



**HAL**  
open science

# CONVEX OPTIMIZATION AND PARSIMONY OF $L_p$ -BALLS REPRESENTATION \*

Jean-Bernard Lasserre

► **To cite this version:**

Jean-Bernard Lasserre. CONVEX OPTIMIZATION AND PARSIMONY OF  $L_p$ -BALLS REPRESENTATION \*. SIAM Journal on Optimization, 2016, SIAM Journal on Optimization, 26 (1), pp.247–273. 10.1137/140983082 . hal-01053891v3

**HAL Id: hal-01053891**

**<https://hal.science/hal-01053891v3>**

Submitted on 5 Feb 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## CONVEX OPTIMIZATION AND PARSIMONY OF $L_p$ -BALLS REPRESENTATION\*

JEAN B. LASSERRE†

**Abstract.** In the family of unit balls with constant volume we look at the ones whose algebraic representation has some extremal property. We consider the family of nonnegative homogeneous polynomials of even degree  $p$  whose sublevel set  $\mathbf{G} = \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$  (a unit ball) has the same fixed volume and want to find in this family the polynomial that minimizes either the parsimony-inducing  $\ell_1$ -norm or the  $\ell_2$ -norm of its vector of coefficients. Equivalently, among all degree- $p$  polynomials of constant  $\ell_1$ - or  $\ell_2$ -norm, which one minimizes the volume of its level set  $\mathbf{G}$ ? We first show that in both cases this is a convex optimization problem with a unique optimal solution  $g_1^*$  or  $g_2^*$ , respectively. We also show that  $g_1^*$  is the  $L_p$ -norm polynomial  $\mathbf{x} \mapsto \sum_{i=1}^n x_i^p$ , thus recovering a parsimony property of the  $L_p$ -norm polynomial via  $\ell_1$ -norm minimization. This once again illustrates the power and versatility of the  $\ell_1$ -norm relaxation strategy in optimization when one searches for an optimal solution with parsimony properties. Next we show that  $g_2^*$  is not sparse at all (and thus differs from  $g_1^*$ ) but is still a sum of  $p$ -powers of linear forms. In fact, and surprisingly, for  $p = 2, 4, 6, 8$ , we show that  $g_2^* = (\sum_i x_i^2)^{p/2}$ , whose level set is the Euclidean (i.e., the  $L_2$ -norm) ball. We also characterize the unique optimal solution of the same problem where one searches for a sum of squares homogeneous polynomial that minimizes the (parsimony-inducing) nuclear norm of its associated (positive semidefinite) Gram matrix, hence aiming at finding a solution which is a sum of a few squares only. Again for  $p = 2, 4$  the optimal solution is  $(\sum_i x_i^2)^{p/2}$ , whose level set is the Euclidean ball, and when  $p \in 4\mathbb{N}$ , this is also true when  $n$  is sufficiently large. Finally, we also extend these results to generalized homogeneous polynomials, which include  $L_p$ -norms when  $0 < p$  is rational.

**Key words.** convex optimization, computational geometry,  $L_p$ -balls, parsimony-inducing norms

**AMS subject classification.** 90C22

**DOI.** 10.1137/140983082

**1. Introduction.** It is well known that the shape of the Euclidean unit ball  $\mathbf{B}_2 = \{\mathbf{x} : \sum_{i=1}^n x_i^2 \leq 1\}$  has spectacular geometric properties with respect to other shapes. For instance, the sphere has the smallest surface area among all surfaces enclosing a given volume, and it encloses the largest volume among all closed surfaces with a given surface area; Hilbert and Cohn-Vossen [8] even describe eleven geometric properties of the sphere.

But  $\mathbf{B}_2$  has also another spectacular (nongeometric) property related to its algebraic representation which is obvious even to people with only a little background in mathematics: Namely, its defining polynomial  $\mathbf{x} \mapsto g^{(2)}(\mathbf{x}) := \sum_{i=1}^n x_i^2$  cannot be simpler. Indeed, among all quadratic homogeneous polynomials  $\mathbf{x} \mapsto g(\mathbf{x}) = \sum_{i \leq j} g_{ij} x_i x_j$  that define a bounded ball  $\{\mathbf{x} : g(\mathbf{x}) \leq 1\}$ ,  $g^{(2)}$  is the one that minimizes the “cardinality norm”  $\|g\|_0 := \#\{(i, j) : g_{ij} \neq 0\}$  (which actually is not a norm). Only  $n$  coefficients of  $g^{(2)}$  do not vanish, and there cannot be fewer than  $n$  nonzero coefficients to define a bounded ball  $\{\mathbf{x} : g(\mathbf{x}) \leq 1\}$ . The same is true for the  $L_d$ -unit ball  $\mathbf{B}_d = \{\mathbf{x} : \sum_{i=1}^n x_i^d \leq 1\}$  and its defining polynomial  $\mathbf{x} \mapsto g^{(d)}(\mathbf{x}) := \sum_i x_i^d$  for any even integer  $d > 2$ , when compared to any other homogeneous polynomial  $g$  of

\*Received by the editors August 19, 2014; accepted for publication (in revised form) October 1, 2015; published electronically January 26, 2016. This work was partly supported by a PGMO grant of the Fédération Mathématique Jacques Hadamard (Paris).

<http://www.siam.org/journals/siopt/26-1/98308.html>

†LAAS-CNRS and Institute of Mathematics, University of Toulouse, 7 Av. Colonel Roche, 31031 Toulouse Cédex 4, France (lasserre@laas.fr).

degree  $d$  whose sublevel set  $\{\mathbf{x} : g(\mathbf{x}) \leq 1\}$  has finite Lebesgue volume (and so  $g$  is necessarily nonnegative). Indeed, again  $\|g^{(d)}\|_0 = n$ ; i.e., out of potentially  $\binom{n+d-1}{d}$  coefficients of  $g^{(d)}$ , only  $n$  do not vanish. In other words,  $g^{(d)}$  is an optimal solution of the optimization problem

$$(1.1) \quad \mathbf{P}_0 : \inf_g \{ \|g\|_0 : \text{vol}(\mathbf{G}) \leq \rho_d; \quad g \in \mathbf{Hom}_d \},$$

where the minimum is taken over all homogeneous polynomials of degree  $d$  and  $\rho_d$  denotes the Lebesgue volume of the  $L_d$ -unit ball  $\mathbf{B}_d$ .

So a natural question arises: *In view of the many “geometric properties” of the unit ball  $\mathbf{B}_d$ , is the “algebraic sparsity” of its representation  $\{\mathbf{x} : \sum_i x_i^d \leq 1\}$  a coincidence or does it also correspond to a certain extremal property on all unit balls with the same volume?*

Thus we are interested in the following optimization problem in computational geometry and with an algebraic flavor.

*Given an even integer  $d$ , determine the homogeneous polynomial  $g_1^*$  (resp.,  $g_2^*$ ) of degree  $d$  whose  $\ell_1$ -norm  $\|g_1^*\|_1$  (resp.,  $\ell_2$ -norm  $\|g_2^*\|_2$ ) of its vector of coefficients is the minimum among all degree- $d$  homogeneous polynomials with the same (fixed) volume of their sublevel set  $\mathbf{G} = \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$ . That is, solve*

$$(1.2) \quad \inf_g \{ \|g\|_{p=1,2} : \text{vol}(\mathbf{G}) = \rho_d; \quad g \text{ homogeneous of degree } d \}.$$

(Notice that in (1.2) the constraint  $\text{vol}(\mathbf{G}) = \rho_d$  implies that  $g$  is nonnegative, and one may also replace the constraint  $\text{vol}(\mathbf{G}) = \rho_d$  with  $\text{vol}(\mathbf{G}) \leq \rho_d$ .) In particular, we have the following: *Can the parsimony property of the  $L_d$ -unit balls be recovered from (1.2) with the parsimony-inducing  $\ell_1$ -norm  $\|g\|_1$  (instead of minimizing the nasty function  $\|\cdot\|_0$  in (1.1))?*

By homogeneity, this problem also has the equivalent formulation: *Among all homogeneous polynomials  $g$  of degree  $d$  and with constant norm  $\|g\|_1 = 1$  (or  $\|g\|_2 = 1$ ) find the one with level set  $\mathbf{G}$  of minimum volume.*

One goal of this paper is to prove that (1.2) is a convex optimization problem with a *unique optimal solution*  $g_1^* = g^{(d)}$ . In addition,  $g^{(d)}$  cannot be an optimal solution of (1.2) when one minimizes the  $\ell_2$ -norm  $\|g\|_2$  (except when  $d = 2$ ). This illustrates in this context of computational geometry that, again, the sparsity-inducing  $\ell_1$ -norm does a perfect job in the relaxation (1.2) (with  $\|\cdot\|_1$ ) of problem (1.1) with  $\|\cdot\|_0$ . This convex “relaxation trick” in (nonconvex)  $\ell_0$ -optimization has been used successfully in several important applications; see, e.g., Candès, Romberg, and Tao [4], Donoho [5], and Donoho and Elad [6] in compressed sensing applications and Recht, Fazel, and Parrilo [15] for matrix applications (where the small-rank induced *nuclear norm* is the matrix analogue of the  $\ell_1$ -norm). For more details on optimization with sparsity constraints and/or sparsity-induced penalties, the interested reader is referred to Beck and Eldar [3] and Bach et al. [2].

To address our problem we consider the following framework: Let  $\mathbf{Hom}_d \subset \mathbb{R}[\mathbf{x}]_d$  be the vector space of homogeneous polynomials of even degree  $d$ , and given  $g \in \mathbf{Hom}_d$ , let  $\mathbf{g} = (g_\alpha)$  be its vector of coefficients, i.e.,

$$\mathbf{x} \mapsto g(\mathbf{x}) = \sum_{\alpha} g_{\alpha} \mathbf{x}^{\alpha} \left( = \sum_{\alpha} g_{\alpha} \mathbf{x}_1^{\alpha_1} \cdots \mathbf{x}_n^{\alpha_n} \right), \quad \sum_i \alpha_i = d,$$

with standard  $\ell_1$ -norm  $\|g\|_1 = |\mathbf{g}| = \sum_{\alpha} |g_{\alpha}|$ . With any  $g \in \mathbf{Hom}_d$  is associated its sublevel set  $\mathbf{G} \subset \mathbb{R}^n$  defined by

$$(1.3) \quad \mathbf{G} := \{ \mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) \leq 1 \}, \quad g \in \mathbf{Hom}_d.$$

In particular, with  $\mathbf{x} \mapsto g^{(d)}(\mathbf{x}) := \sum_{i=1}^n x_i^d$ , the associated sublevel set is nothing less than the standard  $L_d$ -unit ball

$$\mathbf{B}_d = \left\{ \mathbf{x} : \sum_{i=1}^n x_i^d \leq 1 \right\} =: \{ \mathbf{x} : \|\mathbf{x}\|_d^d \leq 1 \},$$

whose Lebesgue volume  $\text{vol}(\mathbf{B}_d)$  is denoted  $\rho_d$ . (When  $g \in \mathbf{Hom}_d$  is convex, then  $\mathbf{x} \mapsto g(\mathbf{x})$  defines a norm  $\|\mathbf{x}\|_g := g(\mathbf{x})^{1/d}$  with  $\mathbf{G}$  as associated unit ball.)

**Contribution.** (a) In the first contribution we prove that the optimization problem

$$(1.4) \quad \mathbf{P}_1 : \inf_g \{ \|g\|_1 : \text{vol}(\mathbf{G}) \leq \rho_d; \quad g \in \mathbf{Hom}_d \}$$

has a unique optimal solution  $g_1^*$  which is the  $L_d$ -norm polynomial  $g^{(d)}$ . Observe that  $g_1^*$  has the minimal number  $n$  of coefficients over potentially  $s(d) := \binom{n-1+d}{d}$  coefficients. (Indeed, for a polynomial  $g \in \mathbf{Hom}_d$  with  $m < n$  nonzero coefficients, its sublevel set  $\mathbf{G}$  cannot have finite Lebesgue volume.) Therefore the  $L_d$ -norm polynomial  $g^{(d)}$  associated with the unit ball  $\mathbf{B}_d$  is the “sparsest” solution among all  $g \in \mathbf{Hom}_d$  such that  $\text{vol}(\mathbf{G}) \leq \text{vol}(\mathbf{B}_d)$ . In particular,  $g^{(d)}$  not only solves problem  $\mathbf{P}_1$  but also solves the nonconvex optimization problem  $\mathbf{P}_0$  of which  $\mathbf{P}_1$  is a “convex relaxation.” But this is also equivalent to stating that among all homogeneous polynomials  $g$  of degree  $d$  with constant  $\ell_1$ -norm, the  $L_d$ -norm polynomial  $g^{(d)}$  is the one whose associated  $L_d$ -unit ball  $\mathbf{B}_d$  has minimum volume (i.e.,  $\text{vol}(\mathbf{B}_d) \leq \text{vol}(\mathbf{G})$ ). (Notice that, in fact,  $\mathbf{P}_0$  is easy and even easier than  $\mathbf{P}_1$ .)

(b) In our second contribution we consider the  $\ell_2$ -norm version of (1.4):

$$(1.5) \quad \mathbf{P}_2 : \inf_g \{ \|g\|_2 : \text{vol}(\mathbf{G}) \leq \rho_d; \quad g \in \mathbf{Hom}_d \},$$

with weighted Euclidean norm  $g \mapsto \|g\|_2$  defined by

$$\|g\|_2^2 := \sum_{|\alpha|=d} c_{\alpha} g_{\alpha}^2, \quad g \in \mathbf{Hom}_d, \quad \text{where } c_{\alpha} := \frac{d!}{\alpha_1! \cdots \alpha_n!},$$

with  $g$  now written as  $g(\mathbf{x}) = \sum_{\alpha} c_{\alpha} g_{\alpha} \mathbf{x}^{\alpha}$ . (This norm on forms has very specific properties, as discussed in, e.g., Reznick [16].)

We then show that  $\mathbf{P}_2$  also has a unique optimal solution  $g_2^*$ , but in contrast to the sparse optimal solution  $g_1^* = g^{(d)} = \sum_i x_i^d$  of problem  $\mathbf{P}_1$ , the optimal solution  $g_2^*$  of  $\mathbf{P}_2$  is not sparse. Indeed, as we will see in section 4, all coefficients of  $g_2^*$  corresponding to monomials  $\mathbf{x}^{2\beta}$  with  $|\beta| = d/2$  are nonzero. In addition,  $g_2^*$  is a particular sum of squares (SOS) polynomial as it is a sum of  $d$ -powers of linear forms. (Notice that  $g_1^*$  is also a (very particular and simple) sum of  $d$ -powers of linear forms.) In fact, one proves the rather surprising result that if  $d = 2, 4, 6$ , and  $8$ , the unique optimal solution of  $\mathbf{P}_2$  is  $g_2^* = (\sum_i x_i^2)^{d/2}$ , whose level set is the Euclidean ball  $\mathbf{B}_2$ . We conjecture that the result also holds for arbitrary even  $d$ .

(c) We also consider the SOS version of  $\mathbf{P}_1$ ; that is, one now searches for a degree- $d$  SOS homogeneous polynomial  $g_{\mathbf{Q}}(\mathbf{x}) = \mathbf{v}_{d/2}(\mathbf{x})\mathbf{Q}\mathbf{v}_{d/2}(\mathbf{x})$ ,  $\mathbf{Q} \succeq 0$  (where  $\mathbf{v}_{d/2}(\mathbf{x}) = (\mathbf{x}^\alpha)$ ,  $|\alpha| = d/2$ ). That is, one characterizes the unique optimal solution  $g_3^* = g_{\mathbf{Q}^*}$  associated with optimal solutions  $\mathbf{Q}^*$  of the optimization problem

$$(1.6) \quad \mathbf{P}_3 : \quad \inf_{\mathbf{Q} \succeq 0} \{ \text{trace}(\mathbf{Q}) : \text{vol}(\mathbf{G}_{\mathbf{Q}}) \leq \rho_d; \mathbf{Q} \succeq 0 \}.$$

In this matrix context,  $\text{trace}(\mathbf{Q})$  is the parsimony-inducing *nuclear norm* of  $\mathbf{Q}$ , and solving  $\mathbf{P}_3$  aims at finding an optimal solution  $\mathbf{Q}^*$  with small rank, which translates into a homogeneous polynomial  $g_3^* = g_{\mathbf{Q}^*}$  which is a sum of few squares (again, a parsimony property). For  $d = 2$  we prove that the optimal solution of  $\mathbf{P}_3$  is  $g_3^* = g^{(2)}$ . However, for even integers  $d \geq 4$ , the polynomial  $g^{(d)}$  is not an optimal solution of  $\mathbf{P}_3$ . In fact, for  $d = 4$ ,  $g_3^*(\mathbf{x}) = \theta(\sum_i x_i^2)^{d/2}$  (with  $\theta$  a scaling factor to make  $\text{vol}(\{\mathbf{x} : g_3^*(\mathbf{x}) \leq 1\}) = \rho_d$ ). The same result holds for any  $d$  of the form  $d = 4p$  (for some integer  $p \geq 1$ ), provided that  $n$  is sufficiently large. Notice that in this case  $(\sum_i x_i^2)^{d/2}$  is the *single square*  $((\sum_i x_i^2)^p)^2$  (again a parsimony property), and so the associated optimal solution  $\mathbf{Q}^*$  has rank 1.

(d) Finally we also show that results in (a) and (b) extend to the case of  $L_d$ -unit balls with other values of  $d$  (including  $d = 1$  and  $d$  rational), in which case one now deals with positively homogeneous “generalized polynomials” (instead of homogeneous polynomials), and one has to define an appropriate finite-dimensional analogue of  $\mathbf{Hom}_d$ . This includes the interesting case of the  $L_1$ -unit ball  $\{\mathbf{x} : \sum_i |x_i| \leq 1\}$  and, when  $d < 1$ , balls which are not associated with norms.

To conclude, this paper should be viewed as a contribution of convex optimization to show that the usual representation of  $L_d$ -balls has extremal properties with respect to some well-known norms and, in particular, parsimony-inducing norms.

## 2. Notation, definitions, and preliminary results.

**2.1. Notation and definitions.** Let  $\mathbb{R}[\mathbf{x}]$  denote the ring of real polynomials in the variables  $\mathbf{x} = (x_1, \dots, x_n)$ , and let  $\mathbb{R}[\mathbf{x}]_d$  be the vector space of real polynomials of degree at most  $d$ . Similarly, let  $\Sigma[\mathbf{x}] \subset \mathbb{R}[\mathbf{x}]$  denote the convex cone of real polynomials that are SOS polynomials, and let  $\Sigma[\mathbf{x}]_d \subset \Sigma[\mathbf{x}]$  denote its subcone of SOS polynomials of degree at most  $d$ . Denote by  $\mathcal{S}^m$  the space of  $m \times m$  real symmetric matrices. For a given matrix  $\mathbf{A} \in \mathcal{S}^m$ , the notation  $\mathbf{A} \succeq 0$  (resp.,  $\mathbf{A} \succ 0$ ) means that  $\mathbf{A}$  is positive semidefinite (PSD) (resp., positive definite); i.e., all of its eigenvalues are real and nonnegative (resp., positive). For every real  $d \geq 1$ , let  $\|\mathbf{x}\|_d := (\sum_{i=1}^n |x_i|^d)^{1/d}$ .

A polynomial  $p \in \mathbb{R}[\mathbf{x}]_d$  is homogeneous if  $p(\lambda\mathbf{x}) = \lambda^d p(\mathbf{x})$  for all  $\mathbf{x} \in \mathbb{R}^n$ ,  $\lambda \in \mathbb{R}$ . A function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is positively homogeneous of degree  $d \in \mathbb{R}$  if  $f(\lambda\mathbf{x}) = \lambda^d f(\mathbf{x})$  for all  $0 \neq \mathbf{x} \in \mathbb{R}^n$ ,  $\lambda > 0$ . For instance,  $\mathbf{x} \mapsto |\mathbf{x}|$  is not homogeneous but is positively homogeneous of degree 1.

Let  $\mathbb{N}_d^n := \{(\alpha_1, \dots, \alpha_n) : \sum_i \alpha_i = d\}$ ,  $s(d) := \binom{n-1+d}{d}$ , and let  $\mathbf{Hom}_d \subset \mathbb{R}[\mathbf{x}]_d$  be the vector space of homogeneous polynomials of degree  $d$ . Also let  $|\alpha| := \sum_i \alpha_i$  for all  $\alpha \in \mathbb{N}^n$ . Every homogeneous polynomial  $g \in \mathbf{Hom}_d$  can be identified with its vector of coefficients  $\mathbf{g} = (g_\alpha) \in \mathbb{R}^{s(d)}$  in the usual canonical basis of monomials; i.e., one writes

$$\mathbf{x} \mapsto g(\mathbf{x}) := \sum_{\alpha \in \mathbb{N}_d^n} g_\alpha \mathbf{x}^\alpha \left( = \sum_{\alpha \in \mathbb{N}_d^n} g_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} \right),$$

and therefore  $\mathbf{Hom}_d$  can be identified with  $\mathbb{R}^{s(d)}$ . Denote by  $\mathbf{G} \subset \mathbb{R}^n$  the sublevel set  $\mathbf{G} := \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$  associated with every  $g \in \mathbf{Hom}_d$ .

**2.2. Some preliminary results.** Some results of this section are already contained in Lasserre [10], but to make the paper as self-contained as possible, they are restated (and sometimes with additional information).

We first characterize the set of homogeneous polynomials of degree  $d$  whose associated level set  $\mathbf{G}$  has finite Lebesgue volume.

**THEOREM 2.1.** *Let  $\mathbf{P}[\mathbf{x}]_d \subset \mathbf{Hom}_d$  be the set of homogeneous polynomials of degree  $d$  whose associated level set  $\mathbf{G}$  has finite Lebesgue volume. Then the following hold:*

- (a)  $\mathbf{P}[\mathbf{x}]_d \neq \emptyset$  if and only if  $d$  is even, and every  $g \in \mathbf{P}[\mathbf{x}]_d$  is nonnegative. Moreover,  $\mathbf{P}[\mathbf{x}]_d$  is a convex cone,<sup>1</sup> and  $0 \notin \mathbf{P}[\mathbf{x}]_d$ .
- (b) For  $g \in \mathbf{P}[\mathbf{x}]_d$  the following assertions are equivalent:
  - (i)  $g$  belongs to the interior of  $\mathbf{P}[\mathbf{x}]_d$ .
  - (ii)  $g$  is strictly positive on  $\mathbb{R}^n \setminus \{0\}$ .
  - (iii) The level set  $\mathbf{G} = \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$  is compact.

*Proof.* (a) Let  $g \in \mathbf{Hom}_d$  with  $d$  odd. Then necessarily there exists  $\mathbf{x}_0$  such that  $g(\mathbf{x}_0) < 0$ , and so  $g(\mathbf{x}) < 0$  for all  $\mathbf{x} \in \mathbf{B}(\mathbf{x}_0, \rho) := \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_0\|_2 < \rho\}$  for some  $\rho > 0$ . Hence  $\mathbf{G}$  contains the cone  $\{\lambda \mathbf{x} : \mathbf{x} \in \mathbf{B}(\mathbf{x}_0, \rho), \lambda \geq 0\}$  generated by  $\mathbf{B}(\mathbf{x}_0, \rho)$ , and so  $\mathbf{G}$  does not have a finite volume. Therefore  $\mathbf{P}[\mathbf{x}]_d = \emptyset$  whenever  $d$  is odd. Next, let  $d$  be even,  $g \in \mathbf{P}[\mathbf{x}]_d$ , and suppose that  $g(\mathbf{x}_0) < 0$  for some  $\mathbf{x}_0$ . Then by the previous argument,  $\text{vol}(\mathbf{G}) = +\infty$ . Therefore  $g \geq 0$  whenever  $g \in \mathbf{P}[\mathbf{x}]_d$ .

Next let  $g_1, g_2 \in \mathbf{P}[\mathbf{x}]_d$  with respective associated level sets  $\mathbf{G}_1$  and  $\mathbf{G}_2$ , and let  $g = \lambda g_1 + (1 - \lambda)g_2$ ,  $\lambda \in (0, 1)$ , with level set  $\mathbf{G} = \{\mathbf{x} : \lambda g_1(\mathbf{x}) + (1 - \lambda)g_2(\mathbf{x}) \leq 1\}$ . Write  $\mathbf{G} = \Theta_1 \cup \Theta_2$ , with

$$\Theta_1 := \mathbf{G} \cap \{\mathbf{x} : g_1(\mathbf{x}) \geq g_2(\mathbf{x})\}; \quad \Theta_2 := \mathbf{G} \cap \{\mathbf{x} : g_1(\mathbf{x}) < g_2(\mathbf{x})\},$$

and observe that  $\Theta_1 \subset \mathbf{G}_2$ , whereas  $\Theta_2 \subset \mathbf{G}_1$ . Therefore  $\text{vol}(\mathbf{G}) \leq \text{vol}(\mathbf{G}_1) + \text{vol}(\mathbf{G}_2) < +\infty$ , which proves that  $g \in \mathbf{P}[\mathbf{x}]_d$ , and so  $\mathbf{P}[\mathbf{x}]_d$  is convex. Finally,  $g = 0 \notin \mathbf{P}[\mathbf{x}]_d$  because  $\mathbf{G} = \{\mathbf{x} : 0 \leq 1\} = \mathbb{R}^n$  (hence  $\text{vol}(\mathbf{G}) = \infty$ ).

(b) (i) $\Rightarrow$ (ii). Suppose that  $g \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$  and  $g(\mathbf{x}_0) = 0$  for some  $\mathbf{x}_0 \in \mathbb{R}^n \setminus \{0\}$ . Consider the polynomial  $g_\epsilon(\mathbf{x}) := g(\mathbf{x}) - \epsilon \|\mathbf{x}\|^d$  with  $\epsilon > 0$  fixed, arbitrary. Then  $g_\epsilon \in \mathbf{Hom}_d$ , and  $g_\epsilon(\mathbf{x}_0) < 0$ . Therefore by (a),  $g_\epsilon \notin \mathbf{P}[\mathbf{x}]_d$  for every  $\epsilon > 0$ , which in turn implies that  $g \notin \text{int}(\mathbf{P}[\mathbf{x}]_d)$ .

(ii) $\Rightarrow$ (iii).  $g(\mathbf{x}) > 0$  on  $\mathbb{R}^n \setminus \{0\}$  implies  $g(\mathbf{x}) > \delta > 0$  on  $\mathbb{S}^{n-1}$  (the unit sphere of  $\mathbb{R}^n$ ) because  $g$  is continuous and  $\mathbb{S}^{n-1}$  is compact. Next if  $\mathbf{x} \in \mathbf{G}$ , then by homogeneity of  $g$ ,  $\mathbf{x}/\|\mathbf{x}\| \in \mathbb{S}^{n-1}$ , and so

$$1 \geq g(\mathbf{x}) \geq \|\mathbf{x}\|^d g(\mathbf{x}/\|\mathbf{x}\|) \geq \|\mathbf{x}\|^d \delta.$$

Therefore  $\|\mathbf{x}\| \leq \delta^{-1/d}$ , which shows that  $\mathbf{G}$  is bounded (hence compact).

(iii) $\Rightarrow$ (i). If  $\mathbf{G}$  is compact, then  $g > 0$  on  $\mathbb{R}^n \setminus \{0\}$ . Indeed, if  $g(\mathbf{x}_0) \leq 0$  for some  $\mathbf{x}_0 \in \mathbb{R}^n \setminus \{0\}$ , then  $g(\lambda \mathbf{x}_0) \leq 0 \leq 1$  for all  $\lambda > 0$ . But then  $\mathbf{G}$  would not be compact as it contains the half-line  $\{\lambda \mathbf{x}_0 : \lambda > 0\}$ . Hence there is some  $\delta > 0$  such that  $g(\mathbf{x}) > \delta \|\mathbf{x}\|^d$  for all  $\mathbf{x}$ . We next construct a constant  $\rho > 0$  such that for any  $f \in \mathbf{Hom}_d$  satisfying  $\|f - g\|_1 < \rho$  (with  $\|f\|_1 = \sum_\alpha |f_\alpha|$ ), its level set  $\mathbf{F}$  is bounded, and thus  $f \in \mathbf{P}[\mathbf{x}]_d$ , implying that  $g \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$ . Let  $C > 0$  be a constant such that  $|\mathbf{x}^\alpha| < C$  for all  $\mathbf{x} \in \mathbb{S}^{n-1}$  and  $|\alpha| = d$ . By homogeneity of  $f$  and  $g$ ,

$$f(\mathbf{x}) - g(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^n} (f_\alpha - g_\alpha) \mathbf{x}^\alpha \leq C \|f - g\|_1 \|\mathbf{x}\|^d.$$

<sup>1</sup>In Rockafellar [17, p. 13] a convex cone does not necessarily contain the origin (whereas some authors add this restriction in the definition of a convex cone).

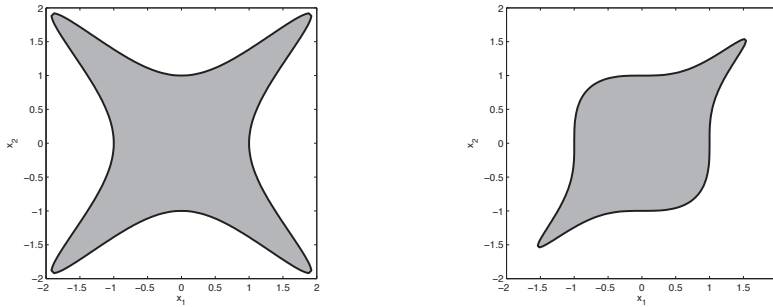


FIG. 1.  $\mathbf{G}$  with  $g(\mathbf{x}) = x_1^4 + x_2^4 - 1.925 x_1^2 x_2^2$  (left) and  $g(\mathbf{x}) = x_1^6 + x_2^6 - 1.925 x_1^3 x_2^3$  (right).

Choose  $\rho > 0$  such that  $\rho < \delta/C$ . Then for every  $\mathbf{x} \in \mathbf{F}$ ,

$$\delta \|\mathbf{x}\|^d \leq g(\mathbf{x}) = g(\mathbf{x}) - f(\mathbf{x}) + f(\mathbf{x}) \leq C \|f - g\|_1 \|\mathbf{x}\|^d + 1 \leq 1 + \rho C \|\mathbf{x}\|^d,$$

implying that  $\|\mathbf{x}\|^d \leq 1/(\delta - \rho C)$ . Therefore  $\mathbf{F}$  is compact, which in turn implies that  $f \in \mathbf{P}[\mathbf{x}]_d$ .  $\square$

It is important to realize that the sublevel set  $\mathbf{G}$  need not be convex. See Figure 1, which displays two examples of nonconvex sets  $\mathbf{G}$ .

*Example 1. A noncompact set  $\mathbf{G}$  with finite volume.* The polynomial  $\mathbf{x} \mapsto g(\mathbf{x}) := x_1^4 x_2^2 + x_1^2 x_2^4$  is in  $\mathbf{P}[\mathbf{x}]_6$  but not in  $\text{int}(\mathbf{P}[\mathbf{x}]_6)$ . (Notice that  $g$  is positive on  $(\mathbb{R} \setminus \{0\})^n$  but not on  $\mathbb{R}^n \setminus \{0\}$ , and its level set  $\mathbf{G}$  is not compact as it contains the four axes.) To prove  $\text{vol}(\mathbf{G}) < \infty$ , by symmetry it suffices to prove that the smaller set,

$$\mathbf{G}_+ := \{\mathbf{x} \geq 0 : x_1^2 x_2^4 + x_1^4 x_2^2 \leq 1\},$$

has finite volume.<sup>2</sup> Now observe that

$$\mathbf{G}_+ \subset \{\mathbf{x} \geq 0 : x_1^2 x_2^4 \leq 1; x_1^4 x_2^2 \leq 1\}$$

and

$$\mathbf{G}_+ \subset \underbrace{\mathbf{G}_+ \cap \{\mathbf{x} \geq 0 : x_1 \geq 1\}}_{A_1} \cup \underbrace{\mathbf{G}_+ \cap \{\mathbf{x} \geq 0 : x_2 \geq 1\}}_{A_2} \cup \underbrace{\mathbf{G}_+ \cap \{\mathbf{x} \geq 0 : \mathbf{x} \leq 1\}}_{A_3}.$$

Clearly  $A_3$  is compact and hence has finite volume. Next

$$A_1 \subset \{\mathbf{x} \geq 0 : x_1^4 x_2^2 \leq 1; x_1 \geq 1\} \subset \{\mathbf{x} \geq 0 : x_1^2 x_2 \leq 1; x_1 \geq 1\}.$$

Hence

$$\text{vol}(A_1) \leq \int_1^\infty \left( \int_0^{1/x_1^2} dx_2 \right) dx_1 = \int_1^\infty \frac{1}{x_1^2} dx_1 = 1,$$

and similarly  $\text{vol}(A_2) \leq 1$ . The shape of the associated level set  $\mathbf{G}$  is displayed in Figure 2.

<sup>2</sup>This simple proof was suggested to us by Prof. Tien Son Pham.



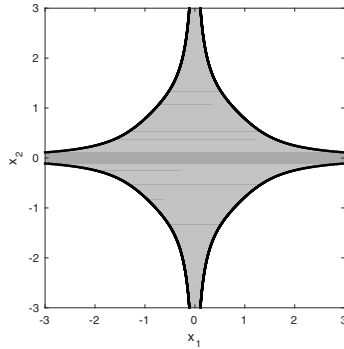


FIG. 2. A noncompact set  $\mathbf{G}$  with  $g(\mathbf{x}) = x_1^4 x_2^2 + x_1^2 x_2^4$ .

We will also need the following result of independent interest, already proved in [10]. We provide a brief sketch. In particular, formula (2.2) below appeared earlier in Morozov and Shakirov [14, 13] with a different proof. We use the same technique based on Laplace transform as in Lasserre [9] and Lasserre and Zeron [12] for providing closed form expressions for a certain class of integrals. Recall that the Gamma function  $\Gamma$  is defined by

$$z \mapsto \Gamma(z) := \int_0^\infty t^{z-1} \exp(-t) dt, \quad z \in \mathbb{C}; \Re(z) > 0,$$

and  $z\Gamma(z) = \Gamma(z + 1)$  for all  $z$  (and  $\Gamma(1 + n) = n!$  for all  $n \in \mathbb{N} \setminus \{0\}$ ). A classical reference for the Laplace transform is Widder [18].

**THEOREM 2.2.** *Let  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  be a nonnegative positively homogeneous function of degree  $0 \neq d \in \mathbb{R}$  such that  $\text{vol}(\{\mathbf{x} : h(\mathbf{x}) \leq 1\}) < \infty$ . Then for every  $\alpha \in \mathbb{N}^n$ ,*

$$(2.1) \quad \int_{\{\mathbf{x}:h(\mathbf{x})\leq 1\}} \mathbf{x}^\alpha d\mathbf{x} = \frac{1}{\Gamma(1 + (n + |\alpha|)/d)} \int_{\mathbb{R}^n} \mathbf{x}^\alpha \exp(-h(\mathbf{x})) d\mathbf{x},$$

whenever either integral is finite (for instance, if  $h \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$ ).

In particular, when  $d$  is an even integer, let  $v : \mathbf{P}[\mathbf{x}]_d \rightarrow \mathbb{R}_+$  be the function

$$(2.2) \quad g \mapsto v(g) := \text{vol}(\mathbf{G}) = \frac{1}{\Gamma(1 + n/d)} \int_{\mathbb{R}^n} \exp(-g(\mathbf{x})) d\mathbf{x}.$$

Then the function  $v$  is nonnegative, strictly convex, and homogeneous of degree  $-n/d$ . Moreover, if  $g \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$ , then  $v$  is differentiable<sup>3</sup> and

$$(2.3) \quad \frac{\partial v(g)}{\partial g_\alpha} = -\frac{n + d}{d} \int_{\mathbf{G}} \mathbf{x}^\alpha d\mathbf{x}, \quad \alpha \in \mathbb{N}_d^n,$$

$$(2.4) \quad \int_{\mathbf{G}} g(\mathbf{x}) d\mathbf{x} = \frac{n}{n + d} \int_{\mathbf{G}} d\mathbf{x}.$$

The proof is postponed to the proofs section, section 7.

Given Theorem 2.2, the function  $g \mapsto v(g) = \text{vol}(\mathbf{G})$  is a strictly convex function which is differentiable on the whole interior  $\text{int}(\mathbf{P}[\mathbf{x}]_d)$  of its domain  $\mathbf{P}[\mathbf{x}]_d$ . For

<sup>3</sup>For the same statement in [10],  $g \in \mathbf{P}[\mathbf{x}]_d$  should be corrected to  $g \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$ , as here.



instance, the function  $v$  is not differentiable at the point  $g \notin \text{int}(\mathbf{P}[\mathbf{x}]_d)$ , where  $\mathbf{x} \mapsto g(\mathbf{x}) = x_1^4 x_2^2 + x_1^2 x_2^4$ .

We also have the following.

LEMMA 2.3. *Let  $v : \mathbf{Hom}_d \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  be the function defined by  $v(g) := \int_{\mathbf{G}} d\mathbf{x}$ , so that*

$$g \mapsto v(g) := \begin{cases} \text{vol}(\mathbf{G}) & \text{if } g \in \mathbf{P}[\mathbf{x}]_d, \\ +\infty & \text{otherwise.} \end{cases}$$

The function  $v$  is strictly convex and lower semicontinuous (l.s.c.) on its domain  $\mathbf{P}[\mathbf{x}]_d$ . Hence for every  $a \in \mathbb{R}_+$ , the set  $\{g \in \mathbf{Hom}_d : v(g) \leq a\}$  is a closed convex set.

*Proof.* Let  $(g_n) \subset \mathbf{P}[\mathbf{x}]_d$  be such that  $\mathbf{g}_n \rightarrow \mathbf{g}$  (i.e.,  $g_{n,\alpha} \rightarrow g_\alpha$  for all  $\alpha \in \mathbb{N}_d^n$ ) as  $n \rightarrow \infty$ . Then the pointwise convergence  $g_n(\mathbf{x}) \rightarrow g(\mathbf{x})$  as  $n \rightarrow \infty$  holds for all  $\mathbf{x} \in \mathbb{R}^n$ . Therefore by Theorem 2.2,

$$\liminf_{n \rightarrow \infty} v(g_n) = \frac{1}{\Gamma(1 + n/d)} \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^n} \exp(-g_n(\mathbf{x})) d\mathbf{x},$$

and by Fatou's lemma<sup>4</sup> [1],

$$\begin{aligned} \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^n} \exp(-g_n(\mathbf{x})) d\mathbf{x} &\geq \int_{\mathbb{R}^n} \left( \liminf_{n \rightarrow \infty} \exp(-g_n(\mathbf{x})) \right) d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \exp(-g(\mathbf{x})) d\mathbf{x}, \end{aligned}$$

which shows that  $v$  is l.s.c. (and by Theorem 2.2,  $v$  is strictly convex). Therefore its sublevel set  $\{g \in \mathbf{Hom}_d : v(g) \leq a\}$  is closed and convex for every  $a \in \mathbb{R}_+$ .  $\square$

**3. The  $\ell_1$ -norm formulation.** With  $d \in \mathbb{N}$  a fixed even integer and  $g \in \mathbf{Hom}_d$  written as

$$\mathbf{x} \mapsto g(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_d^n} g_\alpha \mathbf{x}^\alpha, \quad \mathbf{x} \in \mathbb{R}^n,$$

let  $\mathbf{B}_d \subset \mathbb{R}^n$  be the  $L_d$ -unit ball and  $\rho_d$  its Lebesgue volume; i.e.,

$$(3.1) \quad \mathbf{B}_d = \left\{ \mathbf{x} : \sum_{i=1}^n x_i^d \leq 1 \right\} \quad \text{and} \quad \rho_d := \text{vol}(\mathbf{B}_d) = \int_{\mathbf{B}_d} d\mathbf{x}.$$

Let  $\|g\|_1 := \sum_{\alpha \in \mathbb{N}_d^n} |g_\alpha|$ , and consider the optimization problem  $\mathbf{P}_1$ :

$$(3.2) \quad \mathbf{P}_1 : \quad \inf_g \{ \|g\|_1 : v(g) \leq \rho_d; \quad g \in \mathbf{Hom}_d \}.$$

In fact, one may replace the constraint " $v(g) \leq \rho_d$ " with " $v(g) = \rho_d$ ." That is, among all degree- $d$  homogeneous polynomials  $g$  whose level sets  $\mathbf{G}$  have the same Lebesgue volume as the  $L_d$ -unit ball  $\mathbf{B}_d$ , find one that minimizes the  $\ell_1$ -norm of its coefficients. Indeed, the following holds.

<sup>4</sup>Fatou's lemma states that if  $g_n : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $n \in \mathbb{N}$ , is a sequence of nonnegative measurable functions, then  $\liminf_{n \rightarrow \infty} \int_{\mathbb{R}^n} g_n d\mathbf{x} \geq \int_{\mathbb{R}^n} (\liminf_n g_n(\mathbf{x})) d\mathbf{x}$ .

PROPOSITION 3.1.

$$(3.3) \quad \mathbf{P}_1 : \quad \text{opt}_1 = \inf_g \{ \|g\|_1 : v(g) \leq \rho_d; \quad g \in \mathbf{Hom}_d \} \\ = \inf_g \{ \|g\|_1 : v(g) = \rho_d; \quad g \in \mathbf{Hom}_d \},$$

and  $\mathbf{P}_1$  is a convex optimization problem.

*Proof.* As  $v$  is positively homogeneous of degree  $-n/d$ ,  $v(\lambda g) = \lambda^{-n/d}v(g)$  whenever  $\lambda > 0$ , and so one may replace the constraint  $v(g) = \rho_d$  with the inequality constraint  $v(g) \leq \rho_d$ . Indeed, if  $v(g) < \rho_d$ , then  $v(\lambda g) = \rho_d$  if  $\lambda^{n/d} = v(g)/\rho_d$  (so that  $\lambda < 1$ ), and therefore  $\|\lambda g\|_1 < \|g\|_1$ , which shows that a better solution  $g' = \lambda g$  is obtained with  $v(g') = \rho_d$ . Therefore  $\mathbf{P}_1$  reduces to minimizing a convex function (the  $\ell_1$ -norm) on the closed convex set  $\{g \in \mathbf{Hom}_d : v(g) \leq \rho_d\}$  (see Lemma 2.3), a convex optimization problem.  $\square$

THEOREM 3.2. *Let  $d \geq 2$  be an even integer. Then problem  $\mathbf{P}_1$  in (3.3) has a unique optimal solution  $g_1^* \in \mathbf{P}[\mathbf{x}]_d$ , which is the  $L_d$ -norm polynomial*

$$\mathbf{x} \mapsto g^{(d)}(\mathbf{x}) = \sum_{i=1}^n x_i^d \quad (= \|\mathbf{x}\|_d^d),$$

and moreover,

$$(3.4) \quad \text{vol}(\mathbf{B}_d) = \int_{\mathbf{B}_d} d\mathbf{x} = \frac{2^n \Gamma(1/d)^n}{n d^{n-1} \Gamma(n/d)}; \quad \int_{\mathbf{B}_d} x_i^d d\mathbf{x} = \frac{\text{vol}(\mathbf{B}_d)}{n+d}, \quad i = 1, \dots, n.$$

The proof is postponed to section 7.

**An alternative formulation.** We may also consider the alternative but equivalent formulation

$$(3.5) \quad \mathbf{P}'_1 : \quad \rho' = \inf_g \{ v(g) : \|g\|_1 \leq n; \quad g \in \mathbf{Hom}_d \}.$$

PROPOSITION 3.3. *Let  $\mathbf{P}_1$  and  $\mathbf{P}'_1$  be as in (3.3) and (3.5), respectively. Then  $\mathbf{P}'_1$  and  $\mathbf{P}_1$  have the same unique optimal solution  $\mathbf{x} \mapsto g^{(d)}(\mathbf{x}) = \sum_{i=1}^n x_i^d$ .*

*Proof.* Again, as  $v(\lambda g) = \lambda^{-n/d}v(g)$ , we may consider only those  $h \in \mathbf{Hom}_d$  with  $\|h\|_1 = n$ . Suppose that  $h$  is an optimal solution of  $\mathbf{P}'_1$  with  $v(h) = \rho' < \rho_d$  and  $\|h\|_1 = n$ . Then take  $\tilde{g} = kh$  with  $k^{-n/(d)}\rho' = \rho_d$  so that  $k < 1$ . Then  $v(\tilde{g}) = \rho_d$ , and  $\|\tilde{g}\|_1 = k\|h\|_1 = kn < n$ . But this implies that  $\tilde{g}$  would be a better solution for  $\mathbf{P}_1$  than  $g^{(d)}$ , a contradiction. Therefore  $\rho' \geq \rho_d$ , and in fact  $\rho' = \rho_d$ , as  $g^{(d)}$  is feasible for  $\mathbf{P}'_1$  with  $v(g^{(d)}) = \rho_d$ . Next, observe that  $\mathbf{P}'_1$  is a convex optimization problem with a strictly convex objective function; hence an optimal solution is unique.  $\square$

Therefore problem  $\mathbf{P}_1$  has the equivalent formulation: *Among all homogeneous polynomials  $g \in \mathbf{Hom}_d$  with  $\|g\|_1 = 1$ , find the one with minimum associated volume.* By Theorem 2.2 the  $L_d$ -unit ball has minimum volume.

**4. The  $\ell_2$ -norm formulation.** For every  $\alpha \in \mathbb{N}^n$  introduce the multinomial coefficient  $c_\alpha := \frac{(\sum_i \alpha_i)!}{\alpha_1! \dots \alpha_n!}$ . Recall that  $\mathbb{N}_d^n := \{\alpha \in \mathbb{N}^n : |\alpha| = d\}$  and  $s_d = \binom{n-1+d}{d}$ . We now write

$$\mathbf{x} \mapsto p(\mathbf{x}) := \sum_{\alpha \in \mathbb{N}_d^n} c_\alpha p_\alpha \mathbf{x}^\alpha, \quad p \in \mathbf{Hom}_d,$$

for some vector  $\mathbf{p} = (p_\alpha) \in \mathbb{R}^{s_d}$  (called Bernstein coefficients of  $p$ ), and equip  $\mathbf{Hom}_d$  with the scalar product

$$\langle p, q \rangle_d := \sum_{\alpha \in \mathbb{N}_d^n} c_\alpha p_\alpha q_\alpha, \quad p, q \in \mathbf{Hom}_d,$$

with associated norm  $\|p\|_{2,d}^2 = \langle p, p \rangle_d$ . Next, denote by  $\mathcal{P}_d \subset \mathbf{Hom}_d$  the convex cone of homogeneous polynomials of degree  $d$  which are nonnegative, and let  $\mathcal{C}_d \subset \mathbf{Hom}_d$  be the convex cone of (positive) sums of  $d$ -powers of linear forms. Then  $\mathcal{C}_d$  is the dual cone of  $\mathcal{P}_d$ ; i.e.,  $\mathcal{P}_d^* = \mathcal{C}_d$ ; see, e.g., Reznick [16].

As in section 3, let  $\rho_d = \text{vol}(\mathbf{B}_d)$ , and consider the following optimization problem:

$$(4.1) \quad \mathbf{P}_2 : \quad \text{opt}_2 = \inf_g \{ \|g\|_{2,d}^2 : v(g) \leq \rho_d; \quad g \in \mathbf{Hom}_d \},$$

a (weighted)  $\ell_2$ -norm analogue of  $\mathbf{P}_1$  in (3.2). In view of Lemma 2.3, the feasible set is a closed convex set, and so problem  $\mathbf{P}_2$  is a convex optimization problem.

**THEOREM 4.1.** *Problem  $\mathbf{P}_2$  in (4.1) has a unique optimal solution  $g_2^* \in \mathbf{P}[\mathbf{x}]_d$ . If  $g_2^* \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$ , then its vector of coefficients  $\mathbf{g}_2^* = (g_{2,\alpha}^*) \in \mathbb{R}^{s(d)}$  satisfies*

$$(4.2) \quad g_{2,\alpha}^* = \text{opt}_2 \frac{n+d}{n} \cdot \frac{\int_{\mathbf{G}^*} \mathbf{x}^\alpha d\mathbf{x}}{\int_{\mathbf{G}^*} d\mathbf{x}} \quad \forall \alpha \in \mathbb{N}_d^n,$$

where  $\mathbf{G}^* = \{\mathbf{x} : g_2^*(\mathbf{x}) \leq 1\}$  and  $\text{vol}(\mathbf{G}^*) = \rho_d$ . Therefore,  $\mathbf{g}_2^* = (g_{2,\alpha}^*)$ ,  $\alpha \in \mathbb{N}_d^n$ , is an element of  $\mathcal{C}_d = \mathcal{P}_d^*$ . More precisely,

$$(4.3) \quad \mathbf{x} \mapsto g_2^*(\mathbf{x}) = \text{opt}_2 \frac{n+d}{n} \cdot \frac{\int_{\mathbf{G}^*} (\mathbf{z} \cdot \mathbf{x})^d d\mathbf{z}}{\int_{\mathbf{G}^*} d\mathbf{z}},$$

and in fact there exist  $\mathbf{z}_i \in \mathbb{R}^n$ ,  $\theta_i > 0$ ,  $i = 1, \dots, s$ , with  $s \leq \binom{n-1+d}{d} + 1$ , such that

$$(4.4) \quad \mathbf{x} \mapsto g_2^*(\mathbf{x}) = \text{opt}_2 \frac{n+d}{n} \sum_{i=1}^s \theta_i (\mathbf{z}^i \cdot \mathbf{x})^d.$$

Conversely, if  $g \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$  satisfies (4.2) (with  $\text{opt}_2 := \|g\|_{2,d}^2$ ), then  $g$  is the unique optimal solution of  $\mathbf{P}_2$ .

*Proof.* That  $\mathbf{P}_2$  has a unique optimal solution  $g_2^*$  follows exactly from the same arguments as for  $\mathbf{P}_1$ . As Slater's condition also holds for  $\mathbf{P}_2$ , if  $v$  is differentiable at  $g_2^*$ , then by the Karush–Kuhn–Tucker (KKT)-optimality conditions, there exists  $\lambda^* \geq 0$  such that

$$2g_{2,\alpha}^* c_\alpha = -\lambda^* \frac{\partial v(g_2^*)}{\partial g_\alpha} = \lambda^* \frac{n+d}{d} \int_{\mathbf{G}^*} c_\alpha \mathbf{x}^\alpha d\mathbf{x} \quad \forall \alpha \in \mathbb{N}_d^n,$$

where we have used Theorem 2.2. Therefore, multiplying each side with  $g_{2,\alpha}^*$ , summing up, and using that  $v$  is positively homogeneous yields

$$2\|g_2^*\|_{2,d}^2 = 2\text{opt}_2 = -\lambda^* \langle \nabla v(g_2^*), g_2^* \rangle = \lambda^* \frac{n}{d} v(g_2^*) = \lambda^* \frac{n}{d} \rho_d.$$

Hence  $\lambda^* = 2\text{opt}_2 d / (n\rho_d)$ , and  $g_{2,\alpha}^* = \text{opt}_2 \frac{n+d}{n\rho_d} \int_{\mathbf{G}^*} \mathbf{x}^\alpha d\mathbf{x}$ , which is (4.2).

But then (4.2) means that  $(g_{2,\alpha}^*) \in \mathcal{P}_d^* = \mathcal{C}_d$ , which yields (4.3). Finally, (4.4) follows from Caratheodory's theorem.<sup>5</sup>

Finally, for the converse statement, as Slater's condition holds and  $v$  is differentiable at  $g \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$ , the KKT conditions (4.2) are also sufficient conditions of optimality.  $\square$

So both optimal solutions  $g_1^*$  of  $\mathbf{P}_1$  and  $g_2^*$  of  $\mathbf{P}_2$  are sums of  $d$ -powers of linear forms. But  $g_2^*$  does not have the parsimony property, as shown below.

**COROLLARY 4.2.** *If  $d \geq 4$ , then the polynomial  $\mathbf{x} \mapsto g^{(d)}(\mathbf{x}) = \sum_{i=1}^n x_i^d$  (optimal solution of  $\mathbf{P}_1$ ) cannot be an optimal solution of  $\mathbf{P}_2$ .*

*Proof.* Suppose that the unique optimal solution  $g_1^* = g^{(d)}$  of  $\mathbf{P}_1$  is also the unique optimal solution of  $\mathbf{P}_2$ . As  $v$  is differentiable at  $g = g_1^*$ , then by the characterization (4.2) of the unique optimal solution of  $\mathbf{P}_2$ , every coefficient  $g_{1,2\beta}^*$  with  $2|\beta| = d$  must be strictly positive, a contradiction.  $\square$

Corollary 4.2 states that the optimal solution of  $\mathbf{P}_2$  does not have a parsimony property as it has at least  $\binom{n+d/2-1}{d/2}$  nonzero coefficients!

The only case where the optimal solution of  $\mathbf{P}_1$  also solves  $\mathbf{P}_2$  is the quadratic case  $d = 2$ . Indeed, straightforward computation shows that (4.2) is satisfied by the polynomial  $g_1^* = g^{(d)}$  of Theorem 3.2.

*Example 2.* Let  $n = 2$  and  $d = 4$ . By uniqueness of the minimizer we may guess that the optimal solution  $g_2^* \in \mathbf{Hom}_d$  is symmetric and of the form

$$\mathbf{x} \mapsto g_2^*(\mathbf{x}) = g_{40}^* (x_1^4 + x_2^4) + 6 g_{22}^* x_1^2 x_2^2,$$

with  $\text{opt}_2 = 2(g_{40}^*)^2 + (g_{22}^*)^2$  and

$$g_{40}^* = 3 \text{opt}_2 \frac{\int_{\mathbf{G}^*} x_1^4 d\mathbf{x}}{\int_{\mathbf{G}^*} d\mathbf{x}}; \quad g_{22}^* = 3 \text{opt}_2 \frac{\int_{\mathbf{G}^*} x_1^2 x_2^2 d\mathbf{x}}{\int_{\mathbf{G}^*} d\mathbf{x}},$$

with  $\mathbf{G}^* = \{\mathbf{x} : g_2^*(\mathbf{x}) \leq 1\}$ . Observe that by homogeneity, an optimal solution  $g_2^*$  of  $\mathbf{P}_2$  is also optimal when we replace  $\rho_d$  with any constant  $a > 0$ ; only the optimal value changes, and the characterization (4.2) remains the same with the new optimal value  $\text{opt}_2$ . After several numerical trials we conjecture that

$$\mathbf{x} \mapsto g_2^*(\mathbf{x}) \approx x_1^4 + x_2^4 + 2 x_1^2 x_2^2 = (x_1^2 + x_2^2)^2,$$

i.e.,  $\mathbf{g}_2^* = (1, 0, 1/3, 0, 1)$ , is an optimal solution. But then observe that

$$\mathbf{G}^* = \{\mathbf{x} : g_2^*(\mathbf{x}) \leq 1\} = \{\mathbf{x} : (x_1^2 + x_2^2)^2 \leq 1\} = \mathbf{B}_2!$$

That is,  $g_2^*$  is another representation of the unit sphere  $\mathbf{B}_2$  by a homogeneous polynomial of degree 4 instead of quadratics:

$$\int_{\mathbf{G}^*} d\mathbf{x} \approx 3.1415926; \quad \int_{\mathbf{G}^*} x_1^4 d\mathbf{x} \approx 0.392699; \quad \int_{\mathbf{G}^*} x_1^2 x_2^2 d\mathbf{x} \approx 0.130899.$$

With  $a := \int_{\mathbf{G}^*} d\mathbf{x}$ , (4.2) yields (up to  $10^{-8}$ )

$$3 \text{opt}_2 \cdot \frac{\int_{\mathbf{G}^*} x_1^4 d\mathbf{x}}{\int_{\mathbf{G}^*} d\mathbf{x}} = 1 = g_{40}^*; \quad 3 \text{opt}_2 \cdot \frac{\int_{\mathbf{G}^*} x_1^2 x_2^2 d\mathbf{x}}{\int_{\mathbf{G}^*} d\mathbf{x}} = \frac{1}{3} = g_{22}^*.$$

<sup>5</sup>Every point in the convex hull of a set  $S \subset \mathbb{R}^n$  is a convex combination of at most  $n + 1$  points of  $S$ ; see, e.g., [17, p. 153].

Observe also that

$$(x_1^2 + x_2^2)^2 = \frac{1}{6}(x_1 + x_2)^4 + \frac{1}{6}(x_1 - x_2)^4 + \frac{2}{3}(x_1^4 + x_2^4),$$

i.e., a sum of 4-powers of linear forms as predicted by Theorem 4.1.

In fact, we have the following theorem.

**THEOREM 4.3.** *If  $d = 2, 4, 6, 8$ , then the unique optimal solution of  $\mathbf{P}_2$  is  $g_2^*(\mathbf{x}) = \theta (\sum_{i=1}^n x_i^2)^{d/2}$  (for some scaling factor  $\theta > 0$ ) whose level set  $\mathbf{G}$  is homothetic to the unit ball  $\mathbf{B}_2$ .*

The proof is postponed to section 7.

In other words, when  $d = 2, 4, 6$ , and  $8$ , surprisingly, the Euclidean unit ball  $\mathbf{B}_2 = \{\mathbf{x} : \sum_i x_i^2 \leq 1\}$  (which has the equivalent representation  $\{\mathbf{x} : (\sum_{i=1}^n x_i^2)^{d/2} \leq 1\}$ ) solves problem  $\mathbf{P}_2$ ! In other words, its associated homogeneous polynomial,  $\mathbf{x} \mapsto g_2^*(\mathbf{x}) := \|\mathbf{x}\|_2^d = \sum_{\alpha} c_{\alpha} g_{\alpha} \mathbf{x}^{\alpha}$ , minimizes the (weighted) Euclidean norm  $\|g\|_{2,d}^2 = \sum_{\alpha} c_{\alpha} g_{\alpha}^2$  among all homogeneous polynomials  $g$  of degree  $d$  whose unit ball  $\mathbf{G}$  has the same volume.

**5. The SOS formulation.** As we have seen, optimal solutions of both the  $\ell_1$ -norm and  $\ell_2$ -norm formulations are (positive) sums of  $d$ -powers of linear forms, hence sums of squares of  $d/2$ -forms since  $d$  is an even integer. Therefore we now restrict the discussion to homogeneous polynomials in  $\mathbf{Hom}_d$  that are SOS, i.e., polynomials of the form

$$\mathbf{x} \mapsto g_{\mathbf{Q}}(\mathbf{x}) = \mathbf{v}_{d/2}(\mathbf{x})^T \mathbf{Q} \mathbf{v}_{d/2}(\mathbf{x}),$$

where  $\mathbf{v}_{d/2}(\mathbf{x}) = (\mathbf{x}^{\alpha})$ ,  $\alpha \in \mathbb{N}_{d/2}^n$ , and  $\mathbf{Q}$  is some real PSD symmetric matrix ( $\mathbf{Q} \succeq 0$ ) of size  $s(d/2) = \binom{n-1+d/2}{d/2}$ . If we denote by  $\mathcal{S}_d$  the space of real symmetric matrices of size  $s(d/2)$ , there is not a one-to-one correspondence between  $g \in \mathbf{Hom}_d$  and  $\mathbf{Q} \in \mathcal{S}_d$ , as several  $\mathbf{Q}$  may produce the same polynomial  $g_{\mathbf{Q}}$ .

Given  $0 \preceq \mathbf{Q} \in \mathcal{S}_d$ , denote by  $\mathbf{G}_{\mathbf{Q}}$  the sublevel set  $\{\mathbf{x} : g_{\mathbf{Q}}(\mathbf{x}) \leq 1\}$  associated with  $g_{\mathbf{Q}} \in \mathbf{Hom}_d$ . Denote by  $\hat{v} : \mathcal{S}_d \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  the function

$$\mathbf{Q} \mapsto \hat{v}(\mathbf{Q}) = v(g_{\mathbf{Q}}) = \text{vol}(\mathbf{G}_{\mathbf{Q}}).$$

Observe that  $\hat{v}$  inherits properties of  $v$ , and in particular,  $\hat{v}$  is positively homogeneous of degree  $-n/d$ , convex, and l.s.c. (but not strictly convex). Moreover, when  $g_{\mathbf{Q}} \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$ , then  $\hat{v}$  is differentiable at  $\mathbf{Q}$ , and its gradient reads as

$$\begin{aligned} \nabla \hat{v}(\mathbf{Q}) &= \frac{1}{\Gamma(1+n/d)} \nabla \int_{\mathbb{R}^n} \exp(-\langle \mathbf{v}_{d/2}(\mathbf{x}), \mathbf{v}_{d/2}(\mathbf{x})^T, \mathbf{Q} \rangle) d\mathbf{x} \\ &= \frac{1}{\Gamma(1+n/d)} \int_{\mathbb{R}^n} -\mathbf{v}_{d/2}(\mathbf{x}) \mathbf{v}_{d/2}(\mathbf{x})^T \exp(-\langle \mathbf{v}_{d/2}(\mathbf{x}), \mathbf{v}_{d/2}(\mathbf{x})^T, \mathbf{Q} \rangle) d\mathbf{x} \\ &= \frac{-\Gamma(1+(n+d)/d)}{\Gamma(1+n/d)} \int_{\mathbf{G}_{\mathbf{Q}}} \mathbf{v}_d(\mathbf{x}) \mathbf{v}_{d/2}(\mathbf{x})^T d\mathbf{x} \\ &= \frac{-(n+d)}{d} \int_{\mathbf{G}_{\mathbf{Q}}} \mathbf{v}_{d/2}(\mathbf{x}) \mathbf{v}_{d/2}(\mathbf{x})^T d\mathbf{x}, \end{aligned}$$

where we have used Theorem 2.2 and the identity  $z\Gamma(z) = \Gamma(z+1)$ .

Thus the natural analogue for  $\mathbf{Q}$  of the  $\ell_1$ -norm  $\|g\|_1$  for  $g \in \mathbf{Hom}_d$  is now the nuclear norm of  $\mathbf{Q}$ , which (as  $\mathbf{Q} \succeq 0$ ) reduces to  $\langle \mathbf{I}, \mathbf{Q} \rangle = \text{trace}(\mathbf{Q})$ . It is well

known that optimizing the nuclear norm on convex problems with matrices induces a parsimony effect; namely, an optimal solution is expected to have a small rank. In our context,  $\mathbf{Q}$  having a small rank means that  $g_{\mathbf{Q}}$  can be written as a sum of a small number of squares. However, when expanded in the monomial basis,  $g_{\mathbf{Q}}$  may have many nonzero coefficients, and so its  $\ell_1$ -norm  $\|g_{\mathbf{Q}}\|_1$  may not be small.

So in the same spirit as for the  $\ell_1$ - and the  $\ell_2$ -norm, we now consider the optimization problem

$$(5.1) \quad \mathbf{P}_3 : \quad \text{opt}_3 = \inf_{\mathbf{Q} \in \mathcal{S}_d} \{ \langle \mathbf{I}, \mathbf{Q} \rangle : \hat{v}(\mathbf{Q}) \leq \rho_d; \quad \mathbf{Q} \succeq 0 \}.$$

We show that there exists a unique homogeneous polynomial  $g_3^* \in \mathbf{Hom}_d$  such that  $g_3^* = g_{\mathbf{Q}^*}$  for every optimal solution  $\mathbf{Q}^*$  of  $\mathbf{P}_3$ .

**THEOREM 5.1.** *Problem  $\mathbf{P}_3$  has an optimal solution, and there exists a unique  $g_3^* \in \mathbf{Hom}_d$  such that  $g_3^* = g_{\mathbf{Q}^*}$  for every optimal solution  $\mathbf{Q}^* \in \mathcal{S}_d$  of  $\mathbf{P}_3$ . Next, we have the following:*

(a) *Assume that  $g_3^* \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$ . Then for every optimal solution  $\mathbf{Q}^* \in \mathcal{S}_d$  of  $\mathbf{P}_3$ ,*

$$(5.2) \quad \Psi := \mathbf{I} - \frac{(n+d) \langle \mathbf{I}, \mathbf{Q}^* \rangle}{n \rho_d} \int_{\mathbf{G}_{\mathbf{Q}^*}} \mathbf{v}_{d/2}(\mathbf{x}) \mathbf{v}_{d/2}(\mathbf{x})^T d\mathbf{x} \succeq 0,$$

where  $\mathbf{x} \mapsto g_{\mathbf{Q}^*}(\mathbf{x}) := \mathbf{v}_{d/2}(\mathbf{x})^T \mathbf{Q}^* \mathbf{v}_{d/2}(\mathbf{x}) = g_3^*(\mathbf{x})$  and  $\langle \mathbf{Q}^*, \Psi \rangle = 0$ .

(b) *Conversely, suppose that  $0 \preceq \mathbf{Q}^* \in \mathcal{S}_d$  satisfies (5.2),  $g_{\mathbf{Q}^*} \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$ , and  $\hat{v}(\mathbf{Q}^*) = \rho_d$ . Then  $\mathbf{Q}^*$  is an optimal solution of  $\mathbf{P}_3$  and  $g_3^* = g_{\mathbf{Q}^*}$ .*

Finally, if  $d > 2$ , the polynomial  $\mathbf{x} \mapsto g^{(d)}(\mathbf{x}) = \sum_{i=1}^n x_i^d$  cannot be an optimal solution of  $\mathbf{P}_3$ . For  $d = 2, 4$  (as well as for all  $d \in 4\mathbb{N}$ , provided that  $n$  is sufficiently large), the optimal solution  $g_3^*$  is the polynomial  $\mathbf{x} \mapsto \theta (\sum_{i=1}^n x_i^2)^{d/2}$  with  $\theta = (\rho_2/\rho_d)^{d/n}$ ; i.e.,  $\mathbf{G}^*$  is homothetic to the Euclidean ball  $\mathbf{B}_2$ .

The proof is postponed to section 7.

The fact that the  $L_d$ -unit ball is not an optimal solution of  $\mathbf{P}_3$  is not a surprise, as the parsimony-inducing norm  $\text{trace}(\mathbf{Q})$  (when  $\mathbf{Q} \succeq 0$ ) aims at finding a polynomial  $g_{\mathbf{Q}} \in \mathbf{Hom}_d$  which can be written as an SOS with as few terms as possible in the sum. On the other hand, the sparsity-inducing norm  $\|g\|_1$  aims at finding a polynomial  $g \in \mathbf{Hom}_d$  with as few monomials as possible when  $g$  is expanded in the monomial basis. These are two conflicting criteria, and indeed the single square  $g_3^*(\mathbf{x}) = (\sum_i x_i^2)^{d/2}$  which minimizes the nuclear norm in problem  $\mathbf{P}_3$  contains a lot of monomials!

Again the Euclidean ball  $\mathbf{B}_2$  exhibits a rather strong property: Namely, its representing quadratic polynomial  $g(\mathbf{x}) := \|\mathbf{x}\|^2 (= g^{(2)}(\mathbf{x}))$  is such that that  $g^{d/2} \in \mathbf{Hom}_d$  is the unique optimal solution of problem  $\mathbf{P}_2$  for the  $\ell_2$ -norm, and of problem  $\mathbf{P}_3$  for the parsimony-inducing nuclear norm of its associated Gram matrix  $\mathbf{Q}$  (when  $d \in 4\mathbb{N}$  and  $n$  is sufficiently large)!

**6. Extension to generalized polynomials.** In this section  $d$  is now a (positive) rational with  $L_d$ -unit ball  $\{\mathbf{x} : \sum_{i=1}^n |x_i|^d \leq 1\}$ . Even though the function  $\mathbf{x} \mapsto \sum_{i=1}^n |x_i|^d$  is now a “generalized polynomial” and not a polynomial, it is still a nonnegative positively homogeneous function of degree  $d$  for which Theorem 2.2 applies. On the other hand, the vector space of positively homogeneous functions of degree  $d$  is not finite-dimensional, and so for optimization purposes we need to define an appropriate finite-dimensional analogue of  $\mathbf{Hom}_d$ .

We will use the notation  $|\mathbf{x}| \in \mathbb{R}_+^n$  for the vector  $(|x_1|, \dots, |x_n|)$ , and  $|\mathbf{x}|^\alpha$  for the generalized monomial  $|x_1|^{\alpha_1} \cdots |x_n|^{\alpha_n}$ , whenever  $\alpha \in \mathbb{Q}_+^n$ .

DEFINITION 6.1. Let  $0 < d \in \mathbb{Q}$ . Define the space  $\mathcal{C}_d$  as

$$(6.1) \quad \mathcal{C}_d := \left\{ \sum_{\alpha \in \mathbb{Q}_+^n} g_\alpha |\mathbf{x}|^\alpha : g_\alpha \in \mathbb{R}; \quad |\alpha| \left( := \sum_{i=1}^n \alpha_i \right) = d \right\},$$

where only finitely many coefficients  $g_\alpha$  are nonzero. Then

$$\|g\|_1 = \sum_{\alpha} |g_\alpha|; \quad \|g\|_2^2 = \sum_{\alpha} c_\alpha g_\alpha^2.$$

The space  $\mathcal{C}_d$  is a real infinite-dimensional vector space, and each element of  $\mathcal{C}_d$  is a positively homogeneous function of degree  $d$ .

DEFINITION 6.2. With  $q \in \mathbb{N}$ , let  $\mathbb{Z}_q^n \subset \mathbb{R}^n$  be the lattice  $\{\mathbf{z} \in \mathbb{R}^n : q\mathbf{z} \in \mathbb{Z}^n\}$ . With  $0 < d \in \mathbb{Z}_q$ , denote by  $\mathcal{N}_{dq}^n$  the finite set  $\{\alpha \in \mathbb{Z}_q^n : \alpha \geq 0; \sum_{i=1}^n \alpha_i = d\}$  of cardinality  $m(d, q)$ , and denote by  $\mathbf{Hom}_d^q \subset \mathcal{C}_d$  the vector space of functions  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  defined by

$$(6.2) \quad \mathbf{Hom}_d^q := \left\{ \sum_{\alpha \in \mathcal{N}_{dq}^n} g_\alpha |\mathbf{x}|^\alpha : (g_\alpha) \in \mathbb{R}^{m(d, q)} \right\},$$

which is a finite-dimensional vector space.

For instance, with  $n = 2$  and  $d = 1$ , one has  $\mathbf{B}_1 := \{\mathbf{x} \in \mathbb{R}^2 : \sum_{i=1}^2 |x_i| \leq 1\}$ , and with  $0 < q \in \mathbb{N}$ ,

$$\mathcal{N}_{1q}^2 = \left\{ \left( \frac{k}{q}, \frac{q-k}{q} \right) : k = 0, \dots, q \right\}; \quad m(d, q) = q + 1,$$

and  $g \in \mathbf{Hom}_d^q$  can be written as

$$\mathbf{x} \mapsto g(\mathbf{x}) = \sum_{k=0}^q g_k |x_1|^{\frac{k}{q}} |x_2|^{\frac{q-k}{q}}, \quad \mathbf{x} \in \mathbb{R}^2,$$

for some vector  $\mathbf{g} = (g_k) \in \mathbb{R}^{q+1}$ .

Obviously  $\mathbf{Hom}_d^q$  is a vector space of dimension  $m(d, q)$ , and every function in  $\mathbf{Hom}_d^q$  is positively homogeneous of degree  $d$ . As we did in section 3, with  $g \in \mathbf{Hom}_d^q$  associate the level set  $\mathbf{G} := \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$ , and whenever  $\mathbf{G}$  has finite volume, let  $v(g) := \text{vol}(\mathbf{G})$ ,  $g \in \mathbf{Hom}_d^q$ . It follows that  $v$  is positively homogeneous of degree  $-n/d$ . Therefore (2.2)–(2.4) in Theorem 2.2 hold for  $v$ . Finally, again let

$$\mathbf{x} \mapsto g^{(d)}(\mathbf{x}) := \sum_{i=1}^n |x_i|^d; \quad \mathbf{B}_d = \{\mathbf{x} : g^{(d)}(\mathbf{x}) \leq 1\}; \quad \rho_d = \text{vol}(\mathbf{B}_d),$$

and consider the finite-dimensional optimization problem

$$(6.3) \quad \mathbf{P}_{1q} : \quad \text{opt}_{1q} = \inf_g \{ \|g\|_1 : v(g) \leq \rho_d; \quad g \in \mathbf{Hom}_d^q \}.$$

When  $d < 1$ , the unit ball  $\mathbf{B}_d$  is not convex and is not associated with a norm, as can be seen in Figure 3, where  $d = 1/2$ .



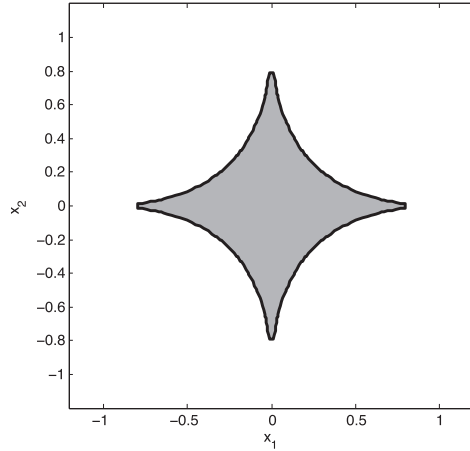


FIG. 3. The  $L_{1/2}$ -unit ball  $\{\mathbf{x} : \sqrt{|x_1|} + \sqrt{|x_2|} \leq 1\}$ .

However, we have the following analogue of Theorem 3.2.

THEOREM 6.3. Let  $0 < q \in \mathbb{N}$  and  $0 < d \in \mathbb{Z}_{q/2}$ . The generalized polynomial

$$(6.4) \quad \mathbf{x} \mapsto g_1^{(d)}(\mathbf{x}) = \sum_{i=1}^n |x_i|^d$$

is the unique optimal solution of Problem  $\mathbf{P}_{1q}$  in (6.3), and moreover,

$$(6.5) \quad \text{vol}(\mathbf{B}_d) = \frac{2^n \Gamma(1/d)^n}{n d^{n-1} \Gamma(n/d)}; \quad \int_{\mathbf{G}^*} |x_i|^d d\mathbf{x} = \frac{\text{vol}(\mathbf{B}_d)}{n+d}, \quad i = 1, \dots, n.$$

*Proof.* The proof is almost a verbatim copy of that of Theorem 3.2 except that we now have to deal with generalized moments  $\int_{\mathbf{G}} |\mathbf{x}|^\alpha d\mathbf{x}$ ,  $\alpha \in \mathcal{N}_{dq}^n$ , instead of standard monomial moments  $\int_{\mathbf{G}} \mathbf{x}^\alpha d\mathbf{x}$ ,  $\alpha \in \mathbb{N}^n$ . But the crucial fact that we exploit is that  $v$  is strictly convex and Theorem 2.2 holds for  $v$ . As in the proof of Theorem 3.2, to show that  $g^{(d)}$  in (6.4) satisfies the KKT-optimality conditions, we need only prove that

$$\int_{\mathbf{B}_d} |\mathbf{x}|^\alpha d\mathbf{x} \leq \int_{\mathbf{B}_d} |x_1|^d d\mathbf{x} \quad \forall \alpha \in \mathcal{N}_{dq}^n.$$

Define the Hankel-type moment matrix  $\mathbf{M}$  to be the real symmetric matrix with rows and columns indexed by  $\alpha \in \mathcal{N}_{\frac{d}{2}q}^n$  and with entries

$$\mathbf{M}(\alpha, \beta) := \int_{\mathbf{B}_d} |\mathbf{x}|^{(\alpha+\beta)} d\mathbf{x}, \quad \alpha, \beta \in \mathcal{N}_{\frac{d}{2}q}^n.$$

Equivalently, letting  $\mathcal{N}_{\frac{d}{2}q}^n = \{\beta \in \mathbb{N}^n : \sum_i \beta_i = dq/2\}$  and reindexing rows and columns of  $\mathbf{M}$  with  $\tilde{\alpha} := q\alpha \in \mathcal{N}_{\frac{d}{2}q}^n$ ,

$$\mathbf{M}(\tilde{\alpha}, \tilde{\beta}) := \int_{\mathbf{B}_d} (|\mathbf{x}|^{1/q})^{\tilde{\alpha}+\tilde{\beta}} d\mathbf{x} =: y_{\tilde{\alpha}+\tilde{\beta}}, \quad \tilde{\alpha}, \tilde{\beta} \in \mathcal{N}_{\frac{d}{2}q}^n.$$

Define  $\mathbf{y} = (y_{\tilde{\alpha}})$ ,  $\tilde{\alpha} \in \mathbb{N}_{dq}^n$ , and  $\tilde{X} = |\mathbf{x}|^{1/q}$ . Observe that from (6.2) one may write

$$\mathbf{Hom}_d^q := \left\{ \sum_{\tilde{\alpha} \in \mathbb{N}_{dq}^n} g_{\alpha} (|\mathbf{x}|^{\frac{1}{q}})^{\tilde{\alpha}} : (g_{\alpha}) \in \mathbb{R}^{m(d,q)} \right\}.$$

Let  $L_{\mathbf{y}} : \mathbf{Hom}_d^q \rightarrow \mathbb{R}_+$  be the linear mapping defined by

$$g \mapsto L_{\mathbf{y}}(g) := \sum_{\tilde{\alpha} \in \mathbb{N}_{dq}^n} g_{\alpha} y_{\tilde{\alpha}} = \sum_{\tilde{\alpha} \in \mathbb{N}_{dq}^n} g_{\alpha} \int_{\mathbf{B}_d} \tilde{X}^{\tilde{\alpha}} d\mathbf{x} = \int_{\mathbf{B}_d} g(\mathbf{x}) d\mathbf{x}.$$

By an adaptation of Lemma 4.3 in Lasserre and Netzer [11] to the present homogeneous context, one has

$$|y_{\tilde{\alpha}}| \leq \sup_{i=1, \dots, n} L_{\mathbf{y}}(\tilde{X}_i^{dq}) = \sup_{i=1, \dots, n} \int_{\mathbf{B}_d} |x_i|^d d\mathbf{x} \left( = \int_{\mathbf{B}_d} |x_1|^d d\mathbf{x} \right) \quad \forall \tilde{\alpha} \in \mathbb{N}_{dq}^n.$$

Indeed, in Lemma 4.3 of [11] one uses only the Hankel structure of the moment matrix  $\mathbf{M}$  and its positive definiteness. Therefore for every  $\alpha \in \mathcal{N}_{dq}^n$ ,

$$\int_{\mathbf{B}_d} |\mathbf{x}|^{\alpha} d\mathbf{x} = \int_{\mathbf{B}_d} \tilde{X}^{\tilde{\alpha}} d\mathbf{x} = y_{\tilde{\alpha}} \leq \int_{\mathbf{B}_d} |x_1|^d d\mathbf{x},$$

and so, as in the proof of Theorem 3.2,  $g^{(d)}$  satisfies the KKT-optimality conditions.  $\square$

We obtain the following even more general extension of Theorem 3.2.

**COROLLARY 6.4.** *For every  $0 < d \in \mathbb{Q}$  the generalized polynomial*

$$\mathbf{x} \mapsto g^{(d)}(\mathbf{x}) = \sum_{i=1}^n |x_i|^d$$

*is the unique optimal solution of*

$$\mathbf{P}_1 : \quad \text{opt}_1 = \inf_g \{ \|g\|_1 : v(g) \leq \rho_d; \quad g \in \mathcal{C}_d \},$$

*and (6.5) holds.*

*Proof.* Let  $0 < d \in \mathbb{Q}$ , and suppose that there exists  $g \in \mathcal{C}_d$  such that  $\text{vol}(\mathbf{G}) = \rho_d$  and  $\|g\|_1 < n$ . Write  $d = p_0/q_0$  with  $0 < p_0, q_0 \in \mathbb{N}$ . For each nonzero coefficient  $g_{\alpha}$  one has  $\alpha_i = p_i(\alpha)/q_i(\alpha)$  for some integers  $0 < p_i(\alpha), q_i(\alpha)$ . Let  $q = 2q'$  with  $q' \in \mathbb{N}$  being the least common multiple of  $\{q_0, (q_i(\alpha)), i = 1, \dots, n, g_{\alpha} \neq 0\}$ . Then  $d \in \mathbb{Z}_{q/2}$ , and  $g \in \mathbf{Hom}_d^q$ . Therefore by Theorem 6.3,  $\|g\|_1 \geq \|g^{(d)}\|_1 = n$  (where  $g^{(d)}(\mathbf{x}) = \sum_{i=1}^n |x_i|^d$ ), in contradiction with our assumption  $\|g\|_1 < n$ . Hence  $\|g\|_1 \geq n$  for all  $g \in \mathcal{C}_d$  such that  $v(g) \leq \rho_d$ . Therefore  $g^{(d)}$  is an optimal solution of  $\mathbf{P}_1$ . Uniqueness again follows from the strict convexity of  $v$ .  $\square$

Then again the parsimony property of the  $L_d$ -unit ball  $\mathbf{B}_d$  can be retrieved by minimizing the  $\ell_1$ -norm over all nonnegative generalized polynomials  $g \in \mathcal{C}_d$  whose associated unit ball  $\mathbf{G}$  has finite volume.

Next, concerning the  $\ell_2$ -norm, with  $0 < q \in \mathbb{N}$ , an analogue of problem (4.1) now reads as

$$(6.6) \quad \mathbf{P}_{2q} : \quad \text{opt}_2 = \inf_g \{ \|g\|_2^2 : v(g) \leq \rho_d; \quad g \in \mathbf{Hom}_d^q \},$$

and we have the following analogue of Theorem 4.1.

**THEOREM 6.5.** *With  $0 < q \in \mathbb{N}$  and  $0 < d \in \mathbb{Z}_q$ , problem  $\mathbf{P}_{2q}$  in (6.6) has a unique optimal solution  $g_2^* \in \mathbf{Hom}_d^q$  whose vector of coefficients  $\mathbf{g}_2^* = (g_{2,\alpha}^*) \in \mathbb{R}^{m(d,q)}$  satisfies*

$$(6.7) \quad g_{2,\alpha}^* = \text{opt}_2 \frac{n+d}{n} \cdot \frac{\int_{\mathbf{G}^*} |\mathbf{x}|^\alpha d\mathbf{x}}{\int_{\mathbf{G}^*} d\mathbf{x}} \quad \forall \alpha \in \mathcal{N}_{dq}^n,$$

where  $\mathbf{G}^* = \{\mathbf{x} : g_2^*(\mathbf{x}) \leq 1\}$  and  $\text{vol}(\mathbf{G}^*) = \rho_d$ .

We omit the proof as it is again a verbatim copy of that of Theorem 4.1. But in contrast to the case of polynomials in Theorem 4.1, in the optimal solution  $g_2^*$  of  $\mathbf{P}_{2q}$ , all coefficients  $(g_{2,\alpha}^*)$ ,  $\alpha \in \mathcal{N}_{dq}^n$ , are nonzero! This follows from (6.7) and the fact that all generalized moments  $\int_{\mathbf{G}^*} |\mathbf{x}|^\alpha d\mathbf{x}$ ,  $\alpha \in \mathcal{N}_{dq}^n$ , are nonzero! For instance, with  $d = 1/2$  and  $q = 8$ ,

$$0 < \int_{\mathbf{G}} |x_1|^{1/2} d\mathbf{x}; \quad 0 < \int_{\mathbf{G}} |x_1|^{1/8} |x_2|^{3/8} d\mathbf{x}; \quad 0 < \int_{\mathbf{G}} |x_1|^{1/4} |x_2|^{1/4} d\mathbf{x};$$

$$0 < \int_{\mathbf{G}} |x_1|^{3/8} |x_2|^{1/8} d\mathbf{x}; \quad 0 < \int_{\mathbf{G}} |x_2|^{1/2} d\mathbf{x}.$$

Hence the unique optimal solution  $g_2^*$  of  $\mathbf{P}_{2q}$  is not sparse. Even more, with fixed  $0 < d \in \mathbb{Q}$ , the larger is  $q$ , the more complicated is  $g_2^*$ . Therefore an analogue of Corollary 6.4 for the  $\ell_2$ -norm cannot exist.

### 7. Proofs.

**7.1. Proof of Theorem 2.2.** For  $\alpha \in \mathbb{N}^n$ , let  $v_\alpha : \mathbb{R}_+ \rightarrow \mathbb{R}$  be the function  $y \mapsto v_\alpha(y) := \int_{\{\mathbf{x}:h(\mathbf{x})\leq y\}} \mathbf{x}^\alpha d\mathbf{x}$ , assumed to be finite. Observe that  $v_\alpha(y) = 0$  whenever  $y < 0$ . Next, with  $\lambda > 0$ ,

$$v_\alpha(\lambda y) = \int_{\{\mathbf{x}:h(\mathbf{x})\leq \lambda y\}} \mathbf{x}^\alpha d\mathbf{x}$$

$$= \lambda^{(n+|\alpha|)/d} \int_{\{\mathbf{x}:h(\lambda^{-1/d}\mathbf{x})\leq y\}} (\lambda^{-1/d}\mathbf{x})^\alpha d(\lambda^{-1/d}\mathbf{x}) = \lambda^{(n+|\alpha|)/d} v_\alpha(y);$$

that is,  $v_\alpha$  is positively homogeneous of degree  $(n + |\alpha|)/d$ . Therefore  $v_\alpha(y) = y^{(n+|\alpha|)/d} v_\alpha(1)$ . It remains to determine  $v_\alpha(1)$ . So let  $\mathcal{L}[v_\alpha] : \mathbb{C} \rightarrow \mathbb{C}$  be the Laplace transform  $\mathcal{L}[v_\alpha]$  of the function  $v_\alpha$ , i.e.,

$$\lambda \mapsto \mathcal{L}[v_\alpha](\lambda) := \int_0^\infty \exp(-\lambda y) v_\alpha(y) dy \quad (\lambda \in \mathbb{C}; \Re(\lambda) > 0)$$

$$= v_\alpha(1) \int_0^\infty y^{(n+|\alpha|)/d} \exp(-\lambda y) dy$$

$$= v_\alpha(1) \frac{\Gamma(1 + (n + |\alpha|)/d)}{\lambda^{1+(n+|\alpha|)/d}} \quad (\lambda \in \mathbb{C}; \Re(\lambda) > 0).$$

On the other hand, whenever  $\lambda \in \mathbb{R}_+$ ,

$$\begin{aligned} \mathcal{L}[v_\alpha](\lambda) &= \int_0^\infty \exp(-\lambda y) \left( \int_{\{\mathbf{x}: h(\mathbf{x}) \leq y\}} \mathbf{x}^\alpha \, d\mathbf{x} \right) dy \\ &= \int_{\mathbb{R}^n} \mathbf{x}^\alpha \left( \int_{h(\mathbf{x})}^\infty \exp(-\lambda y) \, dy \right) d\mathbf{x} \quad (\text{by the Fubini–Tonelli theorem [1]}) \\ &= \frac{1}{\lambda} \int_{\mathbb{R}^n} \mathbf{x}^\alpha \exp(-\lambda h(\mathbf{x})) \, d\mathbf{x} \\ &= \frac{1}{\lambda^{1+(n+|\alpha|)/d}} \int_{\mathbb{R}^n} \mathbf{x}^\alpha \exp(-h(\mathbf{x})) \, d\mathbf{x} \quad (\text{by homogeneity}) \\ &= \frac{\Gamma(1+(n+|\alpha|)/d)}{\lambda^{1+(n+|\alpha|)/d}} \underbrace{\left( \frac{\int_{\mathbb{R}^n} \mathbf{x}^\alpha \exp(-h(\mathbf{x})) \, d\mathbf{x}}{\Gamma(1+(n+|\alpha|)/d)} \right)}_c, \end{aligned}$$

and so  $v_\alpha(1) = c$ , which yields (2.1).

**Strict convexity.** The function  $v$  is convex on  $\mathbf{P}[\mathbf{x}]_d$  because with  $g_1, g_2 \in \mathbf{P}[\mathbf{x}]_d$  and  $\lambda \in (0, 1)$ ,

$$\begin{aligned} v(\lambda g_1 + (1-\lambda)g_2) &= \frac{1}{\Gamma(1+n/d)} \int \exp(-\lambda g_1(\mathbf{x}) - (1-\lambda)g_2(\mathbf{x})) \, d\mathbf{x} \\ &\leq \frac{1}{\Gamma(1+n/d)} \int (\lambda \exp(-g_1(\mathbf{x})) + (1-\lambda) \exp(-g_2(\mathbf{x}))) \, d\mathbf{x} \\ &= \lambda \frac{\int \exp(-g_1) \, d\mathbf{x}}{\Gamma(1+n/d)} + (1-\lambda) \frac{\int \exp(-g_2) \, d\mathbf{x}}{\Gamma(1+n/d)} \\ &= \lambda v(g_1) + (1-\lambda)v(g_2), \end{aligned}$$

where we have used convexity of the function  $t \mapsto \exp(-t)$ . Next the strict convexity of  $v$  follows from the strict convexity of  $t \mapsto \exp(-t)$  and the fact that

$$\begin{aligned} \{\mathbf{x} : \exp(-\lambda g_1(\mathbf{x}) - (1-\lambda)g_2(\mathbf{x})) &= \lambda \exp(-g_1(\mathbf{x})) + (1-\lambda) \exp(-g_2(\mathbf{x}))\} \\ &= \{\mathbf{x} : g_1(\mathbf{x}) = g_2(\mathbf{x})\}, \end{aligned}$$

and this set has zero Lebesgue measure.

**Differentiability.** When  $g \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$ , one takes partial derivatives under the integral sign, which in this context is allowed. Indeed, for every  $\alpha \in \mathbb{N}_d^n$  with  $|\alpha| = d$ , let  $(e_\alpha) \subset \mathbb{R}^{s(d)}$  be the standard unit vectors of  $\mathbb{R}^{s(d)}$ , and let  $c := \Gamma(1+n/d)^{-1}$ . Then as  $g \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$ , for every  $t > 0$  sufficiently small,  $\mathbf{x} \mapsto g(\mathbf{x}) + t\mathbf{x}^\alpha \in \mathbf{P}[\mathbf{x}]_d$ , and so

$$\frac{v(g + te_\alpha) - v(g)}{t} = c \int_{\mathbb{R}^n} \exp(-g) \underbrace{\left( \frac{\exp(-t\mathbf{x}^\alpha) - 1}{t} \right)}_{\psi(t, \mathbf{x})} \, d\mathbf{x}$$

is well defined and finite. Notice that for every  $\mathbf{x}$ , by convexity of the function  $t \mapsto \exp(-t\mathbf{x}^\alpha)$ ,

$$\lim_{t \downarrow 0} \psi(t, \mathbf{x}) = \inf_{t > 0} \psi(t, \mathbf{x}) = \exp(-t\mathbf{x}^\alpha)'|_{t=0} = -\mathbf{x}^\alpha,$$

because for every  $\mathbf{x}$ , the function  $t \mapsto \psi(t, \mathbf{x})$  is nondecreasing; see, e.g., Rockafellar [17, Theorem 23.1]. Hence, the one-sided directional derivative  $v'(g; e_\alpha)$  in the direction  $e_\alpha$  satisfies

$$\begin{aligned} v'(g; e_\alpha) &= \lim_{t \downarrow 0} \frac{v(g + te_\alpha) - v(g)}{t} = \lim_{t \downarrow 0} c \int_{\mathbb{R}^n} \exp(-g) \psi(t, \mathbf{x}) \, d\mathbf{x} \\ &= c \int_{\mathbb{R}^n} \exp(-g) \lim_{t \downarrow 0} \psi(t, \mathbf{x}) \, d\mu(\mathbf{x}) = c \int_{\mathbb{R}^n} -\mathbf{x}^\alpha \exp(-g) \, d\mathbf{x}, \end{aligned}$$

where the third equality follows from the Extended Monotone Convergence Theorem [1, Theorem 1.6.7]. Indeed for all  $t < t_0$  with  $t_0$  sufficiently small, the function  $\psi(t, \cdot)$  is bounded above by  $\psi(t_0, \cdot)$  and  $\int_{\mathbb{R}^n} \exp(-g) \psi(t_0, \mathbf{x}) \, d\mu < \infty$ .

Similarly, for every  $t > 0$

$$\frac{v(g - te_\alpha) - v(g)}{t} = c \int_{\mathbb{R}^n} \exp(-g) \underbrace{\frac{\exp(t\mathbf{x}^\alpha) - 1}{t}}_{\xi(t, \mathbf{x})} \, d\mathbf{x},$$

and by convexity of the function  $t \mapsto \exp(t\mathbf{x}^\alpha)$ ,

$$\lim_{t \downarrow 0} \xi(t, \mathbf{x}) = \inf_{t > 0} \xi(t, \mathbf{x}) = \exp(t\mathbf{x}^\alpha)'|_{t=0} = \mathbf{x}^\alpha.$$

Therefore, with exactly the same arguments as before,

$$\begin{aligned} v'(g; -e_\alpha) &= \lim_{t \downarrow 0} \frac{v(g - te_\alpha) - v(g)}{t} \\ &= c \int_{\mathbb{R}^n} \mathbf{x}^\alpha \exp(-g) \, d\mathbf{x} = -v'(g; e_\alpha), \end{aligned}$$

and so

$$\frac{\partial v(g)}{\partial g_\alpha} = -c \int_{\mathbb{R}^n} \mathbf{x}^\alpha \exp(-g) \, d\mathbf{x}$$

for every  $\alpha$  with  $|\alpha| = d$ . Finally, to obtain (2.3) one combines the above with (2.1) and the identity  $\Gamma(z + 1) = z\Gamma(z)$ .

Similar arguments can be used for the Hessian  $\nabla^2 v(g)$ , which yields (2.4). Finally, to get (2.4) observe that  $v$  is a positively homogeneous function of degree  $-n/d$ , and so Euler's identity  $\langle \nabla v(g), g \rangle = -n v(g)/d$  for homogeneous functions yields

$$-\frac{n}{d} \int_{\mathbf{G}} d\mathbf{x} = \langle g, \nabla v(g) \rangle = -\frac{n+d}{d} \int_{\mathbf{G}} g(\mathbf{x}) \, d\mathbf{x}. \quad \square$$

**7.2. Proof of Theorem 3.2.** Problem  $\mathbf{P}_1$  has an optimal solution  $g_1^* \in \mathbf{Hom}_d$ . Indeed, let  $(g_k)$ ,  $k \in \mathbb{N}$ , be a minimizing sequence with  $\|g_k\|_1 \rightarrow \text{opt}_1 \geq 0$  as  $k \rightarrow \infty$ . Hence the sequence  $(g_n)$  is  $\ell_1$ -norm bounded, and therefore there are a subsequence  $(k_t)_{t \in \mathbb{N}}$  in  $\mathbb{N}$  and a polynomial  $g_1^* \in \mathbf{Hom}_d$  such that for every  $\alpha \in \mathbb{N}_d^n$ ,  $g_{k_t, \alpha} \rightarrow g_{1, \alpha}^*$  as  $t \rightarrow \infty$  (and so  $\|g_1^*\|_1 = \text{opt}_1$ ). Next, by Lemma 2.3,  $v$  is l.s.c., and so

$$\rho_d \geq \liminf_{t \rightarrow \infty} v(g_{k_t}) \geq v(g_1^*),$$

which proves that  $g_1^*$  is feasible for  $\mathbf{P}_1$  and thus is an optimal solution of  $\mathbf{P}_1$ .

We next prove in two steps that  $g_1^* = g^{(d)}$  is the unique optimal solution of  $\mathbf{P}_1$ : First we show uniqueness (due to the strict convexity of the function  $v$ ). Then we provide an equivalent formulation of  $\mathbf{P}_1$  for which we can apply the KKT-optimality conditions and show that  $g^{(d)}$  satisfies those optimality conditions.

• Uniqueness. Suppose that  $\mathbf{P}_1$  has another optimal solution  $h \in \mathbf{Hom}_d$  (hence such that  $h \neq g_1^*$  and  $\|h\|_1 = \|g_1^*\|_1 = \text{opt}_1$ ). As we have seen, necessarily  $v(h) = v(g_1^*) = \rho_d = \text{vol}(\mathbf{B}_d)$ . Let  $h_\lambda := \lambda h + (1 - \lambda)g_1^* \in \mathbf{Hom}_d$ ,  $\lambda \in (0, 1)$ , so that  $v(h_\lambda) \leq \rho_d$  because the feasible set is convex. By strict convexity of  $v$ ,

$$v(h_\lambda) < \lambda v(h) + (1 - \lambda)v(g_1^*) = \rho_d,$$

and by taking  $\kappa^{-n/d}v(h_\lambda) = \rho_d$  (so that  $\kappa < 1$ ), we exhibit another feasible solution  $\tilde{g} := \kappa h_\lambda \in \mathbf{Hom}_d$  with  $\ell_1$ -norm  $\|\tilde{g}\|_1 = \kappa \text{opt}_1 < \text{opt}_1$ , in contradiction with  $\text{opt}_1$  being the optimal value of  $\mathbf{P}_1$ . Hence  $g_1^*$  is the unique optimal solution of  $\mathbf{P}_1$ .

• KKT conditions. Next, Problem  $\mathbf{P}_1$  has the equivalent formulation

$$(7.1) \quad \begin{aligned} & \inf_{\lambda, g_\alpha} \sum_{\alpha \in \mathbb{N}_d^n} \lambda_\alpha \\ & \text{subject to } v(g) \leq \rho_d; \quad \lambda_\alpha - g_\alpha \geq 0; \quad \lambda_\alpha + g_\alpha \geq 0 \quad \forall \alpha \in \mathbb{N}_d^n, \end{aligned}$$

which is a convex optimization problem for which Slater's condition holds. Indeed, let  $g \in \mathbf{Hom}_d$  be an arbitrary feasible solution, and let  $h := 2g$  so that  $v(h) = 2^{-n/d}v(g) < v(g) \leq \rho_d$ . Then set  $\lambda_\alpha := 2|h_\alpha|$  so that  $\lambda_\alpha > \pm h_\alpha$  for all  $\alpha$ , and therefore  $(\lambda, h)$  is a strictly feasible solution.

Hence at an optimal solution  $(g^*, \lambda)$ , if  $v$  is differentiable at  $g^*$ , then the (necessary) KKT-optimality conditions state that

$$(7.2) \quad \begin{aligned} 1 - u_\alpha - v_\alpha &= 0 \quad \forall \alpha \in \mathbb{N}_d^n, \\ u_\alpha - v_\alpha + \theta \frac{\partial v(g^*)}{\partial g_\alpha} &= 0 \quad \forall \alpha \in \mathbb{N}_d^n, \\ \lambda_\alpha, u_\alpha, v_\alpha, \theta &\geq 0 \quad \forall \alpha \in \mathbb{N}_d^n, \\ u_\alpha (\lambda_\alpha - g_\alpha^*) &= 0; \quad v_\alpha (\lambda_\alpha + g_\alpha^*) = 0 \quad \forall \alpha \in \mathbb{N}_d^n, \\ v(g^*) &\leq \rho_d; \quad \theta (1 - v(g^*)) = 0 \end{aligned}$$

for some dual variables  $(\mathbf{u}, \mathbf{v}, \theta)$ . The meaning of the above optimality conditions is clear. Indeed, at an optimal solution  $(g^*, \lambda)$  we must have  $\lambda_\alpha = |g_\alpha^*|$  for all  $\alpha$ . Moreover, from the complementarity conditions one also has  $u_\alpha v_\alpha = 0$  whenever  $g_\alpha^* \neq 0$ . In addition, from the two first equations and the fact that  $1 = u_\alpha + v_\alpha = |u_\alpha - v_\alpha|$ ,  $\lambda_\alpha = |g_\alpha^*|$ , all moments  $\int_{\mathbf{G}^*} \mathbf{x}^\alpha d\mathbf{x}$  must be equal whenever  $g_\alpha^* \neq 0$ .

We next show that  $\mathbf{x} \mapsto g_1^*(\mathbf{x}) := \sum_{i=1}^n x_i^d = g^{(d)}(\mathbf{x})$  satisfies (7.2) and so is the unique optimal solution of (7.1). First observe that the set  $\mathbf{B}_d = \{\mathbf{x} : g^{(d)}(\mathbf{x}) \leq 1\}$  is compact, and so by Theorem 2.1,  $g^{(d)} \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$ , and by Theorem 2.2,  $v$  is differentiable at  $g^{(d)}$ . Recall also that by Theorem 2.2,  $\partial v(g^{(d)})/\partial g_\alpha = \frac{n+d}{d} \int_{\mathbf{B}_d} \mathbf{x}^\alpha d\mathbf{x}$ . Choose

$$\theta := \frac{d}{n+d} \left( \int_{\mathbf{B}_d} x_1^d d\mathbf{x} \right)^{-1}; \quad u_\alpha = \theta \frac{n+d}{d} \int_{\mathbf{B}_d} \mathbf{x}^\alpha d\mathbf{x};$$

and  $\lambda_\alpha^* = g_\alpha^{(d)}$  for all  $\alpha \in \mathbb{N}_d^n$ . Then set  $v_\alpha = 0$  whenever  $g_\alpha^{(d)} \neq 0$ , and  $v_\alpha = 1 - u_\alpha$  otherwise. (Notice that  $u_\alpha = 1$  whenever  $\mathbf{x}^\alpha = x_i^d$  for some  $i$ .) Hence  $u_\alpha \geq 0$  for all

$\alpha \in \mathbb{N}_d^n$  (because  $u_\alpha = 0$  whenever some  $\alpha_i$  is odd). Finally,  $v_\alpha = (1 - u_\alpha) \geq 0$  for all  $\alpha \in \mathbb{N}_d^n$  because

$$\left| \int_{\mathbf{B}_d} \mathbf{x}^\alpha \, d\mathbf{x} \right| \leq \int_{\mathbf{B}_d} x_1^d \, d\mathbf{x} \quad \forall \alpha \in \mathbb{N}_d^n.$$

Therefore,  $(g^{(d)}, \boldsymbol{\lambda}^*, \mathbf{u}, \mathbf{v}, \theta)$  satisfies the (necessary) KKT-optimality conditions (7.2), and as Slater's condition holds and (7.1) is convex, the KKT-optimality conditions are also sufficient. Indeed, the Lagrangian  $(g, \boldsymbol{\lambda}) \mapsto L(g, \boldsymbol{\lambda}, \mathbf{u}, \mathbf{v}, \theta)$  defined by

$$L(g, \boldsymbol{\lambda}, \mathbf{u}, \mathbf{v}, \theta) := \sum_{\alpha} \lambda_{\alpha} - u_{\alpha}(\lambda_{\alpha} - g_{\alpha}) - v_{\alpha}(\lambda_{\alpha} + g_{\alpha}) + \theta(v(g) - \rho_d)$$

is convex, and  $\nabla L(g^{(d)}, \boldsymbol{\lambda}^*, \mathbf{u}, \mathbf{v}, \theta) = 0$ . Hence by convexity, for any feasible solution  $(g, \boldsymbol{\lambda})$  of (7.1),

$$\sum_{\alpha} \lambda_{\alpha}^* = L(g^{(d)}, \boldsymbol{\lambda}^*, \mathbf{u}, \mathbf{v}, \theta) \leq L(g, \boldsymbol{\lambda}, \mathbf{u}, \mathbf{v}, \theta) \leq \sum_{\alpha} \lambda_{\alpha};$$

i.e.,  $(g^{(d)}, \boldsymbol{\lambda}^*)$  is an optimal solution of (7.1). Equivalently,  $g^{(d)}$  is the unique optimal solution of  $\mathbf{P}_1$ . Next observe that

$$\begin{aligned} \text{opt}_1 = \|g^{(d)}\|_1 &= n = \sum_{\alpha \in \mathbb{N}_d^n} \lambda_{\alpha}^* = \sum_{\alpha \in \mathbb{N}_d^n} (u_{\alpha} - v_{\alpha}) g_{\alpha}^{(d)} \\ &= -\theta \langle \nabla v(g^{(d)}), g^{(d)} \rangle = \theta \frac{n}{d} v(g^{(d)}), \end{aligned}$$

from which we deduce

$$\int_{\mathbf{B}_d} x_i^d \, d\mathbf{x} = \frac{1}{n+d} \int_{\mathbf{B}_d} d\mathbf{x}$$

for  $i = 1, \dots, n$ , which is (3.4). Finally, for the numerical value of  $\text{vol}(\mathbf{B}_d)$  in (3.4), see Lemma A.1.  $\square$

**7.3. Proof of Theorem 4.3.** In (4.1) one may replace  $\rho_d$  with the volume of the Euclidean ball  $\mathbf{B}_2$  since by homogeneity, the optimal solution is the same (up to a scaling factor). By Theorem 4.1 it is enough to prove that (4.2) holds for  $\mathbf{x} \mapsto g(\mathbf{x}) = (\sum_i x_i^2)^{d/2}$ . We will use the fact that all moments of the Lebesgue measure on  $\mathbf{B}_2$  can be obtained in closed form.

- The case  $d = 2$ . Then  $\text{opt}_2 = n$ , and

$$A := \int_{\mathbf{B}_2} x_1^2 \, d\mathbf{x} \left( \int_{\mathbf{B}_2} d\mathbf{x} \right)^{-1} = \frac{1}{(n+2)} \quad (\text{by (3.4)}),$$

so that

$$\text{opt}_2 \frac{n+d}{n} A = n \frac{(n+2)}{n} \frac{1}{(n+2)} = 1,$$

which is indeed the coefficient  $1/c_{\alpha}$  of  $x_i^2$  in  $g_2^*(\mathbf{x}) = \sum_{\alpha} c_{\alpha} g_{2,\alpha}^* \mathbf{x}^{\alpha}$ .

• The case  $d = 4$ . Let  $\mathbf{x} \mapsto g_2^*(\mathbf{x}) := (\sum_{i=1}^n x_i^2)^2$ . It is enough to prove that (4.2) holds. Since  $g_2^*(\mathbf{x}) = \sum_i x_i^4 + 2 \sum_{i < j} x_i^2 x_j^2$ , we have  $\text{opt}_2 = (n + \frac{n(n-1)}{2} \cdot 6(\frac{2}{6})^2) = n(n+2)/3$ . Moreover,

$$A := \int_{\mathbf{B}_2} x_i^4 \, d\mathbf{x} \left( \int_{\mathbf{B}_2} d\mathbf{x} \right)^{-1} = \frac{3}{(n+4)(n+2)}.$$



Therefore with  $d = 4$  and  $\alpha = (4, 0, \dots, 0)$ ,

$$\text{opt}_2 \frac{n+d}{n} A = \frac{n(n+2)}{3} \frac{n+4}{n} \frac{3}{(n+2)(n+4)} = 1,$$

which is the coefficient  $1/c_\alpha$  of  $x_i^4$  in  $g_2^*(\mathbf{x})$ . Similarly, with  $\alpha = (2, 2, 0, \dots, 0)$ ,

$$B := \int_{\mathbf{B}_2} x_i^2 x_j^2 d\mathbf{x} \left( \int_{\mathbf{B}_2} d\mathbf{x} \right)^{-1} = \frac{1}{(n+4)(n+2)}$$

so that

$$\text{opt}_2 \frac{n+4}{n} B = \frac{n(n+2)}{3} \frac{n+4}{n} \frac{1}{(n+2)(n+4)} = \frac{1}{3},$$

which is the coefficient  $2/c_\alpha$  of  $x_i^2 x_j^2$  in  $g_2^*(\mathbf{x})$ .

- The case  $d = 6$ . Let  $\mathbf{x} \mapsto g_2^*(\mathbf{x}) := (\sum_{i=1}^n x_i^2)^3$ , i.e.,

$$g_2^*(\mathbf{x}) = \sum_{i=1}^n x_i^6 + 3 \sum_{i \neq j} x_i^4 x_j^2 + 6 \sum_{i < j < k} x_i^2 x_j^2 x_k^2.$$

Again it is enough to prove that (4.2) holds. First one obtains  $\text{opt}_2 = n(n+2)(n+4)/15$ . Then

$$A := \int_{\mathbf{B}_2} x_i^6 d\mathbf{x} \left( \int_{\mathbf{B}_2} d\mathbf{x} \right)^{-1} = \frac{15}{(n+6)(n+4)(n+2)},$$

and therefore with  $d = 6$  and  $\alpha = (6, 0, \dots, 0)$ ,

$$\text{opt}_2 \frac{n+d}{n} A = \frac{n(n+2)(n+4)}{15} \frac{n+6}{n} \frac{15}{(n+2)(n+4)(n+6)} = 1,$$

which is the coefficient  $1/c_\alpha$  of  $x_i^6$  in  $g_2^*(\mathbf{x})$ . Similarly, with  $\alpha = (4, 2, 0, \dots, 0)$ ,

$$B = \int_{\mathbf{G}^*} x_i^4 x_j^2 d\mathbf{x} \left( \int_{\mathbf{G}^*} d\mathbf{x} \right)^{-1} = \frac{3}{(n+2)(n+4)(n+6)},$$

so that

$$\text{opt}_2 \frac{n+d}{n} B = \frac{n(n+2)(n+4)}{15} \frac{n+6}{n} \frac{3}{(n+2)(n+4)(n+6)} = \frac{1}{5},$$

which is the coefficient  $3/c_\alpha$  of  $x_i^4 x_j^2$  in  $g_2^*(\mathbf{x})$ .

- The case  $d = 8$ . Let  $\mathbf{x} \mapsto g_2^*(\mathbf{x}) := (\sum_{i=1}^n x_i^2)^4$ . Again, it is enough to prove that (4.2) holds. Since

$$g_2^*(\mathbf{x}) = \sum_{i=1}^n x_i^8 + 4 \sum_{i \neq j} x_i^6 x_j^2 + 6 \sum_{i < j} x_i^4 x_j^4 + 12 \sum_{i \neq j \neq k} x_i^4 x_j^2 x_k^2 + 24 \sum_{i \neq j \neq k \neq \ell} x_i^2 x_j^2 x_k^2 x_\ell^2,$$

we obtain  $\text{opt}_2 = n(n+2)(n+4)(n+6)/105$ . Next,

$$A := \int_{\mathbf{B}_2} x_1^8 d\mathbf{x} \left( \int_{\mathbf{B}_2} d\mathbf{x} \right)^{-1} = \frac{105}{(n+8)(n+6)(n+4)(n+2)},$$

and therefore with  $d = 8$  and  $\alpha = (8, 0, \dots, 0)$ , one obtains  $(n + 8)\text{opt}_2 A/n = 1$ , which is the coefficient  $1/c_\alpha$  of  $x_i^8$  in  $g_2^*(\mathbf{x})$ . Similarly, with  $\alpha = (6, 2, 0, \dots, 0)$ ,

$$A := \int_{\mathbf{B}_2} x_i^6 x_j^2 d\mathbf{x} \left( \int_{\mathbf{B}_2} d\mathbf{x} \right)^{-1} = \frac{15}{(n + 8)(n + 6)(n + 4)(n + 2)},$$

so that  $(n + 8)\text{opt}_2 A/n = 1/7$ , which is the coefficient  $4/c_\alpha$  of  $x_i^6 x_j^2$  in  $g_2^*(\mathbf{x})$ . With  $\alpha = (4, 4, 0, \dots, 0)$ ,

$$A := \int_{\mathbf{B}_2} x_i^4 x_j^4 d\mathbf{x} \left( \int_{\mathbf{B}_2} d\mathbf{x} \right)^{-1} = \frac{9}{(n + 8)(n + 6)(n + 4)(n + 2)},$$

so that  $(n + 8)\text{opt}_2 A/n = 3/35$ , which is the coefficient  $6/c_\alpha$  of  $x_i^4 x_j^4$  in  $g_2^*(\mathbf{x})$ .

With  $\alpha = (4, 2, 2, 0, \dots, 0)$ ,

$$A := \int_{\mathbf{B}_2} x_i^4 x_j^2 x_\ell^2 d\mathbf{x} \left( \int_{\mathbf{B}_2} d\mathbf{x} \right)^{-1} = \frac{3}{(n + 8)(n + 6)(n + 4)(n + 2)},$$

so that  $(n + 8)\text{opt}_2 A/n = 1/35$ , which is the coefficient  $12/c_\alpha$  of  $x_i^4 x_j^2 x_\ell^2$  in  $g_2^*(\mathbf{x})$ .

Finally, with  $\alpha = (2, 2, 2, 2, 0, \dots, 0)$ ,

$$A := \int_{\mathbf{B}_2} x_i^2 x_j^2 x_k^2 x_\ell^2 d\mathbf{x} \left( \int_{\mathbf{B}_2} d\mathbf{x} \right)^{-1} = \frac{1}{(n + 8)(n + 6)(n + 4)(n + 2)},$$

so that  $(n + 8)\text{opt}_2 A/n = 1/105$ , which is the coefficient  $24/c_\alpha$  of  $x_i^2 x_j^2 x_k^2 x_\ell^2$  in  $g_2^*(\mathbf{x})$ .  $\square$

**7.4. Proof of Theorem 5.1.** Let  $(0 \preceq \mathbf{Q}_k)$ ,  $k \in \mathbb{N}$ , be a minimizing sequence of  $\mathbf{P}_3$ ; i.e.,  $\langle \mathbf{I}, \mathbf{Q}_k \rangle \rightarrow \text{opt}_3$  as  $k \rightarrow \infty$ . As  $\sup_k \langle \mathbf{I}, \mathbf{Q}_k \rangle \leq \langle \mathbf{I}, \mathbf{Q}_1 \rangle$ , the sequence  $(\mathbf{Q}_k)$  is norm-bounded in  $\mathcal{S}_d$ . Therefore it has a converging subsequence  $\mathbf{Q}_{k_j} \rightarrow \mathbf{Q}^* \in \mathcal{S}_d$  as  $j \rightarrow \infty$ , with  $\mathbf{Q}^* \succeq 0$ . Then  $\langle \mathbf{I}, \mathbf{Q}^* \rangle = \text{opt}_3$ , and as  $\hat{v}$  is l.s.c.,

$$\rho_d \geq \liminf_{j \rightarrow \infty} \hat{v}(\mathbf{Q}_{k_j}) \geq \hat{v}(\mathbf{Q}^*),$$

so that  $\mathbf{Q}^*$  is an optimal solution of  $\mathbf{P}_3$ . Let  $g_3^* = g_{\mathbf{Q}^*} \in \mathbf{Hom}_d$  be the homogeneous polynomial associated with an optimal solution of  $\mathbf{Q}^*$  of  $\mathbf{P}_3$ , and suppose that there exists another optimal solution  $0 \preceq \mathbf{H} \in \mathcal{S}_d$  of  $\mathbf{P}_3$  such that  $g_{\mathbf{H}} \neq g_3^*$ . Then we have  $v(g_3^*) = \hat{v}(\mathbf{Q}^*) = \hat{v}(\mathbf{H}) = v(g_{\mathbf{H}}) = \rho_d$ . For  $\lambda \in (0, 1)$  fixed, arbitrary, the polynomial  $h_\lambda := \lambda g_3^* + (1 - \lambda)g_{\mathbf{H}}$  satisfies  $v(h_\lambda) < \rho_d$  because  $g_3^* \neq g_{\mathbf{H}}$ . Moreover,

$$h_\lambda(\mathbf{x}) := \lambda g_3^*(\mathbf{x}) + (1 - \lambda)g_{\mathbf{H}}(\mathbf{x}) = \mathbf{v}_{d/2}(\mathbf{x})^T \underbrace{(\lambda \mathbf{Q}^* + (1 - \lambda)\mathbf{H})}_{\mathbf{H}_\lambda \succeq 0} \mathbf{v}_{d/2}(\mathbf{x}),$$

and so  $\mathbf{H}_\lambda$  is also an optimal solution of  $\mathbf{P}_3$  since  $\langle \mathbf{I}, \lambda \mathbf{Q}^* + (1 - \lambda)\mathbf{H} \rangle = \text{opt}_3$ . But then as we have already shown before, by homogeneity of  $v$ , there is  $0 < \theta < 1$  such that  $v(\theta h_\lambda) = \hat{v}(\theta \mathbf{H}_\lambda) = \rho_d$ . Hence  $(0 \preceq) \theta \mathbf{H}_\lambda$  is a feasible solution of  $\mathbf{P}_3$  with associated cost  $\langle \mathbf{I}, \theta \mathbf{H}_\lambda \rangle = \theta \langle \mathbf{I}, \mathbf{H}_\lambda \rangle = \theta \text{opt}_3 < \text{opt}_3$ , a contradiction. Therefore  $g_3^* \in \mathbf{Hom}_d$  is the unique solution associated with all optimal solutions of  $\mathbf{P}_3$ .

(a) Next, assume that  $g_3^* \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$  so that  $v$  is differentiable at  $g_3^*$  and so is  $\hat{v}$  at  $\mathbf{Q}^*$ . Slater's condition obviously holds for  $\mathbf{P}_3$ , which is a convex optimization

problem. Therefore the KKT-optimality conditions read as

$$(7.3) \quad \mathbf{I} + \lambda \nabla \hat{v}(\mathbf{Q}^*) = \Psi \succeq 0,$$

$$(7.4) \quad \langle \mathbf{Q}^*, \Psi \rangle = 0,$$

$$(7.5) \quad \hat{v}(\mathbf{Q}^*) \leq \rho_d; \quad \lambda(\hat{v}(\mathbf{Q}^*) - \rho_d) = 0$$

for some dual variables  $(\lambda, \Psi) \in \mathbb{R}_+ \times \mathcal{S}_d$ . By homogeneity one must have  $\hat{v}(\mathbf{Q}^*) = \rho_d$ . By Euler's identity for homogeneous functions,  $\langle \nabla \hat{v}(\mathbf{Q}^*), \mathbf{Q}^* \rangle = -n\hat{v}(\mathbf{Q}^*)/d$ . Therefore using (7.3) and (7.4), one obtains  $\text{opt}_3 = \langle \mathbf{I}, \mathbf{Q}^* \rangle = \lambda n \rho_d / d$ ; that is,  $\lambda = d \text{opt}_3 / (n \rho_d)$ . Hence

$$\Psi = \mathbf{I} - \frac{(n+d) \langle \mathbf{I}, \mathbf{Q}^* \rangle}{n \rho_d} \int_{\mathbf{G}_{\mathbf{Q}^*}} \mathbf{v}_{d/2}(\mathbf{x}) \mathbf{v}_{d/2}(\mathbf{x})^T d\mathbf{x} \succeq 0,$$

which is (5.2).

(b) Conversely, assume that  $0 \preceq \mathbf{Q}^* \in \mathcal{S}_d$ ,  $g_{\mathbf{Q}^*} \in \text{int}(\mathbf{P}[\mathbf{x}]_d)$ ,  $\hat{v}(\mathbf{Q}^*) = \rho_d$ , and (5.2) holds. Then let

$$\lambda := \frac{d \langle \mathbf{I}, \mathbf{Q}^* \rangle}{n \rho_d}; \quad \Psi := \mathbf{I} - \frac{(n+d) \langle \mathbf{I}, \mathbf{Q}^* \rangle}{n \rho_d} \int_{\mathbf{G}_{\mathbf{Q}^*}} \mathbf{v}_{d/2}(\mathbf{x}) \mathbf{v}_{d/2}(\mathbf{x})^T d\mathbf{x}.$$

Obviously  $\lambda \geq 0$ ,  $\Psi \succeq 0$ , and

$$\begin{aligned} \langle \mathbf{Q}^*, \Psi \rangle &= \langle \mathbf{I}, \mathbf{Q}^* \rangle \left[ 1 - \frac{n+d}{n \rho_d} \int_{\mathbf{G}_{\mathbf{Q}^*}} \langle \mathbf{Q}^*, \mathbf{v}_{d/2}(\mathbf{x}) \mathbf{v}_{d/2}(\mathbf{x})^T \rangle d\mathbf{x} \right] \\ &= \langle \mathbf{I}, \mathbf{Q}^* \rangle \left[ 1 - \frac{n+d}{n \rho_d} \int_{\mathbf{G}_{\mathbf{Q}^*}} g_{\mathbf{Q}^*}(\mathbf{x}) d\mathbf{x} \right] \\ &= \langle \mathbf{I}, \mathbf{Q}^* \rangle \left[ 1 - \frac{n+d}{n \rho_d} \frac{n}{n+d} \int_{\mathbf{G}_{\mathbf{Q}^*}} d\mathbf{x} \right] \quad (\text{by Theorem 2.2}) \\ &= 0, \end{aligned}$$

which shows that the triplet  $(\mathbf{Q}^*, \lambda, \Psi)$  satisfies the KKT-optimality conditions (7.3)–(7.5) and so is an optimal solution of  $\mathbf{P}_3$ ; therefore  $g_3^* = g_{\mathbf{Q}^*}$ .

Finally, of course, when  $d = 2$ , it is immediate to check that  $g_3^* = \sum_i x_i^2$ . On the other hand, if  $d > 2$ , the polynomial  $\mathbf{x} \mapsto g^{(d)}(\mathbf{x}) := \sum_{i=1}^n x_i^d$  (so that  $\mathbf{G} = \mathbf{B}_d$ ) cannot be the optimal solution of  $\mathbf{P}_3$ . Among all  $\mathbf{Q} \succeq 0$  such that  $\sum_i x_i^d = \mathbf{v}_{d/2}(\mathbf{x})^T \mathbf{Q} \mathbf{v}_{d/2}(\mathbf{x})$ , the one that minimizes  $\text{trace}(\mathbf{Q})$  has all of its entries equal to zero except the diagonal entries corresponding to columns associated with the monomials  $x_i^{d/2}$ ,  $i = 1, \dots, n$ , and  $\text{trace}(\mathbf{Q}) = n$ . By Lemma A.1, observe that

$$\frac{\int_{\mathbf{G}} x_i^d d\mathbf{x}}{\int_{\mathbf{G}} d\mathbf{x}} = \frac{\int_{\mathbf{B}_d} x_i^d d\mathbf{x}}{\rho_d} = \frac{1}{n+d},$$

and so (5.2) cannot hold because, for instance, the northwest and southeast corner elements of the matrix

$$\begin{aligned} \Psi &:= \mathbf{I} - \frac{(n+d) \langle \mathbf{I}, \mathbf{Q} \rangle}{n \rho_d} \int_{\mathbf{G}} \mathbf{v}_{d/2}(\mathbf{x}) \mathbf{v}_{d/2}(\mathbf{x})^T d\mathbf{x} \\ &= \mathbf{I} - (n+d) \frac{\int_{\mathbf{G}} \mathbf{v}_{d/2}(\mathbf{x}) \mathbf{v}_{d/2}(\mathbf{x})^T d\mathbf{x}}{\int_{\mathbf{G}} d\mathbf{x}} \end{aligned}$$

vanish, whereas the northeast and southwest corner elements are nonzero, in contradiction with  $\mathbf{A} \succeq 0$ .

Next, if  $d = 4$ , consider the quartic form  $\mathbf{x} \mapsto (\sum_i x_i^2)^2$ , which is the single square  $(\mathbf{q}^T \mathbf{v}_2(\mathbf{x}))^2$ , where the entries of the vector  $\mathbf{q} \in \mathbb{R}^{s(2)}$  corresponding to monomials  $x_i^2$  are equal to 1, and zero otherwise. Hence  $\mathbf{Q} = \mathbf{q}\mathbf{q}^T \succeq 0$  with  $\text{trace}(\mathbf{Q}) = n$ . To avoid the presence of a scaling factor, in problem  $\mathbf{P}_3$  in (5.1), one may and will replace  $\rho_d$  with  $\rho_2$  (the volume of the Euclidean unit ball). In the proof of Theorem 4.3 we have already seen that

$$\int_{\mathbf{G}_{\mathbf{Q}}} x_i^4 d\mathbf{x} = \frac{3\rho_2}{(n+2)(n+4)}; \quad \int_{\mathbf{G}_{\mathbf{Q}}} x_i^2 x_j^2 d\mathbf{x} = \frac{\rho_2}{(n+4)(n+2)}.$$

Therefore, in view of its specific form, the matrix  $\Psi$  in (5.2) is PSD because  $1 - 3/(n+2) \geq 0$ ,  $1 - 1/(n+2) \geq 0$ , and  $((n-1)/(n+2))^2 \geq 1/(n+2)^2$ . For instance, with  $n = 3$ ,

$$\Psi = \begin{bmatrix} \frac{n-1}{n+2} & 0 & 0 & \frac{-1}{n+2} & 0 & \frac{-1}{n+2} \\ 0 & \frac{n+1}{n+2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{n+1}{n+2} & 0 & 0 & 0 \\ \frac{-1}{n+2} & 0 & 0 & \frac{n-1}{n+2} & 0 & \frac{-1}{n+2} \\ 0 & 0 & 0 & 0 & \frac{n+1}{n+2} & 0 \\ \frac{-1}{n+2} & 0 & 0 & \frac{-1}{n+2} & 0 & \frac{n-1}{n+2} \end{bmatrix} \succeq 0.$$

From the definition of  $\mathbf{Q}$ ,  $\langle \Psi, \mathbf{Q} \rangle = 0$  so that  $\mathbf{Q}$  is an optimal solution of  $\mathbf{P}_3$ ; that is, by uniqueness,  $g_3^*(\mathbf{x}) = (\sum_i x_i^2)^2$ .

Finally, consider the case  $d = 8$ . Again, to avoid the presence of a scaling factor, in problem  $\mathbf{P}_3$  in (5.1), one may and will replace  $\rho_d$  with  $\rho_2$  (the volume of the Euclidean unit ball). The degree-8 form

$$\mathbf{x} \mapsto \left( \sum_i x_i^2 \right)^4 = \left( \left( \sum_i x_i^2 \right)^2 \right)^2 = \left( \sum_i x_i^4 + 2 \sum_{i < j} x_i^2 x_j^2 \right)^2$$

is the single square  $(\mathbf{q}^T \mathbf{v}_4(\mathbf{x}))^2$ , where the entries of the vector  $\mathbf{q} \in \mathbb{R}^{s(4)}$  are equal to 1 (resp., 2) when they correspond to monomials  $x_i^4$  (resp.,  $x_i^2 x_j^2$ ), and zero otherwise. Hence  $\mathbf{Q} = \mathbf{q}\mathbf{q}^T \succeq 0$ , and with  $\Psi$  as in (5.2), one has  $\langle \Psi, \mathbf{Q} \rangle = 0$ . To prove the result it remains to show that  $\Psi \succeq 0$ . Observe that  $\text{trace}(\mathbf{Q}) = \|\mathbf{q}\|^2 = n + 4n(n-1)/2 = O(n^2)$ . Next, from the proof of Theorem 4.3,

$$\frac{\int_{\mathbf{B}_2} \mathbf{x}^\alpha d\mathbf{x}}{\int_{\mathbf{B}_2} d\mathbf{x}} = \frac{s_\alpha}{(n+8)(n+6)(n+4)(n+2)} = O(n^{-4}) \quad \forall \alpha \in \mathbb{N}^{s(d)},$$

where depending on  $\alpha$ ,  $s_\alpha$  is 105, 15, 9, 3, or 1. Hence the entries of the matrix

$$(7.6) \quad \Delta := \frac{(n+d)\langle \mathbf{I}, \mathbf{Q} \rangle}{n} \frac{\int_{\mathbf{B}_2} \mathbf{v}_{d/2}(\mathbf{x}) \mathbf{v}_{d/2}(\mathbf{x})^T d\mathbf{x}}{\int_{\mathbf{B}_2} d\mathbf{x}}$$

are all  $O(n^{-2})$ , and so  $\Psi = \mathbf{I} - \Delta \succeq 0$  whenever  $n$  is sufficiently large.

Finally, the same limiting argument also holds for all  $d = 4p$ ,  $p \in \mathbb{N}$ . Indeed, one may repeat the above argument when  $d = 8$  by considering the degree- $d$  form

$g(\mathbf{x})^2$  with  $g(\mathbf{x}) := (\sum_i x_i^2)^p$ . Writing  $g(\mathbf{x}) = \mathbf{q}^T \mathbf{v}_{2p}(\mathbf{x})$  for an appropriate vector  $\mathbf{q} \in \mathbb{R}^{s(2p)}$ , one has  $\mathbf{Q} = \mathbf{q}\mathbf{q}^T \succeq 0$  and  $\text{trace}(\mathbf{Q}) = O(n^p)$ . On the other hand, we also have

$$\frac{\int_{\mathbf{B}_2} \mathbf{x}^\alpha d\mathbf{x}}{\int_{\mathbf{B}_2} d\mathbf{x}} = \frac{s_\alpha}{(n+d)(n+d-2)\cdots(n+2)} = O(n^{-d/2}) \quad \forall \alpha \in \mathbb{N}^{s(d)},$$

and so all entries of the matrix  $\Delta$  in (7.6) are  $O(n^{-d/4})$ . Therefore for  $n$  sufficiently large,  $\Psi = \mathbf{I} - \Delta \succeq 0$  and so  $g_3^*(\mathbf{x}) = (\sum_i x_i^2)^{d/2}$ .  $\square$

### Appendix A.

LEMMA A.1. *Let  $d$  be a positive real, and let  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  be the function*

$$\mathbf{x} \mapsto g(\mathbf{x}) := \sum_{i=1}^n |x_i|^d; \quad \mathbf{G} := \{ \mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) \leq 1 \}.$$

Then

$$(A.1) \quad \int_{\mathbf{G}} d\mathbf{x} = \frac{2^n}{n d^{n-1}} \frac{\Gamma(1/d)^n}{\Gamma(n/d)}; \quad \int_{\mathbf{G}} |x_i|^d d\mathbf{x} = \frac{2^n}{n(n+d) d^{n-1}} \frac{\Gamma(1/d)^n}{\Gamma(n/d)}$$

for all  $i = 1, \dots, n$ .

*Proof.* The function  $g$  is positively homogeneous of degree  $d$ . Observe that by Theorem 2.2,  $\int_{\mathbf{G}} d\mathbf{x} = \frac{1}{\Gamma(1+n/d)} \int_{\mathbb{R}^n} \exp(-g(\mathbf{x})) d\mathbf{x}$ , and again by Theorem 2.2,

$$\int_{\mathbb{R}^n} \exp(-g(\mathbf{x})) d\mathbf{x} = \left( \int_{\mathbb{R}} \exp(-|t|^d) dt \right)^n = \left( \Gamma\left(1 + \frac{1}{d}\right) \int_{|t|^d \leq 1} dt \right)^n = \left( \frac{2\Gamma(1/d)}{d} \right)^n,$$

where we have used the identity  $x\Gamma(x) = \Gamma(1+x)$ . This yields the result on the left-hand side of (A.1). Similarly, for every  $i = 1, \dots, n$ ,

$$\begin{aligned} \int_{\mathbb{R}^n} |x_i|^d \exp(-g(\mathbf{x})) d\mathbf{x} &= \left( \int_{\mathbb{R}} |t|^d \exp(-|t|^d) dt \right) \left( \int_{\mathbb{R}} \exp(-|t|^d) dt \right)^{n-1} \\ &= \left( \Gamma\left(1 + \frac{d+1}{d}\right) \int_{|t|^d \leq 1} |t|^d dt \right) \left( \Gamma\left(1 + \frac{1}{d}\right) \int_{|t|^d \leq 1} dt \right)^{n-1} \\ &= \left( \frac{2}{d} \Gamma\left(\frac{d+1}{d}\right) \right) \left( \frac{2}{d} \Gamma\left(\frac{1}{d}\right) \right)^{n-1} \\ &= \frac{2^n}{d^{n+1}} \Gamma\left(\frac{1}{d}\right)^n. \end{aligned}$$

Therefore,

$$\int_{\mathbf{G}} |x_i|^d d\mathbf{x} = \frac{1}{\Gamma(1 + \frac{n+d}{d})} \int_{\mathbb{R}^n} |x_i|^d \exp(-g(\mathbf{x})) d\mathbf{x} = \frac{2^n}{n(n+d) d^{n-1}} \frac{\Gamma(\frac{1}{d})^n}{\Gamma(\frac{n}{d})}. \quad \square$$

**Acknowledgment.** We wish to thank an anonymous referee whose comments and suggestions were very helpful in improving the initial version of this paper.

## REFERENCES

- [1] R. B. ASH, *Real Analysis and Probability*, Academic Press, Boston, 1972.
- [2] F. BACH, R. JENATTON, J. MAIRAL, AND G. OBOZINSKI, *Optimization with sparsity-inducing penalties*, Found. Trends Mach. Learn., 4 (2012), pp. 1–106.
- [3] A. BECK AND Y. C. ELDAR, *Sparsity constrained nonlinear optimization: Optimality conditions and algorithms*, SIAM J. Optim., 23 (2013), pp. 1480–1509.
- [4] E. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.
- [5] D. L. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.
- [6] D. L. DONOHO AND M. ELAD, *Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell^1$  minimization*, Proc. Natl. Acad. Sci. USA, 100 (2003), pp. 2197–2202.
- [7] E. FREITAG AND R. BUSAM, *Complex Analysis*, 2nd ed., Springer-Verlag, Berlin, 2009.
- [8] D. HILBERT AND S. COHN-VOSSEN, *Geometry and the Imagination*, 2nd ed., Chelsea, New York, 1952.
- [9] J. B. LASSERRE, *A quick proof for the volume of  $n$ -balls*, Amer. Math. Monthly, 108 (2001), pp. 768–769.
- [10] J. B. LASSERRE, *A generalization of Löwner-John’s ellipsoid theorem*, Math. Program., 152 (2015), pp. 559–591.
- [11] J. B. LASSERRE AND T. NETZER, *SOS approximations of nonnegative polynomials via simple high degree perturbations*, Math. Z., 256 (2007), pp. 99–112.
- [12] J. B. LASSERRE AND E. S. ZERON, *Solving a class of multivariable integration problems via Laplace techniques*, Appl. Math. (Warsaw), 28 (2001), pp. 391–405.
- [13] A. MOROZOV AND SH. SHAKIROV, *Introduction to integral discriminants*, J. High Energy Phys., 12 (2009), 002.
- [14] A. YU. MOROZOV AND SH. R. SHAKIROV, *New and old results in resultant theory*, Theoret. and Math. Phys., 163 (2010), pp. 587–617.
- [15] B. RECHT, M. FAZEL, AND P. A. PARRILO, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501.
- [16] B. REZNICK, *Sums of even powers of real linear forms*, Mem. Amer. Math. Soc., 96 (1992), no. 463.
- [17] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- [18] D. V. WIDDER, *The Laplace Transform*, Princeton Math. Ser. 6, Princeton University Press, Princeton, NJ, 1941.