An Approximation Bound Analysis for Lasserre's Relaxation in Multivariate Polynomial Optimization

Jiawang Nie*

July 8, 2009

Abstract

Suppose $f(x), g_1(x), \ldots, g_m(x)$ are multivariate polynomials in $x \in \mathbb{R}^n$ and have degrees no greater than $2d$. Consider optimization problem

$$\min \quad f(x) \quad s.t. \quad x \in S = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}.$$

Assume f_{min} (resp., f_{max}) is the minimum (resp., maximum) of $f(x)$ on S. Let f_{sos} be the lower bound of f_{min} given by Lasserre's relaxation of order d. First, we study its approximation performance. Under a suitable condition on $g_1, \ldots, g_m$, we prove that

$$f_{max} - f_{min} \leq f_{max} - f_{sos} \leq Q \cdot (f_{max} - f_{min}).$$

Here Q is a constant depending only on $g_1, \ldots, g_m$. In particular, if S is the unit ball $B(0,1)$, $Q = \mathcal{O}(n^d)$; if S is the hypercube $[-1,1]^n$, $Q = \mathcal{O}(n^{\frac{3d}{2}})$; if S is the discrete set $\{\pm 1\}^n$ or $\{0,1\}^n$, $Q = \mathcal{O}(n^d)$; for general cases, estimates for Q are also given. Second, when $f(x)$ is a form and S is defined by homogeneous polynomial inequalities, we prove a similar approximation bound. Third, when $f, g_1, \ldots, g_m$ have certain sparsity patterns, we prove an approximation bound for a sparse version of Lasserre's relaxation.

Key words approximation bound, Lasserre's relaxation, polynomials, semidefinite programming, sum of squares

AMS subject classification 68Q25, 90C22, 90C60

1 Introduction

Consider the polynomial optimization problem

$$\begin{array}{rl} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ s.t. & g_1(x) \geq 0, \ldots, g_m(x) \geq 0. \end{array} \tag{1.1}$$

Here $f(x), g_1(x), \ldots, g_m(x)$ are all multivariate polynomials in $x = (x_1, \ldots, x_n)$. Problem (1.1) is quite general. When $f(x)$ and $g_i(x)$ are all linear, (1.1) reduces to a linear programming (LP); when $f(x)$ and $g_i(x)$ are all quadratic, (1.1) becomes a quadratically constrained

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

quadratic programming (QCQP). Generally it is quite difficult to solve (1.1). For instance, when $f(x)$ is a nonconvex quadratic function and every $g_i(x)$ is linear, problem (1.1) becomes a nonconvex quadratic programming (QP), which is NP-hard [27]. So problem (1.1) is NP-hard. Thus approximation algorithms are more interesting. Lasserre's relaxation [11] is a typical numerical method for solving (1.1) approximately by using semidefinite programming and sum of squares techniques. Recently there is much work in this area. We refer to [7, 9, 11, 14, 15, 16, 23, 24, 28, 29, 33].

Lasserre's relaxation was originally proposed by Lasserre in his milestone paper [11]. When $f(x), g_1(x), \ldots, g_m(x)$ have degrees no greater than $2d$, Lasserre proposed the following sum of squares (SOS) program to find a lower bound for the minimum f_{min} of (1.1)

$$\begin{array}{rl} \max & \gamma \\ s.t. & f(x) - \gamma = \sigma_0(x) + \sigma_1(x)g_1(x) + \cdots + \sigma_m(x)g_m(x), \\ & \deg(\sigma_0), \deg(\sigma_1 g_1), \ldots, \deg(\sigma_m g_m) \leq 2d, \\ & \sigma_0(x), \sigma_1(x), \ldots, \sigma_m(x) \text{ are SOS.} \end{array} \tag{1.2}$$

In the above, a polynomial is said to be SOS if it is a sum of squares of other polynomials. If a polynomial is SOS, then it must be nonnegative everywhere, but the reverse might not be true. See [32] for a survey on SOS and nonnegative polynomials. Though it is difficult to check the nonnegativity of a polynomial, it is easier to check whether a polynomial is SOS. This is because checking SOS is equivalent to solving a semidefinite programming (SDP) problem, which can be solved efficiently. Problem (1.1) is NP-hard, but the SOS program (1.2) is equivalent to an SDP problem. The integer d in (1.2) is called the relaxation order.

Now we give a short review about the convergence of Lasserre's relaxation (1.2). For convenience, denote $g(x) = (g_1(x), \cdots, g_m(x))$ and

$$S = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}.$$

So S is a basic closed semialgebraic set [1]. For any γ feasible in (1.2), we have

$$f(x) - \gamma = \sigma_0(x) + \sigma_1(x)g_1(x) + \cdots + \sigma_m(x)g_m(x) \geq 0 \quad \forall\, x \in S.$$

Thus, every γ feasible in (1.2) satisfies $f(x) \geq \gamma$ for all $x \in S$. If we denote by $f_{sos,d}$ the optimal value of (1.2), then $f_{min} \geq f_{sos,d}$ for all d. As d increases, the lower bound $f_{sos,d}$ is monotonically increasing. Based on Putinar's Positivstellensatz [31], Lasserre [11] proved $f_{sos,d} \to f_{min}$ as $d \to \infty$ under a so-called *archimedean condition (AC)*, that is, there exist $M > 0$ and SOS polynomials $s_0(x), s_1(x), \ldots, s_m(x)$ such that

$$M - \|x\|_2^2 = s_0(x) + s_1(x)g_1(x) + \ldots + s_m(x)g_m(x).$$

To make AC hold, S must be compact, but the reverse might not be true. However, this does not limit the applications of Lasserre's relaxation very much, because otherwise we can always add a redundant ball condition $M - \|x\|_2^2 \geq 0$ if S is compact. Nie and Schweighofer [24] analyzed the convergence speed of Lasserre's relaxations. Under AC, they proved

$$0 \leq f_{min} - f_{sos,d} \leq \mathcal{O}((\log d)^{-c}) \quad \text{as} \quad d \to \infty,$$

where $c > 0$ is a constant depending on g.

Due to the computational cost of (1.2), people often choose small d in practical applications. This is because (1.2) is very expensive to solve for big d, though its computational complexity is polynomial for any fixed d. So it is interesting to know the approximation performance of (1.2) for a fixed relaxation order d. Suppose $\deg(f) = 2d$ and $\deg(g) \leq 2d$, then the lowest order of Lasserre's relaxation is d. For convenience, just denote by f_{sos} the optimal value of (1.2) for given $f(x)$. We have seen $f_{sos} \leq f_{min}$, but do not know how far away f_{sos} is from f_{min}. Denote by f_{max} the maximum of $f(x)$ on S, which always exists when S is compact. For fixed g, a constant Q is called an approximation bound of (1.2) if it holds

$$f_{max} - f_{sos} \leq Q \cdot (f_{max} - f_{min}). \tag{1.3}$$

Does the above Q exist? If so, how big is Q? Or what conditions make Q exist? To the best knowledge of the author, these questions are known very little.

There has been much work on analyzing the performance of the first order Lasserre's relaxation when $f(x)$ and $g_i(x)$ are quadratic polynomials. Most of them are based on standard semidefinite programming relaxations, and usually assume $f(x)$ is a form (homogeneous polynomial) and the constraints are homogeneous. We refer to [4, 5, 19, 21, 22, 34, 37, 38]. When $f(x)$ and $g_i(x)$ are not quadratic, there exist other approximation methods different from Lasserre's relaxation for solving (1.1). When $f(x)$ is a quartic form and there are only homogeneous quadratic constraints, Luo and Zhang [20] proposed an interesting quadratic SDP relaxation, and analyzed its approximation performance. When $f(x)$ is a bi-quadratic form and S is a bi-sphere, Ling, Nie, Qi and Ye [18] gave some approximation bounds based on a bi-linear SDP relaxation and SOS techniques. When S is a simplex, De Klerk, Laurent and Parrilo [2] proposed some polynomial time approximation schemes (PTASs) via Pólya's theorem or rational grid points, and proved some approximation bounds. Lasserre and Netzer [13] analyzed SOS approximations of nonnegative polynomials via simple high degree perturbations. In [3], De Klerk gave a nice survey on the complexity of optimization over a simplex, hypercube or sphere. However, when $f(x)$ and $g_i(x)$ are general polynomial functions, there is very few work on analyzing the approximation performance of Lasserre's relaxation. Recently, Nie [26] proved some approximation bounds for standard SOS relaxations for minimizing forms over spheres or hypersurfaces. The techniques used there only work when $f(x)$ is a form and there is only a single homogeneous equality constraint. They can be generalized in a nontrivial way to analyze the approximation performance of Lasserre's relaxation (1.2) when $f(x)$ is not a form and there are several inequality constraints.

Contributions. First, we analyze the approximation bound of Lasserre's relaxation (1.2) when S is compact. Let $f(x)$ be an arbitrary polynomial of degree $2d$, and denote by f_{min} (resp., f_{max}) its minimum (resp., maximum) value on S. Under a suitable condition on $g_1, \ldots, g_m$, we show that there exists a constant Q such that

$$1 \leq \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \leq Q. \tag{1.4}$$

The constant Q only depends on the tuple $(g_1, \ldots, g_m)$, and can be estimated numerically. This will be presented in Section 3.

Second, we give explicit estimates for Q in (1.4) for some special problems. It will be shown that: when S is a unit ball $B(0,1)$, $Q = \mathcal{O}(n^d)$; when S is a hypercube $[-1,1]^n$,

$Q = \mathcal{O}(n^{\frac{3d}{2}})$; when S is the discrete set $\{\pm 1\}$ or $\{0,1\}^n$, $Q = \mathcal{O}(n^d)$; when S is a multi-unit ball, $Q = \mathcal{O}(n^d)$. This will be shown in Section 4.

Third, we study homogeneous polynomial optimization, that is, $f(x)$ is a form and S is a polysoid that is defined by homogeneous polynomial inequalities, and $\deg(f) = \deg(g_i) = 2d$. In this case, the SOS polynomials $\sigma_i(x)$ in (1.2) reduce to nonnegative scalars. Under a suitable condition on the homogeneous parts of $g_1, \dots, g_m$, a similar approximation result like (1.4) can also be proven. This will be presented in Section 5.

Last, we study the approximation performance of a sparse version of Lasserre's relaxation. When $f(x)$ and $g_1(x), \dots, g_m(x)$ have certain sparsity patterns, Lasserre's relaxation (1.2) can be modified to approximately solve problem (1.1) of much larger sizes. Under a suitable condition on sparsity patterns, an approximation bound like (1.4) can also be proven. This will be presented in Section 6.

The proofs of these approximation results are heavily based on estimating norms of polynomials and using semidefinite programming properties. So we will first introduce some basics about semidefinite programming, sum of squares, various norms of polynomials and their relations. This will be presented in Section 2.

Notations. The symbol $\mathbb{N}$ (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 $0 < k \in \mathbb{N}$, $[k] = \{1, \dots, k\}$. For $x \in \mathbb{R}^n$, x_i denotes the i-th component of x, that is, $x = (x_1, \dots, x_n)$. The $\mathbb{S}^{n-1}$ denotes the $n-1$ dimensional unit sphere $\{x \in \mathbb{R}^n : x_1^2 + \cdots + x_n^2 = 1\}$. For $\alpha \in \mathbb{N}^n$, denote $|\alpha| = \alpha_1 + \cdots + \alpha_n$, and $\mathrm{supp}(\alpha) = \{1 \le i \le n : \alpha_i \ne 0\}$. For $\alpha, \beta \in \mathbb{N}^n$, denote $\alpha \le \beta$ if every $\alpha_i \le \beta_i$. The symbol $\mathbb{N}_{\le k}$ denotes the multi-index set $\{\alpha \in \mathbb{N}^n : |\alpha| \le k\}$, and $\mathbb{N}(k)$ denotes $\{\alpha \in \mathbb{N}^n : |\alpha| = k\}$. For $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{N}^n$, x^α denotes $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. The $[x]_d$ denotes the vector of all monomials having degrees at most d and ordered graded alphabetically, that is,

$$[x]_d^T = [\,1 \quad x_1 \quad \cdots \quad x_n \quad x_1^2 \quad x_1x_2 \quad \cdots\cdots \quad x_1^d \quad x_1^{d-1}x_2 \quad \cdots\cdots \quad x_n^d\,],$$

and $[x^d]$ denotes the homogeneous part of $[x]_d$ having degree d, that is,

$$[x^d]^T = [\,x_1^d \quad x_1^{d-1}x_2 \quad \cdots\cdots \quad x_n^d\,].$$

The $\mathbb{R}[x]$ denotes the ring of polynomials in $(x_1, \dots, x_n)$ whose coefficients are real numbers, $\mathbb{R}[x]_k$ denotes the subspace of homogeneous polynomials whose degrees are equal to k; $\mathbb{R}[x]_{\le k}$ denotes the subspace of polynomials whose degrees are at most k; $Sfr[x]_{\le k}$ denotes the subspace of square free polynomials whose degrees are at most k; for integers $k_1, \dots, k_\ell$, denote $\mathbb{R}[x]_{k_1,\dots,k_\ell} := \mathbb{R}[x]_{k_1} + \cdots + \mathbb{R}[x]_{k_\ell}$. 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 general set S, $int(S)$ denotes its interior. 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, and $X \succeq 0$ (resp., $X \succ 0$) means $\lambda_{min}(X) \ge 0$ (resp. $\lambda_{min}(X) > 0$). For a general matrix X, $\sigma_{max}(X)$ and $\sigma_{min}(X)$ denote the maximum and minimum singular values of X respectively. For two matrices A and B, $A \otimes B$ denotes the standard Kronecker product of A and B. For any vector $u \in \mathbb{R}^N$, $\|u\|_2 = \sqrt{u^Tu}$ denotes the standard Euclidean norm. For any matrix A, $\|A\|_2$ denotes the maximum singular value of A, and $\|A\|_F$ denotes the Frobinus norm of A, i.e., $\|A\|_F = \sqrt{Trace(A^TA)}$.

2 Sum of squares and norms of polynomials

This section presents some basics in sum of squares, semidefinite programming, various norms of polynomials and their relations.

2.1 Sum of squares and semidefinite programming

For a polynomial $f(x)$ of degree $2d$, there exists a symmetric matrix F such that

$$f(x) = [x]_d^T F [x]_d.$$

The length of $[x]_d$ is $\binom{n+d}{d}$, and the dimension of F is $\binom{n+d}{d} \times \binom{n+d}{d}$. The matrix F is called a *Gram* matrix of $f(x)$ and is not unique if $d > 1$ and $n > 1$. For convenience, we index the columns and rows of F by monomials of degrees at most d, or equivalently by vectors in $\mathbb{N}^n$ whose standard $\|\cdot\|_1$ norms are at most d.

A polynomial $f(x)$ is said to be a sum of squares (SOS) if there exist polynomials $f_1(x), \dots, f_k(x)$ such that $f(x) = f_1(x)^2 + \cdots + f_k(x)^2$. As shown in [28, 29], $f(x)$ is SOS if and only if it has a Gram matrix F which is positive semidefinite, that is,

$$f(x) \text{ is SOS} \quad \iff \quad f(x) = [x]_d^T F [x]_d,\ F \succeq 0.$$

Define constant symmetric matrices A_α such that

$$[x]_d [x]_d^T = \sum_{\alpha \in \mathbb{N}_{\le 2d}} A_\alpha x^\alpha, \qquad \text{where } \mathbb{N}_{\le 2d} = \{\alpha \in \mathbb{N}^n : |\alpha| \le 2d\}. \tag{2.1}$$

If $f(x)$ is given as

$$f(x) = \sum_{\alpha \in \mathbb{N}_{\le 2d}} f_\alpha x^\alpha,$$

then $f(x)$ is SOS if and only if there exists a symmetric matrix X such that

$$A_\alpha \bullet X = f_\alpha \quad \forall\, \alpha \in \mathbb{N}_{\le 2d}, \quad X \succeq 0.$$

In the above $\bullet$ denotes the standard Frobinus inner product in matrix spaces. So checking whether $f(x)$ is SOS can be done by solving a semidefinite programming problem.

The standard form of a semidefinite programming is

$$\begin{aligned} \min \quad & C \bullet X \\ s.t. \quad & A_i \bullet X = b_i,\ i = 1, \dots, m, \\ & X \succeq 0. \end{aligned} \tag{2.2}$$

Here C and $A_1, \dots, A_m$ are constant symmetric matrices. Lasserre [11] showed that the SOS program (1.2) is equivalent to an SDP problem like (2.2). So (1.2) can be solved efficiently. SDP is a very nice convex optimization and has many attractive properties. There is a large amount of work on solving SDP efficiently and its applications. We refer to [36] for more details on the theory, algorithms and applications of semidefinite programming.

2.2 Norms of polynomials

If a polynomial $f(x)$ is given in the form

$$f(x) = \sum_{\alpha \in \mathbb{N}_{\leq 2d}} f_\alpha x^\alpha,$$

we define its 2-norm and G-norm as

$$\|f(x)\|_2 = \left(\sum_{\alpha \in \mathbb{N}_{\leq 2d}} f_\alpha^2 \right)^{1/2}, \quad \|f(x)\|_G = \left(\sum_{\alpha \in \mathbb{N}_{\leq 2d}} \mathfrak{p}(\alpha)^{-1} f_\alpha^2 \right)^{1/2}. \tag{2.3}$$

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

$$\mathfrak{p}(\alpha) = \left| \left\{ (\beta, \nu) \in \mathbb{N}_{\leq d} \times \mathbb{N}_{\leq d} :\, \beta + \nu = \alpha \right\} \right| \leq \binom{|\mathrm{supp}(\alpha)| + d}{d}. \tag{2.4}$$

Obviously, the norms $\|\cdot\|_2$ and $\|\cdot\|_G$ are equivalent and satisfy the relation

$$\binom{3d}{d}^{-1/2} \|f(x)\|_2 \leq \|f(x)\|_G \leq \|f(x)\|_2. \tag{2.5}$$

In view of (2.3), we denote the coefficient vectors

$$f = \left(f_\alpha : \alpha \in \mathbb{N}_{\leq 2d} \right), \quad f_G = \left(\mathfrak{p}(\alpha)^{-1/2} f_\alpha : \alpha \in \mathbb{N}_{\leq 2d} \right), \tag{2.6}$$

and denote by $[x]_{G,2d}$ the column vector of weighted monomials

$$[x]_{G,2d} = \left(\mathfrak{p}(\alpha)^{1/2} x^\alpha : \alpha \in \mathbb{N}_{\leq 2d} \right). \tag{2.7}$$

The components in f, f_G and $[x]_{G,2d}$ are ordered gradedly alphabetically according to their indices. Thus we have $f(x) = f^T[x]_{2d} = f_G^T[x]_{G,2d}$ and $\|f(x)\|_2 = \|f\|_2, \|f(x)\|_G = \|f_G\|_2$. The G-norm $\|f\|_G$ is closely related to Gram matrices of $f(x)$.

Lemma 2.1. *If $f(x) \in \mathbb{R}[x]_{\leq 2d}$, there exists a symmetric matrix W such that*

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

In particular, if $f(x) \in \mathbb{R}[x]_{0,2d}$, there exist $\tau \in \mathbb{R}$ and a symmetric matrix V such that

$$f(x) = \tau + [x^d]^T V [x^d], \quad \tau^2 + \|W\|_F^2 = \|f(x)\|_G^2.$$

Proof. For any matrix W satisfying $f(x) = [x]_d^T W [x]_d$, it must hold

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

Choose a particular W satisfying the above as follows

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

Then W is symmetric and satisfies

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

In particular, if $f(x) \in \mathbb{R}[x]_{0,2d}$, the above W has the zero pattern

$$W = \begin{bmatrix} \tau & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & V \end{bmatrix}.$$

Here τ is the $(1,1)$ entry of W, and V is the block whose rows and columns are indexed by monomials of degree equal to d. Obviously $\|W\|_F^2 = \tau^2 + \|V\|_F^2$. So the lemma follows. □

Another type of useful norms is the so-called L^2-norm. Throughout this paper, we always assume S is compact. So we can define

$$\|f(x)\|_{L^2(S)} = \left(\int_S f(x)^2 d\mu(x)\right)^{1/2}. \tag{2.8}$$

Here μ is the uniform probability measure on S. The $\|f(x)\|_{L^2(S)}$ defined in (2.8) is a norm in $\mathbb{R}[x]$ when S has nonempty interior. This is because if $\|f(x)\|_{L^2(S)} = 0$ then $f(x)$ vanishes on S, so $f(x)$ vanishes in an open set, and hence it must be an identically zero polynomial.

Now we define the so-called *marginal* L^2-norm of polynomials. Throughout this paper, we always assume $n \ge 2d$. Given a subset $\Delta \subset \{1, \ldots, n\}$ with $|\Delta| = 2d$, denote by x_Δ the subvector of x whose indices are in Δ, that is,

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

The restriction $f_\Delta(x_\Delta)$ of $f(x)$ to x_Δ is defined as

$$f_\Delta(x_\Delta) = f(\tilde{x}), \quad \text{where} \quad \tilde{x}_i = \begin{cases} x_i & \text{if } i \in \Delta, \\ 0 & \text{otherwise.} \end{cases} \tag{2.9}$$

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

$$\Omega_{2d} = \big\{\Delta \subset [n] : |\Delta| = 2d\big\}. \tag{2.10}$$

Obviously $|\Omega_{2d}| = \binom{n}{2d}$. We also denote by $g_i(x_\Delta)$ the restriction of $g_i(x)$ to x_Δ, and g_Δ denotes the tuple $(g_1(x_\Delta), \ldots, g_m(x_\Delta))$. The set S_Δ is then similarly defined as

$$S_\Delta = \{x_\Delta : g_1(x_\Delta) \ge 0, \ldots, g_m(x_\Delta) \ge 0\}. \tag{2.11}$$

If S_Δ has nonempty interior, the $L^2(S_\Delta)$-norm can be accordingly defined as

$$\|f_\Delta(x_\Delta)\|_{L^2(S_\Delta)} = \left(\int_{S_\Delta} f_\Delta(x_\Delta)^2 d\mu_\Delta(x_\Delta)\right)^{1/2},$$

where $\mu_\Delta(\cdot)$ is the uniform probability measure on S_Δ. Note that if the origin belongs to the interior of S, then every S_Δ has nonempty interior. When every $int(S_\Delta) \neq \emptyset$, if $f(x) \in \mathbb{R}[x]_{\leq 2d}$, we can define the marginal $L^2(S)$-norm of $f(x)$ as

$$\|f(x)\|_{L^2(S),mg} = \left(\sum_{\Delta \in \Omega_{2d}} \|f_\Delta(x_\Delta)\|^2_{L^2(S_\Delta)} \right)^{1/2}.$$

For every $\Delta \in \Omega_{2d}$, define matrix

$$\mathbf{\Theta}_\Delta(S) = \int_{S_\Delta} [x_\Delta]_{G,2d}[x_\Delta]^T_{G,2d} d\mu_\Delta(x_\Delta). \tag{2.12}$$

Obviously $\mathbf{\Theta}_\Delta(S) \succeq 0$. Let

$$\eta_{2d}(S) = \sqrt{\min_{\Delta \in \Omega_{2d}} \lambda_{min}(\mathbf{\Theta}_\Delta(S))}. \tag{2.13}$$

Note that $\eta_{2d}(S)$ is uniquely determined by the geometry of S but independent of the set of defining polynomials $g_1(x), \ldots, g_m(x)$. For a general semialgebraic set S, the matrix $\mathbf{\Theta}_\Delta(S)$ can be evaluated numerically, e.g., by semidefinite programming methods of Henrion, Lasserre and Savorgnan [6]. Hence $\eta_{2d}(S)$ can also be evaluated numerically. For special S, there might exist explicit formula for $\mathbf{\Theta}_\Delta(S)$, as will be shown later in Section 4.

Lemma 2.2. *If $int(S_\Delta) \neq \emptyset$ for every $\Delta \in \Omega_{2d}$, then $\eta_{2d}(S) > 0$. In particular, if $0 \in int(S)$, then $\eta_{2d}(S) > 0$.*

Proof. It suffices to prove every $\mathbf{\Theta}_\Delta(S)$ is positive definite. Prove by contradiction. Suppose there exist $u \neq 0$ such that $u^T \mathbf{\Theta}_\Delta(S) u = 0$ for some $\Delta \in \Omega_{2d}$. Then

$$u^T \mathbf{\Theta}_\Delta(S) u = \int_{S_\Delta} ([x_\Delta]^T_{G,2d} u)^2 d\mu_\Delta(x_\Delta) = 0.$$

So we obtain

$$[x_\Delta]^T_{G,2d} u = 0 \quad \forall\, x \in S_\Delta.$$

Since S_Δ has nonempty interior, $[x_\Delta]^T_{G,2d} u$ must be an identically zero polynomial. The monomials of $[x_\Delta]_{G,2d}$ are linearly independent, so we must have $u = 0$, which is a contradiction. Thus every $\mathbf{\Theta}_\Delta(S)$ must be positive definite, and $\eta_{2d}(S) > 0$.

If $0 \in int(S)$, then $0 \in int(S_\Delta)$ for every S_Δ. Thus $\eta_{2d}(S) > 0$ follows the first part. □

The norms $\|\cdot\|_{L^2(S),mg}$ and $\|\cdot\|_G$ are related by the following lemma.

Lemma 2.3. *Assume $n \geq 2d$. If $f(x), g_i(x) \in \mathbb{R}[x]_{\leq 2d}$, then $\|f(x)\|_{L^2(S),mg} \geq \eta_{2d}(S)\|f(x)\|_G$.*

Proof. By definitions of $L^2(S_\Delta)$-norm and $\eta_{2d}(S)$, we know

$$\|f_\Delta(x_\Delta)\|^2_{L^2(S_\Delta)} = f^T_{\Delta,G} \mathbf{\Theta}_\Delta(S) f_{\Delta,G} \geq \eta_{2d}(S)^2 \|f_\Delta(x_\Delta)\|^2_G.$$

Here $f_{\Delta,G}$ denotes the vector of weighted coefficients of $f_\Delta(x_\Delta)$ (see (2.6)). By definition of the marginal $L^2(S)$-norm, it holds

$$\|f(x)\|^2_{L^2(S),mg} = \sum_{\Delta \in \Omega_{2d}} \|f_\Delta(x_\Delta)\|^2_{L^2(S_\Delta)} \geq \eta_{2d}(S)^2 \sum_{\Delta \in \Omega_{2d}} \|f_\Delta(x_\Delta)\|^2_G \geq \eta_{2d}(S)^2 \|f(x)\|^2_G.$$

Thus the lemma follows. □

2.3 Relations between $\|\cdot\|_2$ and L^2 norms

Sometimes, we need estimate the relation between $\|\cdot\|_2$ and L^2 norms of polynomials. By definitions (2.3) and (2.8), if we write $f(x) = f^T[x]_{2d}$, then

$$\|f(x)\|^2_{L^2(S)} = f^T \underbrace{\left(\int_S [x]_{2d}[x]_{2d}^T d\mu(x)\right)}_{\mathbf{\Theta}(S)} f, \quad \|f(x)\|_2^2 = f^T f.$$

Hence, $\|f(x)\|_{L^2(S)}$ and $\|f(x)\|_2$ satisfy the relationship

$$\sqrt{\lambda_{min}(\mathbf{\Theta}(S))} \le \frac{\|f(x)\|_{L^2(S)}}{\|f(x)\|_2} \le \sqrt{\lambda_{max}(\mathbf{\Theta}(S))}.$$

Once we know one of $\|f(x)\|_{L^2(S)}$ and $\|f(x)\|_2$, the other can be estimated by using maximum and minimum eigenvalues of $\mathbf{\Theta}(S)$. If S is special like a unit ball or hypercube, $\mathbf{\Theta}(S)$ might be given by explicit formula. If S is general, $\mathbf{\Theta}(S)$ would be evaluated numerically, e.g., by the methods in [6]. Once $\mathbf{\Theta}(S)$ is available, its eigenvalues can be found numerically.

In certain situations, we need estimate the minimum eigenvalue of $\mathbf{\Theta}(S)$ in terms of dimension n and degree $2d$ for some special S, e.g., a hypercube. For this purpose, orthogonal polynomials are very useful. Let $P_k(t)$ be the k-th scaled *Legendre Polynomial* defined as

$$P_k(t) = \sqrt{2k+1} \cdot \frac{1}{2^k k!} \cdot \frac{d^k}{dt^k}\left(t^2-1\right)^k. \tag{2.14}$$

The first few of them are

$$P_0(t) = 1, \quad P_1(t) = \sqrt{3}t, \quad P_2(t) = \sqrt{5}\cdot\frac{1}{2}(3t^2-1), \quad P_3(t) = \sqrt{7}\cdot\frac{1}{2}(5t^3-3t),$$
$$P_4(t) = 3\cdot\frac{1}{8}(35t^4-30t^2+3), \quad P_5(t) = \sqrt{11}\cdot\frac{1}{8}(63t^5-70t^3+15t),$$
$$P_6(t) = \sqrt{13}\cdot\frac{1}{16}(231t^6-315t^4+105t^2-5).$$

It is well known that Legendre Polynomials are orthogonal over $[-1,1]$ with weight function 1, and it holds the relation

$$\int_{-1}^{1} P_i(t)P_j(t)dt = \begin{cases} 2 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{2.15}$$

Thus $\{P_k(t)\}_{k=0}^{\infty}$ forms an orthogonal basis for the space of univariate polynomials. Note $P_k(t)$ has only even terms when k is even, and has only odd terms when k is odd. We refer to [17] for more details about the properties of Legendre Polynomials. The univariate monomial t^k can be expressed as a linear combination of polynomials $P_0(t), P_1(t), \dots, P_k(t)$. A general formula is (see [17])

$$t^k = \sum_{\ell=k,k-2,\dots} \frac{(2\ell+1)k!}{2^{(k-\ell)/2}\left(\frac{1}{2}(k-\ell)\right)!(\ell+k+1)!!}\sqrt{\frac{1}{2\ell+1}}P_\ell(t). \tag{2.16}$$

Let $Sfr(\mathbb{N}_{\leq 2d})$ be the set of square free vectors from $\mathbb{N}_{\leq 2d}$, that is, $Sfr(\mathbb{N}_{\leq 2d}) = \mathbb{N}_{\leq 2d} \cap \{0,1\}^n$. For any $\gamma \in Sfr(\mathbb{N}_{\leq 2d})$, define integer

$$r(\gamma) = \Big\lfloor (2d - |\gamma|)/2 \Big\rfloor.$$

Denote by $[x]_{\ell,even}$ be the column vector of monomials which have even degrees in every component x_i and have total degrees not bigger than ℓ. For instance,

$$[x]_{4,even}^T = \begin{bmatrix} 1 & x_1^2 \cdots x_n^2 & x_1^4 & x_1^2x_2^2 \cdots x_1^2x_n^2 & x_2^4 & x_2^2x_3^2 \cdots x_{n-1}^2x_n^2 & x_n^4 \end{bmatrix}.$$

Lemma 2.4. *There exists a positive constant $\omega(d)$ depending only on d such that for every square free vector $\gamma \in \mathbb{N}_{\leq 2d}$, it holds*

$$\lambda_{min}\left(\frac{1}{2^n}\int_{[-1,1]^n} x^{2\gamma}\left([x]_{2r(\gamma),even}\right)\left([x]_{2r(\gamma),even}\right)^T dx\right) \geq \omega(d)\cdot \binom{n+r(\gamma)}{r(\gamma)}^{-1}.$$

Proof. Obviously $x^{2\gamma}\left([x]_{2r(\gamma),even}\right)\left([x]_{2r(\gamma),even}\right)^T = \left(x^\gamma[x]_{2r(\gamma),even}\right)\left(x^\gamma[x]_{2r(\gamma),even}\right)^T$. Every component of the vector $x^\gamma[x]_{2r(\gamma),even}$ takes the form

$$x^\gamma \cdot x^{2\nu}, \quad \nu \in \mathbb{N}_{\leq r(\gamma)}. \tag{2.17}$$

When $i \in \text{supp}(\gamma)$, the univariate polynomial $x_i^{1+2\nu_i}$ is a linear combination of $P_\ell(x_i)$ with odd $\ell \leq 1+2\nu_i$. When $i \notin \text{supp}(\gamma)$, $x_i^{2\nu_i}$ is a linear combination of $P_\ell(x_i)$ with even $\ell \leq 2\nu_i$. Therefore, the polynomial space spanned by the monomials of the form (2.17) is precisely the space spanned by polynomials of the form $P_\alpha(x) := P_{\alpha_1}(x_1)\cdots P_{\alpha_n}(x_n)$ where

$$(\alpha_1,\ldots,\alpha_n) \in \mathbb{N}_{\leq 2d}(\gamma) := \left\{\alpha \in \mathbb{N}_{\leq 2d}: \text{ each } \alpha_i \text{ is } \begin{array}{ll} \text{odd} & \text{if } i \in \text{supp}(\gamma) \\ \text{even} & \text{if } i \notin \text{supp}(\gamma) \end{array}\right\}. \tag{2.18}$$

Let $[P(x)]_{2d,\gamma}$ be the column vector of all above polynomials $P_\alpha(x)$ whose indices $\alpha \in \mathbb{N}_{\leq 2d}(\gamma)$ and are ordered gradedly lexicographically in α. By formula (2.16), for every $\alpha \in \mathbb{N}_{\leq 2d}(\gamma)$, there exist numbers c_β such that

$$x^\alpha = \sum_{\beta\in\mathbb{N}_{\leq 2d}(\gamma):\, \beta\leq\alpha} c_\beta \cdot P_{\beta_1}(x_1)\cdots P_{\beta_n}(x_n). \tag{2.19}$$

So there is a nonsingular lower triangular matrix L such that

$$x^\gamma[x]_{2r(\gamma),even} = L\cdot[P(x)]_{2d,\gamma},$$

and it holds

$$\begin{aligned} &\frac{1}{2^n}\int_{[-1,1]^n} x^{2\gamma}[x]_{2r(\gamma),even}[x]_{2r(\gamma),even}^T dx = \frac{1}{2^n}\int_{[-1,1]^n} L[P(x)]_{2d,\gamma}[P(x)]_{2d,\gamma}^T L^T dx \\ =&L\left(\frac{1}{2^n}\int_{[-1,1]^n}[P(x)]_{2d,\gamma}[P(x)]_{2d,\gamma}^T dx\right)L^T = LL^T. \end{aligned}$$

In the above, we have used the fact that the middle integral matrix there is the identity. This is because for any two entries $P_\alpha(x)$,$P_\beta(x)$ of $[P(x)]_{2d,\gamma}$ with $\alpha, \beta \in \mathbb{N}_{\le 2d}(\gamma)$, it holds

$$\frac{1}{2^n}\int_{[-1,1]^n} P_\alpha(x)\cdot P_\beta(x)dx = \prod_{i=1}^{n}\frac{1}{2}\int_{[-1,1]} P_{\alpha_i}(x_i)P_{\beta_i}(x_i)dx_i = \begin{cases}1 & \text{if } \alpha=\beta,\\ 0 & \text{if } \alpha\neq\beta.\end{cases}$$

Therefore, it suffices for us to prove there exists $\omega(d) > 0$ such that

$$\sigma_{min}(L) \ge \sqrt{\omega(d)}\cdot\binom{n+r(\gamma)}{r(\gamma)}^{-1/2}.$$

Here $\sigma_{min}(\cdot)$ denotes the smallest singular value of a matrix. So

$$\sigma_{min}(L) = \min_{u\neq 0}\frac{\|Lu\|_2}{\|u\|_2} = \min_{u\neq 0}\frac{\|u\|_2}{\|L^{-1}u\|_2} \ge \frac{1}{\|L^{-1}\|_2} \ge \frac{1}{\|L^{-1}\|_F}.$$

Now we turn to estimating $\|L^{-1}\|_F$. The dimension of L is $K\times K$ with $K = \binom{n+r(\gamma)}{r(\gamma)}$. Since L is nonsingular and lower triangular, there exist nonzero scalars $l_{11},\dots,l_{K,K}$ such that

$$L = \begin{bmatrix} 1 & & & & \\ l_{21} & 1 & & & \\ l_{31} & l_{32} & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ l_{K1} & l_{K2} & \cdots & l_{K,K-1} & 1 \end{bmatrix}\begin{bmatrix} l_{11} & & & & \\ & l_{22} & & & \\ & & l_{33} & & \\ & & & \ddots & \\ & & & & l_{KK} \end{bmatrix}.$$

The inverse L^{-1} has the form

$$L^{-1} = \begin{bmatrix} l_{11}^{-1} & & & & \\ & l_{22}^{-1} & & & \\ & & l_{33}^{-1} & & \\ & & & \ddots & \\ & & & & l_{KK}^{-1} \end{bmatrix}\begin{bmatrix} 1 & & & & \\ -l_{21} & 1 & & & \\ -l_{31} & -l_{32} & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ -l_{K1} & -l_{K2} & \cdots & -l_{K,K-1} & 1 \end{bmatrix}.$$

From the expression (2.19), we know the row of L indexed by $\alpha \in \mathbb{N}_{\le 2d}(\gamma)$ has at most

$$\prod_{i\in \text{supp}(\alpha)} \alpha_i \le (2d)^{2d}$$

nonzero entries. So every row of L^{-1} also has at most $(2d)^{2d}$ nonzero entries. Since the entries l_{ij} of L are independent of n, there exists a positive constant $M = M(d)$ such that

$$\left|L_{ij}^{-1}\right| \le M(d), \quad \forall\, 1\le j\le i\le K.$$

Thus all entries of L^{-1} are bounded by $M(d)$. So we get

$$\|L^{-1}\|_F = \sqrt{\sum_{1\le j\le i\le K}\left|L_{ij}^{-1}\right|^2} \le \sqrt{K\cdot(2d)^{2d}\cdot M(d)^2} = (2d)^d\cdot M(d)\cdot K^{1/2},$$

which then implies

$$\sigma_{min}(L) \ge \frac{1}{\|L^{-1}\|_F} \ge \frac{1}{(2d)^d M(d)}K^{-1/2}.$$

If we set $\omega(d) = \big((2d)^d M(d)\big)^{-2}$, the proof is completed. □

Proposition 2.5. *For every polynomial $f(x) \in \mathbb{R}[x]_{\leq 2d}$, it holds that*

$$\|f(x)\|_{L^2([-1,1]^n)} \geq \sqrt{\omega(d)} \binom{n+d}{d}^{-1/2} \cdot \|f\|_2,$$

where the constant $\omega(d)$ is given in Lemma 2.4.

Proof. For every $f(x) \in \mathbb{R}[x]_{\leq 2d}$, we can write it uniquely as

$$f(x) = \sum_{\gamma \in Sfr(\mathbb{N}_{\leq 2d})} x^\gamma \cdot \left((f^{(\gamma)})^T [x]_{2r(\gamma),even} \right)$$

for some vectors $f^{(\gamma)}$. Each $f_\gamma^T [x]_{2r(\gamma),even}$ has even degree in every x_i, and

$$\|f(x)\|^2_{L^2([-1,1]^n)} = \frac{1}{2^n} \int_{[-1,1]^n} f(x)^2 dx, \quad \|f\|_2^2 = \sum_{\gamma \in Sfr(\mathbb{N}_{\leq 2d})} \|f^{(\gamma)}\|_2^2.$$

For any two different $\alpha, \beta \in Sfr(\mathbb{N}_{\leq 2d})$, it must hold

$$\int_{[-1,1]^n} \left(x^\alpha (f^{(\alpha)})^T [x]_{2r(\alpha),even} \right) \left(x^\beta (f^{(\beta)})^T [x]_{2r(\beta),even} \right) dx = 0.$$

Thus we have

$$\int_{[-1,1]^n} f(x)^2 dx = \sum_{\gamma \in Sfr(\mathbb{N}_{\leq 2d})} \int_{[-1,1]^n} \left(x^\gamma \left((f^{(\gamma)})^T [x]_{2r(\gamma),even} \right) \right)^2 dx.$$

Lemma 2.4 implies

$$\begin{aligned}
&\frac{1}{2^n} \int_{[-1,1]^n} \left(x^\gamma (f^{(\gamma)})^T [x]_{2r(\gamma),even} \right)^2 dx \\
&= (f^{(\gamma)})^T \left(\frac{1}{2^n} \int_{[-1,1]^n} x^{2\gamma} [x]_{2r(\gamma),even} [x]^T_{2r(\gamma),even} dx \right) f^{(\gamma)} \\
&\geq \omega(d) \cdot \binom{n+r(\gamma)}{r(\gamma)}^{-1} \|f^{(\gamma)}\|_2^2.
\end{aligned}$$

Therefore, from the above, we get

$$\begin{aligned}
\|f(x)\|^2_{L^2([-1,1]^n)} &\geq \sum_{\gamma \in Sfr(\mathbb{N}_{\leq 2d})} \omega(d) \cdot \binom{n+r(\gamma)}{r(\gamma)}^{-1} \|f^{(\gamma)}\|_2^2 \\
&\geq \omega(d) \left(\min_{\gamma \in Sfr(\mathbb{N}_{\leq 2d})} \binom{n+r(\gamma)}{r(\gamma)}^{-1} \right) \sum_{\gamma \in Sfr(\mathbb{N}_{\leq 2d})} \|f^{(\gamma)}\|_2^2 \\
&\geq \omega(d) \binom{n+d}{d}^{-1} \|f\|_2^2,
\end{aligned}$$

which completes the proof. □

3 Some general bounds for Lasserre's relaxation

This section analyzes the approximation bound of Lasserre's relaxation (1.2). We assume $\deg(g) \leq 2d$ and S is a nonempty compact set. Thus, for every polynomial $f(x)$, it has a minimum f_{min} and a maximum f_{max} on S.

As we mentioned in Introduction, to make Lasserre's relaxations converge, we need assume $g_1, \ldots, g_m$ satisfy AC. However, even when AC holds, (1.2) may not have a feasible solution, i.e., $f_{sos} > -\infty$ is finite, for a fixed relaxation order d, though an arbitrarily good lower bound can be obtained when d goes to infinity. To guarantee (1.2) is always feasible and $f_{sos} > -\infty$ for all $f(x) \in \mathbb{R}[x]_{\leq 2d}$, we need the following assumption.

Assumption 3.1. *There exist a symmetric positive definite matrix $E \succ 0$ and SOS polynomials $\sigma_1(x), \ldots, \sigma_m(x)$ such that each $\deg(\sigma_i g_i) \leq 2d$ and*

$$\sigma_1(x)g_1(x) + \cdots + \sigma_m(x)g_m(x) = 1 - [x]_d^T E [x]_d.$$

The following proposition shows Assumption 3.1 is sufficient and necessary to guarantee $f_{sos} > -\infty$ for all $f(x) \in \mathbb{R}[x]_{\leq 2d}$.

Proposition 3.2. *Lasserre's relaxation (1.2) is feasible and $f_{sos} > -\infty$ for all $f(x) \in \mathbb{R}[x]_{\leq 2d}$ if and only if Assumption 3.1 holds.*

Proof. " $\Leftarrow$ " Every $f(x) \in \mathbb{R}[x]_{\leq 2d}$ can be written as $f(x) = [x]_d^T F [x]_d$ for some symmetric matrix F. Since E is positive definite, we can choose $\lambda > 0$ big enough such that

$$\sigma_0(x) := f(x) + \lambda [x]_d^T E [x]_d = [x]_d^T (F + \lambda E)[x]_d$$

is SOS. Then choose $\gamma = -\lambda$, and we get the identity

$$f(x) - \gamma = \sigma_0(x) + \lambda\sigma_1(x)g_1(x) + \cdots + \lambda\sigma_m(x)g_m(x).$$

Therefore (1.2) has a feasible solution and $f_{sos} > -\infty$.

" $\Rightarrow$ " Consider the special polynomial $\hat{f}(x) = -[x]_d^T [x]_d$. Since (1.2) is feasible, there exist $\hat{\gamma}$ and SOS polynomials $\hat{\sigma}_0(x), \hat{\sigma}_1(x), \ldots, \hat{\sigma}_m(x)$ such that each $\deg(\hat{\sigma}_i g_i) \leq 2d$ and

$$-[x]_d^T [x]_d - \hat{\gamma} = \hat{\sigma}_1(x)g_1(x) + \cdots + \hat{\sigma}_m(x)g_m(x) + \hat{\sigma}_0(x).$$

For any $u \in S$, the right hand side above must be nonnegative. So $-\hat{\gamma} \geq [u]_d^T [u]_d > 0$ and

$$\frac{1}{-\hat{\gamma}}\Big(\hat{\sigma}_1(x)g_1(x) + \cdots + \hat{\sigma}_m(x)g_m(x)\Big) = 1 - \frac{1}{-\hat{\gamma}}\left([x]_d^T [x]_d + \hat{\sigma}_0(x)\right).$$

Therefore, Assumption 3.1 holds. □

In Assumption 3.1, the choice of SOS polynomials $\sigma_1(x), \ldots, \sigma_m(x)$ and positive definite matrix E may not be unique. In our later approximation analysis, the bigger $\lambda_{min}(E)$ is, the better the obtained approximation bound would be. So we want $\sigma_1(x), \ldots, \sigma_m(x)$ and E

such that $\lambda_{min}(E)$ is as large as possible. Fortunately, the best choice $\sigma_1^*(x), \dots, \sigma_m^*(x)$ and E^* can be found by solving the following SOS program:

$$\begin{array}{rl} \max\limits_{\sigma_1(x),\dots,\sigma_m(x),E} & \lambda_{min}(E) \\ s.t. & \sigma_1(x)g_1(x) + \cdots + \sigma_m(x)g_m(x) = 1 - [x]_d^T E[x]_d, \\ & \sigma_1(x), \dots, \sigma_m(x) \text{ are SOS}, \\ & \deg(\sigma_1 g_1), \dots, \deg(\sigma_m g_m) \le 2d. \end{array} \tag{3.1}$$

Note (3.1) is equivalent to an SDP problem, and can be solved efficiently by numerical methods. Throughout this section, assume $\sigma_1^*(x), \dots, \sigma_m^*(x)$ and E^* are optimal for (3.1). Assumption 3.1 holds if and only if $\lambda_{min}(E^*) > 0$, so it is checkable by solving (3.1).

Let $\mathcal{F}$ be a subspace of $\mathbb{R}[x]_{\le 2d}$. Define a constant associated with $\mathcal{F}$ and S

$$\chi(\mathcal{F}, S) := \max_{p \in \mathcal{F}} \Big\{ \|p\|_G : |p(x)| \le 1 \quad \forall\, x \in S \Big\}. \tag{3.2}$$

When S has nonempty interior, $\chi(\mathcal{F}, S) < \infty$. When S has empty interior, $\chi(\mathcal{F}, S)$ might be infinite for some $\mathcal{F}$. For instance, if S is the unit sphere $\mathbb{S}^{n-1}$ and $\mathcal{F} = \mathbb{R}[x]_{\le 2d}$, then

$$\chi(\mathbb{R}[x]_{\le 2d}, \mathbb{S}^{n-1}) = \infty.$$

This is because for polynomials $p_k(x) = k(1 - \|x\|_2^2)$ it holds

$$\|p_k(x)\|_G \to \infty \text{ as } k \to \infty, \qquad |p_k(x)| \le 1 \quad \forall\, x \in \mathbb{S}^{n-1}.$$

So, if S has empty interior, to make $\chi(\mathcal{F}, S) < \infty$, $\mathcal{F}$ should not intersect the subspace

$$\mathcal{V}(S) = \Big\{ p(x) \in \mathbb{R}[x] : p(x) = 0 \quad \forall\, x \in S \Big\}$$

except at the zero polynomial. In this case, if we minimize nonzero polynomials from $\mathcal{F}$ over S, $\mathcal{F}$ should be chosen to be a subspace of the modulo space $\mathbb{R}[x]/\mathcal{V}(S)$. Obviously, if S has nonempty interior, $\mathcal{V}(S) = \{0\}$ is the singleton of the identically zero polynomial.

Theorem 3.3. *Suppose $\mathcal{F}$ is a subspace of $\mathbb{R}[x]_{\le 2d}$ containing 1, $\chi(\mathcal{F}, S) < \infty$, Assumption 3.1 holds, and the tuple $(\sigma_1^*(x), \dots, \sigma_m^*(x), E^*)$ is optimal for (3.1). Let $f(x) \in \mathcal{F}$, and f_{min} (resp. f_{max}) be its minimum (resp. maximum) on S. If f_{sos} is the optimal value of Lasserre's relaxation (1.2), then it holds*

$$1 \le \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \le \frac{\chi(\mathcal{F}, S)}{\lambda_{min}(E^*)}.$$

Proof. Define the median of $f(x)$ on S as

$$med(f) = \frac{1}{2}(f_{min} + f_{max}) \in [f_{min}, f_{max}].$$

Without loss of generality, assume $f_{min} < med(f) < f_{max}$, because otherwise if $f_{min} = f_{max}$, then $f(x)$ is constant and the theorem is obviously true. Let

$$\tilde{f}(x) = \frac{f(x) - med(f)}{med(f) - f_{min}} \in \mathcal{F}.$$

Then $|\tilde{f}(x)| \leq 1$ for all $x \in S$ and $\left\| \tilde{f}(x) \right\|_G \leq \chi(\mathcal{F}, S)$ by definition (3.2). Now set

$$\theta^* = \frac{\chi(\mathcal{F}, S)}{\lambda_{min}(E^*)} > 0, \qquad \gamma^* = med(f) - \theta^* \big(med(f) - f_{min}\big). \tag{3.3}$$

Thus we have

$$\left\| \frac{1}{\theta^*} \tilde{f}(x) \right\|_G \leq \lambda_{min}(E^*).$$

So, by Lemma 2.1, there exists a Gram matrix W such that

$$\frac{1}{\theta^*} \tilde{f}(x) = [x]_d^T W [x]_d, \quad \|W\|_F \leq \lambda_{min}(E^*),$$

$$\frac{1}{\theta^*} \tilde{f}(x) + [x]_d^T E^* [x]_d = [x]_d^T (W + E^*) [x]_d.$$

Since $\|W\|_2 \leq \|W\|_F \leq \lambda_{min}(E)$, we know $W + E \succeq 0$. Hence the polynomial

$$\sigma_0(x) := (med(f) - f_{min}) \Big(\tilde{f}(x) + \theta^* [x]_d^T E^* [x]_d \Big)$$

must be SOS. Let $\sigma_i(x) = (med(f) - f_{min}) \theta^* \sigma_i^*(x)$ for each i, which are all SOS. From

$$\sigma_1^*(x) g_1(x) + \cdots + \sigma_m^*(x) g_m(x) = 1 - [x]_d^T E^* [x]_d,$$

we have the identity

$$f(x) - \gamma^* = \sigma_0(x) + \sigma_1(x) g_1(x) + \cdots + \sigma_m(x) g_m(x).$$

So $\sigma_0(x), \sigma_1(x), \ldots, \sigma_m(x)$ and γ^* are feasible in (1.2). By optimality of f_{sos}, we have $f_{sos} \geq \gamma^*$. By the choice of γ^* in (3.3), it holds

$$\frac{med(f) - f_{sos}}{med(f) - f_{min}} \leq \frac{\chi(\mathcal{F}, S)}{\lambda_{min}(E^*)}.$$

Since $med(f) \in [f_{min}, f_{max}]$ and $f_{sos} \leq f_{min}$, the above implies the theorem. $\square$

Theorem 3.3 can be applied to get an approximation bound if we can estimate $\chi(\mathcal{F}, S)$ for given $\mathcal{F}$ and S. Define a constant associated to the set S

$$\kappa_{2d}(S) = \min_{p \in \mathbb{R}[x]_{\leq 2d}} \left\{ \|p(x)\|_{L^2(S)} : \|p(x)\|_2 = 1 \right\}. \tag{3.4}$$

If we write $p(x) = p^T [x]_{2d}$, then $\|p(x)\|_{L^2(S)}^2 = p^T \left(\int_S [x]_{2d} [x]_{2d}^T d\mu(x) \right) p$, where $\mu(\cdot)$ is the uniform probability measure on S. So it holds

$$\kappa_{2d}(S) = \sqrt{\lambda_{min} \left(\int_S [x]_{2d} [x]_{2d}^T d\mu(x) \right)}.$$

The constant $\kappa_{2d}(S)$ can be evaluated numerically by methods like in [6].

Proposition 3.4. *If $int\big(S\big) \neq \emptyset$, then $\kappa_{2d}(S) > 0$ and*

$$\chi(\mathbb{R}[x]_{\leq 2d}, S) \leq \frac{1}{\kappa_{2d}(S)}.$$

Proof. Since $int\big(S\big) \neq \emptyset$, for any nonzero polynomial $p(x)$, $\|p(x)\|_{L^2(S)} > 0$. So $\kappa_{2d}(S)$ is the minimum of a positive continuous function defined on a compact set. Thus $\kappa_{2d}(S) > 0$.

Now we prove the second part. Suppose $p(x) \in \mathbb{R}[x]_{\leq 2d}$ is such that $|p(x)| \leq 1$ for all $x \in S$. Then $\|p(x)\|_{L^2(S)} \leq 1$ and (3.4) implies $1 \geq \kappa_{2d}(S)\|p(x)\|_2$. Therefore, we get

$$\|p(x)\|_G \leq \|p(x)\|_2 \leq \frac{1}{\kappa_{2d}(S)},$$

which completes the proof. □

Obviously Theorem 3.3 and Proposition 3.4 imply the following theorem.

Theorem 3.5. *Suppose $int\big(S\big) \neq \emptyset$, Assumption 3.1 holds, and $\sigma_1^*(x), \dots, \sigma_m^*(x), E^*$ are optimal for (3.1). Let $f(x) \in \mathbb{R}[x]_{\leq 2d}$, and f_{min} (resp. f_{max}) be its minimum (resp. maximum) on S. If f_{sos} is the optimal value of Lasserre's relaxation (1.2), then*

$$1 \leq \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \leq \frac{1}{\kappa_{2d}(S)\lambda_{min}(E^*)}.$$

Sometimes it would be quite difficult to estimate $\kappa_{2d}(S)$ for given S. So we need other kind of estimates for $\chi(\mathbb{R}[x]_{\leq 2d}, S)$.

Proposition 3.6. *Assume $n \geq 2d$. If $0 \in int\big(S\big)$, then*

$$\chi(\mathbb{R}[x]_{\leq 2d}, S) \leq \frac{1}{\eta_{2d}(S)}\sqrt{\binom{n}{2d}}.$$

Proof. Suppose $p(x) \in \mathbb{R}[x]_{\leq 2d}$ and $|p(x)| \leq 1$ for all $x \in S$. Since $0 \in int\big(S\big)$, the restrictions $p_\Delta(x_\Delta)$ of $p(x)$ must satisfy

$$|p_\Delta(x_\Delta)| \leq 1 \quad \forall\, x_\Delta \in S_\Delta.$$

By definition of the marginal $L^2(S)$-norm, we have

$$\|p(x)\|^2_{L^2(S),mg} = \sum_{\Delta \in \Omega_{2d}} \int_{S_\Delta} p_\Delta(x_\Delta)^2 d\mu_\Delta(x_\Delta) \leq \sum_{\Delta \in \Omega_{2d}} 1 = \binom{n}{2d}. \tag{3.5}$$

Since $int\big(S\big) \neq \emptyset$, Lemma 2.2 implies $\eta_{2d}(S) > 0$. Therefore, by Lemma 2.3, we get

$$\|p(x)\|_G \leq \frac{1}{\eta_{2d}(S)}\|p(x)\|_{L^2(S),mg}$$

which results in the proposition by applying (3.5). □

The following is then implied by Theorem 3.3 and Proposition 3.6.

Theorem 3.7. *Suppose $n \geq 2d$, $0 \in int\big(S\big)$, Assumption 3.1 holds, and $\sigma_1^*(x), \ldots, \sigma_m^*(x), E^*$ are optimal for (3.1). Let $f(x) \in \mathbb{R}[x]_{\leq 2d}$, and f_{min} (resp. f_{max}) be its minimum (resp. maximum) on S. If f_{sos} is the optimal value of Lasserre's relaxation (1.2), then*

$$1 \leq \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \leq \frac{1}{\eta_{2d}(S)\lambda_{min}(E^*)} \sqrt{\binom{n}{2d}},$$

where $\eta_{2d}(S)$ is given by (2.13).

Remark 3.8. We would like to mention that the $\lambda_{min}(E^*)$ in Theorem 3.7 is closely related to the "radius" of the set S. Let R be the radius of S, that is, $R = \max_{x \in S} \|x\|_2$. From

$$\|x\|_2^{2k} = (x_1^2 + \cdots + x_n^2)^k = \sum_{\alpha \in \mathbb{N}(k)} x^{2\alpha} \frac{k!}{\alpha_1! \cdots \alpha_n!} \leq k! \sum_{\alpha \in \mathbb{N}(k)} x^{2\alpha} = k! \|[x^k]\|_2^2$$

where $N(k) = \{\alpha \in \mathbb{N}^n : |\alpha| = k\}$, we get $1 + \|x\|_2^2 + \cdots + \|x\|_2^{2d} \leq d! \|[x]_d\|_2^2$. Since

$$1 \geq [x]_d^T E^* [x]_d \geq \lambda_{min}(E^*) \|[x]_d\|_2^2 \quad \forall x \in S,$$

it holds

$$\lambda_{min}(E^*) \leq \frac{1}{1 + R^2 + \cdots + R^{2d}}.$$

Therefore, we can see that

$$\frac{1}{\lambda_{min}(E^*)} \geq 1 + R^2 + \cdots + R^{2d}.$$

So the approximation bound given by Theorem 3.7 is at least

$$(1 + R^2 + \cdots + R^{2d}) \eta_{2d}(S)^{-1} \sqrt{\binom{n}{2d}}. \tag{3.6}$$

The polynomial $1 - (1 + R^2 + \cdots + R^{2d})^{-1} [x]_d^T [x]_d$ is nonnegative on S. If fortunately there exist SOS polynomials $s_0(x), s_1(x), \ldots, s_m(x)$ such that every $\deg(s_i g_i) \leq 2d$ and

$$1 - (1 + R^2 + \cdots + R^{2d})^{-1} [x]_d^T [x]_d = s_0(x) + s_1(x) g_1(x) + \cdots + s_m(x) g_m(x),$$

then there exists a symmetric matrix $\hat{E}$ be such that

$$[x]_d^T \hat{E} [x]_d = (1 + R^2 + \cdots + R^{2d})^{-1} [x]_d^T I [x]_d + s_0(x),$$
$$1 - [x]_d^T \hat{E} [x]_d = s_1(x) g_1(x) + \cdots + s_m(x) g_m(x).$$

Since $s_0(x)$ is SOS, we know

$$\frac{1}{\lambda_{min}(E^*)} \leq \frac{1}{\lambda_{min}(\hat{E})} \leq 1 + R^2 + \cdots + R^{2d}.$$

In this case, $\lambda_{min}(E^*) = 1 + R^2 + \cdots + R^{2d}$. The assumption on the existence of above $s_i(x)$ is nontrivial. However, if R is known in advance, we can add the redundant constraint

$$1 - (1 + R^2 + \cdots + R^{2d})^{-1} [x]_d^T [x]_d \geq 0$$

to (1.1), and then have $\lambda_{min}(E^*) = 1 + R^2 + \cdots + R^{2d}$. □

4 Bounds for some special optimization problems

Theorems 3.3, 3.5 and 3.7 give approximation bounds for Lasserre's relaxation (1.2) in terms of different constants. Given a specific problem (1.1), we need estimate $\lambda_{min}(E^*)$ and one of $\chi(\mathcal{F}, S)$, $\kappa_{2d}(S)$ and $\eta_{2d}(S)$ to get an explicit bound. In this section, we show how to do this for some special problems.

4.1 Optimizing polynomials over unit balls

Consider polynomial optimization over the unit ball $B(0,1) = \{x \in \mathbb{R}^n : \|x\|_2 \le 1\}$

$$\min_{x \in B(0,1)} \quad f(x). \tag{4.1}$$

When $f(x) \in \mathbb{R}[x]_{\le 2d}$ has degree $2d$, the d-th Lasserre's relaxation is

$$\begin{array}{rl} \max & \gamma \\ s.t. & f(x) - \gamma = \sigma_0(x) + \sigma_1(x)(1 - \|x\|_2^2), \\ & \deg(\sigma_0) \le 2d,\ \deg(\sigma_1) \le 2d-2, \\ & \sigma_0(x), \sigma_1(x) \text{ are SOS.} \end{array} \tag{4.2}$$

Let $g(x) = 1 - \|x\|_2^2$. We are going to get an approximation bound for (4.2) by applying Theorem 3.7. So $\eta_{2d}\big(B(0,1)\big)$ and $\lambda_{min}(E^*)$ in (3.1) need to be estimated.

First, we estimate $\eta_{2d}\big(B(0,1)\big)$. By definition (2.13), $\eta_{2d}\big(B(0,1)\big) = \sqrt{\lambda_{min}\left(\boldsymbol{\Theta}_\Delta\big(B(0,1)\big)\right)}$ for any $\Delta \in \Omega_{2d}$ because of symmetry. So we just choose $\Delta = \{1, \ldots, 2d\}$. Then

$$\boldsymbol{\Theta}_\Delta\big(B(0,1)\big) = \int_{\|x_\Delta\|_2 \le 1} [x_\Delta]_{G,2d}[x_\Delta]_{G,2d}^T d\mu_\Delta(x_\Delta),$$

where $\mu_\Delta(\cdot)$ is the uniform probability measure on $\{x_\Delta : \|x_\Delta\|_2 \le 1\}$. Thus

$$\boldsymbol{\Theta}_\Delta\big(B(0,1)\big) = \frac{1}{Vol(\|x_\Delta\|_2 \le 1)} \int_{\|x_\Delta\|_2 \le 1} [x_\Delta]_{G,2d}[x_\Delta]_{G,2d}^T dx_\Delta$$

where dx_Δ is the standard Lebesgue measure. For $\alpha = (\alpha_1, \ldots, \alpha_{2d})$, if one α_i is odd, then $\int_{\|x_\Delta\|_2 \le 1} x_\Delta^\alpha dx_\Delta = 0$. If α is an even vector, then

$$\int_{\|x_\Delta\|_2 \le 1} x_\Delta^\alpha dx_\Delta = Area(\mathbb{S}^{2d-1}) \cdot \int_{\|x_\Delta\|_2 = 1} x_\Delta^\alpha d\nu_\Delta(x_\Delta) \cdot \int_0^1 r^{|\alpha| + 2d - 1} dr.$$

In the above, $Area(\mathbb{S}^{2d-1})$ is area of the unit sphere $\mathbb{S}^{2d-1}$, and $\nu_\Delta(\cdot)$ is the uniform probability measure on $\mathbb{S}^{2d-1}$. Note the formulae

$$Area(\mathbb{S}^{2d-1}) = \frac{2\pi^d}{\Gamma(d)}, \quad Vol(\|x_\Delta\|_2 \le 1) = \frac{\pi^d}{\Gamma(1+d)},$$

$$\int_{\|x_\Delta\|_2 = 1} x_\Delta^\alpha d\nu_\Delta(x_\Delta) = \frac{\Gamma(d) \prod_{i=1}^{2d} \Gamma(\beta_i + 1/2)}{\pi^d \Gamma(|\beta| + d)}.$$

$2d$	2	4	6	8
$\eta_{2d}\big(B(0,1)\big)$	0.19204	0.01670	0.00161	0.00004

Table 1: A list of $\eta_{2d}\big(B(0,1)\big)$ for $2d = 2, 4, 6, 8$.

Here $\Gamma(\cdot)$ is the standard Gamma function. So, if $\alpha = 2\beta = 2(\beta_1, \dots, \beta_{2d})$ is even, we get

$$\int_{\|x_\Delta\|_2 \le 1} x_\Delta^\alpha d\mu_\Delta(x_\Delta) = \frac{\Gamma(1+d)\prod_{i=1}^{2d}\Gamma(\beta_i + 1/2)}{\pi^d(|\beta|+d)\Gamma(|\beta|+d)}.$$

The entries of $\mathbf{\Theta}_\Delta\big(B(0,1)\big)$ can be found explicitly, and $\eta_{2d}\big(B(0,1)\big)$ is a constant depending only on d. A list of typical values of $\eta_{2d}\big(B(0,1)\big)$ is in Table 1.

Second, we estimate $\lambda_{min}(E^*)$ in (3.1). For any integer $k \ge 1$, it holds

$$(1 + t + \cdots + t^{k-1})(1-t) = 1 - t^k.$$

Let $s_d(t) = \frac{1}{d+1}\sum_{k=1}^{d}\sum_{j=0}^{k-1} t^j$. Then we get the identity

$$s_d(t)(1-t) = 1 - \frac{1}{d+1}\left(1 + t + \cdots + t^d\right).$$

Plugging t by $\|x\|_2^2$ in the above, we get

$$s_d(\|x\|_2^2)(1-\|x\|_2^2) = 1 - \frac{1}{d+1}\left(1 + \|x\|_2^2 + \cdots + \|x\|_2^{2d}\right).$$

Since $s_d(\|x\|_2^2)$ is SOS and has degree $2d-2$, there exists a symmetric E such that

$$\frac{1}{d+1}\left(1 + \|x\|_2^2 + \cdots + \|x\|_2^{2d}\right) = [x]_d^T E [x]_d, \quad \lambda_{min}(E) \ge \frac{1}{d+1}.$$

So Assumption 3.1 holds for $g(x) = 1 - \|x\|_2^2$, and the optimal $\lambda_{min}(E^*) \ge \frac{1}{d+1}$. Therefore Theorem 3.7 implies the following.

Theorem 4.1. *Assume $n \ge 2d$. Let $f(x) \in \mathbb{R}[x]_{\le 2d}$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) on $B(0,1)$. If f_{sos} is the optimal value of SOS relaxation (4.2), then*

$$1 \le \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \le \frac{d+1}{\eta_{2d}\big(B(0,1)\big)}\sqrt{\binom{n}{2d}}.$$

So f_{sos} is an $\mathcal{O}(n^d)$-approximation of f_{min}.

4.2 Optimizing polynomials over hypercubes

Consider polynomial optimization over the hypercube $[-1,1]^n$

$$\begin{array}{rl} \min\limits_{x\in\mathbb{R}^n} & f(x) \\ s.t. & 1 - x_1^2 \ge 0, \dots, 1 - x_n^2 \ge 0. \end{array} \tag{4.3}$$

When $f(x)$ has degree $2d$, the d-th Lasserre's relaxation for the above is

$$\begin{array}{rl} \max & \gamma \\ s.t. & f(x) - \gamma = \sigma_0(x) + \sigma_1(x)(1-x_1^2) + \cdots + \sigma_n(x)(1-x_n^2), \\ & \deg(\sigma_0) \le 2d,\ \deg(\sigma_1), \ldots, \deg(\sigma_n) \le 2d-2, \\ & \sigma_0(x), \sigma_1(x), \ldots, \sigma_n(x) \text{ are SOS.} \end{array} \tag{4.4}$$

Let $g_1(x) = 1 - x_1^2, \ldots, g_n(x) = 1 - x_n^2$. To get an approximation bound for (4.4) by applying Theorem 3.5, we need estimate $\kappa_{2d}([-1,1]^n)$ and $\lambda_{min}(E^*)$ in (3.1).

First, we estimate $\kappa_{2d}([-1,1]^n)$. For any $p(x) \in \mathbb{R}[x]_{\le 2d}$, Proposition 2.5 implies

$$\|p(x)\|^2_{L^2([-1,1]^n)} \ge \omega(d) \binom{n+d}{d}^{-1} \cdot \|p\|_2^2$$

where the constant $\omega(d)$ is given in Lemma 2.4. By definition (3.4), it then holds

$$\kappa_{2d}([-1,1]^n) \ge \sqrt{\omega(d)} \binom{n+d}{d}^{-1/2}.$$

Second, we estimate $\lambda_{min}(E^*)$. Note all points in $[-1,1]^n$ satisfy

$$\frac{1}{n}\sum_{i=1}^n (1 - x_i^2) = 1 - \frac{1}{n}\|x\|_2^2 \ge 0.$$

For polynomial $s_d(t) = \frac{1}{d+1}\sum_{k=1}^d \sum_{j=0}^{k-1} t^j$, it holds

$$s_d(t)(1-t) = 1 - \frac{1}{d+1}\left(1 + t + \cdots + t^d\right).$$

Plugging t by $\frac{1}{n}\|x\|_2^2$ in the above, we get

$$s_d\left(\frac{1}{n}\|x\|_2^2\right)\left(1 - \frac{1}{n}\|x\|_2^2\right) = 1 - \frac{1}{d+1}\left(1 + \frac{1}{n}\|x\|_2^2 + \cdots + \frac{1}{n^d}\|x\|_2^{2d}\right).$$

There exists E such that

$$\frac{1}{d+1}\left(1 + \frac{1}{n}\|x\|_2^2 + \cdots + \frac{1}{n^d}\|x\|_2^{2d}\right) = [x]_d^T E [x]_d, \quad \lambda_{min}(E) \ge \frac{1}{d+1} n^{-d}.$$

So we get

$$\frac{1}{n} s_d\left(\frac{1}{n}\|x\|_2^2\right)\left(\sum_{i=1}^n (1 - x_i^2)\right) = 1 - [x]_d^T E [x]_d, \quad \lambda_{min}(E) \ge \frac{1}{d+1} n^{-d}.$$

Note $\frac{1}{n} s_d\left(\frac{1}{n}\|x\|_2^2\right)$ is SOS. So Assumption 3.1 holds and the optimal E^* in (3.1) satisfies

$$\lambda_{min}(E^*) \ge \frac{1}{d+1} n^{-d}.$$

The following then follows Theorem 3.5 immediately.

Theorem 4.2. *Let $f(x) \in \mathbb{R}[x]_{\leq 2d}$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) on $[-1,1]^n$. If f_{sos} is the optimal value of SOS relaxation (4.4), then it holds*

$$1 \leq \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \leq \frac{d+1}{\sqrt{\omega(d)}} \binom{n+d}{d}^{1/2} n^d$$

where $\omega(d)$ is given by Lemma 2.4. So f_{sos} is an $\mathcal{O}(n^{\frac{3d}{2}})$-approximation of f_{min}.

Using Theorem 3.7 as in subsection 4.1, we can get an approximation bound $\mathcal{O}(n^{2d})$ for (4.4), which is worse than the bound $\mathcal{O}(n^{\frac{3d}{2}})$ given by Theorem 4.2.

4.3 Optimizing square free polynomials over hypercubes

Consider problem (4.3) for the special case that $f(x)$ there is square free, that is,

$$f(x) = \sum_{\gamma \in Sfr(\mathbb{N}_{\leq 2d})} f_\gamma x^\gamma, \quad \text{where} \quad Sfr(\mathbb{N}_{\leq 2d}) = \mathbb{N}_{\leq 2d} \cap \{0,1\}^n.$$

Denote by $Sfr[x]_{\leq k}$ the space of all square free polynomials having degrees at most k.

Lemma 4.3. *It holds that*

$$\chi(Sfr[x]_{\leq k}, [-1,1]^n) \leq \sqrt{3}^k.$$

Proof. Let $p(x) \in Sfr[x]_{\leq k}$ be such that $|p(x)| \leq 1$ for all $x \in [-1,1]^n$. Write $p(x)$ as

$$p(x) = \sum_{\gamma \in Sfr(\mathbb{N}_{\leq k})} p_\gamma x^\gamma.$$

Simple integration shows that

$$1 \geq \frac{1}{2^n} \int_{[-1,1]^n} p(x)^2 dx = \sum_{\gamma \in Sfr(\mathbb{N}_{\leq k})} p_\alpha^2 \frac{1}{2^n} \int_{[-1,1]^n} x^{2\gamma} dx = \sum_{\gamma \in Sfr(\mathbb{N}_{\leq k})} p_\alpha^2 3^{-|\gamma|} \geq 3^{-k} \|p\|_2^2.$$

Therefore, we have

$$\|p(x)\|_G \leq \|p(x)\|_2 \leq \sqrt{3}^k.$$

By definition (3.2), the lemma follows immediately. □

So Lemma 4.3 implies

$$\chi(Sfr[x]_{\leq 2d}, [-1,1]^n) \leq 3^d.$$

From subsection 4.2, we know $\lambda_{min}(E^*) \geq \frac{1}{d+1} n^{-d}$ for the hypercube $[-1,1]^n$. Thus the following theorem follows Theorem 3.3.

Theorem 4.4. *Let $f(x) \in Sfr[x]_{\leq 2d}$ be a square free polynomial, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) on $[-1,1]^n$. If f_{sos} is the optimal value of (4.4), then*

$$f_{max} - f_{min} \leq f_{max} - f_{sos} \leq (d+1) \cdot (3n)^d (f_{max} - f_{min}).$$

4.4 Optimizing polynomials over ± 1 and $0/1$ sets

Consider the problem of optimizing polynomials over the discrete set $\{\pm 1\}^n$:

$$\min \quad f(x) \quad s.t. \quad x_i^2 = 1,\, i = 1, \ldots, n. \tag{4.5}$$

The constraints of (4.5) are equivalent to

$$g_i(x) := 1 - x_i^2 \geq 0,\, g_{n+i}(x) := -1 + x_i^2 \geq 0,\, i = 1, \ldots, n.$$

When $f(x)$ has degree $2d$, the d-th Lasserre's relaxation for (4.5) is equivalent to

$$\begin{array}{rl} \max & \gamma \\ s.t. & f(x) - \gamma = \sigma(x) + \phi_1(x)(1 - x_1^2) + \cdots + \phi_n(x)(1 - x_n^2), \\ & \deg(\sigma) \leq 2d,\, \deg(\phi_1), \ldots, \deg(\phi_n) \leq 2d - 2, \\ & \sigma(x) \text{ is SOS.} \end{array} \tag{4.6}$$

The approximation bound of (4.6) is summarized as follows.

Theorem 4.5. *Let $f(x) \in \mathbb{R}[x]_{\leq 2d}$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) on $\{\pm 1\}^n$. If f_{sos} is the optimal value of (4.6), then*

$$f_{max} - f_{min} \leq f_{max} - f_{sos} \leq (d+1)n^d (f_{max} - f_{min}).$$

Proof. From subsection 4.2, we know there exists an SOS polynomial $s(x)$ such that

$$s(x)(g_1(x) + \cdots + g_n(x)) = 1 - [x]_d^T E [x]_d, \quad \lambda_{min}(E) \geq \frac{1}{d+1} n^{-d}.$$

So the optimal E^* in (3.1) for the set $\{\pm 1\}^n$ must satisfy $\lambda_{min}(E^*) \geq \frac{1}{d+1} n^{-d}$.

First, assume $f(x)$ is a square free polynomial, i.e., $f(x) \in Sfr[x]_{\leq 2d}$. We claim that

$$\chi(Sfr[x]_{\leq 2d}, \{\pm 1\}^n) \leq 1. \tag{4.7}$$

To see this, suppose $p(x) \in Sfr[x]_{\leq 2d}$ and $|p(x)| \leq 1$ for all $x \in \{\pm 1\}^n$. Write $p(x)$ as

$$p(x) = \sum_{\gamma \in Sfr(\mathbb{N}_{\leq 2d})} p_\gamma x^\gamma.$$

Then we have

$$1 \geq \frac{1}{2^n} \sum_{u \in \{\pm 1\}^n} p(u)^2 = \sum_{\gamma \in Sfr(\mathbb{N}_{\leq 2d})} p_\gamma^2 \cdot \frac{1}{2^n} \sum_{u \in \{\pm 1\}^n} u^{2\gamma} = \sum_{\gamma \in Sfr(\mathbb{N}_{\leq 2d})} p_\gamma^2 = \|p\|_2^2.$$

Hence, $\|p(x)\|_G \leq \|p\|_2 \leq 1$, and (4.7) is true by definition (3.2). So Theorem 3.3 implies $f_{max} - f_{sos} \leq (d+1)n^d (f_{max} - f_{min})$ when $f(x)$ is a square free polynomial.

Second, if $f(x)$ is general, there exists a square free $\hat{f}(x) \in Sfr[x]_{\leq 2d}$ such that

$$f(x) = \hat{f}(x) \quad \forall\, x \in \{\pm 1\}^n.$$

By the previous argument, we also have $f_{max} - f_{sos} \leq (d+1)n^d (f_{max} - f_{min})$. Since $f_{sos} \leq f_{min}$ is always true, the theorem is proven. □

After a linear coordinate transformation, the approximation bound in Theorem 4.5 also holds if we optimize polynomials over $\{0, 1\}^n$ via Lasserre's relaxation. So we get

Theorem 4.6. *Let $f(x) \in \mathbb{R}[x]_{\leq 2d}$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) on $\{0, 1\}^n$. If f_{sos} is the lower bound given by the d-th Lasserre's relaxation, then*

$$f_{max} - f_{min} \leq f_{max} - f_{sos} \leq (d+1)n^d (f_{max} - f_{min}).$$

4.5 Optimizing polynomials over quadratically constrained sets

A quite interesting problem is to minimize polynomials over quadratic constraints, i.e., every $g_i(x) = [x]_1^T Q_i [x]_1$ is a quadratic polynomial. Problem (1.1) then becomes

$$\begin{array}{rl} \min\limits_{x\in\mathbb{R}^n} & f(x) \\ s.t. & [x]_1^T Q_i [x]_1 \geq 0,\ i = 1,\ldots,m. \end{array} \tag{4.8}$$

If Assumption 3.1 holds, Theorem 3.7 can be applied to get an approximation bound. To estimate $\lambda_{min}(E^*)$ in (3.1), there exist simpler methods than solving (3.1) directly.

If we choose $d = 1$, then (3.1) becomes

$$\begin{array}{rl} \max\limits_{\lambda=(\lambda_1,\ldots,\lambda_m),A} & \lambda_{min}(A) \\ s.t. & \lambda_1 \cdot [x]_1^T Q_1 [x]_1 + \cdots + \lambda_m \cdot [x]_1^T Q_m [x]_1 = 1 - [x]_1^T A [x]_1, \\ & \lambda_1,\ldots,\lambda_m \geq 0, \end{array} \tag{4.9}$$

which is then equivalent to the SDP

$$\begin{array}{rl} \max\limits_{\lambda=(\lambda_1,\ldots,\lambda_m),A} & \lambda_{min}(A) \\ s.t. & \lambda_1 Q_1 + \cdots + \lambda_m Q_m + A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \\ & \lambda_1,\ldots,\lambda_m \geq 0. \end{array}$$

In the above, each 0 denotes a zero block of proper dimensions. Generally, (4.9) is much easier to solve than (3.1) does. Let (λ^*, A^*) be an optimal solution of (4.9).

In the sequel, we will prove that if $A^* \succ 0$, then Assumption 3.1 holds. Let $s(t) = \frac{1}{d}\sum_{k=1}^{d}\sum_{j=0}^{k-1} t^j$, then

$$s(t)(1-t) = 1 - \frac{1}{d}\left(t + \cdots + t^d\right).$$

Plugging t by $[x]_1^T A^* [x]_1$ in the above, we get

$$s\big([x]_1^T A^*[x]_1\big)(1 - [x]_1^T A^*[x]_1) = 1 - \frac{1}{d}\Big([x]_1^T A^*[x]_1 + \cdots + ([x]_1^T A^*[x]_1)^d\Big).$$

From (4.9), we know $1 - [x]_1 A^* [x]_1 \geq 0$ for all $x \in S$ and

$$1 \geq [x]_1 A^* [x]_1 \geq \lambda_{min}(A^*)(1 + \|x\|_2^2).$$

So there exists a symmetric E such that

$$\frac{1}{d}\left([x]_1^T A^*[x]_1 + \cdots + ([x]_1^T A^*[x]_1)^d\right) = [x]_d^T E [x]_d, \quad \lambda_{min}(E) \geq \frac{1}{d}\big(\lambda_{min}(A^*)\big)^d.$$

Let $\sigma_i(x) = \lambda_i^* s\big([x]_1^T A^*[x]_1\big)$ for each i, which are all SOS. Then we get

$$\sigma_1(x) g_1(x) + \cdots + \sigma_m(x) g_m(x) = 1 - [x]_d^T E [x]_d.$$

So Assumption 3.1 holds. The optimal E^* in (3.1) for optimization (4.8) satisfies

$$\lambda_{min}(E^*) \geq \frac{1}{d}\big(\lambda_{min}(A^*)\big)^d.$$

Therefore, the following is implied by Theorem 3.7.

Theorem 4.7. *Assume every $g_i(x) = [x]_1^T Q_i [x]_1$ is quadratic and the optimal A^* in (4.9) is positive definite. Let $f(x) \in \mathbb{R}[x]_{\leq 2d}$, and f_{min} (resp., f_{max}) be the minimum (resp., maximum) objective value of (4.8). If f_{sos} is the optimal value of the d-th Lasserre's relaxation for (4.8), then*

$$1 \leq \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \leq \frac{d}{\eta_{2d}(S)\big(\lambda_{min}(A^*)\big)^d} \sqrt{\binom{n}{2d}}.$$

Remark 4.8. In the special case that every $g_i(x) = [x]_1^T Q_i [x]_1$ is concave, that is, the quadratic homogeneous part of $g_i(x)$ is negative semidefinite., we can get $\lambda_{min}(A^*)$ exactly. Let $R = \max_{x \in S} \|x\|_2$ be the radius of S. Then $1 \geq [x]_1^T A^* [x]_1$ for all $x \in S$ and

$$1 \geq \lambda_{min}(A^*)(1 + \|x\|_2^2) \quad \forall x \in S.$$

So we get $\lambda_{min}(A^*) \leq (1 + R^2)^{-1}$. On the other hand, the concave function $1 - (1 + R^2)^{-1}[x]_1^T [x]_1$ is nonnegative on the convex set S. If $int\big(S\big) \neq \emptyset$, there exist $\hat{\lambda}_1 \geq 0, \ldots, \hat{\lambda}_m \geq 0$ such that

$$1 - \frac{1}{1 + R^2}[x]_1^T [x]_1 = \hat{\lambda}_1 [x]_1^T Q_1 [x]_1 + \cdots + \hat{\lambda}_m [x]_1^T Q_m [x]_1. \tag{4.10}$$

Thus $(\hat{\lambda}_1, \ldots, \hat{\lambda}_m, \frac{1}{1+R^2} I_{n+1})$ is feasible for (4.9) and $\lambda_{min}(A^*) \geq (1 + R^2)^{-1}$. So

$$\lambda_{min}(A^*) = (1 + R^2)^{-1}.$$

Therefore, if every $g_i(x)$ is concave, the bound in Theorem 4.7 becomes

$$d \cdot \eta_{2d}(S)^{-1}(1 + R^2)^d \sqrt{\binom{n}{2d}}.$$

If some $g_i(x) = [x]_1^T Q_i [x]_1$ are not concave, then it would be quite difficult to estimate $\lambda_{min}(A^*)$ in (4.9). This is because there might not exist nonnegative scalars $\hat{\lambda}_1, \ldots, \hat{\lambda}_m$ such that (4.10) holds. However, if we know R in advance, a redundant quadratic constraint

$$1 - (1 + R^2)^{-1}[x]_1^T [x]_1 \geq 0$$

can be added to (4.8), and then we have $\lambda_{min}(A^*) = (1 + R^2)^{-1}$. □

Let us end this section with an example.

Example 4.9 (Multi-unit balls)**.** Suppose $x = (x^{(1)}, \ldots, x^{(m)})$ where each $x^{(i)}$ is an n_i-dimensional vector and $n_1 + \cdots + n_m = n$, each $g_i(x) = 1 - \|x^{(i)}\|_2^2$, and every $n_i \geq 2d$. The set S is a so-called multi-unit ball. Note the identity

$$\frac{1}{m+1}(g_1(x) + \cdots + g_m(x)) = \frac{1}{m+1}(m - \|x\|_2^2) = 1 - [x]_1^T \frac{I_{n+1}}{m+1}[x]_1.$$

So we know the optimal A^* of (4.9) satisfies $\lambda_{min}(A^*) \geq \frac{1}{m+1}$. Obviously $\eta_{2d}(S)$ is a positive constant depending only on d. Then Theorem 4.7 implies an approximation bound $\mathcal{O}((mn)^d)$ holds for Lasserre's relaxation (1.2) if we minimize polynomials over a multi-unit ball. □

5 Homogeneous polynomial optimization

Let $f(x), h_1(x), \dots, h_m(x) \in \mathbb{R}[x]_{2d}$ be forms of degree $2d$. Consider the optimization

$$\begin{array}{rl} \min\limits_{x\in\mathbb{R}^n} & f(x) \\ s.t. & h_1(x) \leq 1, \dots, h_m(x) \leq 1. \end{array} \tag{5.1}$$

Denote the tuple $h(x) = (h_1(x), \dots, h_m(x))$. The feasible set in the above is a *polysoid*

$$\mathcal{P}(h) = \{x \in \mathbb{R}^n : h_1(x) \leq 1, \dots, h_m(x) \leq 1\}.$$

Since every $h_i(x)$ is homogeneous, $0 \in int\big(\mathcal{P}(h)\big)$. If we set $g(x) = \big(1-h_1(x), \dots, 1-h_m(x)\big)$, then Lemma 2.2 implies $\eta_{2d}\big(\mathcal{P}(h)\big) > 0$ (see definition (2.13)).

The d-th Lasserre's relaxation for (5.1) is

$$\begin{array}{rl} \max & \gamma \\ s.t. & f(x) - \gamma - s_1 \cdot (1 - h_1(x)) - \cdots - s_m \cdot (1 - h_m(x)) \text{ is SOS,} \\ & s_1 \geq 0, \dots, s_m \geq 0. \end{array} \tag{5.2}$$

Since all $f(x)$ and $h_i(x)$ are forms of degree $2d$, (5.2) is equivalent to

$$\begin{array}{rl} \max & -(s_1 + \cdots + s_m) \\ s.t. & f(x) + s_1 h_1(x) + \cdots + s_m h_m(x) \text{ is SOS,} \\ & s_1 \geq 0, \dots, s_m \geq 0. \end{array} \tag{5.3}$$

Problem (5.3) is equivalent to an SDP problem, and hence can be solved efficiently.

Since (5.1) is a special case of (1.1), we would think of applying Theorem 3.3, 3.5, or 3.7 to get an approximation bound for (5.2). However, since every $h_i(x)$ is a form of degree $2d$, Assumption 3.1 does not hold for polynomials $1 - h_1(x), \dots, 1 - h_m(x)$. This is because for any $s_1 \geq 0, \dots, s_m \geq 0$, the polynomial $s_1(1 - h_1(x)) - \cdots - s_m(1 - h_m(x))$ does not have monomials of degrees between 1 and $2d - 1$. So there is no positive definite E such that

$$s_1\big(1 - h_1(x)\big) - \cdots - s_m\big(1 - g_m(x)\big) = 1 - [x]_d^T E [x]_d.$$

However, in view of (5.3), Assumption 3.1 can be slightly modified as follows.

Assumption 5.1. *There exist $\lambda = (\lambda_1, \dots, \lambda_m) \geq 0$ and a symmetric matrix E such that*

$$\lambda_1 h_1(x) + \cdots + \lambda_m h_m(x) = [x^d]^T E [x^d], \quad \lambda_1 + \cdots + \lambda_m = 1, \quad \lambda_{min}(E) > 0.$$

Denote by f_{sos} the optimal value of (5.2). To guarantee (5.2) is feasible and $f_{sos} > -\infty$ for every form $f(x)$ of degree $2d$, Assumption 5.1 is sufficient and necessary.

Proposition 5.2. *Lasserre's relaxation (5.2) is feasible and $f_{sos} > -\infty$ for every form $f(x) \in \mathbb{R}[x]_{2d}$ of degree $2d$ if and only if Assumption 5.1 holds.*

Proposition 5.2 can be proved by almost the same argument as for Proposition 3.2. So we leave it for interested readers as an exercise.

In Assumption 5.2, the choice of λ and E might not be unique. To get better approximation bounds, we want an E whose $\lambda_{min}(E)$ is as large as possible. Similar to (3.1), the best λ^* and E^* can be found by solving the following SOS program:

$$\begin{array}{rl} \max\limits_{\lambda_1,\dots,\lambda_m,E} & \lambda_{min}(E) \\ s.t. & \lambda_1 h_1(x) + \cdots + \lambda_m h_m(x) = [x^d]^T E [x^d], \\ & \lambda_1 + \cdots + \lambda_m = 1, \lambda_1 \geq 0, \dots, \lambda_m \geq 0. \end{array} \tag{5.4}$$

The SOS program (5.4) is equivalent to an SDP problem, and can be solved efficiently by numerical methods. Let (λ^*, E^*) be an optimal solution of (5.4). Assumption 5.1 holds if and only if $\lambda_{min}(E^*) > 0$, so it is verifiable by solving (5.4).

Let $\mathbb{R}[x]_{0,2d}$ be the space of constant polynomials and forms of degree $2d$, and $\mathcal{F} \subseteq \mathbb{R}[x]_{0,2d}$ be a subspace. Recall that

$$\chi(\mathcal{F}, \mathcal{P}(h)) = \max_{p \in \mathcal{F}} \left\{ \|p(x)\|_G : |p(x)| \leq 1 \quad \forall\, x \in \mathcal{P}(h) \right\}.$$

Theorem 5.3. *Suppose Assumption 5.1 holds, (λ^*, E^*) is optimal for (5.4), and $\mathcal{F}$ is a subspace of $\mathbb{R}[x]_{0,2d}$ containing 1. Let $f(x) \in \mathcal{F}$, and f_{min} (resp. f_{max}) be its minimum (resp. maximum) on the polysoid $\mathcal{P}(h)$. If f_{sos} is the optimal value of (5.2), then*

$$1 \leq \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \leq \left(1 + \frac{1}{\lambda_{min}(E^*)}\right) \chi(\mathcal{F}, \mathcal{P}(h)).$$

Proof. The proof is almost the same as for Theorem 3.3. Here we follow the same approach and only list the distinctive parts. Set

$$med(f) = \frac{1}{2}(f_{min} + f_{max}), \quad \tilde{f}(x) = \frac{f(x) - med(f)}{med(f) - f_{min}} \in \mathcal{F}.$$

Then $|\tilde{f}(x)| \leq 1$ for all $x \in \mathcal{P}(h)$, and it holds $\|\tilde{f}(x)\|_G \leq \chi(\mathcal{F}, \mathcal{P}(h))$. Fix two constants

$$\theta^* = \frac{\chi(\mathcal{F}, \mathcal{P}(h))}{\lambda_{min}(E^*)} > 0, \quad \gamma^* = med(f) - \theta^*(1 + \lambda_{min}(E^*)) \cdot (med(f) - f_{min}).$$

Then $\left\| \frac{1}{\theta^*} \tilde{f}(x) \right\|_G \leq \lambda_{min}(E^*)$ and $\tilde{f}(x) \in \mathbb{R}[x]_{0,2d}$. By Lemma 2.1, there exist τ and V such that

$$\frac{1}{\theta^*} \tilde{f}(x) = \tau + [x^d]^T V [x^d], \quad \sqrt{\tau^2 + \|V\|_F^2} = \left\| \frac{1}{\theta^*} \tilde{f}(x) \right\|_G \leq \lambda_{min}(E^*).$$

Thus we have $\lambda_{min}(V) \leq \|V\|_F \leq \lambda_{min}(E^*)$, $|\tau| \leq \lambda_{min}(E^*)$, and

$$\begin{aligned} \sigma_0(x) &:= (med(f) - f_{min}) \left(\tilde{f}(x) + \theta^* \left(\lambda_{min}(E^*) + [x^d]^T E^* [x^d] \right) \right) \\ &= (med(f) - f_{min}) \theta^* \Big(\tau + \lambda_{min}(E^*) + [x^d]^T (V + E^*) [x^d] \Big) \end{aligned}$$

must be SOS. Set $s_i = (med(f) - f_{min}) \theta^* \lambda_i^* \geq 0$. From

$$\lambda_1^* h_1(x) + \cdots + \lambda_m^* h_m(x) = [x^d]^T E^* [x^d], \quad \lambda_1^* + \cdots + \lambda_m^* = 1,$$

we obtain the identity

$$f(x) - \gamma^* = \sigma_0(x) + s_1 \cdot (1 - h_1(x)) + \cdots + s_m \cdot (1 - h_m(x)).$$

By optimality of f_{sos}, we have $f_{sos} \geq \gamma^*$. Then the choice of γ^* implies

$$\frac{med(f) - f_{sos}}{med(f) - f_{min}} \leq \left(1 + \frac{1}{\lambda_{min}(E^*)}\right) \chi(\mathcal{F}, \mathcal{P}(h)).$$

Since $med(f) \in [f_{min}, f_{max}]$ and $f_{sos} \leq f_{min}$, the theorem follows. □

For any compact polysoid $\mathcal{P}(h)$, we can similarly define a constant like (3.4)

$$\kappa_{2d}\big(\mathcal{P}(h)\big) = \min_{p \in \mathbb{R}[x]_{0,2d}} \big\{ \|p(x)\|_{L^2(\mathcal{P}(h))} : \|p(x)\|_2 = 1 \big\}. \tag{5.5}$$

For a polynomial $p(x) = \begin{bmatrix} 1 & [x^d]^T \end{bmatrix}^T p$, it holds

$$\|p(x)\|^2_{L^2(\mathcal{P}(h))} = p^T \left(\int_{\mathcal{P}(h)} \begin{bmatrix} 1 \\ [x^d] \end{bmatrix} \begin{bmatrix} 1 \\ [x^d] \end{bmatrix}^T d\mu(x) \right) p,$$

where $\mu(\cdot)$ is the uniform probability measure on $\mathcal{P}(h)$. So

$$\kappa_{2d}\big(\mathcal{P}(h)\big) = \sqrt{\lambda_{min} \left(\int_{\mathcal{P}(h)} \begin{bmatrix} 1 \\ [x^d] \end{bmatrix} \begin{bmatrix} 1 \\ [x^d] \end{bmatrix}^T d\mu(x) \right)},$$

and it can be evaluated by numerical methods like in [6]. Since the origin is in the interior of $\mathcal{P}(h)$, we always have $\kappa_{2d}\big(\mathcal{P}(h)\big) > 0$.

Proposition 5.4. *Assume $n \geq 2d$. For any compact polysoid $\mathcal{P}(h)$, it holds*

$$\chi(\mathbb{R}[x]_{0,2d}, \mathcal{P}(h)) \leq \min \left\{ \frac{1}{\kappa_{2d}\big(\mathcal{P}(h)\big)}, \frac{1}{\eta_{2d}\big(\mathcal{P}(h)\big)} \sqrt{\binom{n}{2d}} \right\}.$$

Proof. Let $p(x) \in \mathbb{R}[x]_{0,2d}$ be such that $|p(x)| \leq 1$ for all $x \in \mathcal{P}(h)$. First, obviously $\|p(x)\|_{L^2(\mathcal{P}(h))} \leq 1$ and $1 \geq \kappa_{2d}(\mathcal{P}(h))\|p(x)\|_2$. So we get

$$\|p(x)\|_G \leq \|p(x)\|_2 \leq \frac{1}{\kappa_{2d}\big(\mathcal{P}(h)\big)}.$$

Second, for any $\Delta \in \Omega(2d)$, the restriction $p_\Delta(x_\Delta)$ of $p(x)$ must satisfy

$$|p_\Delta(x_\Delta)| \leq 1 \quad \forall x_\Delta \in S_\Delta.$$

By definition of the marginal $L^2\big(\mathcal{P}(h)\big)$-norm, we have

$$\|p(x)\|^2_{L^2\left(\mathcal{P}(h)\right),mg} = \sum_{\Delta \in \Omega_{2d}} \int_{S_\Delta} p_\Delta(x_\Delta)^2 d\mu_\Delta(x_\Delta) \leq \sum_{\Delta \in \Omega_{2d}} 1 = \binom{n}{2d}.$$

Therefore, Lemma 2.3 implies

$$\|p(x)\|_G \leq \frac{1}{\eta_{2d}\big(\mathcal{P}(h)\big)} \|p(x)\|_{L^2\left(\mathcal{P}(h)\right),mg} \leq \frac{1}{\eta_{2d}\big(\mathcal{P}(h)\big)} \sqrt{\binom{n}{2d}},$$

which completes the proof. □

Obviously Theorem 3.3 and Proposition 5.4 imply the following.

Theorem 5.5. *Suppose $n \geq 2d$, Assumption 5.1 holds and (λ^*, E^*) is optimal for (5.4). Let $f(x)$ be a form of degree $2d$, and f_{min} (resp. f_{max}) be its minimum (resp. maximum) on the polysoid $\mathcal{P}(h)$. If f_{sos} is the optimal value of SOS relaxation (5.2), then it holds*

$$1 \leq \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \leq \left(1 + \frac{1}{\lambda_{min}(E^*)}\right) \cdot \min\left\{\frac{1}{\kappa_{2d}\big(\mathcal{P}(h)\big)}, \frac{1}{\eta_{2d}\big(\mathcal{P}(h)\big)}\sqrt{\binom{n}{2d}}\right\},$$

where $\eta_{2d}\big(\mathcal{P}(h)\big)$ is defined by (2.13).

Remark 5.6. Similar to Remark 3.8, we can estimate $\lambda_{min}(E^*)$ in Theorem 5.5 in terms of the radius of the polysoid $\mathcal{P}(h)$. Let $R = \max_{x \in \mathcal{P}(h)} \|x\|_2$. From

$$1 - [x^d]E^*[x^d] = \lambda_1^*(1 - h_1(x)) + \cdots + \lambda_m^*(1 - h_m(x)) \geq 0 \quad \forall\, x \in \mathcal{P}(h),$$

$$1 \geq [x^d]E^*[x^d] \geq \lambda_{min}(E^*)\|[x^d]\|_2^2 \geq \frac{\lambda_{min}(E^*)}{d!}\|x\|_2^{2d} \quad \forall\, x \in \mathcal{P}(h),$$

we know

$$\lambda_{min}(E^*) \leq \frac{d!}{R^{2d}}.$$

The polynomial $1 - \frac{d!}{R^{2d}}\|[x^d]\|_2^2$ is nonnegative on $\mathcal{P}(h)$. If fortunately there exist scalars $\hat{\lambda}_1 \geq 0, \ldots, \hat{\lambda}_m \geq 0$ and an SOS polynomial $s(x)$ such that

$$s(x) + \frac{d!}{R^{2d}}\|[x^d]\|_2^2 = \hat{\lambda}_1 h_1(x) + \cdots + \hat{\lambda}_m h_m(x), \quad \hat{\lambda}_1 + \cdots + \hat{\lambda}_m = 1,$$

then $\lambda_{min}(E^*) \geq \frac{d!}{R^{2d}}$. It is nontrivial to assume the existence of the above $\lambda_i \geq 0$. However, if R is known in advance, we can add the redundant constraint $1 - \frac{d!}{R^{2d}}\|[x^d]\|_2^2 \geq 0$ to (5.1), and then have $\lambda_{min}(E^*) = \frac{d!}{R^{2d}}$. □

Example 5.7. Suppose x has a partition $(x^{(1)}, \ldots, x^{(m)})$, where each $x^{(i)}$ is n_i-dimensional and $n_1 + \cdots + n_m = n$, and each $h_i(x) = \|x^{(i)}\|_{2d}^{2d}$. Thus each $h_i(x) \leq 1$ defines a unit ball in the $x^{(i)}$-space under the standard $2d$-norm. It is shown in Example 5.4 of [26] that

$$\sum_{i=1}^{n} x_i^{2d} \geq n\binom{n+d-1}{d}^{-1}\|[x^d]\|_2^2.$$

So we get

$$\frac{1}{m}(h_1(x) + \cdots + h_m(x)) = \frac{1}{m}\sum_{i=1}^{n} x_i^{2d} \geq \frac{n}{m}\binom{n+d-1}{d}^{-1}\|[x^d]\|_2^2,$$

and the optimal E^* of (5.4) satisfies

$$\lambda_{min}(E^*) \geq \mathcal{O}(m^{-1}n^{1-d}).$$

The constant $\eta_{2d}(S)$ is also independent of n. So Theorem 5.5 implies (5.2) is an $\mathcal{O}(mn^{2d-1})$ approximation for (5.1) in this special case. □

6 Sparse Lasserre's relaxation

Due to the computational cost of Lasserre's relaxation (1.2), it it important to exploit the sparsity patterns of polynomials in practical applications. There has been a large amount of work on sparse polynomial optimization. Much larger polynomial optimization problems can be solved by exploiting sparsity, as shown by the work [8, 10, 25, 30, 35]. Under a certain condition, Lasserre [12] proved the convergence of a sparse version of Lasserre's relaxation when its order goes to infinity. But there is very few work on analyzing the approximation bound for a fixed relaxation order. This section will address this issue.

Suppose $g_1(x), \dots, g_m(x)$ in (1.1) are sparse polynomials, that is, each $|\mathrm{supp}(g_i)| \ll \binom{n+2d}{2d}$, and $\deg(g) \le 2d$. A typical sparsity pattern is each g_i involves only a few variables. For an index set $I \subset \{1, \dots, n\}$, denote

$$x_I = \big(x_i : i \in I\big),$$

that is, x_I is a subvector of x whose indices belong to I. Denote by $\mathbb{R}[x_I]$ the subring of real polynomials in x_I, and set $\mathbb{R}[x_I]_{\le 2d} = \mathbb{R}[x_I] \cap \mathbb{R}[x]_{\le 2d}$. Suppose $I_1, \dots, I_m \subset \{1, \dots, n\}$ are index sets such that

$$g_1(x) = g_1(x_{I_1}), \dots, g_m(x) = g_m(x_{I_m}).$$

Throughout this section, we assume S is a nonempty compact set, and $n \ge 2d$.

Let $f(x) \in \mathbb{R}[x_{I_1}]_{\le 2d} + \cdots + \mathbb{R}[x_{I_m}]_{\le 2d}$. Consider the sparse polynomial optimization

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ s.t. \quad & g_1(x_{I_1}) \ge 0, \dots, g_m(x_{I_m}) \ge 0. \end{aligned} \tag{6.1}$$

A sparse version of the d-th Lasserre's relaxation is

$$\begin{aligned} \max \quad & \gamma \\ s.t. \quad & f(x) - \gamma = \sum_{i=1}^m \sigma_{0,i}(x_{I_i}) + \sum_{i=1}^m \sigma_i(x_{I_i}) g_i(x_{I_i}), \\ & \deg(\sigma_{0,i}), \deg(\sigma_i g_i) \le 2d, \ i = 1, \dots, m, \\ & \sigma_{0,i}, \sigma_i \text{ are SOS}, \ i = 1, \dots, m. \end{aligned} \tag{6.2}$$

For convenience, we still denote the above optimal value by f_{sos}. Our goal is to estimate a bound Q such that

$$f_{max} - f_{sos} \le Q \cdot (f_{max} - f_{min}).$$

To estimate Q above, we need define the marginal L^2 type norm for sparse polynomials. For $p(x) \in \mathbb{R}[x_{I_1}]_{\le 2d} + \cdots + \mathbb{R}[x_{I_m}]_{\le 2d}$, we say a subset $\Phi \subseteq \Omega_{2d}$ is a *cover* of $p(x)$ if for every $\alpha \in \mathrm{supp}(p)$, there exists $\Delta \in \Phi$ such that $\mathrm{supp}(\alpha) \subseteq \Delta$. Let $\Omega(p)$ be a cover of $p(x)$ whose cardinality is the smallest, that is,

$$\Omega(p) = \underset{\Phi \subseteq \Omega_{2d}}{argmin} \Big\{ |\Phi| : \ \Phi \text{ is a cover of } p(x) \Big\}. \tag{6.3}$$

Like (2.9), denote by $p_\Delta(x_\Delta)$ the restriction of $p(x)$ to x_Δ, and define S_Δ by (2.11). If $S_\Delta \ne \emptyset$, we can define

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

where μ_Δ is the uniform probability measure on S_Δ. If every $int(S_\Delta) \neq \emptyset$, a sparse marginal L^2-norm of $p(x)$ can be defined as

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

To guarantee the sparse Lasserre's relaxation (6.2) is always feasible and $f_{sos} > -\infty$ for every $f(x) \in \mathbb{R}[x_{I_1}]_{\le 2d} + \cdots + \mathbb{R}[x_{I_m}]_{\le 2d}$, we need the following assumption.

Assumption 6.1. *There exist sparse SOS polynomials $\sigma_1(x_{I_1}), \ldots, \sigma_m(x_{I_m})$ and positive definite matrices $E_1, \ldots, E_m$ such that every $\deg(\sigma_i g_i) \le 2d$ and*

$$\sigma_1(x_{I_1})g_1(x_{I_1}) + \cdots + \sigma_m(x_{I_m})g_m(x_{I_m}) = 1 - \sum_{i=1}^m [x_{I_i}]_d^T E_i [x_{I_i}]_d.$$

Assumption 6.1 is sufficient and necessary for guaranteeing (6.2) is always feasible for all certain sparse $f(x)$, as shown by the following proposition.

Proposition 6.2. *The sparse Lasserre's relaxation (6.2) is feasible and $f_{sos} > -\infty$ for all $f(x) \in \mathbb{R}[x_{I_1}]_{\le 2d} + \cdots + \mathbb{R}[x_{I_m}]_{\le 2d}$ if and only if Assumption 6.1 holds.*

Proof. "$\Leftarrow$" Every $f(x) \in \mathbb{R}[x_{I_1}]_{\le 2d} + \cdots + \mathbb{R}[x_{I_m}]_{\le 2d}$ can be written as

$$f(x) = f_1(x_{I_1}) + \cdots + f_m(x_{I_m}).$$

Because every $E_i \succ 0$, there exists $\lambda > 0$ big enough such that all

$$\sigma_{0,i}(x_{I_i}) := f_i(x_{I_i}) + \lambda [x_{I_i}]_d^T E [x_{I_i}]_d$$

are SOS. Letting $\gamma = -\lambda$, we get

$$f(x) - \gamma = \sum_{i=1}^m \sigma_{0,i}(x_{I_i}) + \lambda\sigma_1(x_{I_1})g_1(x_{I_1}) + \cdots + \lambda\sigma_m(x_{I_m})g_m(x_{I_m}).$$

Thus (6.2) is feasible and $f_{sos} \ge \gamma > -\infty$.

"$\Rightarrow$" Consider the special sparse polynomial

$$\hat{f}(x) = -\sum_{i=1}^m [x_{I_i}]_d^T [x_{I_m}]_d \in \mathbb{R}[x_{I_1}]_{\le 2d} + \cdots + \mathbb{R}[x_{I_m}]_{\le 2d}.$$

Since (6.2) is feasible, there exist $\hat{\gamma}$ and $\hat{\sigma}_{0,i}(x_{I_i}), \hat{\sigma}_i(x_{I_i})$ feasible for (6.2) such that

$$-\sum_{i=1}^m [x_{I_i}]_d^T [x_{I_m}]_d - \hat{\gamma} = \sum_{i=1}^m \Big(\hat{\sigma}_{0,i}(x_{I_i}) + \hat{\sigma}_i(x_{I_i}) g_i(x_{I_i}) \Big).$$

For any point $u \in S$, it holds $-\hat{\gamma} \ge \sum_{i=1}^m [u_{I_i}]_d^T [u_{I_m}]_d > 0$. So $-\hat{\gamma} > 0$ and

$$\frac{1}{-\hat{\gamma}} \sum_{i=1}^m \hat{\sigma}_i(x_{I_i}) g_i(x_{I_i}) = 1 - \frac{1}{-\hat{\gamma}} \left(\sum_{i=1}^m \Big([x_{I_i}]_d^T [x_{I_i}]_d + \hat{\sigma}_{0,i}(x_{I_i}) \Big) \right).$$

Therefore, Assumption 6.1 holds. □

The sparse SOS polynomials and positive definite matrices E_i in Assumption 6.1 might not be unique. We want all $\lambda_{min}(E_1), \ldots, \lambda_{min}(E_m)$ are as large as possible. Similar to (3.1), the best $\sigma_i^*(x_{I_i})$ and E_i^* can be found by solving the sparse SOS program:

$$\begin{array}{rl} \max & \min\{\lambda_{min}(E_1), \ldots, \lambda_{min}(E_m)\} \\ s.t. & \sum\limits_{i=1}^{m} \sigma_i(x_{I_i}) g_i(x_{I_i}) = 1 - \sum\limits_{i=1}^{m} [x_{I_i}]_d^T E_i [x_{I_i}]_d, \\ & \sigma_1(x_{I_1}), \ldots, \sigma_m(x_{I_m}) \text{ are SOS}, \\ & \deg(\sigma_1 g_1), \ldots, \deg(\sigma_m g_m) \le 2d. \end{array} \tag{6.4}$$

Note (6.4) is equivalent to an SDP problem, and can be solved efficiently by numerical methods. Let $\sigma_1^*(x_{I_1}), \ldots, \sigma_m^*(x_{I_m})$ and $E_1^*, \ldots, E_m^*$ be optimal in (6.4). Assumption 6.1 holds if and only if $\min\{\lambda_{min}(E_1^*), \ldots, \lambda_{min}(E_m^*)\} > 0$, so it is checkable by solving (6.4).

As before, we still denote by p_{max} and p_{min} the maximum and minimum value of a polynomial $p(x)$ on S respectively. Define two sets of sparse polynomials

$$\mathscr{Z}_{sp}(g) = \left\{ p \in \sum_{i=1}^{m} \mathbb{R}[x_{I_i}]_{\le 2d} : p_{max} + p_{min} = 0 \right\},$$

$$TP_{sp}(g) = \left\{ p \in \sum_{i=1}^{m} \mathbb{R}[x_{I_i}]_{\le 2d} : p(x) + 1 \ge 0 \quad \forall x \in S \right\}.$$

Lemma 6.3. *Let $p(x) \in \sum_{i=1}^{m} \mathbb{R}[x_{I_i}]_{\le 2d}$ and $\Omega(p)$ be defined in (6.3). Suppose $int(S_\Delta) \neq \emptyset$ for all $\Delta \in \Omega(p)$.*
(i) If $p(x) \in TP_{sp}(g) \cap \mathscr{Z}_{sp}(g)$, then $\|p(x)\|_{L^2(S),\Omega(p)} \le \sqrt{|\Omega(p)|}$.
(ii) It holds that $\|p(x)\|_{L^2(S),\Omega(p)} \ge \eta_{\Omega(p)}(S) \|p(x)\|_G$ where

$$\eta_{\Omega(p)}(S) := \sqrt{\min_{\Delta \in \Omega(p)} \lambda_{min}(\boldsymbol{\Theta}_\Delta(S))} > 0. \tag{6.5}$$

In the above $\boldsymbol{\Theta}_\Delta(S)$ is defined by (2.12).

Proof. (i) Fix an arbitrary $p(x) \in TP_{sp}(g) \cap \mathscr{Z}_{sp}(g)$, then $p(x) \ge -1 \, \forall x \in S$, which implies (since $p_{max} = -p_{min}$) $-1 \le p(x) \le 1 \, \forall x \in S$. In particular, we get

$$-1 \le p_\Delta(x_\Delta) \le 1 \quad \forall x_\Delta \in S_\Delta.$$

So we can see that

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

Therefore item (i) holds because

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

(ii) When all $int(S_\Delta) \neq \emptyset$, the matrices $\boldsymbol{\Theta}_\Delta(S)$ are all positive definite, as shown in the proof of Lemma 2.2. So $\eta_{\Omega(p)}(S) > 0$. For every $\Delta \in \Omega(p)$, it holds

$$\|p_\Delta(x_\Delta)\|_{L^2(S_\Delta)}^2 = p_{\Delta,G}^T \boldsymbol{\Theta}_\Delta(S) p_{\Delta,G} \ge \left(\eta_{\Omega(p)}(S)\right)^2 \|p_\Delta(x_\Delta)\|_G^2.$$

Here $p_{\Delta,G}$ is the weighted coefficient vector of $p(x)$ (see definition (2.6)). Then

$$\|p(x)\|^2_{L^2(S),\Omega(p)} = \sum_{\Delta\in\Omega(p)} \|p_\Delta(x_\Delta)\|^2_{L^2(S_\Delta)}$$
$$\geq \big(\eta_{\Omega(p)}(S)\big)^2 \sum_{\Delta\in\Omega(p)} \|p_\Delta(x_\Delta)\|^2_G \geq \big(\eta_{\Omega(p)}(S)\big)^2 \|p(x)\|^2_G,$$

which implies item (ii). □

The approximation bound for sparse Lasserre's relaxation (6.2) is given as below.

Theorem 6.4. *Suppose $0 \in int(S)$, Assumption 6.1 holds, and $\sigma_1^*(x_{I_1}),\ldots,\sigma_m^*(x_{I_m})$ and $E_1^*,\ldots,E_m^*$ are optimal for (6.4). Let $f(x) \in \mathbb{R}[x_{I_1}]_{\leq 2d}+\cdots+\mathbb{R}[x_{I_m}]_{\leq 2d}$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) on S. If f_{sos} is the optimal value of the sparse SOS relaxation (6.2), then it holds*

$$1 \leq \frac{f_{max}-f_{sos}}{f_{max}-f_{min}} \leq \frac{\sqrt{|\Omega(f)|}}{\eta_{\Omega(f)}(S)\cdot\lambda^*},$$

where $\eta_{\Omega(f)}(S)$ is defined in (6.5), $\Omega(f)$ is defined in (6.3), and

$$\lambda^* = \min\{\lambda_{min}(E_1^*),\ldots,\lambda_{min}(E_m^*)\}.$$

Proof. We follow the same approach as in the proof of Theorem 3.3. Set

$$med(f) = \frac{1}{2}(f_{min}+f_{max}) \in [f_{min}, f_{max}], \quad \tilde{f}(x) = f(x) - med(f) \in \mathscr{Z}_{sp}(g).$$

Then $\tilde{f}(x) + (med(f) - f_{min}) \geq 0$ for all $x \in S$ and

$$\frac{1}{med(f)-f_{min}}\tilde{f}(x) + 1 \geq 0 \quad \forall x \in S.$$

By definition of $TP_{sp}(g)$, it must hold

$$\frac{1}{med(f)-f_{min}}\tilde{f}(x) \in TP_{sp}(g) \cap \mathscr{Z}_{sp}(g).$$

Note $f(x)$ and $\tilde{f}(x)$ have the same cover set. Since the origin is in the interior of S, $int\big(S_\Delta\big) \neq \emptyset$ for all $\Delta \in \Omega(p)$. So Lemma 6.3 implies

$$\left\|\frac{1}{med(f)-f_{min}}\tilde{f}(x)\right\|_{L^2(S),\Omega(f)} \leq \sqrt{\Omega(\tilde{f})} = \sqrt{\Omega(f)}. \tag{6.6}$$

Now fix two constants

$$\theta^* = (med(f)-f_{min})\cdot\frac{1}{\eta_{\Omega(f)}(S)\cdot\lambda^*}\sqrt{\Omega(f)} > 0, \qquad \gamma^* = med(f) - \theta^*. \tag{6.7}$$

Then (6.6) implies

$$\left\|\frac{1}{\theta^*}\tilde{f}(x)\right\|_{L^2(S),\Omega(f)} \leq \eta_{\Omega(f)}(S)\cdot\lambda^*.$$

Lemma 6.3 and the above then imply

$$\left\|\frac{1}{\theta^*}\tilde{f}(x)\right\|_G \le \left(\eta_{\Omega(f)}(S)\right)^{-1}\left\|\frac{1}{\theta^*}\tilde{f}(x)\right\|_{L^2(S),\Omega(f)} \le \lambda^*.$$

Since $\frac{1}{\theta^*}\tilde{f}(x) \in \sum_{i=1}^m \mathbb{R}[x_{I_i}]_{\le 2d}$, define polynomials $f_1(x_{I_1}), \ldots, f_m(x_{I_m})$ as follows

$$f_1(x_{I_1}) \text{ is the restriction of } \frac{1}{\theta^*}\tilde{f}(x) \text{ to } x_{I_1},$$

$$f_2(x_{I_2}) \text{ is the restriction of } \frac{1}{\theta^*}\tilde{f}(x) - f_1(x_{I_1}) \text{ to } x_{I_2},$$

$$\vdots$$

$$f_m(x_{I_m}) \text{ is the restriction of } \frac{1}{\theta^*}\tilde{f}(x) - \sum_{i=1}^{m-1} f_i(x_{I_i}) \text{ to } x_{I_m}.$$

Then we get

$$\frac{1}{\theta^*}\tilde{f}(x) = f_1(x_{I_1}) + \cdots + f_m(x_{I_m}),$$

$$\|f_i(x_{I_i})\|_G \le \left\|\frac{1}{\theta^*}\tilde{f}(x)\right\|_G \le \lambda^* \quad \forall\, i = 1, \ldots, m.$$

By Lemma 2.1, there exist symmetric matrices W_i such that

$$f_i(x_{I_i}) = [x_{I_i}]_d^T W_i [x_{I_i}]_d, \quad \|W_i\|_F \le \lambda^*.$$

Thus the polynomials

$$s_{0,i}(x_{I_i}) := f_i(x_{I_i}) + [x_{I_i}]_d^T E_i^* [x_{I_i}]_d = [x_{I_i}]_d^T (W_i + E_i^*)[x_{I_i}]_d$$

must be SOS, because $\|W_i\|_2 \le \|W_i\|_F \le \lambda^* \le \lambda_{min}(E_i^*)$. Let $\sigma_{0,i}(x_{I_i}) = \theta^* s_{0,i}(x_{I_i})$ and $\sigma_i(x_{I_i}) = \theta^* \sigma_i^*(x_{I_i})$ for each i. They are all sparse SOS polynomials. Since

$$1 - \sum_{i=1}^m [x_{I_i}]_d^T E_i^* [x_{I_i}]_d = \sum_{i=1}^m \sigma_i^*(x_{I_i}) g_i(x_{I_i}),$$

we can easily check the identity

$$f(x) - \gamma^* = \sum_{i=1}^m \sigma_{0,i}(x_{I_i}) + \sum_{i=1}^m \sigma_i(x_{I_i}) g_i(x_{I_i})$$

holds. So γ^* is feasible in (6.2). By optimality of f_{sos}, it holds $f_{sos} \ge \gamma^*$. The choice of γ^* in (6.7) implies

$$\frac{med(f) - f_{sos}}{med(f) - f_{min}} \le \frac{1}{\eta_{\Omega(f)}(S)\lambda^*}\sqrt{\Omega(f)}.$$

The theorem now follows the fact that $f_{min} \le med(f) \le f_{max}$ and $f_{sos} \le f_{min}$. □

Example 6.5. Suppose $I_1, \dots, I_m$ are elements of the set

$$\Big\{\{i, \dots, i+k-1\} : 1 \le i \le n-k+1\Big\}.$$

Thus $m = n-k+1$. For each $1 \le i \le n-k+1$ define $g_i(x) = 1 - x_i^2 - \cdots - x_{i+k-1}^2$. The feasible set S is a kind of multi-balls. Assume $k \ge 2d$. Let $f(x) \in \mathbb{R}[x_{I_1}]_{\le 2d} + \cdots + \mathbb{R}[x_{I_m}]_{\le 2d}$ be a sparse polynomial. Then $|\Omega(f)| \le (n-k+1)\binom{k}{2d}$. By definition (6.5), $\eta_{\Omega(f)}(S)$ is a constant independent of n. Using the argument in subsection 4.1, for each $g_i(x)$, there exist an SOS polynomial $s_i(x)$ and a symmetric matrix E_i such that

$$s_i(x)g_i(x) = 1 - [x^{(i)}]_d^T E_i [x^{(i)}]_d, \quad \lambda_{min}(E_i) \ge \frac{1}{d+1}.$$

Here each $x^{(i)} = (x_i, \dots, x_{i+k-1})$. Therefore it holds

$$\frac{1}{n-k+1}\big(s_1(x)g_1(x) + \cdots + s_m(x)g_m(x)\big) = 1 - \frac{1}{n-k+1}\sum_{i=1}^{n-k+1} [x^{(i)}]_d^T E_i [x^{(i)}]_d.$$

So we can see that $\lambda^* \ge \frac{1}{(d+1)(n-k+1)}$. As a result of Theorem 6.4, the sparse Lasserre's relaxation (6.2) has an approximation bound $\mathcal{O}\left((d+1)(n-k+1)^{3/2}\binom{k}{2d}^{1/2}\right)$. □

References

[1] J. Bochnak, M. Coste and M-F. Roy. *Real Algebraic Geometry*, Springer, 1998.

[2] E. de Klerk, M. Laurent, P. Parrilo. A PTAS for the minimization of polynomials of fixed degree over the simplex. *Theoretical Computer Science*, 361(2-3), 210-225.

[3] E. de Klerk. The complexity of optimizing over a simplex, hypercube or sphere: A short survey. *Central European Journal of Operations Research*, 16(2), 111-125.

[4] M. Goemans and D. Williamson. Improved approximation algorithms for Maximum Cut and Satisfiability problems using semidefinite programming. *Journal of ACM*, Vol. 42, (1995), pp. 1115–1145.

[5] S. He, Z. Luo, J. Nie and S. Zhang. Semidefinite relaxation bounds for indefinite homogeneous quadratic optimization. *SIAM J. Optim.*, 19 (2008) 503-523.

[6] D. Henrion, J. B. Lasserre and C. Savorgnan. Approximate volume and integration for basic semialgebraic sets. To appear in *SIAM Review.*

[7] D. Jibetean and M. Laurent. Semidefinite approximations for global unconstrained polynomial optimization. *SIAM J. Optim.* 16 (2005), no. 2, 490–514

[8] S. Kim, M. Kojima and H. Waki. Generalized Lagrangian Duals and Sums of Squares Relaxations of Sparse Polynomial Optimization Problems. *SIAM Journal on Optimization*, Vol. 15 (3) 697-719 (2005).

[9] M. Kojima and L. Tuncel. Discretization and localization in successive convex relaxation methods for nonconvex quadratic optimization. *Math. Program.* 89 (2000), No. 1, Ser. A, 79–111.

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

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

[12] J. Lasserre. Convergent SDP-relaxations in polynomial optimization with sparsity. *SIAM Journal on Optimization*, Vol. 17, No. 3, 822-843, 2006.

[13] J. Lasserre and T. Netzer. SOS approximations of nonnegative polynomials via simple high degree perturbations. *Math. Z.* 256 (2007), no. 1, 99–112.

[14] J. Lasserre. A semidefinite programming approach to the generalized problem of moments. *Math. Program.* 112 (2008), no. 1, Ser. B, 65–92.

[15] M. Laurent. Semidefinite representations for finite varieties. *Math. Program.* 109 (2007), no. 1, Ser. A, 1–26.

[16] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. *Emerging Applications of Algebraic Geometry*, Vol. 149 of *IMA Volumes in Mathematics and its Applications*, M. Putinar and S. Sullivant (eds.), Springer, pp. 157-270, 2009.

[17] Wolfram MathWorld: the web's most extensive mathematics resource. `http://mathworld.wolfram.com/LegendrePolynomial.html`

[18] C. Ling, J. Nie, L. Qi, and Y. Ye. Bi-quadratic optimization over unit spheres and semidefinite programming relaxations. Preprint, 2008.

[19] Z.-Q. Luo, N. Sidiropoulos, P. Tseng and S. Zhang. Approximation bounds for quadratic optimization with homogeneous quadratic constraints. *SIAM Journal on Optimization*, Vol. 18, No. 1 (2007), pp. 1–28.

[20] Z. Luo and S. Zhang. A semidefinite relaxation scheme for multivariate quartic polynomial optimization with quadratic constraints. Technical report, Department of SEEM, Chinese University of Hong Kong, 2008.

[21] A. Nemirovski, C. Roos and T. Terlaky. On maximization of quadratic form over intersection of ellipsoids with common center. *Mathematical Programming*, Vol. 86 (1999), pp. 463–473.

[22] Yu. Nesterov. Semidefinite relaxation and non-convex quadratic optimization. *Optimization Methods and Software*, Vol. 12 (1997), pp. 1–20.

[23] J. Nie, J. Demmel and B. Sturmfels. Minimizing polynomials via sum of squares over the gradient ideal. *Math. Program.* 106 (2006), no. 3, Ser. A, 587–606.

[24] J. Nie and M. Schweighofer. On the complexity of putinar's positivstellensatz. *Journal of Complexity*, 23(2007), pp.135–150.

[25] 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).

[26] J. Nie. Sum of Squares Methods for Minimizing Polynomial Forms over Spheres and Hypersurfaces. *Preprint*, 2009. `http://www.math.ucsd.edu/~njw`.

[27] P.M. Pardalos and S. A. Vavasis. Quadratic programming with one negative eigenvalue is NP-hard. *J. Global Optim.* 1 (1991), no. 1, 15–22.

[28] 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.

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

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

[31] M. Putinar. Positive polynomials on compact semi-algebraic sets, *Ind. Univ. Math. J.* 42 (1993), 203–206.

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

[33] M. Schweighofer. Optimization of polynomials on compact semialgebraic sets. *SIAM J. Optim.* 15 (2005), no. 3, 805–825

[34] P. Tseng. Further results on approximating nonconvex quadratic optimization by semidefinite programming relaxation. *SIAM J. Optim.,* 14 (2003) 268-283.

[35] H. Waki, S. Kim, M. Kojima and M. Muramatsu. Sums of Squares and Semidefinite Programming Relaxations for Polynomial Optimization Problems with Structured Sparsity. *SIAM Journal on Optimization*, Vol.17 (1) 218-242 (2006).

[36] 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.

[37] Y. Ye. Approximating quadratic programming with bound and quadratic constraints . *Mathematical Programming*, Vol. 84 (1999), pp. 219–226.

[38] Y. Ye and S. Zhang. New results on quadratic minimization. *SIAM Journal on Optimization*, Vol. 14, (2003) pp. 245–267.