Sum of Squares Methods for Minimizing Polynomial Forms over Spheres and Hypersurfaces

Jiawang Nie*

January 18, 2011

Abstract

This paper studies the problem of minimizing a homogeneous polynomial (form) $f(x)$ over the unit sphere $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$. The problem is NP-hard when $f(x)$ has degree 3 or higher. Denote by f_{min} (resp., f_{max}) the minimum (resp., maximum) value of $f(x)$ on $\mathbb{S}^{n-1}$. First, when $f(x)$ is an even form of degree $2d$, we study the standard sum of squares (SOS) relaxation for finding a lower bound of the minimum f_{min}:

$$\max \quad \gamma \quad s.t. \quad f(x) - \gamma \cdot \|x\|_2^{2d} \text{ is SOS.}$$

Let f_{sos} be the above optimal value. Then we show that for all $n \geq 2d$

$$1 \leq \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \leq C(d)\sqrt{\binom{n}{2d}}.$$

Here the constant $C(d)$ is independent of n. Second, when $f(x)$ is a multi-form and $\mathbb{S}^{n-1}$ becomes a multi-unit sphere, we generalize the above SOS relaxation and prove a similar bound. Third, when $f(x)$ is sparse, we prove an improved bound depending on its sparsity pattern; when $f(x)$ is odd, we formulate the problem equivalently as minimizing a certain even form, and prove a similar bound. Last, for minimizing $f(x)$ over a hypersurface $H(g) = \{x \in \mathbb{R}^n : g(x) = 1\}$ defined by a positive definite form $g(x)$, we generalize the above SOS relaxation and prove a similar bound.

Key words approximation bound, form, hypersurface, L^2-norm, G-norm, multi-form, polynomial, semidefinite programming, sum of squares

AMS subject classification 65K05, 68Q25, 90C22, 90C59

1 Introduction

Let $f(x)$ be a multivariate homogeneous polynomial (form) in $x \in \mathbb{R}^n$. Consider problem

$$\min_{x \in \mathbb{S}^{n-1}} \quad f(x). \tag{1.1}$$

*Department of Mathematics, University of California, 9500 Gilman Drive, La Jolla, CA 92093, USA. Email: njw@math.ucsd.edu. The research was partially supported by NSF grants DMS-0757212, DMS-0844775 and Hellman Foundation Fellowship.

Here $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ is the $n-1$ dimensional unit sphere. Denote by f_{min} the minimum value of $f(x)$ on $\mathbb{S}^{n-1}$. When $f(x) = f^T x$ is a linear form, $f_{min} = -\|f\|_2$, which can be found easily. When $f(x) = x^T F x$ is a quadratic form, f_{min} is the minimum eigenvalue of the symmetric matrix $\frac{1}{2}(F + F^T)$, which can also be computed efficiently by solving an eigenvalue problem. However, if $\deg(f) > 2$, it is usually very difficult to compute f_{min}. Nesterov [7] showed (1.1) is already NP-hard when $f(x)$ is cubic. So in practical applications, we are more interested in approximation algorithms. The sum of squares (SOS) relaxation is a standard approximation method for solving (1.1).

When $f(x)$ is an even form of degree $2d$, the standard SOS relaxation for (1.1) is

$$\begin{array}{cl} \max & \gamma \\ s.t. & f(x) - \gamma \cdot \|x\|_2^{2d} \text{ is SOS.} \end{array} \tag{1.2}$$

Here a polynomial is said to be SOS if it is a sum of squares of some other polynomials. Denote by f_{sos} the optimal value of (1.2). Obviously, every γ feasible in (1.2) is a lower bound of the minimum f_{min}. This is because if $f(x) - \gamma\|x\|_2^{2d}$ is SOS, then $f(x) - \gamma\|x\|_2^{2d}$ must be globally nonnegative and hence $f(x) \geq \gamma$ for all $x \in \mathbb{S}^{n-1}$. So $f_{sos} \leq f_{min}$. The original problem (1.1) is NP-hard, but SOS relaxation (1.2) is a convex program and can be solved efficiently. In fact, (1.2) is equivalent to a semidefinite programming (SDP) problem.

Note that every form $p(x)$ of degree $2d$ can be written as $p(x) = [x^d]^T P [x^d]$ for a symmetric matrix P. Here $[x^d]$ denotes the column vector of all monomials of degree d ordered lexicographically, that is,

$$[x^d]^T = \begin{bmatrix} x_1^d & x_1^{d-1}x_2 & \cdots & x_1^{d-1}x_n & x_1^{d-2}x_2^2 & \cdots\cdots\cdots & x_n^d \end{bmatrix}.$$

The length of vector $[x^d]$ is $\binom{n+d-1}{d}$. The matrix P is called a Gram matrix of $p(x)$ and it is not unique if $n > 2$ and $d > 1$. For convenience, we index the columns and rows of P by monomials of degree d, or equivalently by n dimensional nonnegative integer vectors whose 1-norm is d. It can be shown [9, 10] that $p(x)$ is SOS if and only if it has a Gram matrix P which is positive semidefinite. Define constant symmetric matrices A_α such that

$$[x^d][x^d]^T = \sum_{\alpha \in \mathbb{N}(2d)} A_\alpha x^\alpha, \quad \text{where } \mathbb{N}(2d) = \{\alpha \in \mathbb{N}^n : |\alpha| = 2d\}. \tag{1.3}$$

Here for $\alpha = (\alpha_1, \ldots, \alpha_n)$, $|\alpha| = \alpha_1 + \cdots + \alpha_n$ and $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, and $\mathbb{N}$ is the set of nonnegative integers. If $p(x)$ is given as

$$p(x) = \sum_{\alpha \in \mathbb{N}(2d)} p_\alpha x^\alpha,$$

then $p(x)$ is SOS if and only if there exists a symmetric matrix X satisfying

$$\begin{array}{rcl} A_\alpha \bullet X & = & p_\alpha \quad \forall\, \alpha \in \mathbb{N}(2d), \\ X & \succeq & 0. \end{array}$$

In the above $X \succeq 0$ (resp., $X \succ 0$) means that X is positive semidefinite (resp., positive definite), and $\bullet$ denotes the standard Frobenius inner product in matrix spaces.

If we write $f(x)$ and $\|x\|_2^{2d}$ as

$$f(x) = \sum_{\alpha \in \mathbb{N}(2d)} f_\alpha x^\alpha, \quad \|x\|_2^{2d} = \sum_{\alpha \in \mathbb{N}(2d)} D_\alpha x^\alpha,$$

then SOS relaxation (1.2) is equivalent to the SDP problem

$$\begin{array}{rl} \max\limits_{\gamma, X} & \gamma \\ s.t. & A_\alpha \bullet X + D_\alpha \gamma = f_\alpha \quad \forall\, \alpha \in \mathbb{N}(2d), \\ & X \succeq 0. \end{array} \tag{1.4}$$

Problem (1.4) can be solved efficiently by numerical methods like interior point algorithms. SDP is a very nice convex optimization and has many attractive properties. There has been much work on designing efficient solvers for SDP and applying SDP in various settings like control and nonconvex optimization. We refer to [14] for more details about the theory, algorithms and applications of semidefinite programming.

Even though the lower bound f_{sos} given by (1.2) might match f_{min} in many situations, as demonstrated by numerical results in [5, 9, 10], we usually can not expect $f_{sos} = f_{min}$. For example, this is the case when $f(x)$ is the so-called *Motzkin* polynomial

$$Mot(x) := x_1^4 x_2^2 + x_1^2 x_2^4 + x_3^6 - 3x_1^2 x_2^2 x_3^2.$$

It is well known that $Mot(x)$ is nonnegative everywhere but not SOS [13]. Thus (1.2) would return a lower bound $f_{sos} < f_{min}$. Blekherman [1] proved a very surprising result: for any fixed even degree bigger than two, there are significantly more nonnegative polynomials than SOS polynomials. So generally we do not have $f_{sos} = f_{min}$. Therefore, it is very interesting to know how well f_{sos} approximates f_{min}. In (1.2), if $f(x) - \gamma\|x\|_2^{2d}$ is replaced by $\|x\|_2^{2N}(f(x) - \gamma\|x\|_2^{2d})$ for an integer N big enough, Faybusovich [2] gave an estimation on $f_{min} - f_{sos}$ based on a result of Reznick [13] regarding degree bounds of uniform denominators in Hilbert's 17th problem. But there is no estimation of $f_{min} - f_{sos}$ when $N = 0$. Generally, how does SOS relaxation (1.2) perform? How large is $f_{min} - f_{sos}$ in the worst case? To the best knowledge of the author, there is very little work on this issue. The motivation of this paper is to analyze the approximation performance of (1.2).

Contributions. First, we discuss the performance of SOS relaxation (1.2). Suppose $f(x)$ is an even form of degree $2d$. Let f_{max} be the maximum value of $f(x)$ on $\mathbb{S}^{n-1}$. Suppose $n \geq 2d$. Then we will show that the lower bound f_{sos} of f_{min} given by (1.2) satisfies

$$1 \leq \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \leq C(d)\sqrt{\binom{n}{2d}}. \tag{1.5}$$

The constant $C(d)$ is independent of n and can be evaluated numerically. Note the first inequality in (1.5) is obvious because $f_{sos} \leq f_{min}$. The second inequality in (1.5) means that f_{sos} is an $\mathcal{O}(n^d)$-approximation of f_{min}. This will be shown in Section 2.

Second, we discuss how to minimize multi-forms (all their terms have fixed degrees in the components of variables) over multi-unit spheres (cross products of lower dimensional unit spheres). This problem is an extension of the bi-quadratic optimization discussed in [6] and is also NP-hard. The SOS relaxation (1.2) can be generalized naturally. We will prove a similar approximation bound like (1.5). This will be presented in Section 3.

Third, SOS relaxation (1.2) might have better performance when $f(x)$ has special features. If $f(x)$ is a sparse form, we will prove an approximation bound better than (1.5), which depends on the sparsity pattern of $f(x)$. When $f(x)$ is an odd form, we can formulate (1.1) equivalently as minimizing a certain even form, and prove an approximation bound based on (1.2). This will be shown in Section 4.

Last, we consider the more general problem of minimizing $f(x)$ over a hypersurface $H(g) = \{x \in \mathbb{R}^n : g(x) = 1\}$, where $g(x)$ is a positive definite form. The SOS relaxation (1.2) can be generalized naturally, and we will prove a similar approximation bound like (1.5). This will be shown in Section 5. Some discussions about bounds will be made in Section 6.

Some notations. $\mathbb{N}$ (resp., $\mathbb{R}$) denotes the set of nonnegative integers (resp., real numbers). For any $t \in \mathbb{R}$, $\lceil t \rceil$ (resp., $\lfloor t \rfloor$) denotes the smallest integer not smaller (resp., the largest integer not bigger) than t. For any $k \in \mathbb{N}$, $[k] = \{1, \ldots, k\}$. The $\mathbb{N}(k)$ denotes the multi-index set $\{\alpha \in \mathbb{N}^n : |\alpha| = k\}$. For any $x \in \mathbb{R}^n$, x_i denotes the i-th component of x, that is, $x = (x_1, \ldots, x_n)$. For any $\alpha \in \mathbb{N}^n$, denote $|\alpha| = \alpha_1 + \cdots + \alpha_n$, and $\mathrm{supp}(\alpha) = \{i \in [n] : \alpha_i \neq 0\}$. For any $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{N}^n$, x^α denotes $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. The $\mathbb{R}[x]$ denotes the ring of real multivariate polynomials in $(x_1, \ldots, x_n)$, and $\mathbb{R}[x]_k$ denotes the subspace of forms of degree k. For nonnegative integers $k_1, \ldots, k_\ell$, denote $\mathbb{R}[x]_{k_1,\ldots,k_\ell} = \mathbb{R}[x]_{k_1} + \cdots + \mathbb{R}[x]_{k_\ell}$. For a polynomial $p(x)$, $\mathrm{supp}(p)$ denotes the support of $p(x)$, i.e., the set of $\alpha \in \mathbb{N}^n$ such that the monomial x^α appears in $p(x)$. For a finite set S, $|S|$ denotes its cardinality. For a matrix A, A^T denotes its transpose. For a symmetric matrix X, $\lambda_{max}(X)$ and $\lambda_{min}(X)$ denote the maximum and minimum eigenvalues of X respectively. For a symmetric matrix X, $X \succeq 0$ (resp., $X \succ 0$) means $\lambda_{min}(X) \geq 0$ (resp., $\lambda_{min}(X) > 0$). For two matrices A and B, $A \otimes B$ denotes the standard Kronecker product of A and B. For any vector $u \in \mathbb{R}^N$, $\|u\|_2 = \sqrt{u^T u}$ denotes the standard Euclidean norm; For matrix A, $\|A\|_2$ denotes the maximum singular value of A, and $\|A\|_F$ denotes the Frobenius norm of A, i.e., $\|A\|_2 = \sqrt{Trace(A^T A)}$.

2 Minimizing general forms

This section analyzes the approximation performance of SOS relaxation (1.2). The basic technique is to estimate the L^2-norm and G-norm of forms. We begin with some definitions of norms.

2.1. Norms of forms

For a form $f(x)$ of degree k given as

$$f(x) = \sum_{\alpha \in \mathbb{N}(k)} f_\alpha x^\alpha,$$

we define its *G-norm* as

$$\|f(x)\|_G = \left(\sum_{\alpha \in \mathbb{N}(k)} \mathfrak{p}(\alpha)^{-1} f_\alpha^2 \right)^{1/2}. \tag{2.1}$$

Here $\mathfrak{p}(\alpha)$ denotes the partition number of the exponent α, that is,

$$\mathfrak{p}(\alpha) = \left| \left\{ (\beta, \eta) \in \mathbb{N}(\lceil k/2 \rceil) \times \mathbb{N}(\lfloor k/2 \rfloor) : \beta + \eta = \alpha \right\} \right|. \tag{2.2}$$

In view of (2.1), denote by f_G the column vector of weighted coefficients of $f(x)$

$$f_G = \left(\mathfrak{p}(\alpha)^{-1/2} f_\alpha : \alpha \in \mathbb{N}(k) \right), \tag{2.3}$$

and denote by $[x^k]_G$ the column vector of weighted monomials

$$[x^k]_G = \left(\mathfrak{p}(\alpha)^{1/2} x^\alpha : \alpha \in \mathbb{N}(k) \right). \tag{2.4}$$

The entries in f_G and $[x^k]_G$ are ordered lexicographically according to their indices. Thus $f(x) = f_G^T [x^k]_G$ and $\|f(x)\|_G = \|f_G\|_2$. The reason that we call this norm G-norm is the close relationship between $\|\cdot\|_G$ and Gram matrices.

Lemma 2.1. *If a form $f(x)$ has degree $2d$, there exists a symmetric W such that*

$$f(x) = [x^d]^T W [x^d], \quad \|W\|_F = \|f(x)\|_G.$$

Proof. For any matrix W satisfying $f(x) = [x^d]^T W [x^d]$, the following holds

$$f_\alpha = \sum_{(\beta,\eta) \in \mathbb{N}(d) \times \mathbb{N}(d) : \beta + \eta = \alpha} W_{\beta,\eta} \quad \forall\, \alpha \in \mathbb{N}(2d).$$

Now we choose W as follows

$$W(\beta, \eta) = \mathfrak{p}(\alpha)^{-1} f_\alpha \quad \forall\, (\beta, \eta) \in \mathbb{N}(d) \times \mathbb{N}(d) : \beta + \eta = \alpha.$$

The above W is a symmetric matrix. Its Frobenius norm is

$$\|W\|_F^2 = \sum_{\alpha \in \mathbb{N}(2d)} \sum_{\substack{(\beta,\eta) \in \mathbb{N}(d) \times \mathbb{N}(d) \\ \beta + \eta = \alpha}} (\mathfrak{p}(\alpha)^{-1} f_\alpha)^2 = \sum_{\alpha \in \mathbb{N}(2d)} (\mathfrak{p}(\alpha)^{-1} f_\alpha)^2 \mathfrak{p}(\alpha) = \|f(x)\|_G^2.$$

So the lemma is proved. □

Useful in our approximation analysis are the L^2 type norms. Define

$$\|f(x)\|_{L^2} = \left(\int_{\mathbb{S}^{n-1}} f(x)^2 d\mu(x) \right)^{1/2}. \tag{2.5}$$

Here μ is the uniform probability measure on $\mathbb{S}^{n-1}$. We also need to define a so-called *marginal* L^2-norm. Throughout out this section, assume $n \geq k$. Given a subset $\Delta \subset \{1, \ldots, n\}$ with $|\Delta| = k \leq n$, denote by x_Δ the subvector of x whose indices are in Δ, that is,

$$x_\Delta = (x_i : i \in \Delta).$$

For $f(x) \in \mathbb{R}[x]_k$, denote by $f_\Delta(x_\Delta)$ the restriction of $f(x)$ to x_Δ, that is,

$$f_\Delta(x_\Delta) = f(\tilde{x}), \quad \text{where} \quad \tilde{x}_i = \begin{cases} x_i & \text{if } i \in \Delta, \\ 0 & \text{otherwise.} \end{cases}$$

So $f_\Delta(x_\Delta)$ is a polynomial only in x_Δ. Denote the set

$$\Omega_k = \{\Delta \subset [n] : |\Delta| = k\}. \tag{2.6}$$

Clearly, its cardinality $|\Omega_k| = \binom{n}{k}$. The marginal L^2-norm of $f(x)$ is then defined as

$$\|f(x)\|_{L^2,mg} = \left(\sum_{\Delta \in \Omega_k} \|f_\Delta(x_\Delta)\|_{L^2}^2 \right)^{1/2}. \tag{2.7}$$

The name "marginal" comes from the observation that the $k-1$ dimensional unit sphere $\{x_\Delta : \|x_\Delta\|_2 = 1\}$ is a sub-sphere of $\mathbb{S}^{n-1}$ when we restrict $x_i = 0$ for all $i \notin \Delta$.

For our purpose of approximation analysis, we need to define the constant matrix

$$\mathbf{\Theta}_k = \int_{\|x_\Delta\|_2=1} [x_\Delta^k]_G [x_\Delta^k]_G^T d\mu_\Delta(x_\Delta), \quad \Delta \in \Omega_k. \tag{2.8}$$

Here $\mu_\Delta(x_\Delta)$ is the uniform probability measure on $\mathbb{S}^{k-1}$. For instance,

$$\mathbf{\Theta}_2 = \frac{1}{8} \begin{bmatrix} 3 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}. \tag{2.9}$$

Note that $\mathbf{\Theta}_k$ is independent of the choice of $\Delta \in \Omega_k$, because the monomials of $[x_\Delta^k]_G$

k	2	4	6	8
δ_k	0.5000	0.0559	0.0039	0.0002

Table 1: A list of the constants δ_k.

are ordered lexicographically and the integrals are independent of Δ. The matrix $\mathbf{\Theta}_k$ is positive definite, because the monomials of $[x_\Delta^k]_G$ are linearly independent. Define the positive constant

$$\delta_k = \sqrt{\lambda_{min}(\mathbf{\Theta}_k)} > 0. \tag{2.10}$$

Note that δ_k is independent of n. A list of typical values of δ_k for even k (we are only interested in even k later) is in Table 1. The constant δ_k relates the marginal L^2-norm and G-norm as follows.

Lemma 2.2. *If $f(x) \in \mathbb{R}[x]_k$, then $\|f(x)\|_{L^2,mg} \geq \delta_k \|f(x)\|_G$.*

Proof. By definitions of L^2-norm and δ_k, we know

$$\|f_\Delta(x_\Delta)\|_{L^2}^2 = f_{\Delta,G}^T \mathbf{\Theta}_k f_{\Delta,G} \geq \delta_k^2 \|f_\Delta(x_\Delta)\|_G^2.$$

Here $f_{\Delta,G}$ denotes the vector of weighted coefficients of polynomial $f_\Delta(x_\Delta)$ (see (2.3)). By definition of the marginal L^2-norm, it holds

$$\|f(x)\|_{L^2,mg}^2 = \sum_{\Delta \in \Omega_k} \|f_\Delta(x_\Delta)\|_{L^2}^2 \geq \delta_k^2 \sum_{\Delta \in \Omega_k} \|f_\Delta(x_\Delta)\|_G^2 \geq \delta_k^2 \|f(x)\|_G^2.$$

Taking the square root of the above results in the lemma. □

The marginal L^2-norm of forms can be estimated as follows.

Lemma 2.3. *Suppose $f(x) \in \mathbb{R}[x]_k$ and $k \leq n$. If $|f(x)| \leq 1$ for all $x \in \mathbb{S}^{n-1}$, then*

$$\|f(x)\|_{L^2,mg} \leq \sqrt{\binom{n}{k}}.$$

Proof. For every $\Delta \in \Omega_k$, the condition that $|f(x)| \leq 1$ for all $x \in \mathbb{S}^{n-1}$ implies $|f_\Delta(x_\Delta)| \leq 1$ for all $x_\Delta \in \mathbb{S}^{k-1}$. By definition of the marginal L^2-norm, we get

$$\|f(x)\|_{L^2,mg}^2 = \sum_{\Delta \in \Omega_k} \int_{\mathbb{S}^{k-1}} f_\Delta(x_\Delta)^2 d\mu_\Delta(x_\Delta) \leq \sum_{\Delta \in \Omega_k} \mu_\Delta(\mathbb{S}^{k-1}) = \binom{n}{k},$$

where the last step is because μ_Δ is the uniform probability measure on $\mathbb{S}^{k-1}$. □

2.2. Bound analysis

Now we analyze the performance of SOS relaxation (1.2). The basic technique is to estimate the marginal L^2 and G norms by applying Lemmas 2.2 and 2.3.

Theorem 2.4. *Let $f(x)$ be a form of degree $2d$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) value on the unit sphere $\mathbb{S}^{n-1}$. Suppose $n \geq 2d$. If f_{sos} is the lower bound given by SOS relaxation (1.2), then it holds*

$$1 \leq \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \leq \frac{1}{\delta_{2d}} \sqrt{\binom{n}{2d}},$$

where δ_{2d} is defined in (2.10).

Proof. Let $f_{med} = \frac{1}{2}(f_{min} + f_{max})$ and $\tilde{f}(x) = f(x) - f_{med} \cdot \|x\|_2^{2d}$. Then we have

$$\left| \frac{1}{f_{med} - f_{min}} \tilde{f}(x) \right| \leq 1 \quad \forall x \in \mathbb{S}^{n-1}.$$

By Lemma 2.3, we know

$$\left\| \frac{1}{f_{med} - f_{min}} \tilde{f}(x) \right\|_{L^2,mg} \leq \sqrt{\binom{n}{2d}}. \tag{2.11}$$

Now fix a constant

$$\gamma^* = f_{med} - (f_{med} - f_{min}) \cdot \frac{1}{\delta_{2d}} \sqrt{\binom{n}{2d}}. \tag{2.12}$$

Then the inequality (2.11) implies

$$\left\| \frac{1}{f_{med} - \gamma^*} \tilde{f}(x) \right\|_{L^2,mg} \leq \delta_{2d}.$$

By Lemma 2.2, the above then implies

$$\left\| \frac{1}{f_{med} - \gamma^*} \tilde{f}(x) \right\|_G \leq \delta_{2d}^{-1} \left\| \frac{1}{f_{med} - \gamma^*} \tilde{f}(x) \right\|_{L^2, mg} \leq 1. \tag{2.13}$$

Thus, by Lemma 2.1, there exists a symmetric matrix W such that

$$\frac{1}{f_{med} - \gamma^*} \tilde{f}(x) = [x^d]^T W [x^d], \quad \|W\|_F \leq 1.$$

Let D be the diagonal matrix such that $\|x\|_2^{2d} = [x^d]^T D [x^d]$. Note $\lambda_{min}(D) \geq 1$ and

$$\frac{1}{f_{med} - \gamma^*} \tilde{f}(x) + \|x\|_2^{2d} = [x^d]^T (W + D)[x^d].$$

Since $\|W\|_2 \leq \|W\|_F \leq 1$, we know $W + D \succeq 0$. Hence the form

$$\frac{1}{f_{med} - \gamma^*} \tilde{f}(x) + \|x\|_2^{2d}$$

must be SOS, or equivalently, the form $f(x) - \gamma^* \|x\|_2^{2d}$ is SOS. Since f_{sos} is the optimal value of (1.2), we have $f_{sos} \geq \gamma^*$. By the choice of γ^* in (2.12), the following holds

$$1 \leq \frac{f_{med} - f_{sos}}{f_{med} - f_{min}} \leq \frac{1}{\delta_{2d}} \sqrt{\binom{n}{2d}}.$$

Since $f_{\min} \leq f_{med} \leq f_{max}$, the above immediately implies the theorem. □

3 Minimizing multi-forms over multi-spheres

This section studies the problem of optimizing multi-forms over multi-unit spheres. We first generalize SOS relaxation (1.2) and then analyze its approximation performance.

Suppose $x = (x_{I_1}, \ldots, x_{I_m})$ is partitioned such that every component x_{I_k} is n_k-dimensional and $n_1 + \cdots + n_m = n$. A form $f(x)$ is said to be a multi-form if all its terms have fixed degrees in each component x_{I_k}. We say $f(x)$ is a $(n_1, \ldots, n_m) \times (r_1, \ldots, r_m)$-form if

$$f(x) = \sum_{\substack{\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{N}^{n_1} \times \cdots \times \mathbb{N}^{n_m} \\ |\alpha_1| = r_1, \ldots, |\alpha_m| = r_m}} f_\alpha \cdot (x_{I_1})^{\alpha_1} \cdots (x_{I_m})^{\alpha_m}. \tag{3.1}$$

Here every $(x_{I_k})^{\alpha_k}$ is defined as before. Consider the optimization problem

$$\begin{array}{rl} \min\limits_{x = (x_{I_1}, \ldots, x_{I_m})} & f(x) \\ s.t. & \|x_{I_1}\|_2 = \cdots = \|x_{I_m}\|_2 = 1, \end{array} \tag{3.2}$$

where $f(x)$ is a $(n_1, \ldots, n_m) \times (r_1, \ldots, r_m)$-form. When $m = 1$, (3.2) reduces to (1.1); when $m = 2$ and $r_1 = r_2 = 2$, (3.2) reduces to the so-called bi-quadratic optimization which was studied by Ling, Nie, Qi and Ye [6]. It was shown in [6] that the bi-quadratic optimization is already NP-hard. Thus, the more general problem (3.2) is also NP-hard. If every $r_k = 2d_k$ is even, a natural generalization of SOS relaxation (1.2) is

$$\begin{array}{rl} \max & \gamma \\ s.t. & f(x) - \gamma \cdot \|x_{I_1}\|_2^{2d_1} \cdots \|x_{I_m}\|_2^{2d_m} \text{ is SOS.} \end{array} \tag{3.3}$$

Like (1.2), the relaxation (3.3) is equivalent to an SDP problem.

Define the index set

$$\mathbb{N}^{n_1,\ldots,n_m}_{r_1,\ldots,r_m} = \Big\{\alpha = (\alpha_1,\ldots,\alpha_m) \in \mathbb{N}^{n_1} \times \cdots \times \mathbb{N}^{n_m} : |\alpha_1| = r_1,\ldots,|\alpha_m| = r_m\Big\}.$$

For every $\alpha \in \mathbb{N}^{n_1,\ldots,n_m}_{r_1,\ldots,r_m}$, denote $x^\alpha = (x_{I_1})^{\alpha_1} \cdots (x_{I_m})^{\alpha_m}$. Define the multi-unit sphere

$$\mathbb{S}^{n_1-1,\ldots,n_m-1} = \mathbb{S}^{n_1-1} \times \cdots \times \mathbb{S}^{n_m-1}.$$

Thus $(x_{I_1},\ldots,x_{I_m}) \in \mathbb{S}^{n_1-1,\ldots,n_m-1}$ if and only if every $x_{I_k} \in \mathbb{S}^{n_k-1}$. Let

$$\mathcal{F}^{n_1,\ldots,n_m}_{r_1,\ldots,r_m} = \big\{ f(x) \text{ is a multi-form given by (3.1) } \big\}$$

be a space of multi-forms. For convenience, f_{min} (resp., f_{max}) still denotes the minimum (resp., maximum) value of $f(x)$ on $\mathbb{S}^{n_1-1,\ldots,n_m-1}$, and f_{sos} denotes the optimal value of (3.3).

3.1. Norms of multi-forms

For a multi-form $f(x) \in \mathcal{F}^{n_1,\ldots,n_m}_{r_1,\ldots,r_m}$ given by (3.1), we define its G-norm as

$$\|f(x)\|_G = \left(\sum_{\alpha \in \mathbb{N}^{n_1,\ldots,n_m}_{r_1,\ldots,r_m}} \mathfrak{p}(\alpha)^{-1} f_\alpha^2 \right)^{1/2}. \tag{3.4}$$

In the above, for every $\alpha = (\alpha_1,\ldots,\alpha_m) \in \mathbb{N}^{n_1,\ldots,n_m}_{r_1,\ldots,r_m}$, the partition number $\mathfrak{p}(\alpha)$ is defined to be $\mathfrak{p}(\alpha_1)\cdots\mathfrak{p}(\alpha_m)$, where each individual $\mathfrak{p}(\alpha_k)$ is defined by (2.2). Note $\mathfrak{p}(\alpha)$ is precisely the cardinality of the set

$$\Big\{(\eta,\nu) \in \mathbb{N}^{n_1,\ldots,n_m}_{\lfloor r_1/2\rfloor,\ldots,\lfloor r_m/2\rfloor} \times \mathbb{N}^{n_1,\ldots,n_m}_{\lceil r_1/2\rceil,\ldots,\lceil r_m/2\rceil} : \eta + \nu = \alpha\Big\}.$$

For $f(x) \in \mathcal{F}^{n_1,\ldots,n_m}_{r_1,\ldots,r_m}$, denote

$$f_G = \Big((\mathfrak{p}(\alpha))^{-1/2} f_\alpha : \alpha \in \mathbb{N}^{n_1,\ldots,n_m}_{r_1,\ldots,r_m} \Big), \tag{3.5}$$

$$[x^{r_1,\ldots,r_m}]_G = \Big(\sqrt{\mathfrak{p}(\alpha)} x^\alpha : \alpha \in \mathbb{N}^{n_1,\ldots,n_m}_{r_1,\ldots,r_m} \Big). \tag{3.6}$$

The components of f_G and $[x^{r_1,\ldots,r_m}]_G$ are ordered lexicographically according to their indices. So $f(x) = f_G^T [x^{r_1,\ldots,r_m}]_G$ and $\|f(x)\|_G = \|f_G\|_2$.

Lemma 3.1. *If $f(x) \in \mathcal{F}^{n_1,\ldots,n_m}_{2d_1,\ldots,2d_m}$, then there exists a symmetric matrix W such that*

$$f(x) = [x^{d_1,\ldots,d_m}]^T W [x^{d_1,\ldots,d_m}], \quad \|W\|_F = \|f(x)\|_G.$$

Lemma 3.1 is a natural generalization of Lemma 2.1, and can be proved in almost the same way. So its proof is omitted here.

Similar to general forms, the L^2-norm of $f(x) \in \mathcal{F}^{n_1,\ldots,n_m}_{r_1,\ldots,r_m}$ is defined as

$$\|f(x)\|_{L^2} = \left(\int_{\mathbb{S}^{n_1-1}} \cdots \int_{\mathbb{S}^{n_m-1}} f(x)^2 d\mu_1(x_{I_1}) \cdots d\mu_m(x_{I_m}) \right)^{1/2}. \tag{3.7}$$

Here every $\mu_k(\cdot)$ is the uniform probability measure on $\mathbb{S}^{n_k-1}$. Throughout this section, we always assume $n_i \geq r_i$ for every i. Then the marginal L^2-norm of $f(x)$ can be defined in a similar way as in Section 2. For this purpose, denote

$$\Omega^{n_1,\ldots,n_m}_{r_1,\ldots,r_m} = \Big\{ (\Delta_1, \ldots, \Delta_m) \subset [n_1] \times \cdots \times [n_m] : |\Delta_1| = r_1, \ldots, |\Delta_m| = r_m \Big\}. \tag{3.8}$$

Clearly, $|\Omega^{n_1,\ldots,n_m}_{r_1,\ldots,r_m}| = \binom{n_1}{r_1} \cdots \binom{n_m}{r_m}$. For $\Delta = (\Delta_1, \ldots, \Delta_m) \in \Omega^{n_1,\ldots,n_m}_{r_1,\ldots,r_m}$, $f_\Delta(x_\Delta)$ denotes the restriction of $f(x)$ to

$$x_\Delta = \big((x_{I_1})_{\Delta_1}, \ldots, (x_{I_m})_{\Delta_m} \big).$$

The L^2-norm of $f_\Delta(x_\Delta)$ is defined similarly as in (3.7) by replacing every n_k by r_k. Like general forms, the marginal L^2-norm of $f(x) \in \mathcal{F}^{n_1,\ldots,n_m}_{r_1,\ldots,r_m}$ is then defined as

$$\|f(x)\|_{L^2,mg} = \left(\sum_{\Delta \in \Omega^{n_1,\ldots,n_m}_{r_1,\ldots,r_m}} \|f_\Delta(x_\Delta)\|^2_{L^2} \right)^{1/2}. \tag{3.9}$$

Denote the monomial vector

$$[x^{r_1,\ldots,r_m}_{\Delta_{r_1,\ldots,r_m}}]_G = \left(\sqrt{\mathfrak{p}(\alpha)} x^\alpha : \begin{array}{l} \alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{N}^{n_1,\ldots,n_m}_{r_1,\ldots,r_m} \\ \operatorname{supp}(\alpha_1) \subset \Delta(r_1), \ldots, \operatorname{supp}(\alpha_m) \subset \Delta(r_m) \end{array} \right), \tag{3.10}$$

where $\Delta_{r_1,\ldots,r_m} = (\Delta(r_1), \ldots, \Delta(r_m))$ and every $\Delta(r_k) = [r_k]$. Then define the matrix

$$\mathbf{M}^{r_1,\ldots,r_m} = \int_{\mathbb{S}^{r_1-1}} \cdots \int_{\mathbb{S}^{r_m-1}} [x^{r_1,\ldots,r_m}_{\Delta_{r_1,\ldots,r_m}}]_G [x^{r_1,\ldots,r_m}_{\Delta_{r_1,\ldots,r_m}}]^T_G d\mu_{\Delta(r_1)}(x_{\Delta(r_1)}) \cdots d\mu_{\Delta(r_m)}(x_{\Delta(r_m)}).$$

Here every $\mu_{\Delta(r_k)}(\cdot)$ is the uniform probability measure on $\mathbb{S}^{r_k-1}$. Since the monomials of $[x^{r_1,\ldots,r_m}_{\Delta_{r_1,\ldots,r_m}}]_G$ are linearly independent, $\mathbf{M}^{r_1,\ldots,r_m}$ is positive definite. Define the constant

$$\delta_{r_1,\ldots,r_m} = \sqrt{\lambda_{min}(\mathbf{M}^{r_1,\ldots,r_m})} > 0. \tag{3.11}$$

Lemma 3.2. *If $f(x) \in \mathcal{F}^{n_1,\ldots,n_m}_{r_1,\ldots,r_m}$, then $\|f(x)\|_{L^2,mg} \geq \delta_{r_1,\ldots,r_m} \|f(x)\|_G$.*

Proof. By definition of L^2-norm, we know for every $\Delta \in \Omega^{n_1,\ldots,n_m}_{r_1,\ldots,r_m}$

$$\|f_\Delta(x_\Delta)\|^2_{L^2} = f^T_{\Delta,G} B_\Delta f_{\Delta,G},$$

where B_Δ is the following symmetric matrix

$$B_\Delta = \int_{\mathbb{S}^{r_1-1}} \cdots \int_{\mathbb{S}^{r_m-1}} [x^{r_1,\ldots,r_m}_\Delta]_G [x^{r_1,\ldots,r_m}_\Delta]^T_G d\mu_{\Delta_1}(x_{\Delta_1}) \cdots d\mu_{\Delta_1}(x_{\Delta_1}).$$

Note that $B_\Delta = \mathbf{M}^{r_1,\dots,r_m}$. So we have

$$\|f_\Delta(x_\Delta)\|_{L^2}^2 = f_{\Delta,G}^T \mathbf{M}^{r_1,\dots,r_m} f_{\Delta,G} \geq \delta_{r_1,\dots,r_m}^2 \|f_\Delta(x_\Delta)\|_G^2.$$

Here $f_{\Delta,G}$ denotes the vector of weighted coefficients of $f_\Delta(x_\Delta)$ (see (3.5)). Therefore, by definition of the marginal L^2-norm (3.9), the following holds

$$\|f(x)\|_{L^2,mg}^2 = \sum_{\Delta \in \Omega_{r_1,\dots,r_m}^{n_1,\dots,n_m}} \|f_\Delta(x_\Delta)\|_{L^2}^2 \geq \delta_{r_1,\dots,r_m}^2 \sum_{\Delta \in \Omega_{r_1,\dots,r_m}^{n_1,\dots,n_m}} \|f_\Delta(x_\Delta)\|_G^2 \geq \delta_{r_1,\dots,r_m}^2 \|f(x)\|_G^2.$$

So the lemma is proved. □

Lemma 3.3. *If $f(x) \in \mathcal{F}_{r_1,\dots,r_m}^{n_1,\dots,n_m}$ and $|f(x)| \leq 1$ for all $x \in \mathbb{S}^{n_1-1,\dots,n_m-1}$, then*

$$\|f(x)\|_{L^2,mg} \leq \sqrt{\binom{n_1}{r_1} \cdots \binom{n_m}{r_m}}.$$

Proof. For every $\Delta \in \Omega_{r_1,\dots,r_m}^{n_1,\dots,n_m}$, we have $|f_\Delta(x_\Delta)| \leq 1$ for all $x_\Delta \in \mathbb{S}^{r_1-1,\dots,r_m-1}$. Then, by definition of the marginal L^2-norm in (3.9), the following holds

$$\begin{aligned}
\|f(x)\|_{L^2,mg}^2 &= \sum_{\Delta \in \Omega_{r_1,\dots,r_m}^{n_1,\dots,n_m}} \int_{\mathbb{S}^{n_1-1}} \cdots \int_{\mathbb{S}^{n_m-1}} f_\Delta(x_\Delta)^2 d\mu_{\Delta_1}\big((x_{I_1})_{\Delta_1}\big) \cdots d\mu_{\Delta_m}\big((x_{I_m})_{\Delta_m}\big) \\
&\leq \sum_{\Delta \in \Omega_{r_1,\dots,r_m}^{n_1,\dots,n_m}} 1 = \binom{n_1}{r_1} \cdots \binom{n_m}{r_m}.
\end{aligned}$$

The lemma is proved. □

3.2. Bound analysis

Now we analyze the performance of SOS relaxation (3.3). An approximation bound can be obtained by generalizing the techniques used in the proof of Theorem 2.4.

Theorem 3.4. *Let $f(x) \in \mathcal{F}_{2d_1,\dots,2d_m}^{n_1,\dots,n_m}$ be a multi-form, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) value on the multi-unit sphere $\mathbb{S}^{n_1-1,\dots,n_m-1}$. Suppose $n_i \geq 2d_i$ for every i. If f_{sos} is the optimal value of SOS relaxation (3.3), then*

$$1 \leq \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \leq \frac{1}{\delta_{2d_1,\dots,2d_m}} \sqrt{\binom{n_1}{2d_1} \cdots \binom{n_m}{2d_m}},$$

where $\delta_{2d_1,\dots,2d_m}$ is defined by (3.11).

Proof. The proof is very similar to what we have done in proving Theorem 2.4. Set

$$f_{med} = \frac{1}{2}(f_{min} + f_{max}), \quad \tilde{f}(x) = f(x) - f_{med} \cdot \|x_{I_1}\|_2^{2d_1} \cdots \|x_{I_m}\|_2^{2d_m}.$$

Then the following holds

$$\left| \frac{1}{f_{med} - f_{min}} \tilde{f}(x) \right| \leq 1 \quad \forall\, x \in \mathbb{S}^{n_1-1,\dots,n_m-1}.$$

By Lemma 3.3, we know

$$\left\| \frac{1}{f_{med} - f_{min}} \tilde{f}(x) \right\|_{L^2,mg} \leq \sqrt{\binom{n_1}{2d_1} \cdots \binom{n_m}{2d_m}}.$$

Fix a constant

$$\tau^* = f_{med} - (f_{med} - f_{min}) \cdot \frac{1}{\delta_{2d_1,\ldots,2d_m}} \sqrt{\binom{n_1}{2d_1} \cdots \binom{n_m}{2d_m}}. \tag{3.12}$$

The above then implies

$$\left\| \frac{1}{f_{med} - \tau^*} \tilde{f}(x) \right\|_{L^2,mg} \leq \delta_{2d_1,\ldots,2d_m}.$$

By Lemma 3.2, we get

$$\left\| \frac{1}{f_{med} - \tau^*} \tilde{f}(x) \right\|_{G} \leq \frac{1}{\delta_{2d_1,\ldots,2d_m}} \left\| \frac{1}{f_{med} - \tau^*} \tilde{f}(x) \right\|_{L^2,mg} \leq 1.$$

By Lemma 3.1, there exists a symmetric matrix W such that

$$\frac{1}{f_{med} - \tau^*} \tilde{f}(x) = [x^{d_1,\ldots,d_m}]^T W [x^{d_1,\ldots,d_m}], \quad \|W\|_F \leq 1.$$

Let D be the diagonal matrix satisfying

$$\|x_{I_1}\|_2^{2d_1} \cdots \|x_{I_m}\|_2^{2d_m} = [x^{d_1,\ldots,d_m}]^T D [x^{d_1,\ldots,d_m}].$$

Then we get

$$\frac{1}{f_{med} - \tau^*} \tilde{f}(x) + \|x_{I_1}\|_2^{2d_1} \cdots \|x_{I_m}\|_2^{2d_m} = [x^{d_1,\ldots,d_m}]^T (W + D) [x^{d_1,\ldots,d_m}].$$

Since $\lambda_{min}(D) \geq 1$ and $\|W\|_2 \leq \|W\|_F \leq 1$, we know $W + D \succeq 0$. Hence

$$\frac{1}{f_{med} - \tau^*} \tilde{f}(x) + \|x_{I_1}\|_2^{2d_1} \cdots \|x_{I_m}\|_2^{2d_m}$$

must be SOS, or equivalently, the multi-form

$$f(x) - \tau^* \|x_{I_1}\|_2^{2d_1} \cdots \|x_{I_m}\|_2^{2d_m}$$

is SOS. Since f_{sos} is the optimal value of (3.3), $f_{sos} \geq \tau^*$, and then (3.12) implies

$$1 \leq \frac{f_{med} - f_{sos}}{f_{med} - f_{min}} \leq \frac{1}{\delta_{2d_1,\ldots,2d_m}} \sqrt{\binom{n_1}{2d_1} \cdots \binom{n_m}{2d_m}}.$$

Since $f_{min} \leq f_{med} \leq f_{max}$, the theorem follows. $\square$

The constant $\delta_{2d_1,\dots,2d_m}$ is independent of $(n_1,\dots,n_n)$. Now we estimate it. Note that

$$\mathbf{M}^{2d_1,2d_2,\dots,2d_m} = \mathbf{\Theta}_{2d_1} \otimes \mathbf{\Theta}_{2d_2} \otimes \cdots \otimes \mathbf{\Theta}_{2d_m}.$$

Here $\otimes$ denotes the standard Kronecker product, and each $\mathbf{\Theta}_{2d_i}$ is defined by (2.8). Since the eigenvalues of $A_1 \otimes A_2 \otimes \cdots \otimes A_m$ are products of the eigenvalues of A_i, we have

$$\delta_{2d_1,2d_2,\dots,2d_m} = \delta_{2d_1}\delta_{2d_2}\cdots\delta_{2d_m}.$$

In the special case of bi-quadratic optimization, that is, $m = 2$ and $d_1 = d_2 = 1$, the constant $\delta_{2d_1,\dots,2d_m}$ can be found explicitly. This leads to the following corollary.

Corollary 3.5. *Let $m = 2$ and $d_1 = d_2 = 1$. Suppose $n_1 \geq 2$ and $n_2 \geq 2$. If $f(x) \in \mathcal{F}_{2,2}^{n_1,n_2}$ is a bi-quadratic form, then the optimal value f_{sos} of (3.3) satisfies*

$$1 \leq \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \leq 4\sqrt{\binom{n_1}{2}\binom{n_2}{2}}.$$

Proof. When $m = 2$ and $d_1 = d_2 = 1$, $\mathbf{M}^{2,2} = \mathbf{\Theta}_2 \otimes \mathbf{\Theta}_2$ where $\mathbf{\Theta}_2$ is given in (2.9). Since $\mathbf{\Theta}_2$ has eigenvalues $\frac{1}{4}, \frac{1}{4}, \frac{1}{2}$, we get $\delta_{2,2} = \frac{1}{4}$. Then the corollary follows Theorem 3.4. $\square$

4 Sparse and odd forms

The previous sections analyze the approximation performance of SOS relaxations (1.2) and (3.3). When the forms to be optimized have special features, do they have better performance? This section discusses this issue.

4.1. Sparse forms

In many applications, the forms to be optimized are often sparse. For computational efficiency, it is important to exploit their sparsity patterns. There has been much work in this area, and we refer to [4, 8, 11]. For sparse forms, we can certainly apply (1.2) to get a lower bound, and its quality is estimated by Theorem 2.4. However, the approximation bound in Theorem 2.4 would be improved when $f(x)$ is sparse.

Denote $\mathbb{R}[x]_{0,k} = \mathbb{R}[x]_0 + \mathbb{R}[x]_k$. For $p(x) \in \mathbb{R}[x]_{0,k}$, we can write $p(x) = a + q(x)$ with $a \in \mathbb{R}$ and $q(x) \in \mathbb{R}[x]_k$. Then the G-norm of $p(x)$ is naturally defined as

$$\|p\|_G = \sqrt{a^2 + \|q\|_G^2}.$$

Since a nonzero $p(x) \in \mathbb{R}[x]_{0,k}$ might vanish on the unit sphere, we define its L^2-norm as

$$\|p(x)\|_{L_B^2} = \left(\int_{\|x\|_2 \leq 1} p(x)^2 d\nu(x)\right)^{1/2}.$$

Here ν is now the uniform probability measure on the unit ball $B(0,1) = \{x : \|x\|_2 \leq 1\}$.

For $p(x) \in \mathbb{R}[x]_{0,k}$ and $\Phi \subseteq \Omega_k$, we say Φ is a *cover* of $p(x)$ if for every $\alpha \in \text{supp}(p)$, there is a $\Delta \in \Phi$ such that $\text{supp}(\alpha) \subseteq \Delta$. Denote by $\Omega(p)$ the smallest cover of $p(x)$:

$$\Omega(p) = \underset{\Phi \in \Omega_k}{argmin} \Big\{ |\Phi| : \ \Phi \text{ is a cover of } p(x) \Big\}. \tag{4.1}$$

The cardinality $|\Omega(p)|$ is called the length of $p(x)$. Let $p_\Delta(x_\Delta)$ be the restriction of $p(x)$ to x_Δ. We similarly define

$$\|p_\Delta(x_\Delta)\|_{L_B^2}^2 = \left(\int_{\|x_\Delta\|_2 \le 1} p_\Delta(x_\Delta)^2 d\nu_\Delta(x_\Delta)\right)^{1/2}.$$

The above ν_Δ denotes the uniform probability measure on the sub-unit ball $B_\Delta(0,1) = \{x_\Delta : \|x\|_2 \le 1\}$. For $p(x) \in \mathbb{R}[x]_{0,k}$, its sparse marginal L^2-norm is naturally defined as

$$\|p(x)\|_{L_B^2,\Omega(p)} = \left(\sum_{\Delta\in\Omega(p)} \|p_\Delta(x_\Delta)\|_{L_B^2}^2\right)^{1/2}.$$

As before, we denote by p_{max} (resp., p_{min}) the maximum (resp., minimum) value of $p(x)$ on $\mathbb{S}^{n-1}$. Then we define matrix

$$\mathbf{B}_k = \int_{\|x_\Delta\|_2 \le 1} \begin{bmatrix} 1 \\ [x_\Delta^k]_G \end{bmatrix} \begin{bmatrix} 1 \\ [x_\Delta^k]_G \end{bmatrix}^T d\mu_\Delta(x_\Delta), \quad \Delta \in \Omega_k.$$

Note that $\mathbf{B}_k$ is independent of the choice $\Delta \in \Omega_k$ and $\mathbf{B}_k \succ 0$. Set

$$\zeta_k = \sqrt{\lambda_{min}(\mathbf{B}_k)} > 0. \tag{4.2}$$

The relation between the sparse marginal L^2-norm and G-norm is summarized as follows.

Lemma 4.1. *Let $p(x) \in \mathbb{R}[x]_{0,k}$ and $\Omega(p)$ be its smallest cover.*
(i) If $|p(x)| \le 1$ for all $x \in \mathbb{S}^{n-1}$, then $\|p(x)\|_{L_B^2,\Omega(p)} \le \sqrt{|\Omega(p)|}$.
(ii) It always holds that $\|p(x)\|_{L_B^2,\Omega(p)} \ge \zeta_k \|p(x)\|_G$.

Proof. (i) For every $\Delta \in \Omega_k$, we have $|p_\Delta(x_\Delta)| \le 1$ for all $x_\Delta \in \mathbb{S}^{n-1}$, hence

$$\|p_\Delta(x_\Delta)\|_{L_B^2}^2 = \int_{\|x_\Delta\|_2 \le 1} p_\Delta(x_\Delta)^2 d\mu_\Delta(x_\Delta) \le 1.$$

By definition of the sparse marginal L^2-norm, we get

$$\|p(x)\|_{L_B^2,\Omega(p)} = \sqrt{\sum_{\Delta\in\Omega(p)} \|p_\Delta(x_\Delta)\|_{L_B^2}^2} \le \sqrt{|\Omega(p)|}.$$

(ii) For every $\Delta \in \Omega_k$, $p_\Delta(x_\Delta) = a + q(x_\Delta)$ with $a \in \mathbb{R}$ and $q(x_\Delta) \in \mathbb{R}[x_\Delta]_k$. Then

$$\|p_\Delta(x_\Delta)\|_{L_B^2}^2 = \begin{bmatrix} a \\ q_G \end{bmatrix}^T \mathbf{B}_k \begin{bmatrix} a \\ q_G \end{bmatrix} \ge \zeta_k^2 (a^2 + \|q_G\|_2^2) = \zeta_k^2 \|p_\Delta(x_\Delta)\|_G^2.$$

By definition of the sparse marginal L^2-norm, we have

$$\|p(x)\|_{L_B^2,\Omega(p)}^2 = \sum_{\Delta\in\Omega(p)} \|p_\Delta(x_\Delta)\|_{L_B^2}^2 \ge \zeta_k^2 \sum_{\Delta\in\Omega(p)} \|p_\Delta(x_\Delta)\|_G^2 \ge \zeta_k^2 \|p(x)\|_G^2.$$

So item (ii) follows. □

For minimizing sparse forms, Theorem 2.4 can be improved as follows.

Theorem 4.2. *Let $f(x) \in \mathbb{R}[x]_{2d}$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) value on $\mathbb{S}^{n-1}$. If f_{sos} is the lower bound given by (1.2), then the following holds*

$$1 \le \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \le \frac{2}{\zeta_{2d}} \sqrt{|\Omega(f)|}.$$

Here ζ_{2d} is defined in (4.2), and $\Omega(f)$ is defined in (4.1).

Proof. We follow the same approach for proving Theorem 2.4. Let $f_{med} = \frac{1}{2}(f_{min} + f_{max})$ and $\tilde{f}(x) = f(x) - f_{med}$, then

$$\left| \frac{1}{f_{med} - f_{min}} \tilde{f}(x) \right| \le 1 \quad \forall x \in \mathbb{S}^{n-1}.$$

By Lemma 4.1, we know

$$\left\| \frac{1}{f_{med} - f_{min}} \tilde{f}(x) \right\|_{L_B^2, mg} \le \sqrt{|\Omega(f)|}.$$

Fixing a constant

$$\gamma^* = f_{med} - (f_{med} - f_{min}) \cdot \frac{2}{\zeta_{2d}} \sqrt{|\Omega(f)|}, \tag{4.3}$$

we obtain that

$$\left\| \frac{2}{f_{med} - \gamma^*} \tilde{f}(x) \right\|_{L^2(g), mg} \le \zeta_{2d}.$$

Lemma 4.1 and the above imply that

$$\left\| \frac{2}{f_{med} - \gamma^*} \tilde{f}(x) \right\|_G \le \frac{1}{\zeta_{2d}} \left\| \frac{2}{f_{med} - \gamma^*} \tilde{f}(x) \right\|_{L_B^2, mg} \le 1.$$

Let $a \in \mathbb{R}$ and $p(x) \in \mathbb{R}[x]_{2d}$ be such that

$$\frac{2}{f_{med} - \gamma^*} \tilde{f}(x) = a + p(x), \quad a^2 + \|p(x)\|_G^2 = \left\| \frac{2}{f_{med} - \gamma^*} \tilde{f}(x) \right\|_G^2 \le 1. \tag{4.4}$$

By Lemma 2.1, there exists a symmetric matrix P satisfying

$$p(x) = [x^d]^T P [x^d], \quad \|P\|_F = \|p(x)\|_G.$$

Let D be the diagonal matrix such that $\|x\|_2^{2d} = [x^d]^T D [x^d]$. Then $\lambda_{min}(D) \ge 1$ and

$$\frac{2}{f_{med} - \gamma^*} \tilde{f}(x) + (1 + \|x\|_2^{2d}) = 1 + a + [x^d]^T (P + D) [x^d].$$

Since $\|P\|_2 \le \|P\|_F = \|p(x)\|_G$, (4.4) implies $1 + a \ge 0$ and the form

$$\sigma_1(x) = [x^d]^T (P + D) [x^d]$$

is SOS. By definition of $\tilde{f}(x)$, it holds the identity

$$f(x) - f_{med} + \frac{f_{med} - \gamma^*}{2}(1 + \|x\|_2^{2d}) = \frac{f_{med} - \gamma^*}{2}(1 + a + \sigma_1(x)).$$

In the above, replacing x by $\frac{x}{\|x\|_2}$ and multiplying both sides by $\|x\|_2^{2d}$, we get

$$f(x) - \gamma^* \|x\|_2^{2d} = \sigma(x)$$

where the form $\sigma(x) = \frac{f_{med} - \gamma^*}{2}\left((1+a)\|x\|_2^{2d} + \sigma_1(x)\right)$ is SOS. By the optimality of f_{sos}, we have $f_{sos} \geq \gamma^*$. Then the theorem follows (4.3). $\square$

Example 4.3. Consider sparse forms like

$$f(x) = \sum_{i,j=1}^{n-1} f_{ij} x_i x_{i+1} x_j x_{j+1}.$$

Clearly, $|\Omega(f)| = \binom{n-1}{2}$. Therefore, by Theorem 4.2, to minimize $f(x)$ over $\mathbb{S}^{n-1}$, the SOS relaxation (1.2) gives an $\mathcal{O}(n)$-approximation. $\square$

4.2. Odd forms

A quite general problem is to minimize odd forms over unit spheres. For instance, the stability number of a graph can be expressed in terms of the optimal value of a particular cubic form over the unit sphere, as shown by Nesterov [7]. He actually [7] showed that (1.1) is NP-hard when $\deg(f) = 3$. However, SOS relaxation (1.2) can not be applied directly when $f(x)$ is odd. Fortunately, we can formulate the problem equivalently as minimizing a certain even form in a higher dimensional space.

Suppose $f(x)$ is an odd form of degree $2d-1$. Then we must have $f_{max} + f_{min} = 0$ and $f_{min} \leq 0 \leq f_{max}$. Let $\hat{f}(x,t) = f(x)t$ be a new even form in (x,t) and denote

$$\hat{f}_{min} = \min_{\|x\|_2^2 + t^2 = 1} f(x)t, \qquad \hat{f}_{max} = \max_{\|x\|_2^2 + t^2 = 1} f(x)t.$$

Note the following relations

$$\min_{0 \leq t \leq 1} \min_{\|x\|_2 = \sqrt{1-t^2}} f(x)t = \min_{0 \leq t \leq 1} t \min_{\|x\|_2 = \sqrt{1-t^2}} f(x) =$$

$$\min_{0 \leq t \leq 1} (t(1-t^2)^{d-1/2}) f_{min} = f_{min} \max_{0 \leq t \leq 1} (t(1-t^2)^{d-1/2}) = \frac{1}{\sqrt{2d-1}} \left(1 - \frac{1}{2d}\right)^d f_{min},$$

$$\min_{-1 \leq t \leq 0} \min_{\|x\|_2 = \sqrt{1-t^2}} f(x)t = \min_{0 \leq t \leq 1} t \max_{\|x\|_2 = \sqrt{1-t^2}} f(x) =$$

$$\min_{-1 \leq t \leq 0} (t(1-t^2)^{d-1/2}) f_{max} = f_{max} \min_{-1 \leq t \leq 0} (t(1-t^2)^{d-1/2}) = \frac{1}{\sqrt{2d-1}} \left(1 - \frac{1}{2d}\right)^d f_{min}.$$

Thus we have

$$f_{min} = \sqrt{2d-1} \left(1 - \frac{1}{2d}\right)^{-d} \hat{f}_{min}, \quad f_{max} = \sqrt{2d-1} \left(1 - \frac{1}{2d}\right)^{-d} \hat{f}_{max}.$$

Therefore, minimizing $f(x)$ over $\mathbb{S}^{n-1}$ is equivalent to

$$\min_{\|x\|_2^2+t^2=1} f(x)t. \tag{4.5}$$

Since the form $\hat{f}(x,t) = f(x)t$ is even, SOS relaxation (1.2) can be applied to get a lower bound $\hat{f}_{sos}$ for $\hat{f}_{min}$. Then

$$f_{sos} = \sqrt{2d-1}\left(1-\frac{1}{2d}\right)^{-d}\hat{f}_{sos}$$

is also a lower bound of f_{min}. Observe that

$$|\Omega(\hat{f})| = |\Omega(f)| \le \binom{n}{2d-1}.$$

So Theorem 4.2 immediately implies the following.

Theorem 4.4. *Let $f(x) \in \mathbb{R}[x]_{2d-1}$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) value on $\mathbb{S}^{n-1}$. If f_{sos} is obtained as above through solving (4.5), then*

$$1 \le \frac{f_{max}-f_{sos}}{f_{max}-f_{min}} \le \frac{2}{\zeta_{2d}}\sqrt{|\Omega(f)|}.$$

In particular, if $f(x)$ is dense, then f_{sos} is an $\mathcal{O}(n^{d-1/2})$-approximation of f_{min}.

4.3. Odd multi-forms

Let $f(x) \in \mathcal{F}^{n_1,\ldots,n_m}_{2d_1-1,\ldots,2d_m-1}$ be an odd multi-form, i.e., every term of $f(x)$ has a fixed odd degree in each component x_{I_i}. We want to find a lower bound of its minimum value f_{min} over the multi-unit sphere $\mathbb{S}^{n_1-1,\ldots,n_m-1}$. Suppose $f(x)$ is given as

$$f(x) = \sum_{\alpha\in\mathbb{N}^{n_1,\ldots,n_m}_{2d_1-1,\ldots,2d_m-1}} f_\alpha (x_{I_1})^{\alpha_1}\cdots(x_{I_m})^{\alpha_m}.$$

Introduce new variables $t=(t_1,\ldots,t_m)$, and let $\hat{f}(x,t) = f(x)t_1\ldots t_m$. Then $\hat{f}(x,t)$ has even degrees in every component $\tilde{x}_{I_i} = (x_{I_i}, t_i)$. Consider the even multi-form optimization

$$\begin{array}{ll} \min\limits_{x,t} & \hat{f}(x,t) \\ s.t. & \|x_{I_i}\|_2^2 + t_i^2 = 1,\ i=1,\ldots,m. \end{array} \tag{4.6}$$

Denote the minimum (resp., maximum) objective value in the above by $\tilde{f}_{min}$ (resp., $\tilde{f}_{max}$). As in the preceding subsection, we can similarly prove that

$$f_{min} = \left(\prod_{i=1}^m \frac{\sqrt{2d_i-1}}{(1-1/2d_i)^{d_i}}\right)\tilde{f}_{min}, \quad f_{max} = \left(\prod_{i=1}^m \frac{\sqrt{2d_i-1}}{(1-1/2d_i)^{d_i}}\right)\tilde{f}_{max}.$$

The techniques in the preceding two subsections can be generalized in a natural way to get an approximation bound $\mathcal{O}(\sqrt{|\Omega(f)|})$ for SOS relaxation (3.3) applied to (4.6). When $f(x)$ is dense, the approximation bound is $\mathcal{O}(n_1^{d_1-1/2}\cdots n_m^{d_m-1/2})$. We would like to leave this as an exercise for interested readers.

5 Optimization over hypersurfaces

A more general problem is to optimize homogeneous polynomials over hypersurfaces. For instance, we might minimize a form over the $2d$-sphere $\{x \in \mathbb{R}^n : x_1^{2d} + \cdots + x_n^{2d} = 1\}$. This section will propose an SOS relaxation similar to (1.2), and then analyze its approximation performance. Generalizing the techniques we have used earlier, an approximation bound like in Theorem 2.4 can be obtained.

Let $f(x), g(x)$ be two even forms of degree $2d$. Consider optimization problem

$$\begin{array}{rl} \min\limits_{x\in\mathbb{R}^n} & f(x) \\ s.t. & g(x) = 1. \end{array} \tag{5.1}$$

The feasible set $H(g) = \{x \in \mathbb{R}^n : g(x) = 1\}$ is a hypersurface. When $g(x) = \|x\|_2^{2d}$, (5.1) reduces to (1.1). So problem (5.1) is also NP-hard. A natural SOS relaxation for (5.1) is

$$\begin{array}{rl} \max & \gamma \\ s.t. & f(x) - \gamma \cdot g(x) \text{ is SOS.} \end{array} \tag{5.2}$$

For convenience, we still denote by f_{min} (resp., f_{max}) the minimum (resp., maximum) value of $f(x)$ on $H(g)$, and denote by f_{sos} the maximum objective value of (5.2). It is obvious that $f_{sos} \le f_{min}$. We are interested in estimating how far away f_{sos} is from f_{min}.

When $g(x)$ is a positive definite form, the hypersurface $H(g)$ is compact, and we can define a norm of $p(x)$ as

$$\|p(x)\|_{L^2(g)} = \left(\int_{g(x)=1} p(x)^2 d\mu_g(x) \right)^{1/2}.$$

Here $\mu_g(\cdot)$ is the uniform probability measure on $H(g)$. Suppose $n \ge 2d$. When $p(x)$ has degree $2d$, we can similarly define its marginal L^2-norm as

$$\|p(x)\|_{L^2(g),mg} = \left(\sum_{\Delta\in\Omega_{2d}} \|p_\Delta(x_\Delta)\|^2_{L^2(g_\Delta)} \right)^{1/2}.$$

Here p_Δ and g_Δ are the restrictions of $p(x)$ and $g(x)$ to x_Δ respectively, and

$$\|p_\Delta(x_\Delta)\|_{L^2(g_\Delta)} = \left(\int_{g_\Delta(x_\Delta)=1} p_\Delta(x_\Delta)^2 d\mu_{g_\Delta}(x_\Delta) \right)^{1/2}.$$

The above $\mu_{g_\Delta}(\cdot)$ is the uniform probability measure on $H(g_\Delta) := \{x_\Delta : g_\Delta(x_\Delta) = 1\}$. Similarly, for each $\Delta \in \Omega_{2d}$, define matrix

$$\mathbf{\Theta}_\Delta(g) = \int_{g_\Delta(x_\Delta)=1} [x_\Delta^{2d}]_G [x_\Delta^{2d}]_G^T d\mu_{g_\Delta}(x_\Delta).$$

If $g(x)$ is positive definite, then every $g_\Delta(x_\Delta)$ is also positive definite, and $\mathbf{\Theta}_\Delta(g) \succ 0$, because the monomials of $[x_\Delta^k]_G$ are linearly independent. Define a positive constant

$$\delta(g) = \min_{\Delta\in\Omega_{2d}} \sqrt{\lambda_{min}(\mathbf{\Theta}_\Delta(g))} > 0. \tag{5.3}$$

Note $\delta(g)$ is depending only on g. Like Lemmas 2.3 and 2.2, we can similarly prove

Lemma 5.1. *Let $g(x) \in \mathbb{R}[x]_{2d}$ be a positive definite form.*
(i) If $|p(x)| \leq 1$ for all $x \in H(g)$, then $\|p(x)\|_{L^2(g),mg} \leq \sqrt{\binom{n}{2d}}$.
(ii) If $p(x) \in \mathbb{R}[x]_{2d}$, then $\|p(x)\|_{L^2(g),mg} \geq \delta(g)\|p(x)\|_G$.

The performance of SOS relaxation (5.2) is summarized in the following theorem.

Theorem 5.2. *Assume $g(x) = [x^d]^T E[x^d]$ and E is a symmetric positive definite matrix. Let $f(x) \in \mathbb{R}[x]_{2d}$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) value on the hypersurface $H(g)$. Then the optimal value f_{sos} of (5.2) satisfies*

$$1 \leq \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \leq \frac{1}{\delta(g)\lambda_{min}(E)} \sqrt{\binom{n}{2d}}.$$

Proof. We follow the same approach of proving Theorem 2.4, and only list the distinct parts. Set $f_{med} = \frac{1}{2}(f_{min} + f_{max})$ and $\tilde{f}(x) = f(x) - f_{med} \cdot g(x)$. Then

$$\left| \frac{1}{f_{med} - f_{min}} \tilde{f}(x) \right| \quad \forall x \in H(g).$$

By Lemma 5.1, we know

$$\left\| \frac{1}{f_{med} - f_{min}} \tilde{f}(x) \right\|_{L^2(g),mg} \leq \sqrt{\binom{n}{2d}}.$$

Fixing a constant

$$\gamma^* = f_{med} - (f_{med} - f_{min}) \cdot \frac{1}{\delta(g)\lambda_{min}(E)} \sqrt{\binom{n}{2d}},$$

we can get

$$\left\| \frac{1}{f_{med} - \gamma^*} \tilde{f}(x) \right\|_{L^2(g),mg} \leq \delta(g)\lambda_{min}(E).$$

By Lemma 5.1, the above implies

$$\left\| \frac{1}{f_{med} - \gamma^*} \tilde{f}(x) \right\|_G \leq \delta(g)^{-1} \left\| \frac{1}{f_{med} - \gamma^*} \tilde{f}(x) \right\|_{L^2(g),mg} \leq \lambda_{min}(E).$$

By Lemma 2.1, there exists a symmetric matrix W satisfying

$$\frac{1}{f_{med} - \gamma^*} \tilde{f}(x) = [x^d]^T W [x^d], \quad \|W\|_F \leq \lambda_{min}(E).$$

From $\|W\|_2 \leq \|W\|_F \leq \lambda_{min}(E)$, we know $W + E \succeq 0$ and

$$\frac{1}{f_{med} - \gamma^*} \tilde{f}(x) + g(x) = [x^d]^T (W + E)[x^d]$$

is SOS, or equivalently, the form $f(x) - \gamma^* g(x)$ is SOS. By the optimality of f_{sos}, we know $f_{sos} \geq \gamma^*$. Thus the theorem follows the choice of γ^*. □

Now we finish this section with an example.

Example 5.3. For $g(x) = x_1^{2d} + \cdots + x_n^{2d}$, $H(g)$ is a compact hypersurface of degree $2d$. We show that there exists a symmetric matrix $E \succ 0$ such that

$$x_1^{2d} + \cdots + x_n^{2d} = [x^d]^T E [x^d]. \tag{5.4}$$

Recall the arithmetic-geometric inequality (AGI)

$$y_1 \cdots y_{2d} \leq \frac{1}{2d}(y_1^{2d} + \cdots + y_{2d}^{2d}) \quad \forall\, (y_1, \ldots, y_{2d}) \in \mathbb{R}^{2d}.$$

Hurwitz [3] (also see Reznick [12]) proved a very nice result that the form

$$\frac{1}{2d}(y_1^{2d} + \cdots + y_{2d}^{2d}) - y_1 \cdots y_{2d}$$

is SOS. For every $\alpha \in \mathbb{N}(d)$, it holds

$$x_1^{2\alpha_1} \cdots x_n^{2\alpha_n} \leq \frac{1}{2d}(2\alpha_1 x_1^{2d} + \cdots + 2\alpha_n x_n^{2d}).$$

Then Hurwitz's result implies there exists an sos polynomial $s_\alpha(x)$ such that

$$x^{2\alpha} + s_\alpha(x) = \frac{1}{d}\sum_{i=1}^n \alpha_i x_i^{2d}.$$

Observing the equalities

$$\sum_{\alpha \in \mathbb{N}(d)} \frac{\alpha_1}{d} = \cdots = \sum_{\alpha \in \mathbb{N}(d)} \frac{\alpha_n}{d} = \frac{1}{n}\sum_{\alpha \in \mathbb{N}(d)} \left(\frac{\alpha_1 + \cdots + \alpha_n}{d}\right) = \frac{1}{n}\binom{n+d-1}{d},$$

we get the identity

$$\sum_{\alpha \in \mathbb{N}(d)} \left(x^{2\alpha} + s_\alpha(x)\right) = \frac{1}{n}\binom{n+d-1}{d}\sum_{i=1}^n x_i^{2d},$$

or equivalently

$$\sum_{i=1}^n x_i^{2k} = n\binom{n+d-1}{d}^{-1}\left(s_d(x) + [x^d]^T[x^d]\right).$$

Here $s_d(x) = \sum_{\alpha \in \mathbb{N}(d)} s_\alpha(x)$ is also an SOS form. So there exists a symmetric matrix $S \succeq 0$ such that $s_d(x) = [x^d]^T S [x^d]$. Letting

$$E = n\binom{n+d-1}{d}^{-1}(S + I),$$

we know (5.4) holds with

$$\lambda_{min}(E) \geq n\binom{n+d-1}{d}^{-1} = \mathcal{O}(n^{1-d}).$$

By (5.3), $\delta(g)$ is a constant independent of n. So Theorem 5.2 shows that SOS relaxation (5.2) gives an $\mathcal{O}(n^{2d-1})$-approximation for (5.1) when $g(x) = x_1^{2d} + \cdots + x_n^{2d}$. $\square$

6 Some discussions

For minimizing forms of an even degree $2d$ over the unit sphere $\mathbb{S}^{n-1}$, we basically prove that the SOS relaxation (1.2) has an approximation bound $\mathcal{O}(n^d)$ for any fixed d. A very interesting question is whether this bound is tight or not. To the best knowledge of the author, this question is open. The main difficulty is how to construct a nonnegative but non-SOS form that maximizes the ratio $\frac{f_{max}-f_{sos}}{f_{max}-f_{min}}$. Actually, it is very tricky to explicitly find a nonnegative form that is not SOS. It took about eighty years to construct such an explicit example (Motzkin polynomial) after Hilbert showed the existence of nonnegative forms that are not SOS. Thus it would be very difficult to tell the tightness of an approximation bound for SOS relaxation.

We would like to remark that there is no finite approximation bound when we apply SOS relaxation to find a lower bound for the minimum of a polynomial in the whole space $\mathbb{R}^n$. For example, for $f(x) = x_1^4x_2^2 + x_1^2x_2^4 + x_3^6 - 3x_1^2x_2^2x_3^3$ (Motzkin polynomial), its minimum $f_{min} = 0$, but the standard SOS relaxation (see [5, 9, 10])

$$\max \quad \gamma \quad \text{s.t.} \quad f(x) - \gamma \quad \text{is SOS}$$

is not feasible and $f_{sos} = -\infty$. So there is no guaranteed upper bound for $f_{min} - f_{sos}$.

References

[1] G. Blekherman. There are significantly more nonnegative polynomials than sums of squares. *Israel J. Math.* 153 (2006), 355-380.

[2] L. Faybusovich. Global optimization of homogeneous polynomials on the simplex and on the sphere, in: C. Floudas and P. Pardalos, editors, *Frontiers in Global Optimization*, Kluwer Academic Publishers, 2003.

[3] A. Hurwitz. Über den Vergleich des arithmetischen und des geometrischen. *Mittels. J. Reine Angew. Math.*, 108, 266-268 (1891).

[4] M. Kojima, S. Kim and H. Waki. Sparsity in Sums of Squares of Polynomials. *Mathematical Programming*, Vol.103 (1) 45-62 (2005).

[5] J. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, **11** (2001), No. 3, 796–817.

[6] C. Ling, J. Nie, L. Qi, and Y. Ye. Bi-quadratic optimization over unit spheres and semidefinite programming relaxations. *SIAM Journal on Optimization*, Vol. 20, No. 3, pp. 1286-1310, 2009.

[7] Y. Nesterov. Random walk in a simplex and quadratic optimization over convex polytopes. CORE Discussion Paper, CORE, Catholic University of Louvain, Louvain-la-Neuve, Belgium, 2003.

[8] J. Nie and J. Demmel. Sparse SOS relaxations for minimizing functions that are summations of small polynomials. *SIAM Journal On Optimization*, Vol. 19, No. 4, pp. 1534-1558 (2008).

[9] P. Parrilo and B. Sturmfels. Minimizing polynomial functions, Proceedings of the DIMACS Workshop on Algorithmic and Quantitative Aspects of Real Algebraic Geometry in Mathematics and Computer Science (March 2001), (eds. S. Basu and L. Gonzalez-Vega), American Mathematical Society, 2003, pp. 83–100.

[10] P. Parrilo. Semidefinite Programming relaxations for semialgebraic problems. *Mathematical Programming, Ser. B*, 96 (2003), No. 2, 293–320.

[11] P. Parrilo. Exploiting structure in sum of squares programs. *Proceedings for the 42nd IEEE Conference on Decision and Control*, Maui, Hawaii, 2003.

[12] B. Reznick. Forms derived from the arithmetic-geometric inequality. *Math. Ann.*, 283 (1989), 431-464.

[13] B. Reznick. Some concrete aspects of Hilbert's 17^{th} problem. In *Contemporary Mathematics*, volume 253, pages 251-272. American Mathematical Society, 2000.

[14] H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of semidefinite programming: theory, algorithms, and applications*. International Series in Operations Research & Management Science, 27. Kluwer Academic Publishers, Boston, MA, 2000.