AN APPROXIMATION BOUND ANALYSIS FOR LASSERRE'S RELAXATION IN MULTIVARIATE POLYNOMIAL OPTIMIZATION

JIAWANG NIE

ABSTRACT. Suppose $f(x), g_1(x), \ldots, g_m(x)$ are multivariate polynomials in $x \in \mathbb{R}^n$ and their degrees are at most $2d$. Consider the optimization problem

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

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

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

Here Q is a constant depending only on $g_1, \ldots, g_m$ but not on f. In particular, if S is the unit ball, $Q = \mathcal{O}(n^d)$; if $S = [-1, 1]^n$, $Q = \mathcal{O}(n^{2d})$; if $S = \{\pm 1\}^n$ or $\{0, 1\}^n$, $Q = \mathcal{O}(n^d)$. In these special cases, assume canonical defining polynomials g_i for S are used.

1. INTRODUCTION

Consider the polynomial optimization problem

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

Here $f, g_1, \ldots, g_m$ are all multivariate polynomials in $x := (x_1, \ldots, x_n)$. 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, (1.1) becomes a nonconvex quadratic programming (QP) which is NP-hard [6]. So problem (1.1) is NP-hard. Lasserre's relaxation [2] is a typical approach for solving (1.1) approximately by using semidefinite programming and sum of squares techniques. We refer to [2, 3, 4, 5, 7, 8].

When $f, g_1, \ldots, g_m$ have degrees no greater than $2d$, Lasserre [2] proposed the following sum of squares (SOS) program to find a lower bound for the minimum f_{min} of (1.1):

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

1991 *Mathematics Subject Classification.* 68Q25, 90C22, 90C60.

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

Research was partially supported by NSF grants DMS-0757212 and DMS-0844775.

Here a polynomial is said to be SOS if it is a sum of squares of polynomials. If a polynomial is SOS, then it must be nonnegative everywhere, but the reverse might not be true. We refer to [10] for a survey on SOS and nonnegative polynomials. While it is quite difficult to check nonnegativity, checking SOS is much easier because it is equivalent to a semidefinite programming problem (cf. [7, 8]), which can be solved efficiently. Thus, the SOS program (1.2) would be solved by SDP solvers. The integer d in (1.2) is called the relaxation order.

Here we briefly review the convergence of Lasserre's relaxation (1.2) as d increases for a fixed f. For convenience, denote $g = (g_1, \cdots, g_m)$ and

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

For each 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 [9], Lasserre [2] 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 $s_0, s_1, \dots, s_m$ such that

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

For AC to hold, S must be compact, but the reverse might not be true. However, AC is not a very strong condition, because otherwise we can always add a redundant constraint like $M - \|x\|_2^2 \geq 0$ if S is compact. Nie and Schweighofer [5] analyzed the convergence rate of (1.2). Under AC, they proved

$$0 \leq f_{min} - f_{sos,d} \leq K \cdot (\log d)^{-c} \quad \text{as} \quad d \to \infty, \tag{1.3}$$

where $c > 0$ depends only on g and the constant $K = K(f, g)$ depends on both f and g. The above estimate is a kind of absolute error analysis, and is only in asymptotic sense. Typically, the constants K and c are quite complicated to estimate in practice.

Due to the computational cost of (1.2), people tend to choose small d in practical applications. This is because (1.2) is very expensive to solve for big d (the size of the resulting SDP grows exponentially in d). So, it is interesting to know how good (1.2) approximates (1.1) for a fixed relaxation order d. Suppose $\deg(f) \leq 2d$ and $\deg(g) \leq 2d$. For convenience, just denote by f_{sos} the optimal value of (1.2) for a given f. We have seen that $f_{sos} \leq f_{min}$, 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 $Q = Q(g)$ is called an approximation bound of (1.2) if for every f with $\deg(f) \leq 2d$ it holds that

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

Does the above Q exist? What conditions make Q exist? How big is Q? These questions are the main topics of this paper.

Contributions First, we analyze the approximation bound of Lasserre's relaxation (1.2) when S is compact. Let f_{min} (resp., f_{max}) be the minimum (resp., maximum) value of f on S. Under a suitable condition on $g_1, \dots, g_m$, we show that there exists a constant $Q = Q(g_1, \dots, g_m)$ such that for every f with $\deg(f) \le 2d$ it holds that

$$1 \le (f_{max} - f_{sos})/(f_{max} - f_{min}) \le Q. \tag{1.4}$$

The constant Q only depends on $g_1, \dots, g_m, n, d$ but not on f. This will be presented in Section 3.

Second, we give explicit estimates for Q in (1.4) for special cases of S: when S is a unit ball, $Q = \mathcal{O}(n^d)$; when S is a hypercube $[-1,1]^n$, $Q = \mathcal{O}(n^{2d})$; when S is the boolean set $\{\pm 1\}^n$ or $\{0,1\}^n$, $Q = \mathcal{O}(n^d)$; when S is a multi-unit ball, $Q = \mathcal{O}(n^d)$. (Here, we assume canonical defining polynomials g_i for S are used.) This will be shown in Section 4.

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

Here we make some remarks on the difference between the estimates (1.3) and (1.4). The estimates for Q given in this paper depend on g. As $d \to \infty$, the obtained bound Q typically goes to infinity, which does not imply the convergence of (1.2). This is because the approximation bound and convergence are different aspects of Lasserre's relaxation. The bound Q estimates f_{sos} in the *worst* case that (1.2) would behave when the relaxation order d is *fixed*. The estimate (1.4) holds for *arbitrary* polynomial f of degree $2d$, and can be thought of as a relative error analysis. The smaller Q is, the tighter (1.2) approximates (1.1); the bigger Q is, the looser (1.2) approximates (1.1). This is the reason why we call Q an approximation bound. On the other hand, the convergence of (1.2) is the different issue of whether $f_{sos,d}$ approaches f_{min} as $d \to \infty$ for a *fixed* polynomial f. The estimate (1.3) holds for a *fixed* f, and can be thought of as an absolute error analysis. In convergence analysis, the polynomial f in (1.1) is *fixed*, but the relaxation order d goes to infinity. This issue is important when we want to minimize a fixed f over S as accurate as possible. In summary, the convergence concerns the behavior of a *sequence* of Lasserre's relaxations ($d = 1, 2, \dots$) for solving a *fixed single* polynomial optimization (1.1), while the approximation bound Q concerns the behavior of a *single* Lasserre's relaxation (the order d is fixed) for solving a *class* of polynomial optimization(1.1) (f is arbitrary with degree $2d$). So, (1.3) and (1.4) address different aspects of Lasserre's relaxation.

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 symbol $\mathbb{S}^{n-1}$ denotes the 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 $\mathrm{supp}(\alpha) = \{1 \le i \le n : \alpha_i \ne 0\}$. The symbol $\mathbb{N}_k^n$ denotes the multi-index set $\{\alpha \in \mathbb{N}^n : |\alpha| \le k\}$. 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 symbol $[x]_d$ denotes the following vector of monomials

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

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

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

The symbol $\mathbb{R}[x]$ denotes the ring of real polynomials in $(x_1, \ldots, x_n)$; $\mathbb{R}[x]_k$ denotes the subspace of polynomials of degrees at most k; $\texttt{Sfr}[x]_k$ denotes the subspace of square free polynomials of degrees at most k. For a polynomial p, $\mathrm{supp}(p)$ denotes the support of p, i.e., the set of $\alpha \in \mathbb{N}^n$ such that x^α appears in p. 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. The symbol I_N denotes the N-by-N identity matrix. 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 $\lambda_{min}(X) \geq 0$ (resp. $\lambda_{min}(X) > 0$). For $u \in \mathbb{R}^N$, $\|u\|_2 = \sqrt{u^T u}$ denotes the standard Euclidean norm. For any matrix A, recall the definition of $\|A\|_2$ being the standard operator 2-norm of A and $\|A\|_F$ being the Frobenius norm of A, i.e., $\|A\|_F = \sqrt{\mathrm{Trace}(A^T A)}$. Note that $\|A\|_2$ and $\|A\|_2 \leq \|A\|_F$. In matrix spaces, the bullet $\bullet$ denotes the standard Frobenius inner product, i.e., $A \bullet B = \mathrm{Trace}(A^T B)$.

2. Sum of squares and norms of polynomials

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

2.1. Sum of squares and semidefinite programming. For a polynomial f of degree $2d$, there exists a symmetric matrix F such that

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

The length of the monomial vector $[x]_d$ 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 and is not unique if $d > 1$. For convenience, we index the columns and rows of F by monomials of degrees $\leq d$, or equivalently by vectors in $\mathbb{N}_d^n$.

A polynomial f is said to be a sum of squares (SOS) if there exist polynomials $f_1, \ldots, f_k$ such that $f = f_1^2 + \cdots + f_k^2$. As shown in [7, 8], f is SOS if and only if it has a Gram matrix F which is positive semidefinite, that is,

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

Define constant symmetric matrices A_α satisfying

$$[x]_d [x]_d^T = \sum_{\alpha \in \mathbb{N}_{2d}^n} A_\alpha x^\alpha. \tag{2.1}$$

If $f(x) = \sum_{\alpha \in \mathbb{N}_{2d}^n} f_\alpha x^\alpha$, then f is SOS if and only if

$$A_\alpha \bullet X = f_\alpha \quad \forall\, \alpha \in \mathbb{N}_{2d}^n, \text{ for some matrix } X \succeq 0.$$

So, checking whether f is SOS can be done by solving a semidefinite programming problem.

The standard form of a semidefinite programming is

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

Here C and $A_1, \ldots, A_m$ are constant symmetric matrices. Lasserre [2] 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 the theory, algorithms and applications of semidefinite programming. We refer to [11].

2.2. **Norms of polynomials.** For a polynomial $f(x) = \sum_{\alpha\in\mathbb{N}^n_{2d}} f_\alpha x^\alpha$ of degree $2d$, define its 2-norm and G-norm as

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

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

$$\mathfrak{p}(\alpha) = \left|\left\{(\beta,\nu)\in\mathbb{N}^n_d\times\mathbb{N}^n_d : \beta+\nu=\alpha\right\}\right|. \tag{2.4}$$

Let J be the matrix of all ones, then

$$\mathfrak{p}(\alpha) = A_\alpha \bullet J \le \binom{|\mathrm{supp}(\alpha)|+d}{d} \le \binom{|3d}{d}.$$

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

$$\binom{3d}{d}^{-1/2}\|f\|_2 \,\le\, \|f\|_G \,\le\, \|f\|_2. \tag{2.5}$$

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

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

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

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

The entries of f, f_G and $[x]_{G,2d}$ are in graded alphabetical ordering by their indices. Thus, $f(x) = f^T[x]_{2d} = f_G^T[x]_{G,2d}$ and $\|f\|_G = \|f_G\|_2$. The G-norm $\|f\|_G$ is closely related to Gram matrices of f.

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

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

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

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

Choose a particular W satisfying the above as follows

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

Then W is symmetric and satisfies

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

This completes the proof. □

Another useful norm of polynomials is the L^2-norm. Assume S is compact. Define

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

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

When $n \geq 2d$, we can also define the following notion of *marginal* L^2-norm. Given a subset $\Delta \subset \{1, \ldots, n\}$ with $|\Delta| = 2d$, x_Δ denotes the subvector

$$x_\Delta = (x_{i_1}, \ldots, x_{i_{2d}}) \quad \text{if } \Delta = \{i_1, \ldots, i_{2d}\}.$$

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

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

Clearly, the cardinality $|\Omega_{2d}| = \binom{n}{2d}$. We also denote by $g_{i,\Delta}(x_\Delta)$ the restriction of $g_{i,\Delta}(x)$ to x_Δ. Define S_Δ similarly as

$$S_\Delta = \{x_\Delta : g_{1,\Delta}(x_\Delta) \geq 0, \ldots, g_{m,\Delta}(x_\Delta) \geq 0\}. \tag{2.11}$$

Observe that

$$S_\Delta = \{x_\Delta : (0, \ldots, 0, x_{i_1}, 0, \ldots, 0, x_{i_{2d}}, 0, \ldots) \in S\} \quad \text{if } \Delta = \{i_1, \ldots, i_{2d}\}.$$

If $0 \in int(S)$, then every S_Δ has nonempty interior and $0 \in int(S_\Delta)$. If $int(S_\Delta) \neq \emptyset$, the $L^2(S_\Delta)$-norm of $f_\Delta(x_\Delta)$ can be similarly defined as

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

where $\mu_\Delta(\cdot)$ is the uniform probability measure on S_Δ. When every $int\big(S_\Delta\big) \neq \emptyset$, the marginal $L^2(S)$-norm of f is defined as

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

Define two matrices

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

$$\mathbf{\Theta}_{2d}(S) = \int_S [x]_{G,2d} [x]^T_{G,2d} d\mu(x), \tag{2.13}$$

and two constants associated to S

$$\kappa_{2d}(S) = \min_{p \in \mathbb{R}[x]_{2d}} \big\{ \|p\|_{L^2(S)} : \|p\|_G = 1 \big\}. \tag{2.14}$$

$$\eta_{2d}(S) = \sqrt{\min_{\Delta \in \Omega_{2d}} \lambda_{min}(\mathbf{\Theta}_{\Delta,2d}(S))}. \tag{2.15}$$

If we write $p(x) = p_G^T [x]_{G,2d}$, then

$$\|p\|^2_{L^2(S)} = p_G^T \mathbf{\Theta}_{2d}(S) p_G. \tag{2.16}$$

So, $\kappa_{2d}(S) = \sqrt{\lambda_{min}(\mathbf{\Theta}_{2d}(S))}$. These constants depend only on S, and there are explicit formulae for them when S is special, like a ball or hypercube.

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

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

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

implies $u^T[x_\Delta]_{G,2d} = 0$ for every $x \in S_\Delta$. Since $int(S_\Delta) \neq \emptyset$, $u^T[x_\Delta]_{G,2d}$ must be identically zero and $u = 0$, which is a contradiction.

If $0 \in int(S)$, then $0 \in int(S_\Delta)$ for every Δ. Thus $\eta_{2d}(S) > 0$ follows the above. The proof of $\kappa_{2d}(S) > 0$ is also the same. □

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

Lemma 2.3. *Suppose $n \geq 2d$. If $f \in \mathbb{R}[x]_{2d}$, then*

$$\|f\|_{L^2(S),mg} \;\geq\; \eta_{2d}(S)\|f\|_G.$$

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

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

Here, $f_{\Delta,G}$ denotes the vector of scaled coefficients of f_Δ (see (2.6)). So

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

Thus the lemma follows. □

3. Some general bounds for Lasserre's relaxation

This section estimates the approximation bound of Lasserre's relaxation (1.2). As mentioned in Introduction, to make Lasserre's relaxations converge, one needs to assume the archimedean condition (AC). However, even if AC holds, it is still possible that (1.2) is infeasible (i.e., $f_{sos} = -\infty$) for a given d, though an arbitrarily good lower bound could be obtained if we increase the relaxation order. To guarantee the feasibility of (1.2), we need the following assumption.

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

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

Note that Assumption 3.1 is equivalent to that the constant polynomial 1 lies in the interior of the truncated quadratic module defined as

$$M_{2d}(g_1, \ldots, g_m) \;=\; \left\{ \sigma_0 + \sum_{j=1}^m \sigma_j g_j : \text{each } \; g_j \text{ is SOS and } \; \deg(\sigma_j g_j) \leq 2d \right\}.$$

Assumption 3.1 is sufficient and necessary for (1.2) to have a finite optimal value for every $f \in \mathbb{R}[x]_{2d}$.

Proposition 3.2. *Suppose S is nonempty. Then the following statements are equivalent:*

(i) *Assumption 3.1 holds;*
(ii) *The relaxation (1.2) is feasible for every $f \in \mathbb{R}[x]_{2d}$;*

(iii) *The relaxation (1.2) is feasible for* $f = -[x]_d^T[x]_d$.

Proof. First we prove $(i) \Rightarrow (ii)$. Every $f \in \mathbb{R}[x]_{2d}$ can be written as $f(x) = [x]_d^T F[x]_d$ for some symmetric matrix F. By Assumption 3.1, the matrix E is positive definite, so we can choose $\lambda > 0$ big enough such that the polynomial

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

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

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

Therefore, (1.2) is feasible and its optimal value $f_{sos} \geq \gamma$.

The direction $(ii) \Rightarrow (iii)$ is obvious. Now we prove $(iii) \Rightarrow (i)$. Let $f(x) = -[x]_d^T[x]_d$. Since (1.2) is feasible for this f, there exist $\hat{\gamma}$ and SOS polynomials $\hat{\sigma}_0, \hat{\sigma}_1, \ldots, \hat{\sigma}_m$ such that $\deg(\hat{\sigma}_i g_i) \leq 2d$ for every i and

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

Evaluating at $u \in S$, we obtain that $-\hat{\gamma} \geq [u]_d^T[u]_d \geq 1$ and

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

Hence, Assumption 3.1 holds. □

In Assumption 3.1, the choice of SOS polynomials $\sigma_1, \ldots, \sigma_m$ and positive definite matrix E might not be unique. In our bound analysis, the bigger $\lambda_{min}(E)$ is, the better the obtained bound would be. So we want $\lambda_{min}(E)$ to be as large as possible. Interestingly, the best one could be found by solving the SOS program:

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

Assume $(\sigma_1^*, \ldots, \sigma_m^*, E^*)$ is optimal for (3.1). Assumption 3.1 holds if and only if $\lambda_{min}(E^*) > 0$, so it is checkable by solving the SOS program (3.1).

Let $d = \max\{\lceil \deg(f)/2 \rceil, \lceil \deg(g)/2 \rceil\}$, and $\mathcal{F}$ be a subspace of $\mathbb{R}[x]_{2d}$ containing f. Define a constant associated with $\mathcal{F}$ and S as

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

When $int(S) \neq \emptyset$, $\chi(\mathcal{F}, S) < \infty$ because $\kappa_{2d}(S) > 0$ (cf. Lemma 2.2) and

$$\|p\|_{L^2(S)} \;\geq\; \kappa_{2d}(S) \cdot \|p_G\|_2.$$

When $int(S) = \emptyset$, $\chi(\mathcal{F}, S)$ might be infinite for some $\mathcal{F}$. For instance, if S is the unit sphere $\mathbb{S}^{n-1}$, then $\chi(\mathbb{R}[x]_{2d}, \mathbb{S}^{n-1}) = \infty$ because for $p_k = k(1 - \|x\|_2^2)$ it holds that

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

When S has empty interior, in order to ensure $\chi(\mathcal{F}, S) < \infty$, it suffices to choose $\mathcal{F}$ lying in the orthogonal complement of the subspace space

$$\mathcal{V}(S) = \Big\{ p \in \mathbb{R}[x]_{2d} : p(x) = 0 \quad \forall\, x \in S \Big\}.$$

Obviously, if $int(S) \neq \emptyset$, $\mathcal{V}(S) = \{0\}$ is a singleton.

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

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

The above bound is best if the optimal solution $(\sigma_1^*,\dots,\sigma_m^*,E^*)$ of (3.1) is used.

Proof. Let $med(f)=\frac{1}{2}(f_{min}+f_{max})\in[f_{min},f_{max}]$. First, we consider the case that $f_{min}<med(f)<f_{max}$. Let

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

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

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

Then, we have

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

By Lemma 2.1, there exists a symmetric matrix W satisfying

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

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

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

$$\hat{\sigma}_0(x):=(med(f)-f_{min})\theta^*\Big(\frac{1}{\theta^*}\tilde{f}(x)+[x]_d^TE[x]_d\Big)$$

must be SOS. Let $\hat{\sigma}_i=(med(f)-f_{min})\theta^*\sigma_i$ for every i, which are all SOS. Then we can verify that

$$\hat{\sigma}_1g_1+\cdots+\hat{\sigma}_mg_m=(med(f)-f_{min})\theta^*(1-[x]_d^TE[x]_d),$$

$$f-\gamma^*=\hat{\sigma}_0+\hat{\sigma}_1g_1+\cdots+\hat{\sigma}_mg_m.$$

So, $(\hat{\sigma}_0,\hat{\sigma}_1,\dots,\hat{\sigma}_m,\gamma^*)$ is feasible for (1.2) and its optimal value $f_{sos}\ge\gamma^*$. By the choice of γ^* in (3.3), it holds that

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

Since $med(f)\in[f_{min},f_{max}]$ and $f_{sos}\le f_{min}$, the theorem is true.

Second, we consider the case that $f_{min}=f_{max}$, then $f-f_{min}$ is constantly zero on S. Since $\chi(\mathcal{F},S)<\infty$, $f-f_{min}$ must be the identically zero polynomial, i.e., f is the constant f_{min}. So, $f_{sos}=f_{min}$ and the theorem is clearly true. □

To get a concrete bound by applying Theorem 3.3, we need estimate $\chi(\mathcal{F},S)$ and $\lambda_{min}(E)$.

Theorem 3.4. *Suppose $0\in int\big(S\big)$ and $(\sigma_1,\dots,\sigma_m,E)$ satisfies Assumption 3.1. Let $f\in\mathbb{R}[x]_{2d}$, f_{min} (resp. f_{max}) be its minimum (resp. maximum) on S, and f_{sos} be the optimal value of Lasserre's relaxation (1.2).*

(i) *It holds that*

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

(ii) *If $n \ge 2d$, then*

$$1 \le \frac{f_{max} - f_{sos}}{f_{max} - f_{min}} \le \frac{1}{\eta_{2d}(S)\lambda_{min}(E)} \sqrt{\binom{n}{2d}}.$$

The above bounds are best if the optimal solution $(\sigma_1^*, \ldots, \sigma_m^*, E^*)$ of (3.1) is used.

Proof. (i) Since $int(S) \neq \emptyset$, Lemma 2.2 implies $\kappa_{2d}(S) > 0$. Let $\mathcal{F} = \mathbb{R}[x]_{2d}$. If $p \in \mathcal{F}$ satisfies $|p(x)| \le 1$ for all $x \in S$, then $\|p\|_{L^2(S)} \le 1$ and (2.14) implies $1 \ge \kappa_{2d}(S)\|p\|_G$ and $\chi(\mathcal{F}, S) \le \frac{1}{\kappa_{2d}(S)}$. Then, Theorem 3.3 implies the result.

(ii) Since $0 \in int(S)$, every S_Δ has nonempty interior, and Lemma 2.2 implies $\eta_{2d}(S) > 0$. If $p \in \mathbb{R}[x]_{2d}$ and $|p(x)| \le 1$ for all $x \in S$, the restriction $p_\Delta(x_\Delta)$ of $p(x)$ must also satisfy $|p_\Delta(x_\Delta)| \le 1$ for every $x_\Delta \in S_\Delta$. Thus we have

$$\|p\|^2_{L^2(S),mg} = \sum_{\Delta \in \Omega_{2d}} \int_{S_\Delta} p_\Delta(x_\Delta)^2 d\mu_\Delta(x_\Delta) \le \sum_{\Delta \in \Omega_{2d}} 1 = \binom{n}{2d}. \quad (3.4)$$

Therefore, Lemma 2.3 and the above imply

$$\|p\|_G \le \frac{1}{\eta_{2d}(S)} \sqrt{\binom{n}{2d}} \quad \text{and} \quad \chi(\mathbb{R}[x]_{2d}, S) \le \frac{1}{\eta_{2d}(S)} \sqrt{\binom{n}{2d}}.$$

The result is then implied by Theorem 3.3. □

Generally, it is hard to tell which one of $\frac{1}{\kappa_{2d}(S)}$ and $\frac{1}{\eta_{2d}(S)}\sqrt{\binom{n}{2d}}$ is superior in Theorem 3.4, depending on S. However, typically $\eta_{2d}(S)$ is relatively easier to estimate than $\kappa_{2d}(S)$ does. For instance, when S is a unit ball or hypercube, the constant $\eta_{2d}(S)$ is independent of n and easy to estimate, while $\kappa_{2d}(S)$ is quite difficult to estimate in terms of n, d. This will be shown in Section 4.

Remark 3.5. The optimal value $\lambda_{min}(E^*)$ of (3.1) is closely related to the "radius" of S. Let $R = \max_{x \in S} \|x\|_2$. Observe that

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

$$\|[x]_d\|_2^2 = 1 + \|[x^1]\|_2^2 + \|[x^2]\|_2^2 + \cdots + \|[x^d]\|_2^2 \ge \sum_{k=0}^{d} \frac{\|x\|_2^{2k}}{k!}.$$

Since $1 \ge [x]_d^T E^* [x]_d \ge \lambda_{min}(E^*)\|[x]_d\|_2^2$ for all $x \in S$, the above implies

$$\frac{1}{\lambda_{min}(E^*)} \ge \sum_{k=0}^{d} \frac{R^{2k}}{k!}.$$

On the other hand, from

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

we know the polynomial $r(x) := 1 - \left(1 + R^2 + \cdots + R^{2d}\right)^{-1} [x]_d^T [x]_d$ is nonnegative on S. If there exist SOS polynomials $s_0, s_1, \dots, s_m$ such that every $\deg(s_i g_i) \le 2d$ and

$$r \;=\; s_0 + s_1 g_1 + \cdots + s_m g_m, \tag{3.5}$$

then we can find a positive definite matrix $\hat{E}$ satisfying

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

Since s_0 is SOS, we know

$$\frac{1}{\lambda_{min}(E^*)} \le \frac{1}{\lambda_{min}(\hat{E})} \le \sum_{k=0}^{d} R^{2k} \le d! \sum_{k=0}^{d} \frac{R^{2k}}{k!}.$$

So, if (3.5) holds, then

$$\sum_{k=0}^{d} \frac{R^{2k}}{k!} \le \frac{1}{\lambda_{min}(E^*)} \le d! \sum_{k=0}^{d} \frac{R^{2k}}{k!}.$$

If (3.5) fails but R is known in advance, we can add to (1.1) the redundant constraint $r(x) \ge 0$. Hence, $\frac{1}{\lambda_{min}(E^*)}$ can be estimated by $\sum_{k=0}^{d} \frac{R^{2k}}{k!}$, which is tight within a factor of $d!$. □

4. Bounds for some special optimization problems

Theorems 3.3 and 3.4 give some general bounds for Lasserre's relaxation (1.2) in terms of some constants related to the feasible set S. For special S like a unit ball or hypercube, estimating $\eta_{2d}(S)$ is typically easy, while estimating $\kappa_{2d}(S)$ in terms of n, d is typically quite difficult. So, generally we apply Theorem 3.4 (ii) by estimating $\eta_{2d}(S)$. In this section, we assume $\deg(f) = 2d$ and $\deg(g_i) = 2$.

4.1. **Optimizing polynomials over a unit ball.**

Consider the case that S is the unit ball $\mathbf{B} := \{x \in \mathbb{R}^n : \|x\|_2 \le 1\}$. Then $m = 1$ and $g_1(x) = 1 - \|x\|_2^2$. To get a bound, we estimate $\eta_{2d}(\mathbf{B})$ and $\lambda_{min}(E^*)$ in (3.1).

By definition (2.15), we have $\eta_{2d}(\mathbf{B}) = \sqrt{\lambda_{min}\left(\mathbf{\Theta}_{\Delta,2d}(\mathbf{B})\right)}$ for every $\Delta \in \Omega_{2d}$ because of the symmetry of $\mathbf{B}$. Let $\Delta = \{1, \dots, 2d\}$, then

$$\begin{aligned}
\mathbf{\Theta}_{\Delta,2d}(\mathbf{B}) &= \int_{\|x_\Delta\|_2 \le 1} [x_\Delta]_{G,2d} [x_\Delta]_{G,2d}^T d\mu_\Delta(x_\Delta) \\
&= \frac{1}{Vol(\|x_\Delta\|_2 \le 1)} \int_{\|x_\Delta\|_2 \le 1} [x_\Delta]_{G,2d} [x_\Delta]_{G,2d}^T dx_\Delta,
\end{aligned}$$

where dx_Δ is the standard Lebesgue measure. Observe that

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

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

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

When $\alpha = 2\beta = 2(\beta_1, \ldots, \beta_{2d})$ is an even vector, it holds that (cf. Lemma 8 of [1])

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

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

Here, $\Gamma(\cdot)$ is the standard Gamma function. If at least one entry of α is odd, the integral $\int_{\|x_\Delta\|_2=1} x_\Delta^\alpha d\nu_\Delta(x_\Delta) = 0$. A list of typical values of $\eta_{2d}(\mathbf{B})$ is in Table 1.

$2d$	2	4	6	8
$\eta_{2d}(\mathbf{B})$	0.19204	0.01670	0.00161	0.00004

TABLE 1. A list of $\eta_{2d}(\mathbf{B})$ for $2d = 2, 4, 6, 8$.

Lemma 4.1. *When S is the unit ball $\mathbf{B}$, we have $\eta_{2d}(\mathbf{B})$ is independent of n, Assumption 3.1 holds, and the optimal E^* of (3.1) satisfies $\lambda_{min}(E^*) \ge \frac{1}{d+1}$.*

Proof. In the above we have already seen that $\eta_{2d}(\mathbf{B})$ is independent of n. Now we estimate $\lambda_{min}(E^*)$ in (3.1). For any integer $k \ge 1$, it holds that

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

Let $s_d(t) := \frac{1}{d+1}\sum_{k=1}^d \sum_{j=0}^{k-1} t^j$, then

$$s_d(t)(1-t) = 1 - \frac{1}{d+1}\left(1+t+\cdots+t^d\right). \tag{4.1}$$

Plugging t by $\|x\|_2^2$, 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 satisfying

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

So, Assumption 3.1 holds and the optimal value of (3.1) is at least $\frac{1}{d+1}$. $\square$

Clearly, Theorem 3.4 (ii) and Lemma 4.1 imply the following.

Corollary 4.2. *Assume $n \ge 2d$. Let $f \in \mathbb{R}[x]_{2d}$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) on the unit ball $\mathbf{B}$. If f_{sos} is the optimal value of (1.2), then*

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

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

4.2. **Optimizing polynomials over a hypercube.** Consider the case that S is the hypercube $\mathcal{C} := [-1,1]^n$. Then $m = n$ and

$$g_1(x) \;=\; 1 - x_1^2, \;\ldots,\; g_n(x) \;=\; 1 - x_n^2.$$

To get a bound by Theorem 3.4 (ii), we need to estimate $\eta_{2d}(\mathcal{C})$ and $\lambda_{min}(E^*)$ in (3.1).

By definition (2.15), we know $\eta_{2d}(\mathcal{C}) = \sqrt{\lambda_{min}(\mathbf{\Theta}_{\Delta,2d}(\mathcal{C}))}$ for every $\Delta \in \Omega_{2d}$ since $\mathcal{C}$ is symmetric. Let $\Delta = \{1,\ldots,2d\}$, then

$$\mathbf{\Theta}_{\Delta,2d}(\mathcal{C}) = \int_{[-1,1]^{2d}} [x_\Delta]_{G,2d}[x_\Delta]_{G,2d}^T d\mu_\Delta(x_\Delta).$$

If at least one α_i is odd, the integral $\int_{[-1,1]^{2d}} x_\Delta^\alpha d\mu_\Delta(x_\Delta)$ vanishes. If $\alpha = 2(\beta_1,\ldots,\beta_{2d})$ is even, then

$$\int_{[-1,1]^{2d}} x_\Delta^\alpha d\mu_\Delta(x_\Delta) = \frac{1}{(1+2\beta_1)\cdots(1+2\beta_{2d})}.$$

A list of typical values of $\eta_{2d}(\mathcal{C})$ is in Table 2.

$2d$	2	4	6	8
$\eta_{2d}(\mathcal{C})$	0.2678	0.0671	0.0329	0.0096

TABLE 2. A list of $\eta_{2d}(\mathcal{C})$ for $2d = 2,4,6,8$.

Lemma 4.3. *When S is the hypercube $\mathcal{C} = [-1,1]^n$, we have $\eta_{2d}(\mathcal{C})$ is independent of n, Assumption 3.1 holds, and the optimal E^* of (3.1) satisfies $\lambda_{min}(E^*) \geq \frac{1}{d+1}n^{-d}$.*

Proof. We already observed that $\eta_{2d}(\mathcal{C})$ is independent of n. Now we estimate $\lambda_{min}(E^*)$. For $s_d(t) = \frac{1}{d+1}\sum_{k=1}^{d}\sum_{j=0}^{k-1} t^j$, it holds that $s_d(t)(1-t) = 1 - \frac{1}{d+1}\left(1 + t + \cdots + t^d\right)$. In (4.1), replacing t by $\frac{1}{n}\|x\|_2^2$, we get that

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

Hence, there exists a symmetric matrix E such that

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

The polynomial $\frac{1}{n}s_d\left(\frac{1}{n}\|x\|_2^2\right)$ is SOS. So, Assumption 3.1 holds and the optimal E^* in (3.1) satisfies $\lambda_{min}(E^*) \geq \frac{1}{d+1}n^{-d}$. □

Clearly, Theorem 3.4 (ii) and Lemma 4.3 imply the following corollary.

Corollary 4.4. *Let $f \in \mathbb{R}[x]_{2d}$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) on the hypercube $\mathcal{C} = [-1,1]^n$. If f_{sos} is the optimal value of (1.2), then*

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

So, f_{sos} is an $\mathcal{O}(n^{2d})$-approximation of f_{min} as $n \to \infty$.

4.3. **Optimizing square free polynomials over a hypercube.** Consider the case that S is the hypercube $\mathcal{C} = [-1,1]^n$ and $f(x)$ is square free, that is,

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

Recall that $\mathtt{Sfr}[x]_k$ is the space of square free polynomials of degrees at most k.

Lemma 4.5. *It holds that*

$$\chi(\mathtt{Sfr}[x]_k, \mathcal{C}) \leq \sqrt{3}^k, \qquad \lambda_{min}(E^*) \geq \frac{1}{d+1} n^{-d}.$$

Proof. Let $p \in \mathtt{Sfr}[x]_k$ be such that $|p(x)| \leq 1$ for all $x \in \mathcal{C}$. If we write $p(x) = \sum_{\gamma \in \mathtt{Sfr}(\mathbb{N}^n_k)} p_\gamma x^\gamma$, then a simple integration shows

$$1 \geq \frac{1}{2^n} \int_{\mathcal{C}} p(x)^2 dx = \sum_{\gamma \in \mathtt{Sfr}(\mathbb{N}^n_k)} p_\gamma^2 \frac{1}{2^n} \int_{\mathcal{C}} x^{2\gamma} dx = \sum_{\gamma \in \mathtt{Sfr}(\mathbb{N}^n_k)} p_\gamma^2 3^{-|\gamma|} \geq 3^{-k} \|p\|_2^2.$$

Therefore, we have $\|p\|_G \leq \|p(x)\|_2 \leq \sqrt{3}^k$. By definition (3.2), the first inequality follows immediately. The second one is a consequence of Lemma 4.3. □

Lemma 4.5 implies $\chi(\mathtt{Sfr}[x]_{2d}, \mathcal{C}) \leq 3^d$. Hence, Theorem 3.3 implies the following.

Corollary 4.6. *Let $f \in \mathtt{Sfr}[x]_{2d}$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) on $[-1,1]^n$. If f_{sos} is the optimal value of (1.2), then it holds that*

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

4.4. **Optimizing polynomials over boolean sets.** Consider the case that S is the boolean set $\{\pm 1\}^n$. Then $m = 2n$ and

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

The approximation bound of (1.2) for this case is given as below.

Corollary 4.7. *Let $f \in \mathbb{R}[x]_{2d}$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) on $\{\pm 1\}^n$. If f_{sos} is the optimal value of (1.2), then it holds that*

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

Proof. From the proof of Lemma 4.3, we can find an SOS polynomial $s(x)$ satisfying

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

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 $f \in \mathtt{Sfr}[x]_{2d}$ is square free. We claim that

$$\chi(\mathtt{Sfr}[x]_{2d}, \{\pm 1\}^n) \leq 1. \tag{4.2}$$

To see this, suppose $p \in \mathtt{Sfr}[x]_{2d}$ and $|p(x)| \leq 1$ for all $x \in \{\pm 1\}^n$. If we write $p(x) = \sum_{\gamma \in \mathtt{Sfr}(\mathbb{N}^n_{2d})} p_\gamma x^\gamma$, then

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

Hence, $\|p\|_G \leq \|p\|_2 \leq 1$, and (4.2) is true by definition (3.2). So Theorem 3.3 implies the corollary when f is square free.

Second, if f is not square free, there exists $\hat{f} \in \texttt{Sfr}[x]_{2d}$ such that

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

By the previous argument, the corollary is also true. □

After a linear coordinate transformation, the approximation bound in Corollary 4.7 is also true for (1.2) when $S = \{0,1\}^n$. So we get

Corollary 4.8. *Let $f \in \mathbb{R}[x]_{2d}$, and f_{min} (resp., f_{max}) be its minimum (resp., maximum) on $\{0,1\}^n$. If f_{sos} is the optimal value of (1.2), then*

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

4.5. **Optimizing polynomials over quadratically constrained sets.** Consider the case that every $g_i(x) = [x]_1^T Q_i [x]_1$ is quadratic. If Assumption 3.1 holds, Theorem 3.4 can be applied to get a bound. The optimal value $\lambda_{min}(E^*)$ of (3.1) can be estimated by using Q_i. If we set $d = 1$, (3.1) reduces to

$$\text{(4.3)} \qquad \left\{ \begin{array}{rl} \max\limits_{\lambda=(\lambda_1,\ldots,\lambda_m),A} & \lambda_{min}(A) \\ s.t. & [x]_1^T \Big(A + \lambda_1 Q_1 + \cdots + \lambda_m Q_m \Big) [x]_1 = 1, \\ & \lambda_1, \ldots, \lambda_m \ge 0. \end{array} \right.$$

Let (λ^*, A^*) be an optimal solution of (4.3).

Lemma 4.9. *Suppose every $g_i(x)$ is quadratic and $\lambda_{min}(A^*) > 0$ in (4.3). Then Assumption 3.1 holds, and $\lambda_{min}(E^*) \ge \frac{1}{d}\big(\lambda_{min}(A^*)\big)^d$ for (3.1).*

Proof. 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 $[x]_1^T A^* [x]_1$, we get the identity

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

From (4.3), we know $1 \ge [x]_1 A^* [x]_1 \ge \lambda_{min}(A^*)(1 + \|x\|_2^2)$ for all $x \in S$. So there exists a symmetric matrix E satisfying

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

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

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

Hence, Assumption 3.1 holds, and the optimal value of (3.1) is at least $\frac{1}{d}\big(\lambda_{min}(A^*)\big)^d$. □

Clearly, Theorem 3.4 and Lemma 4.9 imply the following corollary.

Corollary 4.10. *Assume every g_i is quadratic and an optimal A^* in (4.3) is positive definite. Let $f \in \mathbb{R}[x]_{2d}$, and f_{min} (resp., f_{max}) be the minimum (resp., maximum) of $f(x)$ on S. If f_{sos} is the optimal value of (1.2), then it holds that*

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

Remark 4.11. If every $g_i(x) = [x]_1^T Q_i [x]_1$ is concave, we can get $\lambda_{min}(A^*)$ exactly. Let $R = \max_{x\in S} \|x\|_2$ be the radius of S. Then $1 \geq [x]_1^T A^* [x]_1 \geq \lambda_{min}(A^*)(1+\|x\|_2^2)$ for all $x \in S$, and hence $\lambda_{min}(A^*) \leq (1+R^2)^{-1}$. The quadratic function $1-(1+R^2)^{-1}[x]_1^T[x]_1$ is concave and nonnegative on S. If $int\big(S\big) \neq \emptyset$, there exist $(\hat{\lambda}_1,\ldots,\hat{\lambda}_m) \geq 0$ satisfying

$$1-(1+R^2)^{-1}(1+\|x\|_2^2) = \hat{\lambda}_1[x]_1^T Q_1 [x]_1 + \cdots + \hat{\lambda}_m [x]_1^T Q_m [x]_1. \tag{4.4}$$

Thus, $(\hat{\lambda}_1,\ldots,\hat{\lambda}_m, \frac{1}{1+R^2} I_{n+1})$ is feasible for (4.3) and $\lambda_{min}(A^*) \geq (1+R^2)^{-1}$. So

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

If some $g_i(x) = [x]_1^T Q_i [x]_1$ is not concave, it would be quite difficult to estimate $\lambda_{min}(A^*)$ in (4.3) because (4.4) might not hold. However, if R is known in advance, the redundant constraint $1-(1+R^2)^{-1}[x]_1^T[x]_1 \geq 0$ could be added, and we still have $\lambda_{min}(A^*) = (1+R^2)^{-1}$. □

Example 4.12 (Multi-unit ball)**.** Suppose $x = (x^{(1)},\ldots,x^{(m)})$ where each $x^{(i)}$ is an n_i-dimensional vector, $g_i(x) = 1-\|x^{(i)}\|_2^2$, $n_i \geq 2d$, and $n_1+\cdots+n_m = n$. Such a set S is called a multi-unit ball. Observe that

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

So the optimal A^* of (4.3) satisfies $\lambda_{min}(A^*) \geq \frac{1}{m+1}$. Obviously $\eta_{2d}(S) > 0$ depends only on d. Corollary 4.10 implies an approximation bound $\mathcal{O}((mn)^d)$ holds for (1.2). □

References

[1] A. Barvinok. Estimating L^∞ norms by L^{2k} norms for functions on orbits. *Found. Comput. Math.*, 2 (2002), no. 4, 393 - 412.

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

[3] J.B. Lasserre. *Moments, Positive Polynomials and Their Applications*, Imperial College Press, 2009.

[4] 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.

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

[6] 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.

[7] 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.

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

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

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

[11] 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.

Department of Mathematics, University of California, San Diego
E-mail address: `njw@math.ucsd.edu`