achieving membership in the set

$$(13) \quad \mathcal{C}_d(\mathbb{R}^n) = \text{conv}(u^{\otimes d}(u^{\otimes d})^T \mid u \in \mathbb{R}^n, \|u\|_2 = 1),$$

(compare with the cone in Proposition 2.14(iii)), instead of the set  $\mathcal{C}_d(\mathbb{C}^n)$  from (10). The answer is negative, as we will show later in Proposition 4.1 by giving a family of counterexamples.

On the other hand, a real analog of Corollary 2.18 has been shown by Doherty and Wehner [17] when making a stronger assumption on the matrix  $M$ , namely assuming  $M$  is maximally symmetric instead of just doubly symmetric.

**Theorem 2.19** (Doherty and Wehner [17], Real QdF, for maximally symmetric matrices). *Let  $n, d, r \in \mathbb{N}$  with  $n \geq 3$  and  $r > d$ . Let  $M$  be a real symmetric matrix indexed by  $[n]^r$ . Assume  $M$  is maximally symmetric, positive semidefinite and  $\text{Tr}(M) = 1$ . Then, there exists a matrix  $\tau \in \mathcal{C}_d(\mathbb{R}^n)$  such that*

$$\|\text{Tr}_{r-d}(M) - \tau\|_{F1} \leq \frac{c_{n,d}}{r},$$

where  $c_{n,d}$  is a constant depending on  $n, d$ , and  $\|\cdot\|_{F1}$  is defined in Definition 2.15.

Finally, we recall some useful facts about harmonic polynomials, that we will use later to show an improved analysis for the above theorem (see Theorem 3.1).

**2.5. Harmonic polynomials.** A polynomial  $f \in R$  is called *harmonic* if it vanishes under the Laplace operator, i.e., if  $\Delta f := \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2} = 0$ . For  $r \geq 1$ , let  $H_r = \{f \in R_r : \Delta f = 0\}$  denote the set of harmonic polynomials of degree  $r$ . Then, the space  $R_r$  of homogeneous polynomials of degree  $r$  decomposes as the direct sum  $R_r = H_r \oplus sR_{r-2}$ , where  $s = x_1^2 + \dots + x_n^2$  (see Remark 2.20). In particular, this gives the decomposition of  $R_{2r}$  into harmonic subspaces:

$$(14) \quad R_{2r} = H_{2r} \oplus sH_{2r-2} \oplus s^2H_{2r-4} \oplus \dots \oplus s^rH_0.$$

This decomposition arises from the irreducible  $O(\mathbb{R}^n)$ -submodules of the polynomial space. Concretely, this means that every  $f \in R_{2r}$  has a unique decomposition

$$(15) \quad f = \sum_{k=0}^r s^{r-k} f_{2k},$$

where  $f_{2k} \in H_{2k}$  is harmonic of degree  $2k$ . As the polynomial  $s$  equals the constant 1 on the sphere, many authors prefer to write the harmonic decomposition as

$$(16) \quad f = \sum_{k=0}^r f_{2k} \text{ on } \mathbb{S}^{n-1}.$$

**Remark 2.20.** For the irreducibility of the spaces  $H_r$  as  $O(\mathbb{R}^n)$ -modules, we refer to [17, Section 2.4] and the references therein. For the reader's convenience we give a brief argument for the identity  $R_r = H_r \oplus sR_{r-2}$ , the decompositions (14) and (15) are then an immediate consequence. Observe that  $R_r$  may be endowed with the well-known apolar inner product (see, e.g., [2, eq. (8)]):  $\langle f, g \rangle = f(\partial_1, \dots, \partial_n)g$  for  $f, g \in R_r$ , where  $f(\partial_1, \dots, \partial_n)$  is a polynomial expression in partial derivatives. It follows directly from the definition that  $\langle f, sg \rangle = \langle \Delta f, g \rangle$  for any  $f \in R_r$  and  $g \in R_{r-2}$ . Hence,  $H_r$  is the orthogonal complement of  $sR_{r-2}$  and both spaces are  $O(\mathbb{R}^n)$ -invariant.

### 3. IMPROVED REAL QdF THEOREM FOR MAXIMALLY SYMMETRIC MATRICES

In this section, we revisit the real QdF theorem (Theorem 2.19) of Doherty and Wehner [17] for maximally symmetric matrices, and show an improved convergence rate in  $O(1/r^2)$  instead of  $O(1/r)$ . For this, we use the analysis of the sos bounds by Fang and Fawzi [18], who design a tailored linear operator  $\mathbf{K}_r$ , which approximates nonnegative polynomials on the sphere by sums of squares of degree  $2r$ . This linear operator  $\mathbf{K}_r$  is constructed from a polynomial integration kernel. We go one step further and use the adjoint polynomial kernel operator to show the following result.

**Theorem 3.1.** *Let  $M$  be a real symmetric matrix indexed by  $[n]^r$  and consider an integer  $d \in \{1, \dots, r\}$ . Assume  $M$  is maximally symmetric, positive semidefinite and  $\text{Tr}(M) = 1$ . Then, there exists  $\tau \in \mathcal{C}_d(\mathbb{R}^n)$  such that*

$$(17) \quad \|\text{Tr}_{r-d}(M) - \tau\|_1 \leq \|\text{Tr}_{r-d}(M) - \tau\|_{F1} \leq c_d \frac{n^2}{r^2} \text{ for all } r \geq c'_d n,$$

Here,  $c_d$  and  $c'_d$  are the constants depending only on  $d$  from Theorem 3.2 below,  $\|\cdot\|_{F1}$  is as in Definition 2.15,  $\|\cdot\|_1$  is the Schatten 1-norm, and  $\mathcal{C}_d(\mathbb{R}^n)$  is as defined in (13).

**3.1. Key ingredients.** Let us denote  $s = x_1^2 + \dots + x_n^2$  and

$$\mathcal{Q}_{2r} = \{\sigma + u(1 - s) : \sigma \text{ sos, } \deg(\sigma) \leq 2r, u \in \mathbb{R}[x]_{\leq 2r-2}\},$$

known as the quadratic module of the sphere, truncated at degree  $2r$ .

The above improved real QdF theorem for maximally symmetric matrices (Theorem 3.1) can be obtained as a consequence of the following result of Fang and Fawzi, more precisely, from their proof technique.

**Theorem 3.2** (Fang and Fawzi [18]). *Let  $p$  be an  $n$ -variate homogeneous polynomial of degree  $2d$  with  $d \geq 1, n \geq 3$ . Then, there exist constants  $c_d$  and  $c'_d$  only depending on  $d$ , such that, for all  $r \geq c'_d n$ , it holds that*

$$(18) \quad p - p_{\min} + c_d(p_{\max} - p_{\min}) \left(\frac{n}{r}\right)^2 \in \mathcal{Q}_{2r},$$

$$(19) \quad p_{\min} - \text{sos}_r(p) \leq c_d(p_{\max} - p_{\min}) \left(\frac{n}{r}\right)^2.$$

So, the above result shows that any homogeneous polynomial that is nonnegative on the sphere can be well approximated by a higher degree sum of squares and it gives a quantitative analysis for the quality of the approximation. As recalled earlier in Section 2 (Proposition 2.14), the cone of nonnegative polynomials on the sphere is in conic duality with the cone of linear functionals having a representing measure on the sphere. A natural question is whether there is an analog on the dual side of the quantitative result of Fang and Fawzi [18]. So, one may ask whether a linear functional that is nonnegative on sums of squares on the sphere can be approximated in “lower degree” by one that has a representing measure on the sphere. Informally, such a linear functional is given by its positive semidefinite moment matrix and lowering the degree will correspond to taking the partial trace.

As we explain below, one can derive such a result. A crucial reason why such a derivation is possible is that Fang and Fawzi [18] use a *linear* operator  $\mathbf{K}_r$  to approximate a nonnegative polynomial function on the sphere by a sum of squares.

Fang and Fawzi use the harmonic expansion (16), which was introduced in Section 2.5. They define an operator  $\bigoplus_{k=0}^{2r} H_{2k} \rightarrow \bigoplus_{k=0}^{2r} H_{2k}$ , which is invertible,  $O(n)$ -equivariant and acts on sums of harmonic polynomials of even degree. For technical reasons, we prefer to work with the formulation (15) of the harmonic expansion. This way, we stay within the framework of homogeneous polynomials and we obtain an operator  $\mathbf{K}_r: R_{2r} \rightarrow R_{2r}$ , simply by replacing  $\bigoplus_{k=0}^{2r} H_{2k}$  with the isomorphic space  $R_{2r}$ , as in (14). The operator  $\mathbf{K}_r: R_{2r} \rightarrow R_{2r}$  has the following properties:

- (P1)  $\mathbf{K}_r s^r = s^r$ .
- (P2) If  $f \in \mathcal{P}_r$ , then  $\mathbf{K}_r f \in \Sigma_r$ .
- (P3)  $\|\mathbf{K}_r^{-1} f - f\|_{\infty} \leq \varepsilon_r \|f\|_{\infty}$  for any  $f \in s^{r-d} R_{2d}$ .

Ideally, one aims for an operator  $\mathbf{K}_r$ , where  $\varepsilon_r$  is as small as possible. Fang and Fawzi [18] show that for every  $n, d$  and  $r$ , there exists an operator  $\mathbf{K}_r$  with  $\varepsilon_r \leq c_d \left(\frac{n}{r}\right)^2$ , if  $r \geq c'_d n$ . Here,  $c_d$  and  $c'_d$  are some constants, which only depend on  $d$ . The following lemma shows that Theorem 3.2 follows immediately from the existence of a “good” linear operator.

**Lemma 3.3.** *Assume that  $\mathbf{K}_r$  satisfies (P1), (P2) and (P3). For  $p \in R_{2d}$ , we have*

$$p - p_{\min} + \varepsilon_r(p_{\max} - p_{\min}) \in \mathcal{Q}_{2r}.$$

*Proof.* Define the polynomial  $q = (s^{r-d} p - p_{\min} s^r) / (p_{\max} - p_{\min}) \in s^{r-d} R_{2d} \subseteq R_{2r}$ , which satisfies  $q_{\min} = 0$ ,  $q_{\max} = 1$ , and  $\|q\|_{\infty} = 1$ . Set  $q_{\varepsilon} := q + \varepsilon_r s^r \in s^{r-d} R_{2d}$ . It suffices to show that  $q_{\varepsilon} \in \Sigma_r$ , since this implies that  $q_{\varepsilon} \in \mathcal{Q}_{2r}$  (by [14]), which in turn implies the desired relation in the lemma. We have  $\|\mathbf{K}_r^{-1}(q_{\varepsilon}) - q_{\varepsilon}\|_{\infty} = \|\mathbf{K}_r^{-1}(q) - q\|_{\infty} \leq \varepsilon_r$ , using (P1) for the equality and both (P3) and  $\|q\|_{\infty} = 1$  for the inequality. Therefore,  $\mathbf{K}_r^{-1}(q_{\varepsilon}) \geq q_{\varepsilon} - \varepsilon_r s^r = q \geq 0$  on the unit sphere  $\mathbb{S}^{n-1}$ . Thus,  $\mathbf{K}_r^{-1}(q_{\varepsilon}) \in \mathcal{P}_r$ , since  $\mathbf{K}_r^{-1}(q_{\varepsilon})$  is homogeneous. Using (P2), this implies  $q_{\varepsilon} = \mathbf{K}_r(\mathbf{K}_r^{-1}(q_{\varepsilon})) \in \Sigma_r$ .  $\square$

**3.2. The adjoint polynomial kernel operator.** Let  $\mathbf{K}_r: R_{2r} \rightarrow R_{2r}$  be the kernel operator from the previous section. We now consider the adjoint operator

$\mathbf{K}_r^*: R_{2r}^* \rightarrow R_{2r}^*$ . It maps  $L \in R_{2r}^*$  to  $\mathbf{K}_r^* L \in R_{2r}^*$ , defined by

$$(\mathbf{K}_r^* L)(f) = L(\mathbf{K}_r f) \text{ for } f \in R_{2r}.$$

For  $L \in R_{2r}^*$  and  $d \leq r$ , we let  $L|_{2d} \in R_{2d}^*$  denote the “restriction” to  $R_{2d}$ , defined by  $L|_{2d}(f) = L(s^{r-d}f)$  for any  $f \in R_{2d}$ . Before we prove Theorem 3.1, we need the following lemmas.

**Lemma 3.4.** *Let  $L \in R_{2r}^*$  and  $M = L(x^{\otimes r}(x^{\otimes r})^T)$  its (tensor-indexed) moment matrix. For  $d \leq r$ , the moment matrix of  $L|_{2d}$  equals the partial trace  $\text{Tr}_{r-d} M$ .*

*Proof.* Indeed,  $\text{Tr}_{r-d} M = \text{Tr}_{r-d}(L(x^{\otimes r}(x^{\otimes r})^T)) = L(\text{Tr}_{r-d}(x^{\otimes r}(x^{\otimes r})^T))$  (by linearity of  $L$ ). Combining with  $\text{Tr}_{r-d}(x^{\otimes r}(x^{\otimes r})^T) = s^{r-d}x^{\otimes d}(x^{\otimes d})^T$ , we obtain that  $\text{Tr}_{r-d} M = L(s^{r-d}x^{\otimes d}(x^{\otimes d})^T) = L|_{2d}(x^{\otimes d}(x^{\otimes d})^T)$ , where the last identity follows by the definition of  $L|_{2d}$ .  $\square$

**Lemma 3.5.** *The adjoint operator  $\mathbf{K}_r^*$  satisfies the following properties: for any  $L \in R_{2r}^*$ ,*

$$(P1^*) \quad (\mathbf{K}_r^* L)(s^r) = L(s^r).$$

$$(P2^*) \quad \text{If } L \geq 0 \text{ on } \Sigma_r, \text{ then } \mathbf{K}_r^* L \text{ has a representing measure on } \mathbb{S}^{n-1}.$$

$$(P3^*) \quad \|((\mathbf{K}_r^*)^{-1}L - L)|_{2d}\|_1 \leq \varepsilon_r \|L\|_1.$$

*Proof.* (P1\*) follows from (P1). We show (P2\*). For this, assume  $L \geq 0$  on  $\Sigma_r$ . For  $f \in \mathcal{P}_r$ , we have  $\mathbf{K}_r f \in \Sigma_r$  by (P2), and thus  $L(\mathbf{K}_r f) \geq 0$ , which gives  $(\mathbf{K}_r^* L)(f) \geq 0$ . Hence,  $\mathbf{K}_r^* L \geq 0$  on  $\mathcal{P}_r$ . By Proposition 2.14, we conclude that  $\mathbf{K}_r^* L$  has a representing measure on  $\mathbb{S}^{n-1}$ . We now show (P3\*). Let  $f \in s^{r-d}R_{2d}$ . Then,

$$|((\mathbf{K}_r^*)^{-1}L - L)(f)| = |L((\mathbf{K}_r)^{-1}f - f)| \leq \|L\|_1 \cdot \|(\mathbf{K}_r)^{-1}f - f\|_\infty,$$

using duality of the norms for the inequality. Combining with (P3) gives (P3\*).  $\square$

We now prove Theorem 3.1.

*Proof of Theorem 3.1.* Assume  $M$  is a matrix indexed by  $[n]^r$  that is maximally symmetric and positive semidefinite, with  $\text{Tr}(M) = 1$ . Let  $L \in R_{2r}^*$  be the associated linear functional, defined via  $L(x^{\otimes r}(x^{\otimes r})^T) = M$  (recall Lemma 2.12). Then,  $L(s^r) = \text{Tr } M = 1$  and, by Proposition 2.14,  $L \geq 0$  on  $\Sigma_r$ .

Using properties (P1\*) and (P2\*), we can conclude that the linear functional  $\mathbf{K}_r^* L$  has a representing probability measure on  $\mathbb{S}^{n-1}$ . This implies  $\|\mathbf{K}_r^* L\|_1 \leq 1$ . Indeed, if  $f \in R_{2r}$  satisfies  $\|f\|_\infty \leq 1$ , i.e.,  $-1 \leq f \leq 1$  on  $\mathbb{S}^{n-1}$ , then  $|\mathbf{K}_r^* L(f)| \leq 1$  and thus  $\|\mathbf{K}_r^* L\|_1 \leq 1$ . Moreover, applying property (P3\*) to  $\widehat{L} := \mathbf{K}_r^* L$  yields

$$(20) \quad \|(L - \mathbf{K}_r^* L)|_{2d}\|_1 = \|((\mathbf{K}_r^*)^{-1}\widehat{L} - \widehat{L})|_{2d}\|_1 \leq \varepsilon_r \|\widehat{L}\|_1 \leq \varepsilon_r.$$

Recall that if we choose  $\mathbf{K}_r$  to be the operator of Fang and Fawzi, we have  $\varepsilon_r \leq c_d(\frac{n}{r})^2$  for all  $r \geq c'_d n$ . Define the following matrix

$$\tau = (\mathbf{K}_r^* L)|_{2d}(x^{\otimes d}(x^{\otimes d})^T).$$

Then,  $\tau \in \mathcal{C}_d(\mathbb{R}^n)$ , since  $\mathbf{K}_r^* L$  has a representing probability measure on  $\mathbb{S}^{n-1}$ . Note also that  $\text{Tr}_{r-d}(M)$  is the tensor-indexed moment matrix of  $L|_{2d}$ , by Lemma 3.4. Hence, we have

$$\|\text{Tr}_{r-d}(M) - \tau\|_{F1} = \|(L - \mathbf{K}_r^* L)|_{2d}\|_1 \leq \varepsilon_r \leq c_d(\frac{n}{r})^2.$$

using the definition of the norm  $\|\cdot\|_{F1}$  in Definition 2.15 for the equality, and (20) for the inequality. This shows the bound for the F1-norm in (17). Combining with relation (9) in Lemma 2.16 gives the first inequality on the 1-norm in (17), which completes the proof.  $\square$

#### 4. A BANDED REAL QdF THEOREM FOR DOUBLY SYMMETRIC MATRICES

A natural question is whether the quantum de Finetti theorem (Theorem 2.17) extends to *real* doubly symmetric matrices. In view of the description of  $\mathcal{P}_d^*$  from Proposition 2.14, one would like to approximate the partial trace of a real doubly symmetric matrix by an element of the set  $\mathcal{C}_d(\mathbb{R}^n)$  from (13). Unfortunately, this is in general not possible, as we show in Section 4.1. However, in Section 4.2, we present a variation of the QdF theorem that does hold for real doubly symmetric matrices.

**4.1. A counterexample for a real analog of the QdF theorem.** Here, we show that the quantum de Finetti theorem (Theorem 2.17) does not extend to real doubly symmetric matrices. For this, we present a counterexample (for the case  $n = d = 2$ ), which is adapted from the example in [6, p. 21, Eq. (5.2) and the text thereafter] (for infinite QdF representations) to our setting.

For  $r \geq 2$ , define the matrices

$$(21) \quad \rho_r = \frac{1}{2}(uu^*)^{\otimes r} + \frac{1}{2}(\overline{u}u^*)^{\otimes r}, \quad \text{where } u = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \in \mathbb{C}^2.$$

Then,  $\rho_r$  is  $r$ -doubly symmetric, real valued and positive semidefinite, with trace 1. However, as we now show, there cannot exist elements  $\tau_r \in \mathcal{C}_2(\mathbb{R}^n)$  that are asymptotically ‘close’ to the partial traces  $\text{Tr}_{r-2}(\rho_r)$ .

**Proposition 4.1.** *For any  $r \geq 2$ , consider the matrices  $\rho_r$  from (21), that are real,  $r$ -doubly symmetric, positive semidefinite, with trace 1. There do not exist elements  $\tau_r \in \mathcal{C}_2(\mathbb{R}^n)$  ( $r \geq 2$ ) for which  $\lim_{r \rightarrow \infty} \|\text{Tr}_{r-2}(\rho_r) - \tau_r\|_1 = 0$ .*

*Proof.* First, observe that  $\text{Tr}_{r-2}(\rho_r) = \rho_2$  (since  $u$  is a unit vector). Assume for contradiction that such  $\tau_r \in \mathcal{C}_2(\mathbb{R}^n)$  would exist. As the set  $\mathcal{C}_2(\mathbb{R}^n)$  is compact, the sequence  $(\tau_r)_{r \geq 2}$  admits a converging subsequence, converging to  $\tau \in \mathcal{C}_2(\mathbb{R}^n)$ . At the limit, we obtain that  $\rho_2 = \tau \in \mathcal{C}_2(\mathbb{R}^n)$ . This yields a contradiction, since any element of  $\mathcal{C}_2(\mathbb{R}^n)$  is maximally symmetric, while  $\rho_2$  is not maximally symmetric since it is not real-valued.  $\square$

**4.2. A banded real QdF theorem.** So, no analog of the QdF theorem exists, for approximating partial traces of real doubly symmetric matrices by elements of  $\mathcal{C}_d(\mathbb{R}^n)$ . On the other hand, the assumption of Theorem 3.1 requiring a maximally symmetric matrix is very strong. We now state an approximation result that applies to real doubly symmetric matrices, but reaches a weaker conclusion, namely approximation of the maximal symmetrization of the partial trace of  $M$  by a *scaled* element of  $\mathcal{C}_d(\mathbb{R}^n)$  and a weaker approximation rate. So, the assumption of Theorem 4.2 is weaker than in Theorem 3.1, but the approximation guarantee is also weaker. Hence, the two results are incomparable.

**Theorem 4.2** (Banded real QdF for real doubly symmetric matrices). *Let  $M$  be a real symmetric matrix indexed by  $[n]^r$ . Assume  $M$  is doubly symmetric, positive semidefinite and  $\text{Tr}(M) = 1$ . Then, for any integer  $1 \leq d \leq r$ , there exist a scalar  $\alpha \in [\alpha_d, 1]$  and an element  $\tau \in \mathcal{C}_d(\mathbb{R}^n)$  such that*

$$(22) \quad \|\Pi_{\text{MaxSym}}(\text{Tr}_{r-d} M) - \alpha\tau\|_1 \leq \sqrt{N_{n,d}} \frac{4d(n-1)}{r+1}.$$

Here,  $N_{n,d} = \binom{n+d-1}{d}$  is the dimension of  $S^d(\mathbb{R}^n)$  and  $\alpha_d = 2^d / \binom{2d}{d}$ .

The following result is a crucial ingredient in the proof of Theorem 4.2; we extract it from the proof of [29, Theorem 4.1] and adapt it to our needs. For clarity and completeness, we give a sketch of proof in Appendix B.

**Lemma 4.3** (adapted from [29]). *Assume  $\tau \in \mathcal{C}_d(\mathbb{C}^n)$  with  $\text{Tr}(\tau) = 1$ , and define  $\alpha_d = 2^d / \binom{2d}{d}$ ,  $\alpha = \text{Tr}(\Pi_{\text{MaxSym}}(\tau))$ . Then,  $\frac{1}{\alpha} \Pi_{\text{MaxSym}}(\tau) \in \mathcal{C}_d(\mathbb{R}^n)$ , and  $\alpha_d \leq \alpha \leq 1$ .*

Lovitz and Johnson [29] use this result (within the proof of their Theorem 4.1, recalled later as Theorem 5.1) to reduce the analysis of the real polynomial optimization problem to its complex analog, for which there is a well-established spectral hierarchy based on the QdF theorem. We will use these same ingredients in a different way in order to show the above banded real QdF theorem (Theorem 4.2), which is then used in Section 5 to recover the analysis of the spectral bounds, however with different more explicit constants and better dependency on the polynomial.

*Proof of Theorem 4.2.* First, we use the quantum de Finetti theorem, in the form of Corollary 2.18, to pick  $\tau \in \mathcal{C}_d(\mathbb{C}^n)$  such that

$$(23) \quad \|\tau - \text{Tr}_{r-d} M\|_1 \leq \frac{4d(n-1)}{r+1}.$$

We claim:

$$(24) \quad \|\Pi_{\text{MaxSym}}(\text{Tr}_{r-d} M) - \Pi_{\text{MaxSym}}(\tau)\|_1 \leq \sqrt{N_{n,d}} \|\text{Tr}_{r-d}(M) - \tau\|_1.$$

Indeed, for the Hermitian matrix  $A = \text{Tr}_{r-d} M - \tau$ , we have

$$\|\Pi_{\text{MaxSym}}(A)\|_1 \leq \sqrt{N_{n,d}} \|\Pi_{\text{MaxSym}}(A)\|_2 \leq \sqrt{N_{n,d}} \|A\|_2 \leq \sqrt{N_{n,d}} \|A\|_1.$$

The first inequality follows from relation (3) applied to the matrix  $\Pi_{\text{MaxSym}}(A)$ , which has at most  $N_{n,d}$  nonzero eigenvalues since it is maximally symmetric. The second inequality follows from the fact that the Schatten 2-norm coincides with the Frobenius norm and thus it can only decrease under projection on a subspace. The last inequality follows using relation (3) for  $A$ .

Define the matrix  $\tau' := \Pi_{\text{MaxSym}}(\tau)$  and set  $\alpha := \text{Tr}(\tau')$ . Then, by Lemma 4.3,  $\tau'' := \frac{1}{\alpha} \tau' \in \mathcal{C}_d(\mathbb{R}^n)$  and  $\alpha_d \leq \alpha \leq 1$ . Combining (23) and (24), we obtain

$$\|\Pi_{\text{MaxSym}}(\text{Tr}_{r-d} M) - \alpha \tau''\|_1 \leq \sqrt{N_{n,d}} \frac{4d(n-1)}{r+1},$$

which completes the proof.  $\square$

**Remark 4.4.** *Note that in order to reach a good approximation of the matrix  $A := \Pi_{\text{MaxSym}}(\text{Tr}_{r-d} M)$  by  $(\alpha, \tau)$  in Theorem 4.2, the scalar  $\alpha$  must be very close to the trace of  $A$ . To see this, use the following inequality:*

$$\|A - \alpha \tau\|_1 = \max_{\|B\|_\infty=1} \langle A - \alpha \tau, B \rangle \geq \langle A - \alpha \tau, \pm I_n^{\otimes d} \rangle = |\text{Tr} A - \alpha|.$$

Hence, if  $\alpha$  and  $\tau$  satisfy relation (22), then

$$|\text{Tr} A - \alpha| \leq \sqrt{N_{n,d}} \frac{4d(n-1)}{r+1} \leq \varepsilon,$$

where the last inequality holds for any  $\varepsilon > 0$  and any  $r$  large enough. By Theorem 4.2 there exists a scalar  $\alpha'$  satisfying  $\alpha_d \leq \alpha' \leq 1$  and  $|\text{Tr} A - \alpha'| \leq \varepsilon$ . This implies  $\alpha_d - \varepsilon < \text{Tr} A < 1 + \varepsilon$ . In turn, it now follows that any scalar  $\alpha$  with  $\|A - \alpha \tau\|_1 \leq \sqrt{N_{n,d}} \frac{4d(n-1)}{r+1}$  must satisfy  $\alpha_d - 2\varepsilon < \alpha < 1 + 2\varepsilon$ . This shows that it is not possible to improve the approximation of  $A$  in Theorem 4.2 by making the “band” larger.

## 5. THE SPECTRAL HIERARCHY OF LOVITZ AND JOHNSTON

Lovitz and Johnston [29] proposed recently a hierarchy of spectral bounds  $\text{sp}_r(p)$  for the minimum value  $p_{\min}$  of a homogeneous polynomial  $p$  over the unit sphere,<sup>1</sup> see Section 5.1 below for their definition and relationship to the moment-sos bounds. Throughout, let  $p \in R_{2d}$ ,  $s = x_1^2 + \cdots + x_n^2$ , and let

$$(25) \quad Q(p) = \text{MaxSym}(p), \quad M = \text{MaxSym}(s^d)$$

denote the maximally symmetric matrices representing the polynomials  $p$  and  $s^d$ , respectively (as in Definition 2.10). They show the following convergence rate.

**Theorem 5.1.** [29, Theorem 4.3] *Let  $p \in R_{2d}$ , let  $Q(p)$  (resp.,  $M$ ) be the maximally symmetric matrix representing  $p$  (resp.,  $s^d$ ), and  $\alpha_d = 2^d / \binom{2d}{d}$ . For  $r \geq d$ ,*

$$p_{\min} - \text{sp}_r(p) \leq \|Q(p)\|_{\infty} \left(1 + \frac{\lambda_{\max}(M)}{\lambda_{\min}(M)}\right) \frac{1}{\alpha_d} \frac{4d(n-1)}{r+1},$$

Several questions arise that we will address in this section.

- Is it possible to show an improved rate, better than  $O(1/r)$ ? We cannot answer this, but what we can do is show that one cannot hope for a better rate than  $O(1/r^2)$ . For this, we construct an instance whose rate is in  $\Omega(1/r^2)$  (see Theorem 5.8).
- The constant in the error estimate of Theorem 5.1 involves the condition number  $\kappa(M) = \frac{\lambda_{\max}(M)}{\lambda_{\min}(M)}$  whose exact value is not known (see [29] and Remark 5.6 below). Is it possible to have another constant with explicit dependency in terms of  $n, d$ ? This is what we offer in Theorem 5.5, where we replace  $\kappa(M)$  by  $N_{n,d} = \binom{n+d-1}{d}$ . For this, we use the banded real QdF (Theorem 4.2) instead of the QdF (Theorem 2.17) used by Lovitz and Johnston [29] for showing Theorem 5.1.
- Can one show an error estimate in terms of the range of values  $p_{\max} - p_{\min}$  taken by  $p$  over the unit sphere? Such estimate would be in line with the analysis of the moment-sos bounds (in Theorem 3.2) and fit with what is customary in the optimization literature. We offer such an analysis in Theorem 5.7.
- Does the spectral hierarchy have generic finite convergence, as is the case for the moment-sos bounds? We give a negative answer (for degree 4) in Theorem 5.14.

**5.1. The spectral bounds.** We introduce the spectral hierarchy of Lovitz and Johnston [29]. Let  $p \in R_{2d}$  and recall the matrices  $Q(p)$  and  $M$  from (25). For an integer  $r \geq d$ , define the matrices

$$(26) \quad Q_r(p) = \Pi_r(Q(p) \otimes I_n^{\otimes(r-d)})\Pi_r,$$

$$(27) \quad M_r = \Pi_r(M \otimes I_n^{\otimes(r-d)})\Pi_r.$$

So,  $Q_r(p)$  and  $M_r$  are  $r$ -doubly symmetric matrices representing  $s^{r-d}p$  and  $s^r$ , respectively. That is,  $\langle Q_r(p), x^{\otimes r}(x^{\otimes r})^T \rangle = s^{r-d}p$  and  $\langle M_r, x^{\otimes r}(x^{\otimes r})^T \rangle = s^r$  hold. Then, the spectral bound  $\text{sp}_r(p)$  at the  $r$ -th level is defined as the smallest generalized eigenvalue of the matrix  $Q_r(p)$  with respect to  $M_r$ . We gather some equivalent formulations for  $\text{sp}_r(p)$ .

<sup>1</sup>Their hierarchy applies more generally to optimization of a multivariate tensor over a spherical Segre-Veronese variety. We only focus here on their hierarchy for polynomials on the sphere.

**Definition 5.2.** The spectral bound  $\text{sp}_r(p)$  is defined by any of the following equivalent programs:

$$(28) \quad \text{sp}_r(p) = \max_{\nu \in \mathbb{R}} \{\nu : Q_r(p) - \nu M_r \succeq 0\},$$

$$(29) \quad = \min_{v \in S^r(\mathbb{R}^n)} \{v^T Q_r(p) v : v^T M_r v = 1\},$$

$$(30) \quad = \min_X \{\langle Q_r(p), X \rangle : X \text{ } r\text{-doubly symmetric, } X \succeq 0, \langle M_r, X \rangle = 1\}.$$

Lovitz and Johnston [29] show that the parameters  $\text{sp}_r(p)$  provide lower bounds for  $p_{\min}$  that are at most as strong as the sos parameters  $\text{sos}_r(p)$  from (2).

**Proposition 5.3.** [29] We have  $\text{sp}_r(p) \leq \text{sp}_{r+1}(p)$  and  $\text{sp}_r(p) \leq \text{sos}_r(p) \leq p_{\min}$  for  $r \geq d$  and  $p \in R_{2d}$ .

For completeness and clarity, we give the proof in Appendix C. To ease comparison between the spectral and sos bounds, we recall some analogous equivalent formulations of the parameters  $\text{sos}_r(p)$ .

**Lemma 5.4.** The sos bound  $\text{sos}_r(p)$  can be equivalently defined by any of the following programs:

$$(31) \quad \text{sos}_r(p) = \max\{\lambda : \lambda \in \mathbb{R}, s^{r-d}p - \lambda s^r \text{ is sos}\},$$

$$(32) \quad = \min\{L(ps^{r-d}) : L \in R_{2r}^*, L(s^r) = 1, L \geq 0 \text{ on } \Sigma_r\},$$

$$(33) \quad = \min\{\langle Q(p) \otimes I_n^{(r-d)}, A \rangle : A \in \text{MaxSym}_r(\mathbb{R}^n), A \succeq 0, \text{Tr}(A) = 1\}.$$

*Proof.* For the equivalence between (2) and (31), see, e.g., [14]. The equivalence of (31) and (32) follows from conic duality, and use Proposition 2.14 for the equivalence between (32) and (33).  $\square$

**5.2. Convergence analysis of the spectral bounds.** Lovitz and Johnston [29] show their convergence analysis in  $O(1/r)$  (recall Theorem 5.1) using the QdF theorem (Theorem 2.17). In this section, we use our new banded real QdF theorem (Theorem 4.2) to give an alternative proof of this fact. First, in Theorem 5.5, we replace the (unknown) constant  $\lambda_{\max}(M)/\lambda_{\min}(M)$  appearing in the  $O(\cdot)$  notation, by the explicit constant  $N_{n,d}$ . The dependence on  $p$  in both results is via the eigenvalues of the representation matrix  $Q(p)$  (via  $\|Q(p)\|_{\infty}$  in Theorem 5.1 and via the parameter  $\gamma(Q(p))$  from (34) in Theorem 5.5). Second, in Theorem 5.7, we replace this dependence on the spectrum of  $Q(p)$  by the value range  $p_{\max} - p_{\min}$  of  $p$ .

**Theorem 5.5.** Let  $p \in R_{2d}$ . For any  $r \geq d$ , we have

$$p_{\min} - \text{sp}_r(p) \leq \gamma(Q(p)) \cdot \frac{\sqrt{N_{n,d}}}{\alpha_d} \cdot \frac{4d(n-1)}{r+1},$$

where  $\alpha_d = 2^d / \binom{2d}{d}$  is the constant from Theorem 4.2,  $N_{n,d} = \binom{n+d-1}{d}$ , and

$$(34) \quad \gamma(Q(p)) = \max\{\lambda_{\max}(Q(p)) - \text{sp}_d(p), p_{\min} - \lambda_{\min}(Q(p))\}.$$

*Proof.* To simplify notation, set  $Q := Q(p)$ . Let  $v \in S^r(\mathbb{R}^n)$  be an optimal solution of the program (29). The following identity (taken from [29]) holds, useful for later use:

$$(35) \quad \begin{aligned} 0 &= \langle Q_r(p) - \text{sp}_r M_r, vv^T \rangle = \langle (Q - \text{sp}_r M) \otimes I_n^{\otimes r-d}, vv^T \rangle \\ &= \langle Q - \text{sp}_r M, \text{Tr}_{r-d}(vv^T) \rangle \\ &= \langle \Pi_{\text{MaxSym}}(Q - \text{sp}_r M), \text{Tr}_{r-d} vv^T \rangle \\ &= \langle Q - \text{sp}_r M, \Pi_{\text{MaxSym}}(\text{Tr}_{r-d} vv^T) \rangle. \end{aligned}$$

Here, we use the fact that  $vv^T$  is doubly  $r$ -symmetric in the first line (to eliminate the projections  $\Pi_r$  in  $Q_r(p) - \text{sp}_r M_r$ ), the definition of the partial trace in the second line and, in the third line, the fact that  $Q - \text{sp}_r M$  is maximally symmetric.

By Theorem 4.2, there exist  $\tau \in \mathcal{C}_d(\mathbb{R}^n)$  and a scalar  $\alpha \in [\alpha_d, 2]$  such that

$$(36) \quad \|\tau - \alpha^{-1} \Pi_{\text{MaxSym}}(\text{Tr}_{r-d} vv^T)\|_1 \leq \frac{\sqrt{N_{n,d}}}{\alpha_d} \frac{4d(n-1)}{r+1}.$$

By definition,  $p_{\min} = \min_{\rho \in \mathcal{C}_d(\mathbb{R}^n)} \langle Q(p), \rho \rangle \leq \langle Q(p), \tau \rangle$ . We proceed to evaluate the range  $p_{\min} - \text{sp}_r(p)$ , setting  $\text{sp}_r = \text{sp}_r(p)$  for simpler notation:

$$(37) \quad \begin{aligned} p_{\min} - \text{sp}_r &\leq \langle Q, \tau \rangle - \text{sp}_r \\ &= \langle Q - \text{sp}_r M, \tau \rangle \\ &= \langle Q - \text{sp}_r M, \tau - \alpha^{-1} \Pi_{\text{MaxSym}}(\text{Tr}_{r-d} vv^T) \rangle \\ &= \langle Q - \text{sp}_r I_n^{\otimes d}, \tau - \alpha^{-1} \Pi_{\text{MaxSym}}(\text{Tr}_{r-d} vv^T) \rangle \\ &\leq \|Q - \text{sp}_r I_n^{\otimes d}\|_{\infty} \cdot \|\tau - \alpha^{-1} \Pi_{\text{MaxSym}}(\text{Tr}_{r-d} vv^T)\|_1 \end{aligned}$$

The first equality goes as follows:  $\langle M, \tau \rangle = \langle \text{MaxSym}(s^d), \tau \rangle = \langle I_n^{\otimes d}, \tau \rangle = 1$ , since  $\tau$  is maximally symmetric with trace 1. For the second equality, use (35) to see that we added a zero term. For the third equality, use again that  $\tau$  is maximally symmetric, so we can replace  $Q - \text{sp}_r M$  by  $Q - \text{sp}_r I_n^{\otimes d}$ . Finally, the last inequality uses Hölder's inequality (for the Schatten 1-norm and  $\infty$ -norm).

In view of (36), there remains only to show

$$(38) \quad \|Q - \text{sp}_r I_n^{\otimes d}\|_{\infty} \leq \gamma(Q).$$

By definition,  $\|Q - \text{sp}_r I_n^{\otimes d}\|_{\infty} = \max\{\lambda_{\max}(Q) - \text{sp}_r, \text{sp}_r - \lambda_{\min}(Q)\}$ . Note that

$$\text{sp}_d \leq \text{sp}_r \leq p_{\min} \leq p_{\max} \leq \lambda_{\max}(Q).$$

For the right most inequality, say  $p_{\max} = p(x)$  with  $x \in \mathbb{S}^{n-1}$ ; then,  $p_{\max} = p(x) = \langle Q, (xx^T)^{\otimes d} \rangle \leq \lambda_{\max}(Q)$ . Hence,  $\lambda_{\max}(Q) - \text{sp}_r \leq \lambda_{\max}(Q) - \text{sp}_d$  and  $\text{sp}_r - \lambda_{\min}(Q) \leq p_{\min} - \lambda_{\min}(Q)$ , which shows (38), as desired.  $\square$

**Remark 5.6.** In comparison to the factor  $\gamma(Q(p)) \cdot \frac{\sqrt{N_{n,d}}}{\alpha_d}$  appearing in Theorem 5.5, the corresponding error term in Theorem 5.1 is  $\|Q(p)\|_{\infty} \frac{1+\kappa(M)}{\alpha_d}$ , where  $\kappa(M) = \frac{\lambda_{\max}(M)}{\lambda_{\min}(M)}$  and  $M = \text{MaxSym}(s^d)$ . Compared to  $N_{n,d}^{1/2} = \binom{n+d-1}{d}^{1/2}$ , the constant  $\kappa(M)$  is not known explicitly. Lovitz and Johnston [29] observe that numerics suggest  $\kappa(M) = \binom{n/2+d-1}{\lfloor d/2 \rfloor}$ , which, if true, would imply that both  $\sqrt{N_{n,d}}$  and  $\kappa(M)$  are in the same order  $O(n^{d/2})$  for fixed  $d$ .

We next revisit the proof of Theorem 5.5 and replace the dependency on  $p$  by the value range  $p_{\max} - p_{\min}$ . Such a dependency is indeed preferable compared to the parameters  $\|Q(p)\|_{\infty}$  and  $\gamma(Q(p))$ , which do not behave well under shifting  $p$  by a scalar multiple of  $s^d$  (as mentioned in Appendix D).

**Theorem 5.7.** There exist constants  $c_{n,d} > 0$ , only depending on  $n$  and  $d$ , and  $r_0 \in \mathbb{N}$  such that

$$(39) \quad p_{\min} - \text{sp}_r(p) \leq (p_{\max} - p_{\min}) \cdot \frac{c_{n,d}}{r} \quad \text{for all } r \geq r_0.$$

*Proof.* We begin with observing that both  $p_{\min} - \text{sp}_r(p)$  and  $p_{\max} - p_{\min}$  remain invariant under shifting  $p$  by a scalar multiple of  $s^d$ . Hence, the inequality (39)

holds for  $p$  if and only if it holds for  $p + as^d$  for any scalar  $a \in \mathbb{R}$ . So, we may assume

$$(40) \quad p \in \bigoplus_{k=1}^d s^{d-k} H_{2k}.$$

Indeed, if  $p = cs^d + \sum_{k=1}^d s^{d-k} p_{2k}$  with  $p_{2k} \in H_{2k}$  and  $c \in \mathbb{R}$ , is the harmonic decomposition of  $p$  (as in (15)), then we may replace  $p$  by  $p - cs^d$ . Hence, we have

$$p_{\min} < 0 < p_{\max}.$$

Indeed, in view of (40),  $p$  is orthogonal to the constant polynomial in  $H_0$ . That is,  $\int_{\mathbb{S}^{n-1}} p(x) d\mu(x) = 0$ , where  $\mu$  is the Haar measure on the sphere (recall Remark 2.20). Since  $\lambda_{\min}(Q(p)) \leq p_{\min} < 0 < p_{\max} \leq \lambda_{\max}(Q(p))$ , this implies

$$\|Q(p)\|_{\infty} \leq \lambda_{\max}(Q(p)) - \lambda_{\min}(Q(p)).$$

Next, we show that the spectral range of  $Q(p)$  can be bounded in terms of the value range of  $p$ : there exists a constant  $c'_{n,d}$  (depending only on  $n$  and  $d$ ) such that

$$(41) \quad \lambda_{\max}(Q(q)) - \lambda_{\min}(Q(q)) \leq c'_{n,d}(q_{\max} - q_{\min})$$

for all  $q \in \bigoplus_{k=1}^d s^{d-k}$ . For this, consider the function

$$\Phi: \bigoplus_{k=1}^d H_{2k} \rightarrow \mathbb{R}, \quad q \mapsto \frac{\lambda_{\max}(Q(q)) - \lambda_{\min}(Q(q))}{q_{\max} - q_{\min}}.$$

Observe that  $\Phi$  is well-defined, since its domain does not contain constant functions, and we have  $\Phi \geq 1$ . Since  $\Phi$  is continuous and homogeneous of degree zero, it attains its maximum value, denoted  $c'_{n,d}$ , and the inequality (41) follows. We can now conclude the proof.

For this, we use the last line of Equation (37). So, it suffices to upper bound each of the two terms  $\|\tau - \alpha^{-1} \Pi_{\text{MaxSym}}(\text{Tr}_{r-d} vv^T)\|_1$  and  $\|Q - \text{sp}_r I_n^{\otimes d}\|_{\infty}$  (with  $Q = Q(p)$ ). The first term involving the 1-norm can be upper bounded by  $\frac{\sqrt{N_{n,d}}}{\alpha_d} \cdot \frac{4d(n-1)}{r+1}$  as in (36). For the second term involving the  $\infty$ -norm, we use the triangle inequality combined with the fact that  $\|I\|_{\infty} = 1$  and  $|\text{sp}_r| \leq |\text{sp}_d|$  (since  $\text{sp}_d \leq \text{sp}_r \leq p_{\min} < 0$ ), to obtain

$$\|Q - \text{sp}_r I\|_{\infty} \leq \|Q\|_{\infty} + |\text{sp}_d|.$$

Observe that  $\frac{M}{\lambda_{\min}(M)} \geq I$  and  $I \succeq -\frac{Q}{\|Q\|_{\infty}}$  (since  $\|Q\|_{\infty} \geq -\lambda_{\min}(Q)$ ). This implies  $Q + \frac{\|Q\|_{\infty}}{\lambda_{\min}(M)} M \succeq 0$ , and thus  $|\text{sp}_d| \leq \frac{\|Q\|_{\infty}}{\lambda_{\min}(M)}$  by the definition of the parameter  $\text{sp}_d = \text{sp}_d(p)$ . Putting things together and combining with (41), we obtain

$$\|Q - \text{sp}_r I\|_{\infty} \leq \|Q\|_{\infty} \left(1 + \frac{1}{\lambda_{\min}(M)}\right) \leq c'_{n,d} \left(1 + \frac{1}{\lambda_{\min}(M)}\right) (p_{\max} - p_{\min}),$$

giving  $p_{\min} - \text{sp}_r(p) \leq c_{n,d}(p_{\max} - p_{\min})$ , with  $c_{n,d} = c'_{n,d} \left(1 + \frac{1}{\lambda_{\min}(M)}\right) \frac{\sqrt{N_{n,d}}}{\alpha_d} 4d(n-1)$ .  $\square$

**5.3. A complexity lower bound for the spectral hierarchy.** As seen earlier, one can show a convergence rate in  $O(1/r)$  for the spectral hierarchy  $\text{sp}_r(p)$ . Whether one can show a sharper analysis in  $O(1/r^k)$  for some  $k > 1$  is not known. In this section we show that the best one can hope for is  $k = 2$ . For this, we are going to construct an explicit form  $p$  of degree 4 in five variables for which  $\text{sp}_r(p) = \Omega(1/r^2)$ .

We start by picking a quartic form in four variables that is nonnegative but not a sum of squares. A concrete example is given by the *Choi-Lam form*  $p_{CL}$ ,

$$(42) \quad p_{CL}(x) = x_1^2 x_2^2 + x_1^2 x_3^2 + x_2^2 x_3^2 + x_4^4 - 4x_1 x_2 x_3 x_4.$$

For simpler notation, denote by  $\text{sp}_r = \text{sp}_r(p_{CL})$  the spectral bounds, as defined in (28), and by  $\text{sos}_r = \text{sos}_r(p_{CL})$  the moment-sos lower bounds, as in (2). Then,  $\text{sp}_r \leq \text{sos}_r \leq (p_{CL})_{\min} = 0$  for any integer  $r \geq 2$ .

The Choi-Lam form  $p_{CL}$  is famously known to be a nonnegative polynomial, which is not a sum of squares (see [7]). Hence, *no* matrix representation of  $p_{CL}$  is positive semidefinite. This property is stronger than necessary. As it will turn out, our strategy for proving the lower bound  $\Omega(1/r^2)$  for the spectral hierarchy works with any nonnegative quartic form whose maximally symmetric representation is not positive semidefinite.

We use one small trick, which is to choose an unnatural representation of  $p_{CL}$ . While  $p_{CL}$  is intrinsically a polynomial in four variables  $x = (x_1, x_2, x_3, x_4)$ , one can artificially view it as a polynomial  $p$  in five variables  $\underline{x} := (x_1, x_2, x_3, x_4, u)$ , where  $u$  is an additional variable, by setting  $p(\underline{x}) = p_{CL}(x)$ . If, as a 4-variate polynomial,  $p_{CL}$  is represented by a matrix  $Q^{(4)} \in \mathbb{R}^{[4]^2 \times [4]^2}$ , then, as a 5-variate polynomial in  $\underline{x} = (x, u)$ , it is represented by the matrix

$$(43) \quad \begin{pmatrix} Q^{(4)} & 0 \\ 0^T & 0 \end{pmatrix} \in \mathbb{R}^{[5]^2 \times [5]^2}.$$

Note that the same holds for the maximally symmetric representations of  $p_{CL}$  in four variables and of  $p$  as a polynomial in five variables:

$$(44) \quad \text{MaxSym}(p) = \begin{pmatrix} \text{MaxSym}(p_{CL}) & 0 \\ 0^T & 0 \end{pmatrix}.$$

The reason is again that  $p$  is only supported on monomials, which do not depend on the fifth variable  $u$ .

**Theorem 5.8.** *Let  $\text{sp}_r(p)$  denote the spectral bound of order  $r$  for the Choi-Lam form  $p \in \mathbb{R}[x_1, x_2, x_3, x_4, u]$ . Then, we have  $p_{\min} = 0$  and  $\text{sp}_r(p) = \Omega(1/r^2)$ .*

*Proof.* Let the variables be ordered as  $\underline{x} = (x, u)$ , where  $x = (x_1, \dots, x_4)$  and  $u$  is the fifth variable. Let  $r \geq 2$ . Set  $s = x_1^2 + x_2^2 + x_3^2 + x_4^2 + u^2$  and

$$(45) \quad \begin{aligned} Q &:= \text{MaxSym}(p), \quad M := \text{MaxSym}(s^2), \\ Q_r &:= \Pi_r(Q \otimes I_5^{\otimes(r-2)})\Pi_r, \quad M_r := \Pi_r(M \otimes I_5^{\otimes(r-2)})\Pi_r. \end{aligned}$$

So,  $Q$  has the block-form as in (43), (44), where  $Q^{(4)} := \text{MaxSym}(p_{CL}) \in \mathbb{R}^{[4]^2 \times [4]^2}$  and the remaining blocks are filled by zeros. As in (4),  $\Pi_2 \in \mathbb{R}^{[5]^2 \times [5]^2}$  denotes the projection onto the symmetric subspace  $S^2(\mathbb{R}^5)$ . The space  $\mathbb{R}^{[4]^2}$  can be seen as a subspace of  $\mathbb{R}^{[5]^2}$  corresponding to the first four variable indices. Let us write  $\Pi_2^{(4)} \in \mathbb{R}^{[4]^2 \times [4]^2}$  for the restriction of  $\Pi_2$  to the subspace  $\mathbb{R}^{[4]^2}$ .

Recall  $\text{sp}_r = \text{sp}_r(p)$  for short. By the definition, we have  $\text{sp}_r \leq p_{\min} = 0$  and

$$Q_r + |\text{sp}_r| M_r = \Pi_r((Q + |\text{sp}_r| M) \otimes I_5^{\otimes(r-2)})\Pi_r \succeq 0.$$

By using  $M \preceq \lambda_{\max}(M) I_5^{\otimes 2}$ , we obtain that

$$G_r := \Pi_r((Q + \lambda_{\max}(M)|\text{sp}_r| I_5^{\otimes 2}) \otimes I_5^{\otimes(r-2)})\Pi_r \succeq 0.$$

In particular, every principal submatrix of  $G_r$  is positive semidefinite. We will now examine in detail the following principal submatrix

$$(46) \quad H := ((G_r)_{(i_1, i_2, 5, \dots, 5), (j_1, j_2, 5, \dots, 5)})_{i_1, i_2, j_1, j_2 \in \{1, \dots, 4\}} \in \mathbb{R}^{[4]^2 \times [4]^2}$$

of  $G_r$ , where the first two row indices  $i_1, i_2$  and the first two column indices  $j_1, j_2$  each range from 1 to 4, while the other  $r - 2$  row and column indices are set equal to 5.

Recall the notation  $\underline{i}^\sigma := (i_{\sigma(1)}, \dots, i_{\sigma(r)})$  for any  $\sigma \in \mathfrak{S}_r$ . For  $\underline{i}, \underline{j} \in [5]^r$ , using (4), we may expand

$$(G_r)_{\underline{i}, \underline{j}} = \left(\frac{1}{r!}\right)^2 \sum_{\sigma \in \mathfrak{S}_r} \sum_{\tau \in \mathfrak{S}_r} (Q \otimes I_5^{\otimes(r-2)} + \lambda_{\max}(M)|\text{sp}_r|I_5^{\otimes r})_{\underline{i}^\sigma, \underline{j}^\tau}.$$

In order to compute the submatrix  $H$ , we set the last  $r - 2$  indices of both  $\underline{i}$  and  $\underline{j}$  equal to 5 and the other two indices not equal to 5. Hence, on the right hand side, we have to count how many permutations “survive” when we do this, in the sense of giving a nonzero contribution to  $H$ . We may write

$$(47) \quad H = A + \lambda_{\max}(M)|\text{sp}_r|B,$$

where  $A$  is the contribution arising from  $Q \otimes I_5^{\otimes(r-2)}$  and  $B$  is the contribution from  $I_5^{\otimes r}$ . Since  $I_5^{\otimes r}$  is diagonal, the only nonzero entries of  $B$  are at indices of the form  $(i_1 i_2, i_1 i_2)$  and  $(i_1 i_2, i_2 i_1)$ . For later use, notice that  $(\Pi_2)_{i_1 i_2, i_1 i_2} = \frac{1}{2}$  if  $i_1 \neq i_2 \in [4]$ , and  $(\Pi_2)_{i_1 i_2, i_1 i_2} = 1$  for  $i_1 \in [4]$ . We now calculate the matrix  $B$ : for  $i_1, i_2 \in [4]$ ,

$$\begin{aligned} B_{i_1 i_2, i_1 i_2} &= B_{i_1 i_2, i_2 i_1} = \frac{1}{(r!)^2} \sum_{\sigma \in \mathfrak{S}_r} \sum_{\tau \in \mathfrak{S}_r} (I_5^{\otimes r})_{(i_1, i_2, 5, \dots, 5)^\sigma, (i_1, i_2, 5, \dots, 5)^\tau} \\ &= 2(\Pi_2)_{i_1 i_2, i_1 i_2} \frac{(r-2)!}{(r!)^2} \sum_{\sigma \in \mathfrak{S}_r} (I_5^{\otimes r})_{(i_1, i_2, 5, \dots, 5)^\sigma, (i_1, i_2, 5, \dots, 5)^\sigma} \\ &= 2(\Pi_2)_{i_1 i_2, i_1 i_2} \frac{(r-2)!}{r!} = \binom{r}{2}^{-1} (\Pi_2)_{i_1 i_2, i_1 i_2}. \end{aligned}$$

Here, in the second line, we used the fact that  $I_n^{\otimes r}$  is diagonal and that the row and column indices are equal if and only if  $(i_{\sigma(1)}, i_{\sigma(2)}) = (i_{\tau(1)}, i_{\tau(2)})$ . Given any permutation  $\sigma \in \mathfrak{S}_r$ , there are  $(r-2)!$  permutations  $\tau \in \mathfrak{S}_r$  with this property if  $i_1 \neq i_2$ , and  $2(r-2)!$  permutations  $\tau$  with this property if  $i_1 = i_2$ . Hence, there are  $2(r-2)!(\Pi_2)_{i_1 i_2, i_1 i_2}$  permutations  $\tau$  with this property for any given  $\sigma$ . In the third line, we used the fact that all diagonal entries of  $I_n^{\otimes r}$  are equal to 1, so the sum evaluates to  $r!$ . Summarized, we conclude that  $B = \binom{r}{2}^{-1} \Pi_2^{(4)}$ . Now, let us calculate  $A$ : for  $i_1, i_2, j_1, j_2 \in [4]$ ,

$$A_{i_1 i_2, j_1 j_2} = \frac{1}{(r!)^2} \sum_{\sigma \in \mathfrak{S}_r} \sum_{\tau \in \mathfrak{S}_r} (Q \otimes I_5^{\otimes(r-2)})_{(i_1, i_2, 5, \dots, 5)^\sigma, (j_1, j_2, 5, \dots, 5)^\tau}.$$

Recall that  $Q_{i,5} = 0$  if  $i \in \{1, \dots, 4\}$ . Therefore, the only nonzero contributions come from permutations  $\sigma$  and  $\tau$ , which keep all 5-indices ‘to the right’, i.e., at the last  $r - 2$  positions. There are  $2(r-2)!$  permutations  $\sigma$  and  $2(r-2)!$  permutations  $\tau$  with this property. They may be written as  $\sigma = \sigma_2 \circ \sigma_{r-2}$  and  $\tau = \tau_2 \circ \tau_{r-2}$ , respectively, with  $\sigma_2, \tau_2 \in \mathfrak{S}_2$  and  $\sigma_{r-2}, \tau_{r-2} \in \mathfrak{S}_{r-2}$ . So, we obtain

$$\begin{aligned} A_{i_1 i_2, j_1 j_2} &= \frac{((r-2)!)^2}{(r!)^2} \sum_{\sigma_2, \tau_2 \in \mathfrak{S}_2} (Q \otimes I_n^{\otimes(r-2)})_{(i_{\sigma_2(1)}, i_{\sigma_2(2)}, 5, \dots, 5), (j_{\tau_2(1)}, j_{\tau_2(2)}, 5, \dots, 5)} \\ &= \left(\frac{(r-2)!}{r!}\right)^2 \sum_{\sigma_2, \tau_2 \in \mathfrak{S}_2} Q_{i_{\sigma_2(1)} i_{\sigma_2(2)}, j_{\tau_2(1)} j_{\tau_2(2)}} \\ &= \binom{r}{2}^{-2} Q_{i_1 i_2, j_1 j_2}. \end{aligned}$$

In the first line, we used the fact that the  $(r-2)!^2$  permutations of the 5-indices all yield the same indices. In the second line, we used that  $(I_n^{\otimes r-2})_{5\dots 5, 5\dots 5} = 1$ . In the third line, we used that  $Q$  is maximally symmetric, hence in particular doubly symmetric. Thus, the  $2^2 = 4$  permutations of the non-5 indices all yield the same entry. Summarizing, we have shown

$$A = \binom{r}{2}^{-2} Q^{(4)}, \quad B = \binom{r}{2}^{-1} \Pi_2^{(4)}.$$

We conclude that the matrix  $H$  from (46)-(47) can be expressed as

$$\binom{r}{2}^2 H = Q^{(4)} + \lambda_{\max}(M) |\text{sp}_r| \binom{r}{2} \Pi_2^{(4)}.$$

Now, we use that  $H \succeq 0$  and  $\begin{pmatrix} \Pi_2^{(4)} & 0 \\ 0^T & 0 \end{pmatrix} \preceq I_5^{\otimes 2} \preceq \frac{1}{\lambda_{\min}(M)} M$  to obtain

$$0 \preceq \binom{r}{2}^2 \begin{pmatrix} H & 0 \\ 0^T & 0 \end{pmatrix} \preceq Q + |\text{sp}_r| \frac{\lambda_{\max}(M)}{\lambda_{\min}(M)} \binom{r}{2} M.$$

By definition,  $|\text{sp}_2|$  is the smallest value such that

$$Q + |\text{sp}_2| M \succeq 0.$$

Hence, it follows that

$$|\text{sp}_r| \frac{\lambda_{\max}(M)}{\lambda_{\min}(M)} \binom{r}{2} \geq |\text{sp}_2|, \quad \text{and thus} \quad |\text{sp}_r| \geq \frac{\lambda_{\min}(M)}{\lambda_{\max}(M)} |\text{sp}_2| \binom{r}{2}^{-1} = \Omega\left(\frac{1}{r^2}\right),$$

using the fact that  $M \succ 0$  [29] and  $|\text{sp}_2| > 0$  as  $p$  is not sos.  $\square$

As an application, we also obtain a complexity lower bound for the approximation result in Theorem 4.2.

**Corollary 5.9** (Complexity lower bound for the banded real QdF). *There exists a unit vector  $v \in S^r(\mathbb{R}^n)$  such that, for all scalars  $\alpha \in \mathbb{R}$  and  $\tau \in \mathcal{C}_d(\mathbb{R}^n)$ , one has*

$$(48) \quad \|\alpha\tau - \Pi_{\text{MaxSym}}(\text{Tr}_{r-2} vv^T)\|_1 = \Omega(1/r^2).$$

*In other words, the convergence rate in Theorem 4.2 is lower bounded by  $\Omega(1/r^2)$ .*

*Proof.* We may assume that  $\frac{1}{2}\alpha_d \leq \alpha \leq 2$ . For, if not, then Remark 4.4 shows that  $\|\alpha\tau - \Pi_{\text{MaxSym}}(\text{Tr}_{r-d} vv^T)\|_1$  is lower bounded by a constant and thus, (48) holds trivially. Hence,  $\alpha$  is bounded by constants, which only depend on  $d$ . Consider the Choi-Lam form in (42), viewed as a 5-variate polynomial  $p$  (so  $n = 5$ ,  $d = 2$ ), and let  $v \in S^2(\mathbb{R}^5)$  be an optimal solution to the program (29) defining the spectral bound  $\text{sp}_r(p)$ . Consider relation (37) applied to it, whose last line reads

$$\begin{aligned} |\text{sp}_r| &= p_{\min} - \text{sp}_r \leq \|Q(p) - \text{sp}_r I_5^{\otimes d}\|_{\infty} \cdot \|\tau - \alpha^{-1} \Pi_{\text{MaxSym}}(\text{Tr}_{r-2}(vv^T))\|_1 \\ &\leq \gamma(Q(p)) \cdot |\alpha^{-1}| \cdot \|\alpha\tau - \Pi_{\text{MaxSym}}(\text{Tr}_{r-2}(vv^T))\|_1, \end{aligned}$$

using (38) for the last inequality. By Theorem 5.8, we know that  $|\text{sp}_r| = \Omega(1/r^2)$ , which concludes the proof.  $\square$

**5.4. The spectral hierarchy does not have generic finite convergence.** A nice property of the moment-sos hierarchy is that it has *generic* finite convergence, as shown by Huang [21] (and by Nie [33] for generic semialgebraic sets). That is, the property:  $\text{sos}_r(p) = p_{\min}$  for some  $r \geq d$ , holds for “almost all” polynomials  $p$ . More precisely, given  $n, d \in \mathbb{N}$ , the set of polynomials  $p \in \mathbb{R}[x_1, \dots, x_n]_{2d}$  for which this property holds is Zariski-open.

Recall that the Zariski topology for a finite-dimensional vector space  $\mathbb{R}^N$  has as closed sets the sets  $\{y \in \mathbb{R}^N : F_1(y) = \dots = F_m(y) = 0\}$  for some polynomials

$F_1, \dots, F_m \in \mathbb{R}[y_1, \dots, y_N]$  and  $m \in \mathbb{N}$ . We will need the well-known fact that each nonempty Zariski-open set is open and dense with respect to both the Euclidean and the Zariski topology.

On the other hand, we now show that the spectral hierarchy  $\text{sp}_r(p)$  does not have generic finite convergence. For this, consider the set

$$(49) \quad A_{n,d} = \{p \in R_{2d} : \text{sp}_r(p) < p_{\min} \text{ for all } r \geq d\}$$

consisting of the  $n$ -variate degree  $2d$  forms for which the spectral hierarchy *does not* have finite convergence. So, showing generic finite convergence amounts to showing that the set  $A_{n,d}$  is Zariski-closed. We will use later the following easy fact: for any  $p \in R_{2d}$  and  $\lambda \in \mathbb{R}$ ,  $p \in A_{n,d}$  if and only if  $p + \lambda s^d \in A_{n,d}$ .

We will show that in the case  $d = 2$  the set  $A_{n,d}$  is not Zariski-closed, by exhibiting a Euclidean ball that is contained in it. We will show the result first for the case  $n = 2$  (Theorem 5.12) and then extend it to  $n \geq 2$  (Theorem 5.14). For this, we first establish a necessary condition for finite convergence, that depends only on the maximally symmetric matrix  $Q(p)$  representing  $p$  (and holds for any  $n, d$ ).

**Proposition 5.10.** *Let  $p \in R_{2d}$ . Assume that  $\text{sp}_r(p) = p_{\min}$  for some  $r \geq d$ . Then, for each minimizer  $a \in \mathbb{S}^{n-1}$  of  $p$ , we have  $(Q(p) - p_{\min}Q(s^d))a^{\otimes d} = 0$ .*

*Proof.* Assume  $\text{sp}_r(p) = p_{\min}$  for some  $r \geq d$ . Then,  $Q_r(p) - p_{\min}M_r \succeq 0$  holds. For each minimizer  $a \in \mathbb{S}^{n-1}$ , we have

$$(a^{\otimes r})^T(Q_r(p) - p_{\min}M_r)a^{\otimes r} = p(a) - p_{\min} = 0.$$

Since  $Q_r(p) - p_{\min}M_r$  is positive semidefinite, it follows that  $a^{\otimes r}$  lies in the kernel of  $Q_r(p) - p_{\min}M_r$ . From the definition of  $Q_r(p)$ , we conclude that

$$0 = \Pi_r((Q(p) - p_{\min}Q(s^d)) \otimes I_n^{\otimes r-d})a^{\otimes r} = \Pi_r((Q(p) - p_{\min}Q(s^d))a^{\otimes d} \otimes a^{\otimes r-d}).$$

Setting  $v = (Q(p) - p_{\min}Q(s^d))a^{\otimes d}$ , we have  $\Pi_r(v \otimes a^{\otimes r-d}) = 0$ . As we now observe, this implies  $v = 0$ . Indeed, for any  $x \in \mathbb{R}^n$ , we have

$$0 = \langle \Pi_r(v \otimes a^{\otimes r-d}), x^{\otimes r} \rangle = \langle v \otimes a^{\otimes r-d}, x^{\otimes r} \rangle = \langle v, x^{\otimes d} \rangle \langle a, x \rangle^{r-d}.$$

Hence,  $\langle v, x^{\otimes d} \rangle = 0$  for any  $x$  that is not orthogonal to  $a$  and thus for all  $x \in \mathbb{R}^n$  (by continuity). Since  $v$  is a symmetric tensor, this implies in turn that  $v = 0$ , which concludes the proof.  $\square$

This motivates introducing the following sets

$$S_{n,d} = \{p \in R_{2d} : p_{\min} = 0\},$$

$$T_{n,d} = \{p \in S_{n,d} : a^{\otimes d} \in \ker Q(p) \text{ for some } a \in \mathbb{S}^{n-1} \text{ such that } p(a) = 0\}.$$

By Proposition 5.10, if the spectral hierarchy has finite convergence for  $p \in S_{n,d}$ , then  $p \in T_{n,d}$ . In fact,  $p$  belongs to the smaller set, where one would require  $a^{\otimes d} \in \ker Q(p)$  for *all* zeros  $a \in \mathbb{S}^{n-1}$  of  $p$ . However, with the chosen definition of the set  $T_{n,d}$ , one can show that it is a closed set, a crucial property used later to show Theorem 5.12.

**Lemma 5.11.** *The set  $T_{n,d}$  is closed with respect to the Euclidean topology on  $R_{2d}$ .*

*Proof.* Consider a sequence of polynomials  $p_k \in T_{n,d}$  that has a limit  $p \in R_{2d}$ . As each  $p_k$  is nonnegative it follows that  $p$  is nonnegative. For each  $k$  there exists  $a_k \in \mathbb{S}^{n-1}$  such that  $p_k(a_k) = 0$  and  $Q(p_k)a_k^{\otimes d} = 0$ . By compactness of the sphere, the sequence  $(a_k)_k$  has an accumulation point  $a \in \mathbb{S}^{n-1}$ ; replacing the sequence by a subsequence we may assume  $a$  is the limit of  $(a_k)_k$ . Then,  $0 = p_k(a_k)$  converges to  $p(a)$ , showing  $p(a) = 0$  and thus  $p \in S_{n,d}$ . As the matrices  $Q(p_k)$  depend linearly on the coefficients of the polynomials  $p_k$ , it follows that  $0 = Q(p_k)a_k^{\otimes d}$  converges

to  $Q(p)a^{\otimes d}$ . Hence,  $Q(p)a^{\otimes d} = 0$ , which shows  $p \in T_{n,d}$ , which concludes the proof.  $\square$

**Theorem 5.12.** *For  $n = d = 2$ , the set  $A_{n,d}$  from (49) is not Zariski-closed. That is, the spectral hierarchy does not have generic finite convergence.*

*Proof.* Consider the bivariate polynomial

$$(50) \quad f = x_1^4 + 2x_2^4 - x_1^2x_2^2 - 2x_1x_2^3 = (x_1x_2 - x_2^2)^2 + (x_1^2 - x_2^2)^2 \in R_4.$$

So,  $f_{\min} = 0$  and  $f$  has two minimizers  $\pm \frac{1}{\sqrt{2}}(1, 1)$  in the sphere  $\mathbb{S}^1$ . Hence,  $f \in S_{2,2}$ . Moreover,  $f \notin T_{2,2}$ . Indeed, its maximally symmetric matrix representation reads

$$Q(f) = \begin{pmatrix} 1 & 0 & 0 & -1/6 \\ 0 & -1/6 & -1/6 & -1/2 \\ 0 & -1/6 & -1/6 & -1/2 \\ -1/6 & -1/2 & -1/2 & 2 \end{pmatrix}$$

(indexing it by  $x_1^2, x_1x_2, x_2x_1, x_2^2$ ), and we have  $Q(f)a^{\otimes 2} \neq 0$  for  $a = (1, 1)$ . Hence,  $f \in S_{2,2} \setminus T_{2,2}$ . Since the set  $T_{2,2}$  is closed (by Lemma 5.11), there exists  $\varepsilon > 0$  such that the ball  $B_f(\varepsilon)$  is contained in the complement of  $T_{2,2}$ , i.e.,  $B_f(\varepsilon) \cap T_{2,2} = \emptyset$ . Here,  $B_f(\varepsilon) = \{p \in R_4 : \|p - f\| < \varepsilon\}$  denotes the open ball with center  $f$  and radius  $\varepsilon$ , with respect to the  $\infty$ -norm on the sphere. Hence, in view of Proposition 5.10, we have  $B_f(\varepsilon) \cap S_{2,2} \subseteq A_{2,2}$ . We now claim that

$$(51) \quad B_f(\varepsilon/2) \subseteq B_f(\varepsilon) \cap S_{2,2} + \mathbb{R}s^2 \subseteq A_{2,2},$$

where  $B_f(\varepsilon/2)$  is the ball in  $R_4$  centered at  $f$  with radius  $\varepsilon/2$ . From this follows immediately that the set  $A_{2,2}$  is not Zariski-closed, which concludes the proof of the theorem.

The right most inclusion in (51) follows from the fact (observed earlier) that finite convergence of the spectral hierarchy is preserved under shifting by a scalar multiple of the constant polynomial  $s^2$ . We now verify the left most inclusion. For this, let  $p \in B_f(\varepsilon/2)$ , i.e.,  $\|p - f\| < \varepsilon/2$ . By evaluating at a minimizer  $a \in \mathbb{S}^1$  of  $f$ , we get  $|p(a)| < \varepsilon/2$ , implying  $p_{\min} \leq p(a) < \varepsilon/2$ . By evaluating at a minimizer  $b \in \mathbb{S}^1$  of  $p$ , we get  $-p_{\min} \leq f(b) - p(b) < \varepsilon/2$ , and thus  $|p_{\min}| < \varepsilon/2$ . Using the triangle inequality, we have  $\|p - p_{\min}s^2 - f\| \leq \|p - f\| + |p_{\min}| < \varepsilon/2 + \varepsilon/2 = \varepsilon$ . This shows that  $p - p_{\min}s^2 \in B_f(\varepsilon)$ , and thus  $p \in B_f(\varepsilon) \cap S_{2,2} + \mathbb{R}s^2$ , as desired.  $\square$

**Remark 5.13.** *Observe that the proof of Theorem 5.12 relies on the existence of a polynomial  $f \in S_{2,2} \setminus T_{2,2}$ . In other words, in order to show that the spectral hierarchy does not have generic finite convergence for  $n$ -variate forms of degree  $2d$ , it suffices to exhibit one concrete form  $F \in S_{n,d} \setminus T_{n,d}$ . As we next observe, one can easily construct such a form of degree 4 for any  $n \geq 2$ , simply by viewing the bivariate form  $f$  from (50) as an  $n$ -variate form. We expect that one can construct such  $F \in S_{n,d} \setminus T_{n,d}$  and thus show non-genericity of finite convergence of the spectral bounds for any  $d \geq 2$ , we leave showing this to further research.*

**Theorem 5.14.** *For  $n \geq 2$  and  $d = 2$ , the set  $A_{n,d}$  from (49) is not Zariski-closed. That is, the spectral hierarchy does not have generic finite convergence.*

*Proof.* Let  $F$  denote the  $n$ -variate degree 4 form obtained by viewing the bivariate form  $f$  from (50) as an  $n$ -variate form. Then, its maximally matrix representation has a block-form

$$\text{MaxSym}(F) = \begin{pmatrix} \text{MaxSym}(f) & 0 \\ 0 & 0 \end{pmatrix},$$

where  $\text{MaxSym}(f)$  is indexed by the pairs  $(i, j) \in [2]^2$  and the zero columns and rows are indexed by the pairs  $(i, j) \in [n]^2$  with  $i \geq 3$  or  $j \geq 3$  (thus the analog situation

as for the matrices in (44)). Clearly,  $F_{\min} = f_{\min} = 0$  and thus  $F \in S_{n,2}$ . Moreover,  $F \notin T_{n,2}$ . For this, let  $b \in \mathbb{S}^{n-1}$  be a zero of  $F$ ; we show that  $b^{\otimes 2} \notin \ker \text{MaxSym}(F)$ . For the vector  $a = (b_1, b_2)^T$ , we have  $f(a) = F(b) = 0$  and thus  $a$  is a scalar multiple of  $(1, 1)^T$ . Hence,  $\text{MaxSym}(f)a^{\otimes 2} \neq 0$ , which implies  $\text{MaxSym}(F)b^{\otimes 2} \neq 0$ .

As observed in Remark 5.13, the proof of Theorem 5.12 can now be followed mutadis mutandis to show that the set  $A_{n,2}$  is not Zariski-closed, showing that the spectral bounds do not have generic finite convergence for  $n$ -variate degree 4 forms.  $\square$

## 6. FURTHER DISCUSSION

**Connections to the moment problem and polynomial optimization.** Real de Finetti representations are elements of the set  $\mathcal{C}_d(\mathbb{R}^n)$  from (13), i.e., convex combinations of matrices  $(u_i u_i^T)^{\otimes d}$ , where all  $u_i$  are real unit vectors. Elements of  $\mathcal{C}_d(\mathbb{R}^n)$  are maximally symmetric matrices, whereas elements of  $\mathcal{C}_d(\mathbb{C}^n)$  are in general not maximally symmetric. Therefore, the quantum de Finetti theorem cannot be transferred verbatim to real doubly symmetric matrices, as we saw in Proposition 4.1. Nevertheless, we saw that it is possible to make statements about maximally symmetric matrices (Theorem 3.1) and about real doubly symmetric matrices (Theorem 4.2). The relevance of such results to polynomial optimization on the sphere should be clear: Doherty and Wehner [17] use their real QdF (Theorem 2.19) to analyze the sos bounds; Lovitz and Johnston [29] use the QdF (Theorem 2.17) to analyze their spectral bounds; we saw how to analyze these spectral bounds using our real banded QdF (Theorem 4.2).

The *moment problem* addresses a fundamental question in real algebraic geometry and functional analysis (see [24]), namely, when a linear functional on the polynomial space has an (exact or approximate) representing measure, possibly with a given support. Real de Finetti representations address the approximate moment problem for the space of homogeneous polynomials and measures supported *on the sphere*. So, the result of Theorem 3.1 can be roughly paraphrased as follows (by combining with Proposition 2.14): If  $L$  is a linear functional acting on degree  $2r$  forms that is nonnegative on sums of squares and satisfies  $L(s^r) = 1$ , then for  $d \leq r$  there exists a measure  $\mu$  on the sphere whose (low degree) moments approximate well the ‘pseudo-moments’ of  $L$  up to degree  $2d$ , with an error term in  $O(1/r^2)$ . In other words, real de Finetti theorems permit to show quantitative results for the approximate moment problem on the sphere. This question has been addressed recently for the moment problem on general compact semi-algebraic sets by Baldi and Mourrain [3] (see also [4]), who show quantitative results with polynomial dependency of the form  $O(1/r^c)$ , for some unknown constant  $c > 0$ . Showing an explicit polynomial dependency of the form  $O(1/r^2)$  for some highly symmetric sets like the ball or the simplex is the subject of further research.

| Result                 | Scope                        | complexity ub | complexity lb   |
|------------------------|------------------------------|---------------|-----------------|
| Diaconis-Freedman [16] | probability distr.           | $O(1/r)$      | $\Omega(1/r)$   |
| Christandl et. al. [9] | doubly sym. ( $\mathbb{C}$ ) | $O(1/r)$      | $\Omega(1/r)$   |
| Doherty-Wehner [17]    | max. sym. ( $\mathbb{R}$ )   | $O(1/r)$      | <b>none</b>     |
| Theorem 3.1            | max. sym. ( $\mathbb{R}$ )   | $O(1/r^2)$    | <b>none</b>     |
| Theorem 4.2            | doubly sym. ( $\mathbb{R}$ ) | $O(1/r)$      | $\Omega(1/r^2)$ |

TABLE 1. Convergence rate for de Finetti approximation results: in rows 2-4, the second column concerns a matrix  $M$ , which is positive semidefinite with trace 1, and has additional symmetry and entries in  $\mathbb{C}$  or  $\mathbb{R}$  as indicated. The last two columns give upper and lower bounds for the complexity of the convergence rate.

**Complexity questions.** In Table 1 we gather known results on the convergence rate of the various de Finetti type results. As indicated there, no complexity lower bound is known for the result of Theorem 3.1 and it would also be interesting to close the gap between  $O(1/r)$  and  $\Omega(1/r^2)$  for Theorem 4.2. Accordingly, it is also an open question to settle what is the exact convergence rate for the sos lower bounds  $\text{sos}_r(p)$  and the spectral bounds  $\text{sp}_r(p)$ . For the spectral bounds, a complexity upper bound in  $O(1/r)$  is known [29] and a complexity lower bound in  $\Omega(1/r^2)$  is shown in this paper. For the sos bounds, only an upper bound in  $O(1/r^2)$  is known from [18], but *no complexity lower bound* is known as of today. A notable result is for polynomial optimization on the hypercube  $[-1, 1]^n$ , where Baldi and Slot [5] show a convergence analysis for the sos bounds in  $O(1/r)$  and a complexity lower bound in  $\Omega(1/r^8)$ .

**Acknowledgments.** The authors were supported by the Dutch Scientific Council (NWO) grant OCENW.GROOT.2019.015 (OPTIMAL). A. Blomenhofer was also supported by the European Research Council (ERC) under Agreement 818761.

A slightly shortened version of this manuscript (without Appendices A,B and D) has been accepted for publication in *SIAM Journal on Optimization*. We are grateful to the referees for their careful reading and many helpful suggestions that helped us improve the presentation as well as some of the results. In particular, we thank a referee for raising the question about generic finite convergence, which inspires Section 5.4. We also thank the journal editor for providing constructive feedback. Furthermore, we thank Benjamin Lovitz, Markus Schweighofer, and Lucas Slot for useful discussions.

## REFERENCES

- [1] AHMADI, A. A., OLSHEVSKY, A., PARRILO, P. A., AND TSITSIKLIS, J. N. NP-hardness of deciding convexity of quartic polynomials and related problems. *Mathematical Programming* 137(1-2) (2013), 453–476.
- [2] ALDAZ, J. M., AND RENDER, H. A Fischer type decomposition theorem from the apolar inner product. *Analysis and Mathematical Physics* 13, 6 (2023), 91.
- [3] BALDI, L., AND MOURRAIN, B. On the effective Putinar’s Positivstellensatz and moment approximation. *Mathematical Programming* 200, 1 (June 2023), 71–103.
- [4] BALDI, L., MOURRAIN, B., AND PARUSIŃSKI, A. On Łojasiewicz inequalities and the effective Putinar’s Positivstellensatz. *Journal of Algebra* 662 (2025), 741–767.
- [5] BALDI, L., AND SLOT, L. Degree bounds for Putinar’s Positivstellensatz on the hypercube. *SIAM Journal on Applied Algebra and Geometry* 8, 1 (2024), 1–25.
- [6] CAVES, C., AND FUCHS, C. Unknown quantum states: The quantum de Finetti representation. *J. Math. Phys.* 43 (02 2002).
- [7] CHOI, M.-D., AND LAM, T.-Y. Extremal positive semidefinite forms. *Mathematische Annalen* 231, 1 (Feb. 1977), 1–18.

- [8] CHOI, M.-D., LAM, T.-Y., AND REZNICK, B. Sums of squares of real polynomials. *Proceedings of Symposia in Pure Mathematics* 58, 2 (1995), 103–126.
- [9] CHRISTANDL, M., KÖNIG, R., MITCHISON, G., AND RENNER, R. One-and-a-half quantum de Finetti theorems. *Communications in Mathematical Physics* 273, 2 (2007), 473–498.
- [10] DE FINETTI, B. Sulla prosequibili di processi aleatori scambiabili. *Rend. 1st. Mat. Trieste* 1 (1969), 53–67.
- [11] DE KLERK, E. The complexity of optimizing over a simplex, hypercube or sphere: a short survey. *Central European Journal of Operations Research* 16 (2008), 111–125.
- [12] DE KLERK, E., AND LAURENT, M. Worst-case examples for Lasserre’s measure-based hierarchy for polynomial optimization on the hypercube. *Mathematics of Operations Research* 45(1) (2020), 86–98.
- [13] DE KLERK, E., AND LAURENT, M. Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere. *Mathematical Programming* 193 (2022), 665–685.
- [14] DE KLERK, E., LAURENT, M., AND PARRILO, P. A. On the equivalence of algebraic approaches to the minimization of forms on the simplex. In *Positive Polynomials in Control*. Springer, 2005, pp. 121–132.
- [15] DE KLERK, E., LAURENT, M., AND PARRILO, P. A. A PTAS for the minimization of polynomials of fixed degree over the simplex. *Theoretical Computer Science* 361, 2-3 (2006), 210–225.
- [16] DIACONIS, P., AND FREEDMAN, D. Finite Exchangeable Sequences. *The Annals of Probability* 8, 4 (1980), 745 – 764.
- [17] DOHERTY, A. C., AND WEHNER, S. Convergence of SDP hierarchies for polynomial optimization on the hypersphere, 2013. arxiv:1210.5048.
- [18] FANG, K., AND FAWZI, H. The sum-of-squares hierarchy on the sphere and applications in quantum information theory. *Mathematical Programming* 190 (07 2020).
- [19] GRIBLING, S., LAURENT, M., AND STEENKAMP, A. Bounding the separable rank via polynomial optimization. *Linear Algebra and its Applications* 648 (2022), 1–55.
- [20] HAVILAND, E. K. On the momentum problem for distributions in more than one dimension. *American Journal of Mathematics* 58 (1936), 164–168.
- [21] HUANG, L. Optimality conditions for homogeneous polynomial optimization on the unit sphere. *Optimization Letters* 17 (2023), 1263–1270.
- [22] KOLDA, T. G., AND BADER, B. W. Tensor decompositions and applications. *SIAM Review* 51, 3 (2009), 455–500.
- [23] KÖNIG, R., AND RENNER, R. A de Finetti representation for finite symmetric quantum states. *Journal of Mathematical Physics* 46, 12 (12 2005), 122108.
- [24] KONRAD SCHMÜDGEN. The  $K$ -moment problem for compact semi-algebraic sets. *Mathematische Annalen* 289 (1991), 203–206.
- [25] LASSERRE, J. A new look at nonnegativity on closed sets and polynomial optimization. *SIAM Journal on Optimization* 21(3) (2011), 864–885.
- [26] LASSERRE, J.-B. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization* 11(3) (2001), 796–817.
- [27] LASSERRE, J.-B. *An Introduction to Polynomial and Semi-Algebraic Optimization*. Cambridge University Press, 2015.
- [28] LIM, L.-H. Singular values and eigenvalues of tensors: a variational approach. In *1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2005*. (2005), pp. 129–132.
- [29] LOVITZ, B., AND JOHNSTON, N. A hierarchy of eigencomputations for polynomial optimization on the sphere. *Mathematical Programming* (2025). Published online Jul 21, 2025.
- [30] MOTZKIN, T., AND STRAUS, E. Maxima for graphs and a new proof of a theorem of Turán. *Canadian Journal of Mathematics* 17 (1965), 533–540.
- [31] MÜLLER, C. *Spherical Harmonics*. Lecture Notes in Mathematics, vol. 17. Springer, 1966.
- [32] NAVASCUÉS, M., OWARI, M., AND PLENIO, M. B. Power of symmetric extensions for entanglement detection. *Phys. Rev. A* 80 (Nov 2009), 052306.
- [33] NIE, J. Optimality conditions and finite convergence of Lasserre’s hierarchy. *Mathematical Programming* 146, 1 (2014), 97–121.
- [34] NIE, J. *Moment and Polynomial Optimization*. MOS-SIAM Series on Optimization (MO-31). SIAM, 2023.
- [35] NIELSEN, M. A., AND CHUANG, I. L. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010 (first publication in 2000).
- [36] POWERS, V., AND WÖRMANN, T. An algorithm for sums of squares of real polynomials. *Journal of Pure and Applied Algebra* 127 (1998), 99–104.

- [37] RENOU, M.-O., TRILLO, D., WEILENMANN, M., LE, T. P., TAVAKOLI, A., GISIN, N., ACIN, A., AND NAVASCUÉS, M. Quantum theory based on real numbers can be experimentally falsified. *Nature* 600 (2021), 625–629.
- [38] REZNICK, B. Uniform denominators in Hilbert’s seventeenth problem. *Mathematische Zeitschrift* 220, 1 (1995), 75–97.
- [39] REZNICK, B. On the length of binary forms. In *Quadratic and Higher Degree Forms*. Springer, 2013, pp. 207–232.
- [40] RIENER, C., AND SCHWEIGHOFER, M. Optimization approaches to quadrature: New characterizations of Gaussian quadrature on the line and quadrature with few nodes on plane algebraic curves, on the plane and in higher dimensions. *Journal of Complexity* 45 (2018), 22–54.
- [41] SLOT, L. Sum-of-squares hierarchies for polynomial optimization and the Christoffel-Darboux kernel. *SIAM Journal on Optimization* 32(4) (2022).
- [42] SLOT, L., AND LAURENT, M. Improved convergence analysis of Lasserre’s measure-based upper bounds for polynomial minimization on compact sets. *Mathematical Programming* 193 (2022), 831–871.
- [43] SLOT, L., AND LAURENT, M. Sum-of-squares hierarchies for binary polynomial optimization. *Mathematical Programming* 197 (2023), 621–660.
- [44] TCHAKALOFF, V. Formules de cubatures mécaniques à coefficients nonnegatifs. *Bulletin des Sciences Mathématiques* 81 (1957), 123–134.
- [45] VAVASIS, S. Approximation algorithms for concave quadratic programming. C. Floudas and P. Pardalos, Eds., *Recent Advances in Global Optimization*, Princeton University Press, pp. 3–18.
- [46] WATROUS, J. *The Theory of Quantum Information*. Cambridge University Press, 2018.

#### APPENDIX A. SOME EXAMPLES

We give here a few small examples to illustrate the tensor notions that have been introduced in Section 2. First, we give an example to illustrate how permutations work on tensors: consider  $v \in (\mathbb{R}^2)^{\otimes 2}$  and the transposition  $\sigma = (12)$ , then, with respect to the ordering of  $[2]^2$  as 11, 12, 21, 22 (corresponding to the standard basis  $e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2$  of  $(\mathbb{R}^2)^{\otimes 2}$ ), we have

$$v = \begin{pmatrix} v_{11} \\ v_{12} \\ v_{21} \\ v_{22} \end{pmatrix}, \quad v^{(12)} = \begin{pmatrix} v_{11} \\ v_{21} \\ v_{12} \\ v_{22} \end{pmatrix}, \quad \Pi_2(v) = \frac{1}{2}(v + v^{(12)}) = \begin{pmatrix} v_{11} \\ (v_{12} + v_{21})/2 \\ (v_{12} + v_{21})/2 \\ v_{22} \end{pmatrix}.$$

Indeed, we have  $v_{i_1 i_2}^{(12)} = v_{i_2 i_1}$  for any  $i_1, i_2 \in [2]$ , which gives  $v_{11}^{(12)} = v_{11}$ ,  $v_{22}^{(12)} = v_{22}$ ,  $v_{12}^{(12)} = v_{21}$ , and  $v_{21}^{(12)} = v_{12}$ .

As another example, for the tensor  $v = e_1 \otimes e_1 \otimes e_2 \otimes e_2 \in (\mathbb{R}^2)^{\otimes 4}$ , its projection  $\Pi_4(v)$  onto the symmetric subspace  $S^4(\mathbb{R}^2)$  reads

$$(52) \quad \begin{aligned} \Pi_4(e_1 \otimes e_1 \otimes e_2 \otimes e_2) = & \frac{1}{6} (e_1 \otimes e_1 \otimes e_2 \otimes e_2 + e_1 \otimes e_2 \otimes e_1 \otimes e_2 + \\ & e_1 \otimes e_2 \otimes e_2 \otimes e_1 + e_2 \otimes e_1 \otimes e_1 \otimes e_2 + e_2 \otimes e_1 \otimes e_2 \otimes e_1 + e_2 \otimes e_2 \otimes e_1 \otimes e_1), \end{aligned}$$

which can be checked by enumerating the permutations of  $[4]$  and their action on  $v$ .

Consider the matrix  $I_2^{\otimes 2} \in \mathbb{R}^{[2]^2 \times [2]^2}$ , acting on  $(\mathbb{R}^2)^{\otimes 2}$ . It is not doubly symmetric (since, e.g., its (12, 12)-entry equals 1 while its (21, 12)-entry equals 0). Hence, it is also not maximally symmetric. For illustration, we show below its projection  $\Pi_2 I_2^{\otimes 2} \Pi_2 = \Pi_2$  onto the space of doubly symmetric matrices and its projection

$\Pi_{\text{MaxSym}}(I_2^{\otimes 2})$  onto the space of maximally symmetric matrices (as given in relation (2.3)):

$$(53) \quad \Pi_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \Pi_{\text{MaxSym}}(I_2^{\otimes 2}) = \begin{pmatrix} 1 & 0 & 0 & 1/3 \\ 0 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1 \end{pmatrix},$$

where entries are indexed by 11, 12, 21, 22. To obtain the projection  $\Pi_{\text{MaxSym}}(I_2^{\otimes 2})$  onto the space of maximally symmetric matrices, one can proceed as follows:

$$I_2^{\otimes 2} = (e_1 e_1^T + e_2 e_2^T) \otimes (e_1 e_1^T + e_2 e_2^T) = (e_1 \otimes e_1)(e_1 \otimes e_1)^T + (e_2 \otimes e_2)(e_2 \otimes e_2)^T \\ + (e_1 \otimes e_2)(e_1 \otimes e_2)^T + (e_2 \otimes e_1)(e_2 \otimes e_1)^T.$$

Then, the matrix  $(e_i \otimes e_i)(e_i \otimes e_i)^T$  is maximally symmetric for  $i = 1, 2$ . To obtain  $\Pi_{\text{MaxSym}}((e_1 \otimes e_2)(e_1 \otimes e_2)^T)$ , consider the associated tensor  $e_1 \otimes e_2 \otimes e_1 \otimes e_2 \in (\mathbb{R}^2)^{\otimes 4}$ , whose projection onto the symmetric subspace  $S^4(\mathbb{R}^2)$  satisfies  $\Pi_4(e_1 \otimes e_2 \otimes e_1 \otimes e_2) = \Pi_4(e_1 \otimes e_1 \otimes e_2 \otimes e_2)$ , whose explicit value is given in (52). By reshaping this tensor as a matrix we get

$$\Pi_{\text{MaxSym}}((e_1 \otimes e_2)(e_1 \otimes e_2)^T) = \frac{1}{6} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Clearly,  $\Pi_{\text{MaxSym}}((e_2 \otimes e_1)(e_2 \otimes e_1)^T) = \Pi_{\text{MaxSym}}((e_1 \otimes e_2)(e_1 \otimes e_2)^T)$ , and from this one gets the desired expression for  $\Pi_{\text{MaxSym}}(I_2^{\otimes 2})$  in (53) (and (2.3)).

To illustrate the partial trace operation, consider a matrix

$$M = \begin{pmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \dots & \vdots \\ A_{n1} & \dots & A_{nn} \end{pmatrix} \in \mathbb{K}^{[n]^2 \times [n]^2}, \quad \text{where } A_{ij} \in \mathbb{K}^{n \times n}$$

and the entries of  $M$  are indexed by 11, 12, ..., 1n, ..., n1, n2, ..., nn. Then, we have

$$\text{Tr}_1(M) = \begin{pmatrix} \text{Tr}(A_{11}) & \dots & \text{Tr}(A_{1n}) \\ \vdots & \dots & \vdots \\ \text{Tr}(A_{n1}) & \dots & \text{Tr}(A_{nn}) \end{pmatrix} \in \mathbb{K}^{n \times n},$$

where we ‘trace out the second subsystem’ to get  $\text{Tr}_1(M)$ . Recall that the index  $k$  in  $\text{Tr}_k$  denotes the number of subsystems from the right, which are traced out. Hence,  $\text{Tr}_2(M) = \text{Tr}(M)$  is a scalar.

#### APPENDIX B. PROOF OF LEMMA 4.3

The result from Lemma 4.3 is essentially extracted from the proof of [29, Theorem 4.1]. Since it plays a crucial role to show Theorem 4.2, we give a sketch of proof for clarity of exposition. Assume  $\tau \in \mathcal{C}_d(\mathbb{C}^n)$  with  $\text{Tr}(\tau) = 1$ , and set  $\alpha := \text{Tr}(\Pi_{\text{MaxSym}}(\tau))$  and  $\alpha_d = \frac{2^d}{(2d)}$ . We show that  $\frac{1}{\alpha} \Pi_{\text{MaxSym}}(\tau) \in \mathcal{C}_d(\mathbb{R}^n)$  and  $\alpha_d \leq \alpha \leq 1$ .

Clearly, it suffices to show the result for  $\tau = (uu^*)^{\otimes d}$ , where  $u \in \mathbb{C}^n$  is a unit vector. Write  $u = a + ib$ , where  $a, b \in \mathbb{R}^n$  with  $\|a\|^2 + \|b\|^2 = 1$ . By Lemma 2.5,  $\Pi_{\text{MaxSym}}(\tau) = \Pi_{\text{MaxSym}}((aa^T + bb^T)^{\otimes d})$  and thus  $\alpha = \text{Tr}(\Pi_{\text{MaxSym}}((aa^T + bb^T)^{\otimes d}))$ . The next step is to exploit a result of Reznick [39, Corollary 5.6] for showing that  $\frac{1}{\alpha} \Pi_{\text{MaxSym}}((aa^T + bb^T)^{\otimes d}) \in \mathcal{C}_d(\mathbb{R})$ .

Reznick [39] shows that the bivariate homogeneous polynomial  $(t_1^2 + t_2^2)^d$  in variables  $t_1, t_2$  can be written as a sum of powers of linear forms. More precisely, he gives explicit real unit vectors  $v_k \in \mathbb{R}^2$  ( $k = 0, 1, \dots, d$ ) such that

$$(t_1^2 + t_2^2)^d = \frac{\alpha_d \cdot 2^d}{d+1} \sum_{k=0}^d \left\langle v_k, \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \right\rangle^{2d}.$$

By plugging in  $t_1 = a^T x$  and  $t_2 = b^T x$ , one obtains that the degree  $2d$   $n$ -variate form  $((a^T x)^2 + (b^T x)^2)^d = \langle (aa^T + bb^T)^{\otimes d}, (xx^T)^{\otimes d} \rangle$  is a sum of  $2d$ -th powers of linear forms in  $x$ , of the form  $\sum_{k=0}^d (c_k^T x)^{2d} = \sum_{k=0}^d \langle (c_k c_k^T)^{\otimes d}, (xx^T)^{\otimes d} \rangle$  for some  $c_k \in \mathbb{R}^n$  (obtained from  $v_k, a, b$ ). So,  $\langle (a^T x)^2 + (b^T x)^2 \rangle^d, (xx^T)^{\otimes d} = \langle (\sum_{k=0}^d (c_k c_k^T)^{\otimes d}), (xx^T)^{\otimes d} \rangle$ . Hence, the maximally symmetric representation of the form  $((a^T x)^2 + (b^T x)^2)^d$  is equal to  $\sum_{k=0}^d (c_k c_k^T)^{\otimes d}$ . Therefore,  $\Pi_{\text{MaxSym}}((aa^T + bb^T)^{\otimes d}) = \sum_{k=0}^d (c_k c_k^T)^{\otimes d}$ , which thus belongs to the cone generated by  $\mathcal{C}_d(\mathbb{R}^n)$ . So, after normalizing by the trace  $\alpha$ , one obtains that  $\Pi_{\text{MaxSym}}(\tau)/\alpha \in \mathcal{C}_d(\mathbb{R}^n)$ .

To show the lower bound  $\alpha \geq \alpha_d$ , the authors of [29] use the eigenvalue decomposition to write  $aa^T + bb^T = t v v^T + (1-t) w w^T$  with  $t \in [0, 1]$  and orthogonal unit vectors  $v$  and  $w$ . They consider the function  $\varphi_d(t) = \text{Tr}(\Pi_{\text{MaxSym}}((t v v^T + (1-t) w w^T)^{\otimes d}))$ , a polynomial in  $t$ , so that  $\alpha = \varphi_d(t_0)$  for some  $t_0 \in [0, 1]$ . They show that it can be written as a sum  $\varphi_d(t) = \sum_{k=0}^{\lfloor d/2 \rfloor} a_k (t - 1/2)^{2k}$  of even powers of  $(t - 1/2)$  with nonnegative coefficients  $a_k \geq 0$ . From this, it readily follows that the minimum of  $\varphi_d(t)$  for  $t \in [0, 1]$  is attained at  $t = 1/2$ , with  $\varphi_d(1/2) = \text{Tr}(\Pi_{\text{MaxSym}}(I_2^{\otimes d})) = 2^{2d} \binom{2d}{d}^{-1} = \alpha_d \cdot 2^d$  (which also follows from [39]). In the same way, the maximum value of  $\varphi_d(t)$  is attained at  $t \in \{0, 1\}$  and is equal to  $\varphi_d(1) = 1$ . The desired bounds  $\alpha_d \leq \alpha \leq 1$  follow directly.

### APPENDIX C. PROOF OF PROPOSITION 5.3

Let  $p \in R_{2d}$ . Recall the definitions in (26) for the matrices  $Q_r(p)$  and  $M_r$  that represent, respectively, the polynomials  $ps^{r-d}$  and  $s^r$ .

First, we show that  $\text{sp}_r(p) \leq \text{sos}_r(p)$ . The matrix  $Q_r(p) - \text{sp}_r(p)M_r$  represents the polynomial  $s^{r-d}p - \text{sp}_r(p)s^r$  and it is positive semidefinite (by definition of  $\text{sp}_r(p)$ ). Hence, the polynomial  $s^{r-d}p - \text{sp}_r(p)s^r$  is sos, and thus  $\text{sp}_r(p) \leq \text{sos}_r(p)$ , by (31).

We now show that  $\text{sp}_r(p) \leq \text{sp}_{r+1}(p)$ . For this, we use the formulation (30). Let  $Y$  be an optimal solution for the formulation (30) of the parameter  $\text{sp}_{r+1}(p)$ ; so  $Y$  is  $(r+1)$ -doubly symmetric,  $Y \succeq 0$  and  $\langle M_{r+1}, Y \rangle = 1$ . Set  $X = \text{Tr}_1(Y)$ . Then,  $X$  is  $r$ -doubly symmetric and positive semidefinite. We claim that  $\langle M_r, X \rangle = 1$ . Indeed,

$$\begin{aligned} 1 &= \langle M_{r+1}, Y \rangle = \langle \Pi_{r+1}(M \otimes I_n^{\otimes(r+1-d)}) \Pi_{r+1}, Y \rangle = \langle M \otimes I_n^{\otimes(r+1-d)}, \Pi_{r+1} Y \Pi_{r+1} \rangle \\ &= \langle M \otimes I_n^{\otimes(r+1-d)}, Y \rangle = \langle M \otimes I_n^{\otimes(r-d)}, \text{Tr}_1(Y) \rangle = \langle M_r, X \rangle, \end{aligned}$$

where, in the second line, we use the fact that  $Y$  is doubly symmetric for the first equality, the definition of the partial trace for the second equality, and the fact that  $X = \text{Tr}_1(Y)$  is doubly symmetric for the last equality. In the same way, we have

$$\langle Q_{r+1}(p), Y \rangle = \langle Q(p) \otimes I_n^{\otimes(r+1-d)}, Y \rangle = \langle Q(p) \otimes I_n^{\otimes(r-d)}, \text{Tr}_1(Y) \rangle = \langle Q_r(p), X \rangle.$$

Hence,  $X$  is a feasible solution for the formulation (30) of  $\text{sp}_r(p)$  with objective value  $\text{sp}_{r+1}(p)$ , which shows  $\text{sp}_r(p) \leq \text{sp}_{r+1}(p)$ .

APPENDIX D. THE CONSTANT  $\gamma(Q(p))$  IS NOT WELL-BEHAVED

First, we discuss the relevance of establishing a performance analysis in terms of the range of values  $p_{\max} - p_{\min}$  as is done in Theorem 5.7. This is indeed the common practice in continuous optimization, as discussed, e.g., in [45, 15]: given  $\varepsilon > 0$ , a value  $\text{val}_\varepsilon$  is called an  $\varepsilon$ -approximation for the problem of minimizing a function  $p$  on a compact set  $S$  if it satisfies  $|p_{\min} - \text{val}_\varepsilon| \leq \varepsilon(p_{\max} - p_{\min})$ , where  $p_{\min}$  and  $p_{\max}$  are the minimum and maximum values taken by  $p$  on  $S$ . Such definition is motivated by the fact that it enjoys nice invariance properties, like translating the function by an (appropriate) constant or scaling by a positive constant. When  $S$  is the sphere, it would thus be desirable to have an error analysis that depends on the range of values  $p_{\max} - p_{\min}$ . This is precisely what we do in Theorem 5.7. Note that the error analysis for the sos bounds in Theorem 3.2 does indeed involve the range  $p_{\max} - p_{\min}$ .

With this in mind, we investigate the parameter

$$\gamma(Q(p)) = \max\{\lambda_{\max}(Q(p)) - \text{sp}_d(p), p_{\min} - \lambda_{\min}(Q(p))\},$$

which enters the error analysis in Theorem 5.5. We show that this parameter does not behave well under translating  $p$  by a multiple of  $s^d$ : there does not exist a constant  $C_{n,d}$  (depending only on  $n$  and  $d$ ) such that

$$(54) \quad \gamma(Q(p)) \leq C_{n,d}(p_{\max} - p_{\min}) \quad \text{for all } p \in R_{2d}.$$

Hence, one cannot directly derive Theorem 5.7 from Theorem 5.5.

For this, given  $p \in R_{2d}$  and  $a \in \mathbb{R}$ , consider the polynomial  $p(a) := p + as^d \in R_{2d}$ . Then, we have  $p(a)_{\max} - p(a)_{\min} = p_{\max} - p_{\min}$ ,  $\text{sp}_d(p(a)) = \text{sp}_d(p) - a$ , and  $Q(p(a)) = Q(p) + aM$ , where  $M = \Pi_{\text{MaxSym}}(I_n^{\otimes d})$ . We disprove the existence of a constant  $C_{n,d}$  satisfying (54) by showing a lower bound on the parameter  $\gamma(Q(p(a)))$  that is linear and monotone increasing in  $a$  (for all large enough  $a$ ).

We first consider the term  $\lambda_{\max}(Q(p(a))) - \text{sp}_d(p(a))$ . Let  $v$  be a unit eigenvector of  $M$  for  $\lambda_{\max}(M)$ . Then,  $\lambda_{\max}(Q(p(a))) \geq v^T Q(p(a))v = v^T Q(p)v + a\lambda_{\max}(M)$ . Therefore,

$$(55) \quad \begin{aligned} \lambda_{\max}(Q(p(a))) - \text{sp}_d(p(a)) &= \lambda_{\max}(Q(p(a))) - \text{sp}_d(p) - a \\ &\geq a(\lambda_{\max}(M) - 1) + v^T Q(p)v - \text{sp}_d(p). \end{aligned}$$

We now consider the term  $p(a)_{\min} - \lambda_{\min}(Q(p(a)))$ . By taking a unit eigenvector  $w$  of  $M$  for  $\lambda_{\min}(M)$ ,  $\lambda_{\min}(Q(p(a))) \leq w^T Q(p(a))w = w^T Q(p)w + a\lambda_{\min}(M)$  holds, and

$$(56) \quad \begin{aligned} p(a)_{\min} - \lambda_{\min}(Q(p(a))) &= p_{\min} - a - \lambda_{\min}(Q(p(a))) \\ &\geq a(1 - \lambda_{\min}(M)) + p_{\min} - w^T Q(p)w. \end{aligned}$$

Therefore,  $\gamma(Q(p(a))) \geq \varphi(a)$ , where

$$\varphi(a) = \max\{a(\lambda_{\max}(M) - 1) + v^T Q(p)v - \text{sp}_d(p), a(1 - \lambda_{\min}(M)) + p_{\min} - w^T Q(p)w\}$$

is the maximum of the lower bounds in (55)-(56). Since  $\lambda_{\min}(M) \leq 1 \leq \lambda_{\max}(M)$  with  $\lambda_{\max}(M) - \lambda_{\min}(M) > 0$ , we obtain that  $\varphi(a)$  is linear monotone increasing in  $a$  for all large enough  $a$ , which concludes the proof.

Note that one can analogously show that the parameter  $\|Q(p(a))\|_\infty$ , which appears in Theorem 5.1, also has this undesirable feature that it is monotone increasing for large enough  $a$ ; the proof is the same (even simpler).