Notes on the paper: "Convergence of SDP hierarchies for polynomial optimization on the hypersphere", by A.C. Doherty and S. Wehner

Monique Laurent *

March 7, 2019

Abstract

For the problem of maximizing an n-variate polynomial f over the unit sphere $\mathbb{S}^{n-1} \subseteq \mathbb{R}^n$, some hierarchies of lower and upper bounds have been introduced in the literature, that converge to the global optimum of f over $\mathbb{S}^{n-1}$. These hierarchies use sums of squares of polynomials with bounded degree $2r$ for increasing values of $r \in \mathbb{N}$ and they can be expressed as semidefinite programs. When f is homogeneous, Doherty and Wehner [1] proposed a method which allows to analyze simultaneously the quality of these two hierarchies of bounds and to show that their rate of convergence to the global optimum is in $O(1/r)$. Quoting from the abstract of [1], their approach is as follows:

"Our method is inspired by a set of results from quantum information known as quantum de Finetti theorems. In particular, we prove a de Finetti theorem for a special class of real symmetric matrices to establish the existence of approximate representing measures for moment matrix relaxations."

In these notes we give a concise exposition of the results and approach in [1]. In particular, we highlight the links between the formulation used in [1] and more well known existing formulations, and we give full details for the proofs, trying to keep the preliminary background to the minimum necessary. Along the way we also correct a few imprecisions we found in the original paper.

1 Introduction

Throughout we set $V = \mathbb{R}^n$, with standard unit basis $\{e_1, \ldots, e_n\}$. We let $\mathbb{R}[x_1, \ldots, x_n] = \mathbb{R}[x]$ denote the space of n-variate polynomials and, for an

*Centrum Wiskunde & Informatica (CWI), Amsterdam and Tilburg University, monique@cwi.nl

integer $a \in \mathbb{N}$, Σ_a denotes the set of polynomials with degree at most $2a$ that can be written as a sum of squares of polynomials. Moreover we let $\mathbb{N}^n_a$ denote the set of sequences $\underline{i} = (i_1, \ldots, i_n) \in \mathbb{N}^n$ satisfying $|\underline{i}| = i_1 + \ldots + i_n = a$.

We let $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ denote the unit sphere in $\mathbb{R}^n$ and μ denote the probability (Haar) measure on $\mathbb{S}^{n-1}$.

The main result in [1] concerns the convergence analysis of hierarchies of lower and upper SDP based bounds for polynomial optimization over the unit sphere $\mathbb{S}^{n-1}$. Let T be a homogeneous polynomial with degree $2a$. (As indicated in [1] - see Section 1.5 below - the case of odd degree homogeneous polynomials can indeed be reduced to the even degree case.) Consider its maximum and minimum values over the unit sphere:

$$T_{\max} = \max_{x \in \mathbb{S}^{n-1}} T(x), \quad T_{\min} = \min_{x \in \mathbb{S}^{n-1}} T(x).$$

Given an integer $r \geq a$ consider the following parameters:

$$\overline{T}^{(r)} = \min\Big\{t : t - T(x) \in \Sigma_r + \Big(1 - \sum_{i=1}^n x_i^2\Big)\mathbb{R}[x]\Big\},$$

$$\underline{T}^{(r)} = \max\Big\{\int_{\mathbb{S}^{n-1}} T(x)h(x)d\mu(x) : \int_{\mathbb{S}^{n-1}} h(x)d\mu(x) = 1,\ h \in \Sigma_r\Big\},$$

which have been considered in [5, 7], [6], respectively. These provide upper and lower bounds for the global maximum of T:

$$\underline{T}^{(r)} \leq T_{\max} \leq \overline{T}^{(r)}.$$

The main result by Doherty & Wehner [1] is the following convergence analysis[1] of the bounds $\overline{T}^{(r)}$ and $\underline{T}^{(r)}$.

Theorem 1.1. [1, Theorem 7.1] *Assume $n \geq 3$, let $a \in \mathbb{N}$ and let T be an n-variate homogeneous polynomial of degree $2a$. Then, for any integer r such that[2] $r \geq a(2a + n - 2) - n/2$, the following inequality holds:*

$$\overline{T}^{(r)} - \underline{T}^{(r)} \leq \gamma_{n,a} \frac{2a^2(2a + n - 2)}{2r + n}(T_{\max} - T_{\min}).$$

Here $\gamma_{n,a}$ is an absolute constant[3] that depends only on n and a.

In these notes we provide a complete exposition of the proof of this result. We follow the approach in [1], but we try to keep the exposition concise and we make a few small adaptations/corrections along the way.

[1] Our formulation in Theorem 1.1 differs slightly from the formulation of Theorem 7.1 in [1]. Indeed, we use the range $T_{\max} - T_{\min}$ instead of $|T_{\max}|$ and we have an additional constant $\gamma_{n,a}$, which does not appear in [1].

[2] Any such integer satisfies $r > a$.

[3] In [1] the result is presented without such a constant, but we do not see how to conclude the proof without this constant. As we will see later in (5), the constant $\gamma_{n,a}$ arises from comparing the usual Frobenius norm of a matrix with its $\|\cdot\|_{F1}$ norm (see Section 1.3 below).

1.1 Preliminaries

1.1.1 Tensors

Given an integer $a \in \mathbb{N}$, $V^{\otimes a}$ denotes the set of a-tensors $\vec{U} = (U_{i_1 \dots i_a})_{i_1,\dots,i_a \in [n]}$, which can also be expressed as $\vec{U} = \sum_{i_1,\dots,i_a \in [n]} U_{i_1 \dots i_a} e_{i_1} \otimes \ldots \otimes e_{i_a}$. Any permutation σ of $[a]$ acts on $V^{\otimes a}$ by setting

$$\sigma(\vec{U}) = \sum_{i_1,\dots,i_a \in [n]} U_{i_1 \dots i_a} e_{i_{\sigma(1)}} \otimes \ldots \otimes e_{i_{\sigma(a)}}.$$

The tensor $\vec{U}$ is called *symmetric* if $\sigma(\vec{U}) = \vec{U}$ for all permutations $\sigma \in \mathrm{Sym}(a)$ and $\mathrm{Sym}V^{\otimes a}$ denotes the vector space of all symmetric a-tensors acting on $V = \mathbb{R}^n$. We let Π_a denote the orthogonal projection from $V^{\otimes a}$ onto $\mathrm{Sym}V^{\otimes a}$. That is,

$$\Pi_a(\vec{U}) = \frac{1}{a!} \sum_{\sigma \in \mathrm{Sym}(a)} \sigma(\vec{U}).$$

The following notation will be useful. Given an a-tuple $\underline{i} = (i_1, \dots, i_a) \in [n]^a$, we let $\alpha(\underline{i}) = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ denote the n-tuple, where for each $\ell \in [n]$, α_ℓ denotes the number of occurrences of ℓ within the multi-set $\{i_1, \dots, i_a\}$, so that $|\alpha(\underline{i})| = \alpha_1 + \ldots + \alpha_n = a$ (and $x_{i_1} \cdots x_{i_a} = x_1^{\alpha_1} \cdots x_n^{\alpha_n} = x^{\alpha(\underline{i})}$).

Note that the vector $\Pi_a(e_{i_1} \otimes \ldots \otimes e_{i_a})$ depends only on the n-tuple $\alpha(\underline{i})$. Thus the dimension of $\mathrm{Sym}V^{\otimes a}$ is equal to $\binom{n+a}{a}$, the number of ways to select integers $\alpha_1, \dots, \alpha_n \in \mathbb{N}$ such that $\alpha_1 + \ldots + \alpha_n = a$.

As an example, for any vector $x \in V$, the associated a-tensor $x^{\otimes a}$ (obtained by taking the ath tensor product of x) is symmetric: $x^{\otimes a} \in \mathrm{Sym}V^{\otimes a}$. Moreover, such vectors form a linear basis of $\mathrm{Sym}V^{\otimes a}$.

1.1.2 Maximally symmetric matrices

Clearly, any matrix $M \in \mathrm{End}(V^{\otimes a})$, say

$$M = (M_{i_1 \dots i_a, j_1 \dots j_a}) = \sum_{i_1,\dots,i_a,j_1,\dots,j_a \in [n]} M_{i_1 \dots i_a, j_1 \dots j_a} e_{i_1} \otimes \ldots \otimes e_{i_a} (e_{j_1} \otimes \ldots \otimes e_{j_a})^T,$$

corresponds in a unique way to a $2a$-tensor

$$\vec{M} = \sum_{i_1,\dots,i_a,j_1,\dots,j_a \in [n]} M_{i_1 \dots i_a, j_1 \dots j_a} e_{i_1} \otimes \ldots \otimes e_{i_a} \otimes e_{j_1} \otimes \ldots \otimes e_{j_a} \in V^{\otimes 2a}.$$

Following [1] the matrix M is called *maximally symmetric* when the associated $2a$-tensor $\vec{M}$ is symmetric. Note that this implies that M is a symmetric matrix, but being maximally symmetric is a stronger property when $a > 1$. We let $\mathrm{MSym}(V^{\otimes a})$ denote the subspace of maximally symmetric matrices within $\mathrm{End}(V^{\otimes a})$.

Note that the notion of "maximally symmetric matrix" can be seen as the analog of the notion of "moment matrix" in the context of tensors. Indeed, M

is maximally symmetric precisely when, for each $\underline{i}, \underline{j} \in [n]^a$, the $(\underline{i}, \underline{j})$-entry $M_{\underline{i},\underline{j}}$ of M depends *only* on the n-tuple $\alpha(\underline{i}) + \alpha(\underline{j})$. (Recall Section 1.1.1).

By construction there is a one-to-one correspondance $M \mapsto \vec{M}$ between the space $\mathrm{MSym}(V^{\otimes a})$ of maximally symmetric matrices and the space $\mathrm{Sym}V^{\otimes 2a}$ of symmetric $2a$-tensors.

1.1.3 Homogeneous polynomials

Let T be an n-variate homogeneous polynomial of degree $2a$. Say,

$$T(x) = \sum_{\alpha=(\alpha_1,\ldots,\alpha_n)\in\mathbb{N}^n_{2a}} t_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$

One may define the corresponding tensor $\vec{U_T} = \sum_{\alpha\in\mathbb{N}^n_{2a}} t_\alpha e_1^{\otimes\alpha_1} \otimes \ldots \otimes e_n^{\otimes\alpha_n}$, so that we have

$$T(x) = \langle \vec{U_T}, x^{\otimes 2a} \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the usual Euclidean inner product. As $x^{\otimes 2a}$ is a symmetric tensor we also have

$$T(x) = \langle \Pi_{2a}(\vec{U_T}), x^{\otimes 2a} \rangle,$$

where $\Pi_{2a}(\vec{U_T})$ is now a symmetric $2a$-tensor. Hence there is a unique maximally symmetric matrix in $\mathrm{MSym}(V^{\otimes a})$, denoted Z_T, whose associated $2a$-tensor is $\Pi_{2a}(\vec{U_T})$, i.e., such that $\vec{Z_T} = \Pi_{2a}(\vec{U_T})$. Summarizing:

Lemma 1.2. *Any homogeneous n-variate polynomial T with degree $2a$ corresponds in a unique way to a maximally symmetric matrix $Z_T \in \mathrm{MSym}(V^{\otimes a})$ such that*

$$T(x) = {x^{\otimes a}}^T Z_T x^{\otimes a} = \langle \vec{Z_T}, x^{\otimes a} {x^{\otimes a}}^T \rangle.$$

In particular, $Z_T = 0$ if and only if T is the identically zero polynomial.

Given an integer $r \geq a$ consider the polynomial

$$T_r(x) = T(x)\Big(\sum_{i=1}^n x_i^2\Big)^{r-a}$$

which is homogeneous with degree $2r$. As the maximally symmetric matrix corresponding to the polynomial $(\sum_i x_i^2)^r$ is the identity matrix I (of suitable size) it follows that $T_r(x) = {x^{\otimes r}}^T (Z_T \otimes I) x^{\otimes r}$ and thus we have

$$\vec{Z_{T_r}} = \Pi_{2r}\Big(Z_T \vec{\otimes} I\Big).$$

Here is a useful observation that will be used later. Consider a matrix M in $\mathrm{MSym}(V^{\otimes r})$ and let $\mathsf{Tr}_{r-a}(M) \in \mathrm{MSym}(V^{\otimes a})$ be the matrix obtained by taking the partial trace (tracing out $r-a$ copies of V in $V^{\otimes r}$); we have the identities:

$$\langle Z_{T_r}, M \rangle = \langle \vec{Z_{T_r}}, \vec{M} \rangle = \langle \Pi_{2r}(Z_T \vec{\otimes} I), \vec{M} \rangle = \langle Z_T \otimes I, M \rangle = \langle Z_T, \mathsf{Tr}_{r-a}(M) \rangle.$$

1.2 Polynomial optimization over the sphere

We let $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ denote the unit sphere in $\mathbb{R}^n$ and consider the problem of optimizing a homogeneous polynomial T over the sphere:

$$T_{\max} = \max_{x \in \mathbb{S}^{n-1}} T(x), \quad T_{\min} = \min_{x \in \mathbb{S}^{n-1}} T(x).$$

We will recall how to derive lower and upper approximations for the parameter $T_{\max}$. We assume T has even degree $2a$; the case when T has odd degree can indeed be reduced to the even case (see Section 1.5).

1.2.1 Upper bounds

Fix an integer $r \geq a$ and as before set $T_r(x) = T(x)(\sum_i x_i^2)^{r-a}$. Maximizing $T(x)$ over $\mathbb{S}^{n-1}$ is obviously equivalent to maximizing $T_r(x)$ over $\mathbb{S}^{n-1}$. As observed above, we have $T_r(x) = \langle Z_{T_r}, x^{\otimes r} {x^{\otimes r}}^T \rangle$. In order to linearize the non-linear term $x^{\otimes r} {x^{\otimes r}}^T$ let us introduce a matrix variable $M = x^{\otimes r} {x^{\otimes r}}^T$. Then, by construction, M is maximally symmetric and satisfies $M \succeq 0$, $\mathsf{Tr}(M) = 1$. Following [1] this motivates defining the following parameter:

$$\overline{T}^{(r)} = \max\{\langle Z_{T_r}, M \rangle : M \in \mathrm{MSym}(V^{\otimes r}), M \succeq 0, \mathsf{Tr}(M) = 1\}. \tag{1}$$

Clearly we have

$$T_{\max} \leq \overline{T}^{(r)}.$$

As we now observe this parameter in fact coincides with the usual well known sum-of-squares bound, considered in the foundational works [5, 7].

Lemma 1.3. *The above parameter (1) can be equivalently defined as follows:*

$$\overline{T}^{(r)} = \min \Big\{ t : t \Big(\sum_i x_i^2 \Big)^r - T_r(x) \in \Sigma_r \Big\}, \tag{2}$$

$$\overline{T}^{(r)} = \min \Big\{ t : t - T_r(x) \in \Sigma_r + \Big(1 - \sum_i x_i^2 \Big) \mathbb{R}[x] \Big\}, \tag{3}$$

Proof. The equivalence between the two claimed reformulations (2) and (3) is not difficult to see (and can be found in [4]). We show the equivalence between (1) and (2). First we write the program (1) defining $\overline{T}^{(r)}$ in standard primal SDP form. For this let $\{B_j, j \in J\}$ be a basis of the linear space $(\mathrm{MSym}(V^{\otimes r}))^{\perp}$, the orthogonal complement of $\mathrm{MSym}(V^{\otimes a})$ in $\mathrm{End}(V^{\otimes a})$. Then we have

$$\overline{T}^{(r)} = \max\{\langle Z_{T_r}, M \rangle : \langle B_j, M \rangle = 0 \ (j \in J), \mathsf{Tr}(M) = 1, M \succeq 0\}.$$

The dual SDP reads

$$\min\{t : tI + Y - Z_{T_r} \succeq 0, \ t \in \mathbb{R}, \ Y \in (\mathrm{MSym}(V^{\otimes r}))^{\perp}\}.$$

As the primal and dual are both strictly feasible there is no duality gap and the optimum is attained in both programs. Hence it suffices to show that the latter program is equivalent to (2).

Indeed, if (t, Y) is dual feasible then the polynomial ${x^{\otimes r}}^T(tI + Y - Z_{T_r})x^{\otimes r}$ belongs to Σ_r and moreover it is equal to $t(\sum_i x_i^2)^r - T_r(x)$. So this gives a feasible solution to program (2).

Conversely, assume that $t(\sum_i x_i^2)^r - T_r(x)$ belongs to Σ_r for some scalar t. Then there exists a matrix $Z \succeq 0$ such that ${x^{\otimes r}}^T Z x^{\otimes r} = {x^{\otimes r}}^T(tI - Z_{T_r})x^{\otimes r}$ for all $x \in \mathbb{R}^n$. This implies that the matrix $Y := tI - Z_{T_r} - Z$ belongs to $(\mathrm{MSym}(V^{\otimes r}))^{\perp}$. Since $tI - Y - Z_{T_r} \succeq 0$, it follows that $(t, -Y)$ is dual feasible, which concludes the proof. $\square$

1.2.2 Lower bounds

Throughout μ denotes the (Haar) probability measure on the sphere. That is, $d\mu(x) = \frac{1}{\omega_n} d\sigma(x)$, where $d\sigma$ is the area measure on the sphere $\mathbb{S}^{n-1}$ and ω_n is the area of $\mathbb{S}^{n-1}$. Following Lasserre [6] we define the following parameter

$$\underline{T}^{(r)} = \max\Big\{ \int_{\mathbb{S}^{n-1}} T(x)h(x)d\mu(x) : \int_{\mathbb{S}^{n-1}} h(x)d\mu(x) = 1,\ h \in \Sigma_r \Big\}. \quad (4)$$

Then we have

$$\underline{T}^{(r)} \leq T_{\max}.$$

The main result of the paper [1] is to analyze simultaneously the convergence rate of the bounds $\overline{T}^{(r)}$ and $\underline{T}^{(r)}$; namely, in [1] it is shown that

$$\overline{T}^{(r)} - \underline{T}^{(r)} = O\Big(\frac{1}{r}\Big).$$

The key ingredient to show this is Theorem 1.4 below, shown in [1].

1.3 De Finetti theorem - the main technical result

We present here the key technical result of [1] that leads to the convergence analysis of the upper and lower bounds (1) and (4).

Following [1], given a matrix $M \in \mathrm{MSym}(V^{\otimes a})$, define the parameter

$$\begin{aligned} \|M\|_{F1} = \max\{\langle M, Z_F\rangle : \quad & F \text{ homogenous polynomial with degree } 2a, \\ & |F(x)| \leq 1 \text{ on } \mathbb{S}^{n-1}\}. \end{aligned}$$

Hence this parameter can be rewritten as

$$\|M\|_{F1} = \max\{\langle M, Z\rangle : Z \in \mathrm{MSym}(V^{\otimes a}), |{x^{\otimes a}}^T Z x^{\otimes a}| \leq 1 \text{ on } \mathbb{S}^{n-1}\}.$$

This in fact defines a norm on $\mathrm{MSym}(V^{\otimes a})$. To see this note that we have $\|M\|_{F1} \geq \|M\|$, where $\|\cdot\|$ is the usual Frobenius norm (the Euclidean norm).

This follows from the fact that $\|M\| = \max_{\|Z\|\leq 1}\langle M, Z\rangle$ and, using Cauchy-Schwartz inequality, $\|Z\| \leq 1$ implies $|{x^{\otimes a}}^T Z x^{\otimes a}| \leq \|Z\| \leq 1$ for all $x \in \mathbb{S}^{n-1}$. In addition, as all norms on a finite dimensional vector space are equivalent there exists a constant $\gamma_{n,a} \geq 1$ such that

$$\|M\| \leq \|M\|_{F1} \leq \gamma_{n,a}\|M\| \tag{5}$$

for all $M \in \mathrm{MSym}(V^{\otimes a})$.

We can now present the main technical result of [1], which is a de Finetti type result. We refer to [1] for discussion and background information about such results.

Theorem 1.4. [1, Theorem 6.2] *Consider integers $r, n \in \mathbb{N}$ such that $n \geq 3$ and $r \geq a(2a+n-2) - n/2$. Consider a matrix $M \in \mathrm{MSym}(V^{\otimes r})$ such that $M \succeq 0$ and $\mathsf{Tr}(M) = 1$. Define the polynomial $Q_M(x) = {x^{\otimes r}}^T M x^{\otimes r}$, the matrix $M_a = \mathsf{Tr}_{r-a}(M) \in \mathrm{MSym}(V^{\otimes a})$ and the matrix*

$$\widetilde{M}_a = C_{n,r}\int_{\mathbb{S}^{n-1}} Q_M(x) x^{\otimes a} {x^{\otimes a}}^T d\mu(x) \in \mathrm{MSym}(V^{\otimes a}),$$

where the constant $C_{n,r}$ is chosen so that $\mathsf{Tr}(\widetilde{M}_a) = 1$. Then we have

$$\|M_a - \widetilde{M}_a\|_{F1} \leq \gamma_{n,a}\frac{2a^2(2a+n-2)}{2r+n}.$$

Note that our formulation slightly differs from that in [1]: we have a constant $\gamma_{n,a}$, which is not present in [1], and the lowest value on r also slightly differs: we assume $r \geq a(2a+n-2) - n/2$ while [1] assumes $r \geq a^2(2a+n-2) - n/2$.

In the next section we indicate how to derive the convergence analysis of Theorem 1.1 from Theorem 1.4 and we will prove Theorem 1.4 in Section 2. For now let us just give a brief sketch of the key steps.

Assume we are given a matrix M satisfying the assumptions of Theorem 1.4. The starting point is to define its Q-representation: the polynomial $Q_M(x)$ (as in Theorem 1.4), and its P-representation: the polynomial $P_M(x)$ (as in Lemma 2.6), having the property that M can be obtained by integrating along the Haar measure with $P_M(x)$ as (signed) density function. The key fact is that these two polynomials, when expressed in the basis of spherical harmonics, have their low order Fourier coefficients which are very close. Based on this one may define a positive semidefinite matrix $\widetilde{M}_a$ which approximates well the reduced matrix M_a (obtained by taking a partial trace of M). While the matrix M_a relates to the upper bound (1), this matrix $\widetilde{M}_a$ provides a feasible solution to the lower bound (4), which permits a detailed analysis of the range between these two bounds.

1.4 Deriving the convergence analysis of Theorem 1.1

Here we show how to complete the convergence analysis in Theorem 1.1 using Theorem 1.4.

Let M be an optimal solution to the semidefinite program defining $\overline{T}^{(r)}$; so $M \in \mathrm{MSym}(V^{\otimes r})$, $M \succeq 0$ and $\mathsf{Tr}(M) = 1$. This implies $M_a := \mathsf{Tr}_{r-a}(M) \succeq 0$ and $\mathsf{Tr}(M_a) = 1$. By definition, we have:

$$\overline{T}^{(r)} = \langle Z_{T_r}, M \rangle = \langle Z_T, M_a \rangle.$$

Moreover, the polynomial $Q_M(x) := x^{\otimes r T} M x^{\otimes r}$ belongs to Σ_r and, by the choice of the constant $C_{n.r}$, its scaling $h(x) := C_{n,r} Q_M(x)$ belongs to Σ_r and satisfies $\int_{\mathbb{S}^{n-1}} h(x) d\mu(x) = 1$. Hence h is feasible for the program defining the lower bound $\underline{T}^{(r)}$. Using the definition of the matrix $\widetilde{M}_a$ in Theorem 1.4, we thus have the chain of inequalities:

$$\langle Z_T, \widetilde{M}_a \rangle = \int_{\mathbb{S}^{n-1}} T(x) h(x) d\mu(x) \leq \underline{T}^{(r)} \leq T_{\max} \leq \overline{T}^{(r)} = \langle Z_T, M_a \rangle.$$

We now apply Theorem 1.4 to the polynomial

$$F(x) := \frac{T_{\max}(\sum_i x_i^2)^a - T(x)}{T_{\max} - T_{\min}}.$$

Then, $Z_F = \frac{T_{\max} I - Z_T}{T_{\max} - T_{\min}}$. As $\mathsf{Tr}(M_a) = \mathsf{Tr}(\widetilde{M}_a) = 1$ we obtain

$$\langle Z_F, \widetilde{M}_a - M_a \rangle = \frac{\langle Z_T, M_a - \widetilde{M}_a \rangle}{T_{\max} - T_{\min}}.$$

Therefore, we obtain

$$\overline{T}^{(r)} - \underline{T}^{(r)} \leq \langle Z_T, M_a \rangle - \langle Z_T, \widetilde{M}_a \rangle = \langle Z_T, M_a - \widetilde{M}_a \rangle \leq \|M_a - \tilde{M}_a\|_{F1} (T_{\max} - T_{\min}).$$

Now, Theorem 1.4 implies that, for all integers r such that $r \geq a(2a+n-2) - n/2$,

$$\overline{T}^{(r)} - \underline{T}^{(r)} \leq \gamma_{n,a} \frac{2a^2(2a+n-2)}{2r+n} (T_{\max} - T_{\min}).$$

This concludes the proof of Theorem 1.1.

1.5 Reduction to the case of even degree polynomials

As shown in [1] the problem of optimizing an odd degree homogeneous polynomial can be reduced to the even degree case. For this consider an n-variate homogeneous polynomial $T(x)$ with odd degree $2a-1$ and define the $(n+1)$-variate polynomial $\tilde{T}(x_0, x) = x_0 T(x)$, which is homogeneous with even degree $2a$.

Lemma 1.5. *Consider the function $\varphi(t) = \frac{t^{2a-1}}{(1+t^2)^a}$. Then we have*

$$\max_{t \geq 0} \varphi(t) = \sqrt{\frac{(2a-1)^{2a-1}}{(2a)^{2a}}} =: \gamma_a.$$

Moreover, the maximum values of the polynomials $T(x)$ over $\mathbb{S}^{n-1}$ and $\tilde{T}(x_0, x)$ over $\mathbb{S}^n$ are related by

$$\tilde{T}_{\max} = \gamma_a T_{\max}.$$

Proof. The first claim follows using standard calculus. We now show the claim $\tilde{T}_{\max} = \gamma_a T_{\max}$. Indeed, we have

$$\tilde{T}_{\max} = \max_{(x_0,x)\in\mathbb{S}^n} \tilde{T}(x_0, x) = \max_{(x_0,x)\in\mathbb{R}^{n+1}} \frac{\tilde{T}(x_0, x)}{\|(x_0, x)\|^{2a}} = \max_{x_0\in\mathbb{R},x\in\mathbb{R}^n} \frac{x_0 T(x)}{(x_0^2 + \|x\|^2)^a},$$

which, in turn, is equal to $\max_{y\in\mathbb{R}^n} \frac{T(y)}{(1+\|y\|^2)^a} =: C$. The inequality $\tilde{T}_{\max} \geq C$ is clear. We show the reverse inequality: $\tilde{T}_{\max} \leq C$. For this pick $(x_0, x) \in \mathbb{R}^{n+1}$. If $x_0 \neq 0$, set $y = x/x_0$ and note that $\frac{x_0 T(x)}{(x_0^2+\|x\|^2)^a} = \frac{T(y)}{(1+\|y\|^2)^a} \leq C$. The case when $x_0 = 0$ follows using a continuity argument.

Now, by setting $x = y/\|y\|$, the program $C = \max_{y\in\mathbb{R}^n} \frac{T(y)}{(1+\|y\|^2)^a}$ can be rewritten as $C = \max\limits_{t\geq 0,x\in\mathbb{S}^{n-1}} \dfrac{t^{2a-1}T(x)}{(1 + t^2)^a} = \max\limits_{t\geq 0} \varphi(t)T_{\max}$. This shows the desired identity $\tilde{T}_{\max} = \gamma_a T_{\max}$. □

2 Proof of Theorem 1.4

In this section we will give the proof of Theorem 1.4. For this we first need to recall basic facts about spherical harmonic polynomials (we will use the monograph by Dai and Xu [2] as general reference). Then we present the P- and Q-representations for maximally symmetric matrices as considered in [1]. After that we are ready to prove Theorem 1.4.

2.1 Spherical harmonics

let $\mathcal{P}_d^n$ denote the set of real n-variate homogeneous polynomials with degree d. The Laplacian operator is $\Delta = \sum_{i=1}^n \frac{\partial^2}{(\partial x_i)^2}$, which maps $\mathcal{P}_d^n$ to $\mathcal{P}_{d-2}^n$. Then the set of *harmonic polynomials* is

$$\mathcal{H}_d^n = \{p \in \mathcal{P}_d^n : \Delta p = 0\}.$$

Spherical harmonics are the restrictions of harmonic polynomials to the unit sphere. By abuse of notation, $\mathcal{H}_d^n$ also denotes the set of spherical harmonics.

We consider the following inner product on the space $L^2(\mathbb{S}^{n-1}, \mu)$ of square integrable functions on $\mathbb{S}^{n-1}$:

$$\langle f, g\rangle_\mu = \int_{\mathbb{S}^{n-1}} f(x)g(x)d\mu(x).$$

Spherical harmonics of different degrees are orthogonal: $\langle f, g\rangle_\mu = 0$ if $f \in \mathcal{H}_j^n$, $g \in \mathcal{H}_k^n$ and $j \neq k$. The dimension of the space $\mathcal{H}_j^n$ is given by

$$N(n, j) := \dim \mathcal{H}_j^n = \binom{n + j - 1}{j} - \binom{n + j - 3}{j - 2},$$

with $N(n,0) = 1$. Let $\{s_{jm} : m \in [N(n,j)]\}$ denote an orthogonal basis of $\mathcal{H}_j^n$ with respect to the inner product $\langle \cdot, \cdot \rangle_\mu$ for each $j \geq 0$. Then the set $\{s_{jm} : j \in \mathbb{N}, m \in [N(n,j)]\}$ provides a basis of the set of polynomials restricted to the unit sphere. The polynomials in the basis are scaled so that

$$\langle s_{jm}, s_{j'm'} \rangle_\mu = \delta_{j,j'} \delta_{m,m'} \frac{1}{\omega_n}, \quad s_0 = \frac{1}{\sqrt{\omega_n}}$$

where $\omega_n = \frac{2\pi^{n/2}}{\Gamma(n/2)}$ denotes the surface area of $\mathbb{S}^{n-1}$.

Any homogeneous polynomial T with degree $2a$ can be decomposed in the basis of spherical harmonics:

$$T = \sum_{j=0}^{2a} \sum_{m=1}^{N(n,j)} t_{jm} s_{jm},$$

where the scalars t_{jm} are known as the *Fourier coefficients*. Note that $t_{jm} = 0$ for all odd j as T has even degree.

A fundamental property that we will use is the following *Funk-Hecke formula*.

Theorem 2.1. [Funk-Hecke formula][2, Theorem 1.2.9] *Consider a function $\varphi : [-1,1] \to \mathbb{R}$ such that $\int_{-1}^1 |\varphi(t)|(1-t^2)^{(n-3)/2} dt < \infty$ and integers $n \geq 2$, $j \geq 0$. Then there exists a constant $\tilde{\lambda}_j(\varphi)$ such that the following relation holds:*

$$\int_{\mathbb{S}^{n-1}} \varphi(x^T y) f(y) d\mu(y) = \tilde{\lambda}_j(\varphi) f(x) \quad \textit{for all } x \in \mathbb{S}^{n-1} \textit{ and } f \in \mathcal{H}_j^n.$$

The constant $\tilde{\lambda}_j(\varphi)$ is given by

$$\tilde{\lambda}_j(\varphi) = \frac{\omega_{n-1}}{\omega_n} \int_{-1}^1 \varphi(t) \frac{C_j^{\frac{n-2}{2}}(t)}{C_j^{\frac{n-2}{2}}(1)} (1-t^2)^{\frac{n-3}{2}} dt = \frac{\omega_{n-1}}{\omega_n} \int_{-1}^1 \varphi(t) P_j(t) (1-t^2)^{\frac{n-3}{2}} dt.$$

Here, $C_j^{\frac{n-2}{2}}(t)$ denotes the Gegenbauer polynomial of degree j and $P_j(t)$ is its normalization, so that $P_j(1) = 1$ (ignoring dependence on n for simplicity in notation) (see Section 3 for details).

Following [1] we use the application of the Funk-Hecke formula to the function $\varphi(t) = t^{2r}$, in which case one can compute the explicit value of the constants $\tilde{\lambda}_j(\varphi)$.

Proposition 2.2. [Application of Funk-Hecke formula] *Given integers $j, r \in \mathbb{N}$ there exists a constant $\lambda(n,r,j)$ such that the following identity holds:*

$$\int_{\mathbb{S}^{n-1}} (x^T y)^{2r} f(y) d\mu(y) = \frac{\omega_{n-1}}{\omega_n} \lambda(n,r,j) f(x) \quad \textit{for all } x \in \mathbb{S}^{n-1} \textit{ and } f \in \mathcal{H}_j^n.$$

The constant $\lambda(n,r,j)$ is given by

$$\lambda(n,r,j) = \int_{-1}^1 t^{2r} P_j(t) (1-t^2)^{\frac{n-3}{2}} dt. \tag{6}$$

Following [1], for any integers $r, j, m \in \mathbb{N}$ define the following 'spherical harmonic' matrices corresponding to the polynomial s_{jm}:

$$S^r_{jm} := \int_{\mathbb{S}^{n-1}} s_{jm}(x) x^{\otimes r} x^{\otimes r^T} d\mu(x).$$

Using the Funk-Hecke formula we get:

$$\langle S^r_{jm}, S^r_{j'm'} \rangle = \delta_{j,j'} \delta_{m,m'} \frac{\omega_{n-1}}{\omega_n} \lambda(n, r, j).$$

Note that each matrix S^r_{jm} is maximally symmetric. In fact one can use the spherical harmonic matrices to give an explicit description of the maximally symmetric matrix associated to any homogeneous polynomial.

Lemma 2.3. *Let $T(x)$ be a homogeneous polynomial of degree $2a$, with Fourier decomposition $T(x) = \sum_{j,m} t_{jm} s_{jm}(x)$. Its associated maximally symmetric matrix Z_T is given by*

$$Z_T = \frac{\omega_n}{\omega_{n-1} \lambda(n, a, j)} \sum_{j=0, j even}^{2a} \sum_{m=1}^{N(n,j)} t_{jm} S^a_{jm}.$$

Proof. Using the Funk-Hecke formula we obtain

$$x^{\otimes a^T} S^a_{jm} x^{\otimes a} = \int_{\mathbb{S}^{n-1}} s_{jm}(y)(x^T y)^{2a} d\mu(y) = s_{jm}(x) \frac{\omega_{n-1}}{\omega_n} \lambda(n, a, j).$$

It suffices now to sum up over all j, m at both sides and to use the unicity of the associated maximally symmetric matrix Z_T. $\square$

We now collect here the properties of the scalars $\lambda(n, r, j)$ that we will use for the proof of Theorem 1.4. The proofs of these properties are delayed till Section 3. Set

$$\epsilon(n, r, j) := \frac{j(j + n - 2)}{2r + n}.$$

Lemma 2.4. *We have: $\lambda(n, r, j) = 0$ if j is odd or if $j > 2r$, and $\lambda(n, r, j) > 0$ for any even integer $j \leq 2r$.*

Lemma 2.5. *Assume $n \geq 3$ and $r \geq a(2a + n - 2) - n/2$, i.e., $\epsilon(n, 2r, 2a) \leq 1$. Then, for any even integer $j \leq 2a$, we have*

$$0 \leq \frac{\lambda(n, r, 0)}{\lambda(n, r, j)} - 1 \leq \epsilon(n, r, 2a) = \frac{2a(2a + n - 2)}{2r + n}.$$

2.2 P- and Q-representations for maximally symmetric matrices

Given a matrix $M \in \mathrm{MSym}(V^{\otimes r})$, its *Q-representation* is the polynomial

$$Q_M(x) := x^{\otimes r^T} M x^{\otimes r},$$

which is homogeneous with degree $2r$.

Lemma 2.6. [P-representation] [1, Lemma 5.1] *For any matrix $M \in \mathrm{MSym}(V^{\otimes r})$ there exists a polynomial $P_M \in \mathbb{R}[x]$ such that*

$$M = \int_{\mathbb{S}^{n-1}} P_M(x) x^{\otimes r} x^{\otimes r^T} d\mu(x). \tag{7}$$

Proof. Let W denote the subspace consisting of the matrices in $\mathrm{MSym}(V^{\otimes r})$ that admit a P-polynomial representation as in (7). We show that $W^\perp = \{0\}$. For this, assume $M \in \mathrm{MSym}(V^{\otimes r})$ satisfies $\langle M, Z\rangle = 0$ for all $Z \in W$. Then, we have

$$0 = \langle M, \int_{\mathbb{S}^{n-1}} P(x) x^{\otimes r} x^{\otimes r^T} d\mu(x)\rangle = \int_{\mathbb{S}^{n-1}} P(x) Q_M(x) d\mu(x)$$

for all $P \in \mathbb{R}[x]$ and thus for all $P \in L^2(\mathbb{S}^{n-1}, \mu)$ (by density of the polynomials). This implies $Q_M(x) = 0$ on $\mathbb{S}^{n-1}$ and thus $Q_M = 0$. This in turns implies $M \in (\mathrm{MSym}(V^{\otimes r}))^\perp$ and thus $M = 0$. □

Next we indicate the link between the Fourier coefficients of the P- and Q-representations of M. Let us decompose both polynomials $P_M(x)$ and $Q_M(x)$ in the basis of spherical harmonics:

$$P_M(x) = \sum_{j\geq 0} \sum_{m=1}^{N(n,j)} p_{jm}^M s_{jm}(x), \quad Q_M(x) = \sum_{j=0}^{2r} \sum_{m=1}^{N(n,j)} q_{jm}^M s_{jm}(x).$$

Lemma 2.7. [1, Lemma 5.2] *Given $M \in \mathrm{MSym}(V^{\otimes r})$ and integers $j, m \in \mathbb{N}$, the following relation holds:*

$$q_{jm}^M = p_{jm}^M \frac{\omega_{n-1}}{\omega_n} \lambda(n, r, j).$$

Moreover, if $\mathsf{Tr}(M) = 1$ then the constant $C_{n,r}$ appearing in Theorem 1.4 is given by

$$C_{n,r} = \frac{\omega_n}{\omega_{n-1}\lambda(n, r, 0)}.$$

Proof. Using the P-representation (7) for the matrix M we obtain:

$$Q_M(x) = x^{\otimes r^T} \int_{\mathbb{S}^{n-1}} P_M(y) y^{\otimes r} y^{\otimes r^T} d\mu(y) x^{\otimes r} = \int_{\mathbb{S}^{n-1}} P_M(y)(x^T y)^{2r} d\mu(y)$$

$$= \sum_{j,m} p_{jm}^M \int_{\mathbb{S}^{n-1}} s_{jm}(y)(x^T y)^{2r} d\mu(y) = \sum_{j,m} p_{jm}^M \frac{\omega_{n-1}}{\omega_n} \lambda(n, r, j) s_{jm}(x),$$

where we use the Funk-Hecke formula for the last equality. The first claim now follows by equating with the Fourier coefficients of $Q_M(x)$.

By its definition, the constant $C_{n,r}$ is chosen so that

$$C_{n,r} \int_{\mathbb{S}^{n-1}} Q_M(x) d\mu(x) \; = \; 1.$$

On the one hand, we have $\int_{\mathbb{S}^{n-1}} Q_M(x) d\mu(x) = q_0^M / \sqrt{\omega_n}$. On the other hand, we have $1 = \mathsf{Tr}(M) = \int_{\mathbb{S}^{n-1}} P_M(x) d\mu(x) = p_0^M / \sqrt{\omega_n}$. Combining with the fact that $q_0^M = p_0^M \frac{\omega_{n-1}}{\omega_n} \lambda(n, r, 0)$ gives the final claimed value for $C_{n,r}$. □

2.3 Proof of Theorem 1.4

Let $M \in \mathrm{MSym}(V^{\otimes r})$ such that $M \succeq 0$ and $\mathsf{Tr}(M) = 1$. Setting $M_a = \mathsf{Tr}_{r-a}(M)$, we have $M_a \succeq 0$ and $\mathsf{Tr}(M_a) = 1$. Define the matrix

$$\widetilde{M}_a = C_{n,r} \int_{\mathbb{S}^{n-1}} Q_M(x) x^{\otimes a} x^{\otimes a^T} d\mu(x),$$

where $C_{n,r}$ is such that $\mathsf{Tr}(\widetilde{M}_a) = 1$, i.e., $C_{n,r} \int_{\mathbb{S}^{n-1}} Q_M(x) d\mu(x) = 1$.

Let $P_M(x) = \sum_{j,m} p^M_{jm} s_{jm}(x)$ be the P-representation of M, which enables us to decompose the matrix M_a using the spherical harmonic matrices:

$$\begin{aligned} M_a &= \textstyle\int_{\mathbb{S}^{n-1}} P_M(x) x^{\otimes a} x^{\otimes a^T} d\mu(x) \\ &= \textstyle\sum_{j,m} p^M_{jm} \int_{\mathbb{S}^{n-1}} s_{jm}(x) x^{\otimes a} x^{\otimes a^T} d\mu(x) \\ &= \textstyle\sum_{j,m} p^M_{jm} S^a_{jm}. \end{aligned}$$

Set $M^{odd}_a := \sum_{j \text{ odd},m} p^M_{jm} S^a_{jm}$ consisting of all terms for odd j. For any even j we use the relations in Lemma 2.7 to express p^M_{jm} in terms of q^M_{jm} and $C_{n,r}$ and obtain:

$$M_a = M^{odd}_a + \sum_{j \text{ even},m} q^M_{jm} C_{n,r} \frac{\lambda(n,r,0)}{\lambda(n,r,j)} S^a_{jm}. \tag{8}$$

In the same way, by using the Fourier decomposition of $Q_M(x)$ we obtain

$$\widetilde{M}_a = C_{n,r} \int_{\mathbb{S}^{n-1}} Q_M(x) x^{\otimes a} x^{\otimes a^T} d\mu(x) = C_{n,r} \sum_{j \text{ even},m} q^M_{jm} S^a_{jm} = \sum_{j \text{ even},j=0}^{2r} \widetilde{M}^j_a, \tag{9}$$

after setting $\widetilde{M}^j_a = \sum_{m=1}^{N(n,j)} C_{n,r} q^M_{jm} S^a_{jm}$ for each j and noting that $q^M_{jm} = 0$ for all odd j since $Q_M(x)$ has even degree.

Combining relations (8) and (9) we obtain

$$M_a - \widetilde{M}_a = M^{odd}_a + \sum_{j \text{ even}} \widetilde{M}^j_a \left(\frac{\lambda(n,r,0)}{\lambda(n,r,j)} - 1 \right). \tag{10}$$

We can now proceed to complete the proof of Theorem 1.4. Let F be a homogeneous polynomial with degree $2a$ such that $|F(x)| \leq 1$ on $\mathbb{S}^{n-1}$ and let Z_F be its associated maximally symmetric matrix. As F has even degree, its Fourier decomposition involves only spherical harmonics s_{jm} with j even and thus, in view of Lemma 2.3, the associated matrix Z_F is a linear combination of the matrices S^a_{jm} for even j. Hence it is orthogonal to any $S^a_{j'm'}$ with j' odd and thus we can deduce that $\langle Z_F, M^{odd}_a \rangle = 0$. Therefore we obtain

$$\langle Z_F, M_a - \widetilde{M}_a \rangle = \sum_{j \text{ even},j=0}^{2a} \left(\frac{\lambda(n,r,0)}{\lambda(n,r,j)} - 1 \right) \langle Z_F, \widetilde{M}^j_a \rangle. \tag{11}$$

Using Lemma 2.5 combined with Lemma 2.8 below we can conclude the proof of Theorem 1.4. Indeed,

$$|\langle Z_F, M_a - \widetilde{M}_a\rangle| \leq \sum_{j \text{ even}, j=0}^{2a} \left(\frac{\lambda(n,r,0)}{\lambda(n,r,j)} - 1\right) |\langle Z_F, \widetilde{M}_a^j\rangle| \leq a\epsilon(n,r,2a)\gamma_{n,a}.$$

Recall the constant $\gamma_{n,a}$, introduced in (5), so that $\|\cdot\| \leq \|\cdot\|_{F1} \leq \gamma_{n,a}\|.\|$.

Lemma 2.8. *We have* $\|\widetilde{M}_a^j\|_{F1} \leq \gamma_{n,a}$ *for all even* j.

Proof. Note first that $\|\widetilde{M}_a\|_{F1} \leq 1$. Indeed, for any degree $2a$ homogeneous polynomial F such that $|F(x)| \leq 1$ on $\mathbb{S}^{n-1}$, we have

$$|\langle Z_F, \widetilde{M}_a\rangle| \leq C_{n,r} \int_{\mathbb{S}^{n-1}} Q_M(x)|F(x)|d\mu(x) \leq C_{n,r} \int_{\mathbb{S}^{n-1}} Q_M(x)d\mu(x) = 1.$$

This implies $\|\widetilde{M}_a\| \leq 1$. As $\widetilde{M}_a = \sum_j \widetilde{M}_a^j$, where the $\widetilde{M}_a^j$ are pairwise orthogonal, we can conclude that, for all j, $\|\widetilde{M}_a^j\| \leq 1$ and thus $\|\widetilde{M}_a^j\|_{F1} \leq \gamma_{n,a}$. $\square$

3 Bounding the constants $\lambda(n,r,j)$ in Funk-Hecke formula

Here we proceed to show the results from Lemmas 2.4 and 2.5 about the behaviour of the constants $\lambda(n,r,j)$ appearing in Funk-Hecke formula in Proposition 2.2.

First we introduce the normalized Gegenbauer polynomials:

$$P_j(t) = \frac{C_j^{\frac{n-2}{2}}(t)}{C_j^{\frac{n-2}{2}}(1)},$$

so that $P_j(1) = 1$. Here, following relations (B.2.1)-(B.2.2) in [2], $C_j^{\frac{n-2}{2}}(t)$ is the Gegenbauer polynomial, obtained as the following Jacobi polynomial:

$$C_j^{\frac{n-2}{2}}(t) = \frac{(n-2)_j}{\left(\frac{n-1}{2}\right)_j} P_j^{\frac{n-3}{2},\frac{n-3}{2}}(t), \quad C_j^{\frac{n-2}{2}}(1) = \frac{(n-2)_j}{j!},$$

so that

$$P_j(t) = \frac{j!}{\left(\frac{n-1}{2}\right)_j} P_j^{\frac{n-3}{2},\frac{n-3}{2}}(t) = j! \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(j+\frac{n-1}{2}\right)} P_j^{\frac{n-3}{2},\frac{n-3}{2}}(t).$$

Recall that, for a scalar a and an integer $j \geq 0$,

$$(a)_j = a(a+1)\cdots(a+j-1) = \frac{\Gamma(a+j)}{\Gamma(a)}, \tag{12}$$

where the last equality follows using the following property of the Gamma function: $\Gamma(z+1) = z\Gamma(z)$.

Using the "differential definition" of the Jacobi polynomials (see, e.g., relation (B.1.2) in [2]):

$$P_j^{\frac{n-3}{2},\frac{n-3}{2}}(t) = \frac{(-1)^j}{2^j j!}(1-t^2)^{-\frac{n-3}{2}}\Big(\frac{d}{dt}\Big)^j\Big((1-t^2)^{j+\frac{n-3}{2}}\Big)$$

one obtains the following "differential definition" for the normalized Gegenbauer polynomial (see relation (195) in [1]):

$$P_j(t) = \Big(-\frac{1}{2}\Big)^j \frac{\Gamma\Big(\frac{n-1}{2}\Big)}{\Gamma\Big(j+\frac{n-1}{2}\Big)}(1-t^2)^{-\frac{n-3}{2}}\Big(\frac{d}{dt}\Big)^j\Big((1-t^2)^{j+\frac{n-3}{2}}\Big). \tag{13}$$

We now proceed to compute the constant $\lambda(n,r,j)$ from (6):

$$\lambda(n,r,j) := \int_{-1}^{1} t^{2r} P_j(t)(1-t^2)^{\frac{n-3}{2}} dt.$$

Lemma 3.1. [1, Lemma A.1] *Assume $n \geq 3$. We have:*

$$\lambda(n,r,j) = \begin{cases} 0 & \text{if } j \text{ is odd or } j > 2r, \\ \frac{\sqrt{\pi}}{2^{2r}} \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma(2r+1)}{\Gamma\left(r+1-\frac{j}{2}\right)\Gamma\left(r+\frac{n+j}{2}\right)} & \text{if } j \text{ is even and } j \leq 2r. \end{cases}$$

Proof. Using the definition (13) of $P_j(t)$ and integration by parts one gets

$$\int_{-1}^{1} t^{2r} P_j(t)(1-t^2)^{\frac{n-3}{2}} dt = \Big(-\frac{1}{2}\Big)^j \frac{\Gamma\Big(\frac{n-1}{2}\Big)}{\Gamma\Big(\frac{n-1}{2}+j\Big)} \int_{-1}^{1} t^{2r}\Big(\frac{d}{dt}\Big)^j(1-t^2)^{j+\frac{n-3}{2}} dt$$

$$= \Big(\frac{1}{2}\Big)^j \frac{\Gamma\Big(\frac{n-1}{2}\Big)}{\Gamma\Big(\frac{n-1}{2}+j\Big)} \int_{-1}^{1} (\Big(\frac{d}{dt}\Big)^j(t^{2r}))(1-t^2)^{j+\frac{n-3}{2}} dt.$$

Note that $\left(\frac{d}{dt}\right)^j(t^{2r}) = 0$ if $j > 2r$ and, if $j \leq 2r$ then

$$\Big(\frac{d}{dt}\Big)^j(t^{2r}) = (2r)(2r-1)\cdots(2r-j+1)t^{2r-j} = \frac{\Gamma(2r+1)}{\Gamma(2r+1-j)}t^{2r-j}.$$

If j is odd the above integral vanishes. So assume now j is even, $j = 2k$ with $k \leq r$. Changing variable $s = t^2$ we obtain

$$\int_{-1}^{1} t^{2(r-k)}(1-t^2)^{j+\frac{n-3}{2}} dt = \int_0^1 s^{r-k-\frac{1}{2}}(1-s)^{j+\frac{n-3}{2}} ds$$

$$= B\Big(r - \frac{j-1}{2}, j+1+\frac{n-3}{2}\Big) = B\Big(r - \frac{j-1}{2}, j+\frac{n-1}{2}\Big)$$
$$= \frac{\Gamma\Big(r-\frac{j-1}{2}\Big)\Gamma\Big(j+\frac{n-1}{2}\Big)}{\Gamma\Big(r+\frac{j+n}{2}\Big)}.$$

Here $B(x,y)$ is the Beta function, defined by

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt,$$

and we have used the following link to the Gamma function:

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

(See, e.g., [3, Chapter 1.1].) Putting things together we obtain that, for any even integer $j \le 2r$:

$$\lambda(n,r,j) = \int_{-1}^{1} t^{2r}P_j(t)(1-t^2)^{\frac{n-3}{2}}dt$$
$$= \Big(\frac{1}{2}\Big)^j \frac{\Gamma\Big(\frac{n-1}{2}\Big)}{\Gamma\Big(\frac{n-1}{2}+j\Big)} \frac{\Gamma(2r+1)}{\Gamma(2r+1-j)} \frac{\Gamma\Big(r-\frac{j-1}{2}\Big)\Gamma\Big(j+\frac{n-1}{2}\Big)}{\Gamma\Big(r+\frac{j+n}{2}\Big)}$$
$$= \Big(\frac{1}{2}\Big)^j \frac{\Gamma\Big(r-\frac{j-1}{2}\Big)}{\Gamma(2r+1-j)} \frac{\Gamma\Big(\frac{n-1}{2}\Big)\Gamma(2r+1)}{\Gamma\Big(r+\frac{j+n}{2}\Big)}.$$

Now we use the Legendre duplication formula:

$$\frac{\Gamma(z)}{\Gamma(2z)} = 2^{1-2z}\frac{\Gamma\Big(\frac{1}{2}\Big)}{\Gamma\Big(z+\frac{1}{2}\Big)}$$

applied to $z = r - \frac{j-1}{2}$ to simplify the first fraction and get

$$\frac{\Gamma\Big(r-\frac{j-1}{2}\Big)}{\Gamma(2r+1-j)} = 2^{j-2r}\frac{\Gamma\Big(\frac{1}{2}\Big)}{\Gamma\Big(r+1-\frac{j}{2}\Big)}.$$

Using the fact that $\Gamma\Big(\frac{1}{2}\Big) = \sqrt{\pi}$ we obtain:

$$\lambda(n,r,j) = \frac{\sqrt{\pi}}{2^{2r}} \frac{\Gamma\Big(\frac{n-1}{2}\Big)\Gamma(2r+1)}{\Gamma\Big(r+1-\frac{j}{2}\Big)\Gamma\Big(r+\frac{j+n}{2}\Big)}.$$

This completes the proof of Lemma 3.1. $\square$

Corollary 3.2. *For any even $j \le 2r$, $j = 2k$ with $k \le r$, we have*

$$\frac{\lambda(n,r,j)}{\lambda(n,r,0)} = \frac{\Gamma(r+1)\Gamma\Big(r+\frac{n}{2}\Big)}{\Gamma\Big(r+1-\frac{j}{2}\Big)\Gamma\Big(r+\frac{n+j}{2}\Big)} = \prod_{i=0}^{k-1} \frac{r-i}{r+\frac{n}{2}+k-1-i}.$$

Proof. Directly from Lemma 3.1 and simplifying the Gamma functions:

$$\Gamma(r+1) = r(r-1)\cdots(r+1-k)\Gamma(r+1-k),$$
$$\Gamma\Big(r+k+\frac{n}{2}\Big) = \Big(r+\frac{n}{2}+k-1\Big)\cdots\Big(r+\frac{n}{2}\Big)\Gamma\Big(r+\frac{n}{2}\Big).$$

□

Lemma 3.3. *Set $\epsilon(n,r,j) := \frac{j(j+n-2)}{2r+n}$. For even $j \le 2r$ we have:*

$$\frac{\lambda(n,r,j)}{\lambda(n,r,0)} \ge 1 - \frac{1}{2}\epsilon(n,r,j).$$

Proof. For $i \in [k-1]$ we have

$$\frac{r-i}{r+\frac{n}{2}+k-1-i} = 1 - \Big(\frac{n}{2}+k-1\Big)\frac{1}{r+\frac{n}{2}+k-1-i} \ge \frac{r-k+1}{r+\frac{n}{2}},$$

which (using Corollary 3.2) implies

$$\frac{\lambda(n,r,j)}{\lambda(n,r,0)} \ge \Big(\frac{r-k+1}{r+\frac{n}{2}}\Big)^k = \Big(1 - \frac{k-1+\frac{n}{2}}{r+\frac{n}{2}}\Big)^k \ge 1 - k\frac{k-1+\frac{n}{2}}{r+\frac{n}{2}},$$

where for the last inequality we use the fact that $(1-t)^k \ge 1-kt$ for all $t \in [0,1]$. This gives the desired inequality. □

Lemma 3.4. *The parameter $\lambda(n,r,j)$ is decreasing in j (j even).*

Proof. We verify that $\lambda(n,r,2k) > \lambda(n,r,2k+2)$ if $k \le r-1$. Indeed, using Lemma 3.1 we have

$$\frac{\lambda(n,r,2k)}{\lambda(n,r,2k+2)} = \frac{\Gamma(r-k)\Gamma\Big(r+\frac{n}{2}+k+1\Big)}{\Gamma(r+1-k)\Gamma\Big(r+\frac{n}{2}+k\Big)} = \frac{r+\frac{n}{2}+k}{r-k} > 1.$$

□

We can now finish the proof of Lemma 2.5: Assume $n \ge 3$, $\epsilon(n,r,2a) \le 1$, i.e., $r \ge a(2a+n-2) - n/2$. The claim is that, for any even $j \le 2a$, we have

$$\frac{\lambda(n,r,0)}{\lambda(n,r,j)} - 1 \le \epsilon(n,r,2a).$$

Since, by Lemma 3.4, $\lambda(n,r,j) \geq \lambda(n,r,2a)$, it suffices to show that

$$\frac{\lambda(n,r,0)}{\lambda(n,r,2a)} - 1 \leq \epsilon(n,r,2a).$$

By Lemma 3.3, $\frac{\lambda(n,r,2a)}{\lambda(n,r,0)} \geq \frac{2-\epsilon(n,r,2a)}{2}$, which implies

$$\frac{\lambda(n,r,0)}{\lambda(n,r,2a)} - 1 \leq \frac{2}{2-\epsilon(n,r,2a)} - 1 = \frac{\epsilon(n,r,2a)}{2-\epsilon(n,r,2a)} \leq \epsilon(n,r,2a),$$

where the last inequality holds since $\epsilon(n,r,2a) \leq 1$.

Remark: This shows that the quantity $\frac{\lambda(n,r,0)}{\lambda(n,r,2a)} - 1$ is in $O\Big(\frac{1}{r}\Big)$. Note that this is the right rate of convergence. For instance, for $a = 1$, using Corollary 3.2 one finds that $\frac{\lambda(n,r,0)}{\lambda(n,r,2)} - 1 = \frac{n/2}{r}$.

Acknowledgements. We thanks Etienne de Klerk and Lucas Slot for several useful discussions.

References

[1] A.C. Doherty, S. Wehner. Convergence of SDP hierarchies for polynomial optimization on the hypersphere. arXiv:1210.5048v2, 2013.

[2] F. Dai and Y. Xu. *Approximation Theory and Harmonic Analysis on Spheres and Balls.* Springer, 2013.

[3] C.F. Dunkl and Y. Xu. *Orthogonal Polynomials of Several Variables.* Encyclopedia of Mathematics, Cambridge University Press, 2001.

[4] E. de Klerk, M. Laurent, P. Parrilo. On the equivalence of algebraic approaches to the minimization of forms on the simplex. In *Positive Polynomials in Control* (D. Henrion and A. Garulli, eds.), *Lecture Notes on Control and Information Sciences,* Vol. 312, pages 121-133, Springer, 2005.

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

[6] J.-B. Lasserre. A new look at nonnegativity on closed sets and polynomial optimization. SIAM *Journal on Optimization,* 21(3), 864–885, 2011.

[7] P. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization,* PhD thesis, California Institute of Technology, 2000.