OPTIMALITY CONDITIONS AND FINITE CONVERGENCE OF LASSERRE'S HIERARCHY

JIAWANG NIE

ABSTRACT. Lasserre's hierarchy is a sequence of semidefinite relaxations for solving polynomial optimization problems globally. This paper studies the relationship between optimality conditions in nonlinear programming theory and finite convergence of Lasserre's hierarchy. Our main results are: i) Lasserre's hierarchy has finite convergence when the constraint qualification, strict complementarity and second order sufficiency conditions hold at every global minimizer, under the standard archimedean condition; the proof uses a result of Marshall on boundary hessian conditions. ii) These optimality conditions are all satisfied at every local minimizer if a finite set of polynomials, which are in the coefficients of input polynomials, do not vanish at the input data (i.e., they hold in a Zariski open set). This implies that, under archimedeanness, Lasserre's hierarchy has finite convergence generically.

1. INTRODUCTION

Given polynomials f, h_i, g_j in $x \in \mathbb{R}^n$, consider the optimization problem

$$(1.1) \qquad \begin{cases} \min & f(x) \\ s.t. & h_i(x) = 0 \, (i = 1, \ldots, m_1), \\ & g_j(x) \geq 0 \, (j = 1, \ldots, m_2). \end{cases}$$

Let K be the feasible set of (1.1). When $m_1 = 0$ (resp. $m_2 = 0$), there are no equality (resp. inequality) constraints. For convenience, denote $h := (h_1, \ldots, h_{m_1})$, $g := (g_1, \ldots, g_{m_2})$ and $g_0 := 1$. A standard approach for solving (1.1) globally is *Lasserre's hierarchy* of semidefinite programming (SDP) relaxations [11]. It is based on a sequence of SOS type representations for polynomials that are nonnegative on K. To describe Lasserre's hierarchy, we first introduce some notation. Let $\mathbb{R}[x]$ be the ring of polynomials with real coefficients and in $x := (x_1, \ldots, x_n)$. A polynomial $p \in \mathbb{R}[x]$ is said to be SOS if $p = p_1^2 + \cdots + p_k^2$ for $p_1, \ldots, p_k \in \mathbb{R}[x]$. The set of all SOS polynomials is denoted by $\Sigma \mathbb{R}[x]^2$. For each $k \in \mathbb{N}$ ($\mathbb{N}$ is the set of nonnegative integers), denote

$$\langle h \rangle_{2k} := \left\{ \sum_{i=1}^{m_1} \phi_i h_i \, \middle| \, \begin{array}{c} \text{each } \phi_i \in \mathbb{R}[x] \\ \text{and } \deg(\phi_i h_i) \leq 2k \end{array} \right\},$$

$$Q_k(g) := \left\{ \sum_{j=0}^{m_2} \sigma_j g_j \, \middle| \, \begin{array}{c} \text{each } \sigma_j \in \Sigma \mathbb{R}[x]^2 \\ \text{and } \deg(\sigma_j g_j) \leq 2k \end{array} \right\}.$$

The set $\langle h \rangle_{2k}$ is called the $2k$-th *truncated ideal* generated by h, and $Q_k(g)$ is called the k-th *truncated quadratic module* generated by g. Lasserre's hierarchy is the sequence of SOS relaxations ($k \in \mathbb{N}$ is called a relaxation order):

$$(1.2) \qquad\qquad \max \quad \gamma \quad \text{s.t.} \quad f - \gamma \in \langle h \rangle_{2k} + Q_k(g).$$

The SOS program (1.2) is equivalent to a semidefinite program [11]. We refer to [12, 14] for surveys in this area.

Let f_{min} denote the minimum value of (1.1) and f_k denote the optimal value of (1.2). Clearly, $f_k \le f_{min}$ for all k and $\{f_k\}$ is monotonically increasing. Under the archimedean condition (i.e., $R - \Sigma_{i=1}^n x_i^2 \in \langle h \rangle_{2t} + Q_t(g)$ for some $t \in \mathbb{N}$ and $R > 0$), Lasserre obtained the asymptotic convergence $f_k \to f_{min}$ as $k \to \infty$, by using Putinar's Positivstellensatz (cf. Theorem 2.1). When $f_k = f_{min}$ for some k, we say Lasserre's hierarchy has *finite convergence*. When $h(x) = 0$ defines a finite set in the complex space $\mathbb{C}^n$, Laurent [13] proved that Lasserre's hierarchy has finite convergence. Indeed, when $h(x) = 0$ defines a finite set in $\mathbb{R}^n$, the sequence $\{f_k\}$ also has finite convergence to f_{min}, as shown in [22]. There exist examples that Lasserre's hierarchy fails to have finite convergence, e.g., when f is the Motzkin polynomial $x_1^2 x_2^2 (x_1^2 + x_2^2 - 3x_3^2) + x_3^6$ and K is the unit ball [20, Example 5.3]. Indeed, such examples always exist when $\dim(K) \ge 3$ (cf. Scheiderer [27, Prop. 6.1]).

However, in practical applications, Lasserre's hierarchy often has finite convergence, e.g., as shown by numerical experiments in Henrion and Lasserre [9, 10]. The known examples for which finite convergence fails are created in very special ways. Since Lasserre proposed his method in [11], people are intrigued very much by the discrepancy between its theory (only asymptotic convergence is guaranteed theoretically) and its practical performance (in applications we often observe finite convergence). The motivation of this paper is trying to resolve this discrepancy. Our main result is that Lasserre's hierarchy has finite convergence when a finite set of polynomials, which are in the coefficients of f and all h_i, g_j, do not vanish at the input data, under the archimedean condition. This implies that, under archimedeanness, Lasserre's hierarchy has finite convergence generically. (We say a property holds *generically* if it holds in the entire space of input data except a set of Lebsgue measure zero.) To prove this, we need to investigate optimality conditions for (1.1).

We here give a short review of optimality conditions in nonlinear programming theory (cf. [1, Section 3.3]). Let u be a local minimizer of (1.1) and $J(u) = \{j_1, \ldots, j_r\}$ be the index set of active inequality constraints. If the *constraint qualification condition (CQC)* holds at u, i.e., the gradients

$$\nabla h_1(u), \ldots, \nabla h_{m_1}(u), \nabla g_{m_1}(u), \ldots, \nabla g_{j_r}(u)$$

are linearly independent, then there exist Lagrange multipliers $\lambda_1, \ldots, \lambda_{m_1}$ and $\mu_1, \ldots, \mu_{m_2}$ satisfying

$$(1.3) \qquad\qquad \nabla f(u) = \sum_{i=1}^{m_1} \lambda_i \nabla h_i(u) + \sum_{j=1}^{m_2} \mu_j \nabla g_j(u),$$

$$(1.4) \qquad \mu_1 g_1(u) = \cdots = \mu_{m_2} g_{m_2}(u) = 0, \quad \mu_1 \ge 0, \ldots, \mu_{m_2} \ge 0.$$

The equation (1.3) is called the *first order optimality condition (FOOC)*, and (1.4) is called the *complementarity condition*. If it further holds that

$$(1.5) \qquad\qquad \mu_1 + g_1(u) > 0, \ldots, \mu_{m_2} + g_{m_2}(u) > 0,$$

we say the *strict complementarity condition (SCC)* holds at u. Note that strict complementarity is equivalent to $\mu_j > 0$ for every $j \in J(u)$. Let $L(x)$ be the associated Lagrange function

$$L(x) := f(x) - \sum_{i=1}^{m_1} \lambda_i h_i(x) - \sum_{j \in J(u)} \mu_j g_j(x).$$

Clearly, (1.3) implies $\nabla_x L(u) = 0$. The polynomials f, h_i, g_j are infinitely many times differentiable everywhere. Thus, under the constraint qualification condition, the *second order necessity condition (SONC)* holds at u, i.e.,

$$(1.6) \qquad v^T \nabla_x^2 L(u) v \geq 0 \quad \text{for all } v \in G(u)^\perp.$$

Here, $G(x)$ denotes the Jacobian of the active constraining polynomials

$$G(x) = \begin{bmatrix} \nabla h_1(x) & \cdots & \nabla h_{m_1}(x) & \nabla g_{j_1}(x) & \cdots & \nabla g_{j_r}(x) \end{bmatrix}^T$$

and $G(u)^\perp$ denotes the null space of $G(u)$. If it holds that

$$(1.7) \qquad v^T \nabla_x^2 L(u) v > 0 \quad \text{for all } 0 \neq v \in G(u)^\perp,$$

we say the *second order sufficiency condition (SOSC)* holds at u.

We summarize the above as follows. If the constraint qualification condition holds at u, then (1.3), (1.4) and (1.6) are necessary conditions for u to be a local minimizer of f on K, but they are not sufficient. If (1.3), (1.4), (1.5) and (1.7) hold at a point $u \in K$, then u is a strict local minimizer of (1.1). The first order optimality, strict complementarity and second order sufficiency conditions are sufficient for strict local optimality. We refer to [1, Section 3.3].

This paper studies the relationship between optimality conditions and finite convergence of Lasserre's hierarchy. Denote $\mathbb{R}[x]_d := \{p \in \mathbb{R}[x] : \deg(p) \leq d\}$ and $[m] := \{1, \ldots, m\}$. Our main conclusions are the following two theorems.

Theorem 1.1. *Suppose the archimedean condition holds for the polynomial tuples h and g in (1.1). If the constraint qualification, strict complementarity and second order sufficiency conditions hold at every global minimizer of (1.1), then Lasserre's hierarchy of (1.2) has finite convergence.*

Theorem 1.2. *Let $d_0, d_1, \ldots, d_{m_1}, d_1', \ldots, d_{m_2}'$ be positive integers. Then there exist a finite set of polynomials $\varphi_1, \ldots, \varphi_L$ (cf. Condition 4.3), which are in the coefficients of polynomials $f \in \mathbb{R}[x]_{d_0}$, $h_i \in \mathbb{R}[x]_{d_i}$ ($i \in [m_1]$), $g_j \in \mathbb{R}[x]_{d_j'}$ ($j \in [m_2]$), such that if $\varphi_1, \ldots, \varphi_L$ do not vanish at the input polynomial, then the constraint qualification, strict complementarity and second order sufficiency conditions hold at every local minimizer of (1.1).*

The proof of Theorem 1.1 uses a result of Marshall on boundary hessian conditions [15, 17], and the proof of Theorem 1.2 uses elimination theory in computational algebra. Theorem 1.2 implies that these classical optimality conditions hold in a Zariski open set in the space of input polynomials with given degrees. The paper is organized as follows. Section 2 presents some backgrounds in the field; Section 3 is mostly to prove Theorem 1.1; Section 4 is mostly to prove Theorem 1.2; Section 5 makes some discussions.

2. Preliminary

2.1. **Notation.** The symbol $\mathbb{R}$ (resp., $\mathbb{C}$) denotes the set of real (resp., complex) numbers. A polynomial is called a form if it is homogeneous. For $f \in \mathbb{R}[x]$, $\widetilde{f}$ denotes the homogenization of f, i.e., $\widetilde{f}(\tilde{x}) = x_0^{\deg(f)} \cdot f(x/x_0)$ with $\tilde{x} := (x_0, x_1, \ldots, x_n)$. The symbol $\|\cdot\|_2$ denotes the standard 2-norm. For a symmetric matrix X, $X \succeq 0$ (resp., $X \succ 0$) means X is positive semidefinite (resp. positive definite). The determinant of a square matrix A is $\det A$. The $N \times N$ identity matrix is denoted as I_N. If p is a polynomial in x, ∇p (resp., $\nabla^2 p$) denotes the gradient (resp., Hessian) of p with respect to x; if p has variables in addition to x, $\nabla_x p$ (resp., $\nabla_x^2 p$) denotes the gradient (resp., Hessian) of p with respect to x. For $p_1, \ldots, p_r \in \mathbb{R}[x]$, $Jac(p_1, \ldots, p_r)|_u$ denotes the Jacobian of $(p_1, \ldots, p_r)$ at u, i.e., $Jac(p_1, \ldots, p_r)|_u = (\partial p_i(u)/\partial x_j)_{1 \le i \le r, 1 \le j \le n}$.

2.2. **Some basics in real algebra.** Here we give a short review on elementary real algebra. More details can be found in [2, 3].

An ideal I of $\mathbb{R}[x]$ is a subset such that $I \cdot \mathbb{R}[x] \subseteq I$ and $I + I \subseteq I$. Given $p_1, \ldots, p_m \in \mathbb{R}[x]$, $\langle p_1, \cdots, p_m \rangle$ denotes the smallest ideal containing all p_i, which is the set $p_1 \cdot \mathbb{R}[x] + \cdots + p_m \cdot \mathbb{R}[x]$. A variety is a subset of $\mathbb{C}^n$ that consists of common zeros of a set of polynomials. A real variety is the intersection of a variety and the real space $\mathbb{R}^n$. Given a polynomial tuple $p = (p_1, \ldots, p_r)$, denote

$$V(p) := \{v \in \mathbb{C}^n : p_1(v) = \cdots = p_r(v) = 0\},$$
$$V_{\mathbb{R}}(p) := \{v \in \mathbb{R}^n : p_1(v) = \cdots = p_r(v) = 0\}.$$

Every set $T \subset \mathbb{R}^n$ is contained in a real variety. The smallest one containing T is called the *Zariski* closure of T, and is denoted by $Zar(T)$. In the Zariski topology on $\mathbb{R}^n$, the real varieties are closed sets, and the complements of real varieties are open sets. Denote $I(T) := \{q \in \mathbb{R}[x] : q(u) = 0 \,\forall\, u \in T\}$, which is an ideal in $\mathbb{R}[x]$ and is called the *vanishing ideal* of T.

Let $h = (h_1, \ldots, h_{m_1})$ and $g = (g_1, \ldots, g_{m_2})$ be the polynomial tuples as in (1.1), and K be the feasible set of (1.1). Recall the definitions of $\langle h \rangle_{2k}$ and $Q_k(g)$ in the Introduction. Clearly, the union $\cup_{k \in \mathbb{N}} \langle h \rangle_{2k}$ is the ideal $\langle h \rangle := \langle h_1, \ldots, h_{m_1} \rangle$. The union $Q(g) := \cup_{k \in \mathbb{N}} Q_k(g)$ is called the *quadratic module* generated by g. The set $\langle h \rangle + Q(g)$ is called *archimedean* if $R - \|x\|_2^2 \in \langle h \rangle + Q(g)$ for some $R > 0$. Clearly, if $p \in \langle h \rangle + Q(g)$, then p is nonnegative on K, while the converse is not always true. However, if p is positive on K and $\langle h \rangle + Q(g)$ is archimedean, then $p \in \langle h \rangle + Q(g)$. This is called *Putinar's Positivstellensatz*.

Theorem 2.1 (Putinar, [23]). *Let K be the feasible set of (1.1). Suppose $\langle h \rangle + Q(g)$ is archimedean. If $p \in \mathbb{R}[x]$ is positive on K, then $p \in \langle h \rangle + Q(g)$.*

2.3. **The boundary hessian condition.** Let K be the feasible set of (1.1) and $h = (h_1, \ldots, h_{m_1})$. Let u be a local minimizer of (1.1), and ℓ be the local dimension of $V_{\mathbb{R}}(h)$ at u (cf. [2, §2.8]). We first state a condition about parameterizing K around u locally, which was proposed by Marshall.

Condition 2.2 (Marshall,[17]). *i) The point u on $V_{\mathbb{R}}(h)$ is nonsingular and there exists a neighborhood $\mathcal{O}$ of u such that $V_{\mathbb{R}}(h) \cap \mathcal{O}$ is parameterized by uniformizing parameters $t_1, \ldots, t_\ell$; ii) there exist $1 \le \nu_1 < \cdots < \nu_r \le m_2$, such that $t_j = g_{\nu_j}$ $(j = 1, \ldots, r)$ on $V_{\mathbb{R}}(h) \cap \mathcal{O}$ and $K \cap \mathcal{O}$ is defined by $t_1 \ge 0, \ldots, t_r \ge 0$.*

The following condition was introduced by Marshall [15, 17] in studying Putinar type representation for nonnegative polynomials, and it is called the *boundary hessian condition* (BHC).

Condition 2.3 (Marshall,[15, 17]). *Assume Condition 2.2 holds. Expand f locally around u as $f = f_0 + f_1 + f_2 + \cdots$, with every f_i being homogeneous of degree i in $t_1, \ldots, t_\ell$. The linear form $f_1 = a_1 t_1 + \cdots + a_r t_r$ for some positive constants $a_1 > 0, \ldots, a_r > 0$, and the quadratic form $f_2(0, \ldots, 0, t_{r+1}, \ldots, t_\ell)$ is positive definite in $(t_{r+1}, \ldots, t_\ell)$.*

If K is compact and the boundary hessian condition holds at every global minimizer, then (1.1) has finitely many global minimizers. (See the proof of Theorem 9.5.3 in [16].) Marshall proved the following important result.

Theorem 2.4. *(Marshall, [16, Theorem 9.5.3]) Let $V = V_{\mathbb{R}}(h)$ and f_{min} be the minimum of (1.1). If $\langle h \rangle + Q(g)$ is archimedean and the boundary hessian condition holds at every global minimizer of (1.1), then $f - f_{min} \in I(V) + Q(g)$.*

In the above, if $I(V) = \langle h \rangle$ (i.e., $\langle h \rangle$ is *real*, [2, §4.1]), then $f - f_{min} \in \langle h \rangle + Q(g)$. Theorem 2.4 can also be found in Scheiderer's survey [28, Theorem 3.1.7].

2.4. **Resultants and discriminants.** Here, we review some basics of resultants and discriminants. We refer to [4, 5, 19, 29] for more details.

Let $f_1, \ldots, f_n$ be forms in $x = (x_1, \ldots, x_n)$. The resultant $Res(f_1, \ldots, f_n)$ is a polynomial, in the coefficients of $f_1, \ldots, f_n$, having the property that

$$Res(f_1, \ldots, f_n) = 0 \quad \Longleftrightarrow \quad \exists\, 0 \neq u \in \mathbb{C}^n,\ f_1(u) = \cdots = f_n(u) = 0.$$

The discriminant of a form f is defined as

$$\Delta(f) := Res\left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right).$$

So, it holds that

$$\Delta(f) = 0 \quad \Longleftrightarrow \quad \exists\, 0 \neq u \in \mathbb{C}^n,\ \nabla f(u) = 0.$$

Both $Res(f_1, \ldots, f_n)$ and $\Delta(f)$ are homogeneous, irreducible and have integer coefficients.

Discriminants and resultants are also defined for nonhomogeneous polynomials. If one of $f_0, f_1, \ldots, f_n$ is not a form in x, then $Res(f_0, f_1, \ldots, f_n)$ is defined to be $Res(\widetilde{f_0}, \ldots, \widetilde{f_n})$, where each $\widetilde{f_i}$ is the homogenization of f_i. Similarly, if f is not a form, then $\Delta(f)$ is defined to be $\Delta(\widetilde{f})$.

Discriminants are also defined for several polynomials [19]. Let $f_1, \ldots, f_m$ be forms in x of degrees $d_1, \ldots, d_m$ respectively, and $m \leq n - 1$. Suppose at least one $d_i > 1$. The discriminant of $f_1, \ldots, f_m$, denoted by $\Delta(f_1, \ldots, f_m)$, is a polynomial in the coefficients of $f_1, \ldots, f_m$, having the property that $\Delta(f_1, \ldots, f_m) = 0$ if and only if there exists $0 \neq u \in \mathbb{C}^n$ satisfying

$$(2.1) \qquad f_1(u) = \cdots = f_m(u) = 0, \quad \text{rank}\begin{bmatrix} \nabla f_1(u) & \cdots & \nabla f_m(u) \end{bmatrix} < m.$$

If one of $f_1, \ldots, f_m$ is nonhomogeneous and $m \leq n$, then $\Delta(f_1, \ldots, f_m)$ is defined to be $\Delta(\widetilde{f_1}, \ldots, \widetilde{f_m})$. In the nonhomogeneous case, $\Delta(f_1, \ldots, f_m) = 0$ if there exists $u \in \mathbb{C}^n$ satisfying (2.1) (cf. [19]).

We conclude this section with an elimination theorem for general homogeneous polynomial systems.

Theorem 2.5. *(Elimination Theory, [7, Theorem 5.7A]) Let $f_1, \ldots, f_r$ be homogeneous polynomials in $x_0, \ldots, x_n$, having indeterminate coefficients a_{ij}. Then there is a set $g_1, \ldots, g_t$ of polynomials in the a_{ij}, with integer coefficients, which are homogeneous in the coefficients of each f_i separately, with the following property: for any field k, and for any set of special values of the $a_{ij} \in k$, a necessary and sufficient condition for the f_i to have a common zero different from $(0, \ldots, 0)$ is that the a_{ij} are a common zero of the polynomials g_j.*

3. Optimality conditions and Finite Convergence

This section is to prove Theorem 1.1. It is based on the following theorem.

Theorem 3.1. *Let u be a local minimizer of (1.1). If the constraint qualification, strict complementarity and second order sufficiency conditions hold at u, then f satisfies the boundary hessian condition at u.*

Proof. Let $J(u) := \{j_1, \ldots, j_r\}$ be the index set of inequality constraints that are active at u. For convenience, we can generally assume $u = 0$, up to a shifting. Since the constraint qualification condition holds at 0, the gradients

$$\nabla h_1(0), \ldots, \nabla h_{m_1}(0), \nabla g_{j_1}(0), \ldots, \nabla g_{j_r}(0)$$

are linearly independent. The origin 0 is a nonsingular point of the real variety $V_{\mathbb{R}}(h)$, because the gradients $\nabla h_1(0), \ldots, \nabla h_{m_1}(0)$ are linearly independent. Up to a linear coordinate transformation, we can further assume that

$$(3.1) \qquad \begin{cases} \begin{bmatrix} \nabla g_{j_1}(0) & \cdots & \nabla g_{j_r}(0) \end{bmatrix} &= \begin{bmatrix} I_r \\ 0 \end{bmatrix}, \\ \begin{bmatrix} \nabla h_1(0) & \cdots & \nabla h_{m_1}(0) \end{bmatrix} &= \begin{bmatrix} 0 \\ I_{m_1} \end{bmatrix}. \end{cases}$$

Let $\ell := n - m_1$, which is the local dimension of $V_{\mathbb{R}}(h_1, \ldots, h_{m_1})$ at 0 (cf. [2, Prop. 3.3.10]). Define a function $\varphi(x) := (\varphi_I(x), \varphi_{II}(x), \varphi_{III}(x)) : \mathbb{R}^n \to \mathbb{R}^n$ as

$$(3.2) \qquad \varphi_I(x) = \begin{bmatrix} g_{j_1}(x) \\ \vdots \\ g_{j_r}(x) \end{bmatrix}, \quad \varphi_{II}(x) = \begin{bmatrix} x_{r+1} \\ \vdots \\ x_\ell \end{bmatrix}, \quad \varphi_{III}(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_{m_1}(x)) \end{bmatrix}.$$

Clearly, $\varphi(0) = 0$, and the Jacobian of φ at 0 is the identity matrix I_n. Thus, by the implicit function theorem, in a neighborhood $\mathcal{O}$ of 0, the equation $t = \varphi(x)$ defines a smooth function $x = \varphi^{-1}(t)$. So, $t = (t_1, \ldots, t_n)$ can serve as a coordinate system for $\mathbb{R}^n$ around 0 and $t = \varphi(x)$. In the t-coordinate system and in the neighborhood $\mathcal{O}$, $V_{\mathbb{R}}(h_1, \ldots, h_{m_1})$ is defined by linear equations $t_{\ell+1} = \cdots = t_n = 0$, and $K \cap \mathcal{O}$ can be equivalently described as

$$t_1 \geq 0, \ldots, t_r \geq 0, \quad t_{\ell+1} = \cdots = t_n = 0.$$

Let $\lambda_i (i \in [m_1])$ and $\mu_j (j \in [m_2])$ be the Lagrange multipliers satisfying (1.3)-(1.4). Define the Lagrange function

$$L(x) := f(x) - \sum_{i=1}^{m_1} \lambda_i h_i(x) - \sum_{k=1}^{r} \mu_{j_k} g_{j_k}(x).$$

Note that $\nabla_x L(0) = 0$. In the t-coordinate system, define functions

$$F(t) := f(\varphi^{-1}(t)), \quad \widehat{L}(t) := L(\varphi^{-1}(t)) = F(t) - \sum_{i=\ell+1}^{n} \lambda_{i-\ell} t_i - \sum_{k=1}^{r} \mu_{j_k} t_k.$$

Clearly, $\nabla_x L(0) = 0$ implies $\nabla_t \widehat{L}(0) = 0$. So, it holds that

$$\frac{\partial F(0)}{\partial t_k} = \mu_{j_k} \, (k = 1, \ldots, r),$$

$$\frac{\partial F(0)}{\partial t_k} = 0 \, (k = r + 1, \ldots, \ell),$$

$$\frac{\partial F(0)}{\partial t_k} = \lambda_{k-\ell} \, (k = \ell + 1, \ldots, n).$$

Expand $F(t)$ locally around 0 as

$$F(t) = f_0 + f_1(t) + f_2(t) + f_3(t) + \cdots$$

where each f_i is a form in t of degree i. Clearly, we have

$$f_1(t) = \mu_{j_1} t_1 + \cdots + \mu_{j_r} t_r \quad \text{on} \quad t_{\ell+1} = \cdots = t_n = 0.$$

For $t_{r+1}, \ldots, t_\ell$ near zero, it holds that

$$F(0, \ldots, 0, t_{r+1}, \ldots, t_\ell, 0, \ldots, 0) = \widehat{L}(0, \ldots, 0, t_{r+1}, \ldots, t_\ell, 0, \ldots, 0) =$$

$$L\big(\varphi^{-1}(0, \ldots, 0, t_{r+1}, \ldots, t_\ell, 0, \ldots, 0)\big).$$

Denote $x(t) := \varphi^{-1}(t) = (\varphi_1^{-1}(t), \ldots, \varphi_n^{-1}(t))$. For all i, j, we have

$$\frac{\partial^2 \widehat{L}(t)}{\partial t_i \partial t_j} = \sum_{1 \le k, s \le n} \frac{\partial^2 L(x(t))}{\partial x_k \partial x_s} \frac{\partial \varphi_k^{-1}(t)}{\partial t_i} \frac{\partial \varphi_s^{-1}(t)}{\partial t_j} + \sum_{1 \le k \le n} \frac{\partial L(x(t))}{\partial x_k} \frac{\partial^2 \varphi_k^{-1}(t)}{\partial t_i \partial t_j}.$$

Evaluating the above at $x = t = 0$, we get (note $\nabla_x L(0) = 0$)

$$\frac{\partial^2 \widehat{L}(0)}{\partial t_i \partial t_j} = \sum_{1 \le k, s \le n} \frac{\partial^2 L(0)}{\partial x_k \partial x_s} \frac{\partial \varphi_k^{-1}(0)}{\partial t_i} \frac{\partial \varphi_s^{-1}(0)}{\partial t_j}.$$

Note that $Jac(\varphi)|_0 = Jac(\varphi^{-1})|_0 = I_n$. So, for all $r + 1 \le i, j \le \ell$, we have

$$(3.3) \qquad \frac{\partial^2 f_2}{\partial t_i \partial t_j}\bigg|_{t=0} = \frac{\partial^2 F}{\partial t_i \partial t_j}\bigg|_{t=0} = \frac{\partial^2 \widehat{L}}{\partial t_i \partial t_j}\bigg|_{t=0} = \frac{\partial^2 L}{\partial x_i \partial x_j}\bigg|_{x=0}.$$

The strict complementarity condition (1.5) implies that $\mu_{j_1} > 0, \ldots, \mu_{j_r} > 0$. So, the coefficients of the linear form $\mu_{j_1} t_1 + \cdots + \mu_{j_r} t_r$ are all positive. The second order sufficiency condition (1.7) implies that the sub-Hessian

$$\left(\frac{\partial^2 L(0)}{\partial x_i \partial x_j} \right)_{r+1 \le i, j \le \ell}$$

is positive definite. By (3.3), the quadratic form f_2 is positive definite in $(t_{r+1}, \ldots, t_\ell)$. Therefore, f satisfies the boundary hessian condition at 0. $\qquad \square$

Now, we give the proof of Theorem 1.1.

Proof of Theorem 1.1. By Theorem 3.1, we know the boundary hessian condition is satisfied at every global minimizer of f on K, when the constraint qualification, strict complementarity and second order sufficiency conditions hold. Then, by Theorem 2.4 of Marshall, we know there exists $\sigma_1 \in Q(g)$ such that

$$f - f_{min} \equiv \sigma_1 \qquad \mod \quad I(V_{\mathbb{R}}(h)).$$

Let $\hat{f} := f - f_{min} - \sigma_1$. Then $\hat{f}$ vanishes identically on $V_{\mathbb{R}}(h)$. By Real Nullstellensatz (cf. [2, Corollary 4.1.8]), there exist $\ell \in \mathbb{N}$ and $\sigma_2 \in \Sigma\mathbb{R}[x]^2$ such that

$$\hat{f}^{2\ell} + \sigma_2 \in \langle h \rangle.$$

Let $c > 0$ be big enough such that $s(t) := 1 + t + ct^{2\ell}$ is an SOS univariate polynomial in t (cf. [22, Lemma 2.1]). For each $\epsilon > 0$, let

$$\sigma_\epsilon := \epsilon s\left(\hat{f}/\epsilon\right) + c\epsilon^{1-2\ell}\sigma_2 + \sigma_1.$$

Then, one can verify that

$$\phi_\epsilon := f - (f_{\min} - \epsilon) - \sigma_\epsilon = -c\epsilon^{1-2\ell}(\hat{f}^{2\ell} + \sigma_2) \in \langle h \rangle.$$

Clearly, there exists $k_0 \in \mathbb{N}$ such that $\sigma_\epsilon \in Q_{k_0}(g)$ and $\phi_\epsilon \in \langle h \rangle_{2k_0}$ for all $\epsilon > 0$. So, for every $\epsilon > 0$, $\gamma = f_{min} - \epsilon$ is feasible in (1.2) for the order k_0. Hence, $f_{k_0} \geq f_{min} - \epsilon$. Since $\epsilon > 0$ can be arbitrary, we get $f_{k_0} \geq f_{min}$. Recall that $f_k \leq f_{min}$ for all k and $\{f_k\}$ is monotonically increasing. Hence, we get $f_k = f_{min}$ for all $k \geq k_0$, i.e., Lasserre's hierarchy has finite convergence. $\qquad\square$

Theorem 3.1 shows that the constraint qualification, strict complementarity and second order sufficiency conditions imply the boundary hessian condition. Typically, to check the boundary hessian condition by its definition, one needs to construct a local parametrization for the feasible set K and verify some sign conditions, which would be very inconvenient in applications. However, checking optimality conditions is generally much more convenient, because it does not need a parametrization and only requires some elementary linear algebra operations. This is an advantage of optimality conditions over the boundary hessian condition. We show this in the following example.

Example 3.2. Consider the optimization problem:

$$\begin{cases} \min & x_1^6 + x_2^6 + x_3^6 + 3x_1^2x_2^2x_3^2 - x_1^4(x_2^2 + x_3^2) - x_2^4(x_3^2 + x_1^2) - x_3^4(x_1^2 + x_2^2) \\ s.t. & x_1^2 + x_2^2 + x_3^2 = 1. \end{cases}$$

The objective is the Robinson form which is nonnegative but not SOS (cf. [24]). The minimum $f_{min} = 0$, and the global minimizers are

$$\frac{1}{\sqrt{3}}(\pm 1, \pm 1, \pm 1), \frac{1}{\sqrt{2}}(\pm 1, \pm 1, 0), \frac{1}{\sqrt{2}}(\pm 1, 0, \pm 1), \frac{1}{\sqrt{2}}(0, \pm 1, \pm 1).$$

The unit sphere is smooth, so the constraint qualification condition holds at every feasible point. There is no inequality constraint, so strict complementarity is automatically satisfied. It can be verified that the second order sufficiency condition (1.7) holds on all the global minimizers. For instance, at $u = \frac{1}{\sqrt{3}}(1,1,1)$,

$$\nabla_x^2 L(u) = \frac{4}{9}\left(3 \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^T\right), \quad G(u)^\perp = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^\perp.$$

Clearly, (1.7) is satisfied at u. By Theorem 1.1, Lasserre's hierarchy for this problem has finite convergence. A numerical experiment by `GloptiPoly 3` [8] verified that $f_5 = f_{min} = 0$, modulo computer round-off errors. $\qquad\square$

In Theorem 1.1, none of the optimality conditions there can be dropped. We show counterexamples as follows.

Example 3.3. (a) Consider the optimization problem:
$$\begin{cases} \min & 3x_1 + 2x_2 \\ s.t. & x_1^2 - x_2^2 - (x_1^2 + x_2^2)^2 \geq 0, \ x_1 \geq 0. \end{cases}$$
It can be shown that the origin 0 is the unique global minimizer. The constraint qualification condition fails at 0, and the first order optimality condition (1.3) fails. The feasible set has nonempty interior, so the SOS program (1.2) achieves its optimal value (cf. [11]). Lasserre's hierarchy for this problem does not have finite convergence, which is implied by Proposition 3.4 in the below.
(b) Consider the optimization problem:
$$\begin{cases} \min & x_1 x_2 + x_1^3 + x_2^3 \\ s.t. & x_1 \geq 0, x_2 \geq 0, 1 - x_1 - x_2 \geq 0. \end{cases}$$
Clearly, 0 is the unique global minimizer. The constraint qualification condition holds at 0. The Lagrange multipliers are all zeros. The second order sufficiency condition (1.7) also holds at 0 because the null space $G(0)^\perp = \{0\}$. However, the strict complementarity condition fails at 0. Lasserre's hierarchy for this problem, does not have finite convergence, as shown by Scheiderer [26, Remark 3.9].
(c) Consider the optimization problem:
$$\begin{cases} \min & x_1^4 x_2^2 + x_1^2 x_2^4 + x_3^6 - 3x_1^2 x_2^2 x_3^2 + \epsilon(x_1^2 + x_2^2 + x_3^2)^3 \\ s.t. & 1 - x_1^2 - x_2^2 - x_3^2 \geq 0. \end{cases}$$
For every $\epsilon > 0$, 0 is the unique global minimizer, and the constraint qualification and strict complementarity conditions hold at 0. However, the second order sufficiency condition fails at 0. For $\epsilon > 0$ sufficiently small, Lasserre's hierarchy for this optimization problem does not have finite convergence, as shown by Marshall [15, Example 2.4]. $\qquad\square$

The first order optimality condition (1.3) is necessary for Lasserre's hierarchy to have finite convergence. This is summarized as follows.

Proposition 3.4. *Suppose (1.2) achieves its optimal value. If the first order optimality condition (1.3) fails at a global minimizer of (1.1), then Lasserre's hierarchy cannot have finite convergence.*

Proof. Suppose otherwise $f_k = f_{min}$ for some k. Since (1.2) achieves its optimum,
$$f - f_{min} = \sum_{i=1}^{m_1} \phi_i h_i + \sum_{j=0}^{m_2} \sigma_j g_j$$
for some $\phi_i \in \mathbb{R}[x]$ and $\sigma_j \in \Sigma\mathbb{R}[x]^2$. Let u be a global minimizer of (1.1). Note that every $h_i(u) = 0$ and $g_j(u)\sigma_j(u) = 0$. Differentiate the above with respect to x and evaluate it at u, then we get
$$\nabla f(u) = \sum_{i=1}^{m_1} \phi_i(u)\nabla h_i(u) + \sum_{j=0}^{m_2} (\sigma_j(u)\nabla g_j(u) + g_j(u)\nabla\sigma_j(u)).$$

Since every σ_j is SOS, $g_j(u)\sigma_j(u) = 0$ implies $g_j(u)\nabla\sigma_j(u) = 0$. Hence,

$$\nabla f(u) = \sum_{i=1}^{m_1} \phi_i(u)\nabla h_i(u) + \sum_{j=0}^{m_2} \sigma_j(u)\nabla g_j(u).$$

But this means that (1.3) holds at u, which is a contradiction. So Lasserre's hierarchy cannot have finite convergence. □

In Proposition 3.4, the assumption that (1.2) achieves its optimal value cannot be dropped. (This assumption is satisfied if K has nonempty interior, cf. [11].) As a counterexample, consider the simple problem

$$\min \quad x \qquad s.t. \quad -x^2 \geq 0.$$

The global minimizer is 0. The first order optimality condition fails at 0, but Lasserre's hierarchy has finite convergence ($f_k = f_{min} = 0$ for all $k \geq 1$).

4. Zariski Openness of Optimality Conditions

This section is mostly to prove Theroem 1.2. For this purpose, we need some results on generic properties of critical points.

4.1. **Generic properties of critical points.** Given polynomials $p_0 \in \mathbb{R}[x]_{d_0}, \ldots, p_k \in \mathbb{R}[x]_{d_k}$ with $k \leq n$, consider the optimization problem

$$(4.1) \qquad \min_{x \in \mathbb{R}^n} \quad p_0(x) \quad s.t. \quad p_1(x) = \cdots = p_k(x) = 0.$$

Its Karush-Kuhn-Tucker (KKT) system is defined by the equations

$$(4.2) \qquad \nabla_x p_0(x) - \sum_{i=1}^{k} \lambda_i \nabla_x p_i(x) = 0, \quad p_1(x) = \cdots = p_k(x) = 0.$$

Every (x, λ) satisfying (4.2) is called a *critical pair*, and such x is called a *critical point*. Let

$$(4.3) \qquad \mathcal{K}(p) := \left\{ x \in \mathbb{C}^n \,\middle|\, \begin{array}{c} \operatorname{rank}\begin{bmatrix} \nabla_x p_0(x) & \nabla_x p_1(x) & \cdots & \nabla_x p_k(x) \end{bmatrix} \leq k \\ p_1(x) = \cdots = p_k(x) = 0 \end{array} \right\}$$

be the *KKT variety* of (4.1). Clearly, every critical point belongs to $\mathcal{K}(p)$.

First, we discuss when does $\mathcal{K}(p)$ intersect the variety $q(x) = 0$ of a polynomial $q \in \mathbb{R}[x]_{d_{k+1}}$, i.e., when does the polynomial system

$$(4.4) \qquad \begin{cases} \operatorname{rank}\begin{bmatrix} \nabla_x p_0(x) & \nabla_x p_1(x) & \cdots & \nabla_x p_k(x) \end{bmatrix} \leq k \\ p_1(x) = \cdots = p_k(x) = 0, \quad q(x) = 0 \end{cases}$$

have a solution in $\mathbb{C}^n$? For a generic p, $\mathcal{K}(p)$ is a finite set (cf. [18, Prop. 2.1]), and it does not intersect $q(x) = 0$ if q is also generic. Consider the homogenization in $x := (x_1, \ldots, x_n)$ of the polynomial system (4.4):

$$(4.5) \qquad \begin{cases} \operatorname{rank}\begin{bmatrix} \nabla_x \widetilde{p_0}(\tilde{x}) & \nabla_x \widetilde{p_1}(\tilde{x}) & \cdots & \nabla_x \widetilde{p_k}(\tilde{x}) \end{bmatrix} \leq k, \\ \widetilde{p_1}(\tilde{x}) = \cdots = \widetilde{p_k}(\tilde{x}) = \widetilde{q}(\tilde{x}) = 0. \end{cases}$$

Its variable is $\tilde{x} := (x_0, \ldots, x_n)$. When $k < n$, the matrix in (4.5) has rank $\leq k$ if and only if all its maximal minors vanish; when $k = n$, the rank condition in (4.5) is always satisfied and can be dropped. Thus, in either case, (4.5) can be equivalently defined by some homogeneous polynomial equations, say,

$$M_1(\tilde{x}) = \cdots = M_\ell(\tilde{x}) = 0.$$

Note that the coefficients of every M_i are also homogeneous in the ones of each of $p_0, \ldots, p_k, q$. By Theorem 2.5, there exist polynomials

$$R_1(p_0, \ldots, p_k; q), \ldots, R_t(p_0, \ldots, p_k; q)$$

in the coefficients of $p_0, \ldots, p_k, q$ such that

- every $R_i(p_0, \ldots, p_k; q)$ has integer coefficients and is homogeneous in the coefficients of each of $p_0, p_1, \ldots, p_k, q$;
- the system (4.5) has a solution $0 \neq \tilde{x} \in \mathbb{C}^{n+1}$ if and only if

$$R_1(p_0, \ldots, p_k; q) = \cdots = R_t(p_0, \ldots, p_k; q) = 0.$$

Define the polynomial $\mathscr{R}(p_0, \ldots, p_k; q)$ as

$$(4.6) \qquad \mathscr{R}(p_0, \ldots, p_k; q) := R_1(p_0, \ldots, p_k; q)^2 + \cdots + R_t(p_0, \ldots, p_k; q)^2.$$

Note that $\mathscr{R}(p_0, \ldots, p_k; q)$ is a polynomial in the coefficients of the tuple

$$(p_0, \ldots, p_k, q) \in \mathbb{R}[x]_{d_0} \times \cdots \times \mathbb{R}[x]_{d_k} \times \mathbb{R}[x]_{d_{k+1}}.$$

Combining the above, we can get the following proposition.

Proposition 4.1. *Let $p_0 \in \mathbb{R}[x]_{d_0}, \ldots, p_k \in \mathbb{R}[x]_{d_k}, q \in \mathbb{R}[x]_{d_{k+1}}$, and $\mathscr{R}$ be as defined in (4.6). Then (4.5) has a solution $0 \neq \tilde{x} \in \mathbb{C}^{n+1}$ if and only if $\mathscr{R}(p_0, \ldots, p_k; q) = 0$. In particular, if $\mathscr{R}(p_0, \ldots, p_k; q) \neq 0$, then (4.4) has no solution in $\mathbb{C}^n$.*

We would like to remark that the polynomial $\mathscr{R}$ in (4.6) does not vanish identically in $(p_0, \ldots, p_k, q) \in \mathbb{R}[x]_{d_0} \times \cdots \times \mathbb{R}[x]_{d_k} \times \mathbb{R}[x]_{d_{k+1}}$, for any given positive degrees $d_0, \ldots, d_k, d_{k+1}$. A proof for this fact is given in the Appendix.

Second, we discuss when the KKT system (4.2) is nonsingular. Denote

$$L_p(x, \lambda) := p_0(x) - \sum_{i=1}^{k} \lambda_i p_i(x).$$

The polynomial system (4.2) is nonsingular if and only if the square matrix

$$H_p(x, \lambda) := \begin{bmatrix} \nabla_x^2 L_p(x, \lambda) & Jac(p_1, \ldots, p_k)|_x^T \\ Jac(p_1, \ldots, p_k)|_x & 0 \end{bmatrix}$$

is nonsingular at every critical pair (x, λ). If every p_i is generic, there are only finitely many critical pairs, and (4.2) is nonsingular if $\det H_p(x, \lambda)$ does not vanish on them.

The matrix $H_p(x, \lambda)$ is singular if and only if there exists $(0, 0) \neq (y, \nu) \in \mathbb{R}^n \times \mathbb{R}^k$ such that

$$(4.7) \qquad \nabla_x^2 L_p(x, \lambda) y + Jac(p_1, \ldots, p_k)|_x^T \nu = 0, \quad y \in \bigcap_{i=1}^{k} \nabla p_i(x)^{\perp}.$$

When $Jac(p_1, \ldots, p_k)|_x$ has full rank k, the existence of a pair $(y, \nu) \neq (0, 0)$ satisfying (4.7) is equivalent to the existence of a pair (y, ν) with $y \neq 0$ satisfying (4.7). When (4.2) is nonsingular, there is no $y \neq 0$ satisfying (4.7) for any critical pair (x, λ). Write $\nu = (\nu_1, \ldots, \nu_k)$, then (4.2) and (4.7) together are equivalent to

$$(4.8) \qquad \begin{cases} \begin{bmatrix} \nabla_x p_0 \\ (\nabla_x^2 p_0) y \end{bmatrix} - \sum_{i=1}^{k} \lambda_i \begin{bmatrix} \nabla_x p_i \\ (\nabla_x^2 p_i) y \end{bmatrix} + \sum_{i=1}^{k} \nu_i \begin{bmatrix} 0 \\ \nabla_x p_i \end{bmatrix} = 0, \\ p_1(x) = \cdots = p_k(x) = (\nabla_x p_1)^T y = \cdots = (\nabla_x p_k)^T y = 0. \end{cases}$$

Define the $(2n) \times (2k+1)$ matrix

$$P(x, y) := \begin{bmatrix} \nabla_x p_0 & \cdots & \nabla_x p_k \\ (\nabla_x^2 p_0)y & \cdots & (\nabla_x^2 p_k)y & \nabla_x p_1 & \cdots & \nabla_x p_k \end{bmatrix}.$$

Clearly, every pair (x, y) in (4.8) satisfies

$$(4.9) \qquad \begin{cases} \operatorname{rank} P(x, y) \le 2k, & p_1(x) = \cdots = p_k(x) = 0, \\ (\nabla_x p_1)^T y = \cdots = (\nabla_x p_k)^T y = 0. \end{cases}$$

If the vectors

$$\begin{bmatrix} \nabla_x p_i \\ (\nabla_x^2 p_i)y \end{bmatrix}, \begin{bmatrix} 0 \\ \nabla_x p_i \end{bmatrix} \qquad (i = 1, \dots, k)$$

are linearly independent, (4.8) and (4.9) are equivalent. Consider the homogenization in x of (4.9):

$$(4.10) \qquad \begin{cases} \operatorname{rank} \widetilde{P}(\tilde{x}, y) \le 2k, & \widetilde{p_1}(\tilde{x}) = \cdots = \widetilde{p_k}(\tilde{x}) = 0, \\ (\nabla_x \widetilde{p_1})^T y = \cdots = (\nabla_x \widetilde{p_k})^T y = 0. \end{cases}$$

In the above, $\tilde{x} := (x_0, \dots, x_n)$ and

$$\widetilde{P}(\tilde{x}, y) := \begin{bmatrix} \nabla_x \widetilde{p_0} & \cdots & \nabla_x \widetilde{p_k} \\ (\nabla_x^2 \widetilde{p_0})y & \cdots & (\nabla_x^2 \widetilde{p_k})y & \nabla_x \widetilde{p_1} & \cdots & \nabla_x \widetilde{p_k} \end{bmatrix}.$$

When $k = n$, we always have $\operatorname{rank} \widetilde{P}(\tilde{x}, y) \le 2k$ and the rank condition in (4.10) can be dropped. When $k < n$, we can replace $\operatorname{rank} \widetilde{P}(\tilde{x}, y) \le 2k$ by the vanishing of all maximal minors of $\widetilde{P}(\tilde{x}, y)$. In either case, (4.10) could be equivalently defined by some polynomial equations, say,

$$N_1(\tilde{x}, y) = \cdots = N_r(\tilde{x}, y) = 0.$$

Note that all $N_1, \dots, N_r$ are homogeneous in both $\tilde{x}$ and y, and their coefficients are also homogeneous in the ones of each of $p_0, p_1, \dots, p_k$. By applying Theorem 2.5 twice (first in $\tilde{x}$ and then in y), there exist polynomials $D_i(p_0, p_1, \dots, p_k)$ ($i = 1, \dots, s$), in the coefficients of $p_0, p_1, \dots, p_k$, such that

- every $D_i(p_0, p_1, \dots, p_k)$ has integer coefficients and is homogeneous in the coefficients of each of $p_0, p_1, \dots, p_k$;
- there exist $0 \ne \tilde{x} \in \mathbb{C}^{n+1}$ and $0 \ne y \in \mathbb{C}^n$ satisfying (4.10) if and only if

$$D_1(p_0, p_1, \dots, p_k) = \cdots = D_s(p_0, p_1, \dots, p_k) = 0.$$

Define the polynomial $\mathscr{D}(p_0, p_1, \dots, p_k)$ as

$$(4.11) \qquad \mathscr{D}(p_0, p_1, \dots, p_k) := D_1(p_0, p_1, \dots, p_k)^2 + \cdots + D_s(p_0, p_1, \dots, p_k)^2.$$

Note that $\mathscr{D}(p_0, \dots, p_k)$ is a polynomial in the coefficients of the tuple

$$(p_0, \dots, p_k) \in \mathbb{R}[x]_{d_0} \times \cdots \times \mathbb{R}[x]_{d_k}.$$

Combining the above, we can get the following proposition.

Proposition 4.2. *Let $p_0 \in \mathbb{R}[x]_{d_0}, \dots, p_k \in \mathbb{R}[x]_{d_k}$ and $\mathscr{D}$ be as defined in (4.11). Then (4.10) has a solution $(\tilde{x}, y) \in \mathbb{C}^{n+1} \times \mathbb{C}^n$ with $\tilde{x} \ne 0, y \ne 0$ if and only if $\mathscr{D}(p_0, \dots, p_k) = 0$. In particular, if $\mathscr{D}(p_0, \dots, p_k) \ne 0$, then (4.2) is a nonsingular system.*

The following special cases are useful to illustrate Proposition 4.2.

- (Every $\deg(p_i) = 1$.) Let $p_i = a_i^T x + b_i$ for $i = 0, \ldots, k$. If $k < n$ and $a_0, a_1, \ldots, a_k$ are linearly independent, then $\widetilde{P}$ is a constant matrix of rank $2k + 1$. If $k = n$ and $a_1, \ldots, a_n$ are linearly independent, then there is no $y \neq 0$ satisfying $\nabla_x \widetilde{p_i}^T y = 0$ for $i = 1, \ldots, n$. So, if every p_i is generic, then (4.10) has no complex solution $(\tilde{x}, y)$ with $\tilde{x} \neq 0, y \neq 0$.

- ($k = 0$, i.e., (4.1) has no constraints.) The system (4.10) is then reduced to

$$(4.12) \qquad \nabla_x \widetilde{p_0}(\tilde{x}) = 0, \quad \left(\nabla_x^2 \widetilde{p_0}(\tilde{x})\right) y = 0.$$

If $\deg(p_0) = 1$ and p_0 is nonzero, $\nabla_x \widetilde{p_0}(\tilde{x}) = 0$ has no complex solution. If $\deg(p_0) = 2$ and $p_0 = x^T A x + 2b^T x + c$ with $\det(A) \neq 0$, there is no $y \neq 0$ satisfying $\left(\nabla_x^2 \widetilde{p_0}(\tilde{x})\right) y = 0$. When $\deg(p_0) \geq 3$, by the definition of discriminants for several polynomials (cf. §2.4), (4.12) has a complex solution $(\tilde{x}, y)$ with $\tilde{x} \neq 0, y \neq 0$ if and only if

$$\Delta\left(\frac{\partial \widetilde{p_0}}{\partial x_1}, \ldots, \frac{\partial \widetilde{p_0}}{\partial x_n}\right) = 0.$$

So, if p_0 is generic, there are no $\tilde{x} \neq 0, y \neq 0$ satisfying (4.12).

The above observations can be simply implied by Proposition 4.2.

In Proposition 4.2, one might naturally think of replacing $\mathscr{D}$ by

$$(4.13) \qquad \Delta(\nabla_x p_0 - Jac(p_1, \ldots, p_k)|_x^T \lambda, p_1, \ldots, p_k),$$

which is the discriminant for the set of polynomials defining (4.2), by considering $\lambda_1, \ldots, \lambda_k$ as new variables, in addition to x. However, this approach is problematic. The main issue is that the discriminantal polynomial in (4.13) might be identically zero, e.g., when $\deg(p_0) \leq \max_{1 \leq i \leq k} \deg(p_i)$. For convenience, consider the simple case $n > k = 1$ and $a := \deg(p_1) - \deg(p_0) \geq 0$. By definition of discriminants for several polynomials (cf. §2.4), the discriminant in (4.13) vanishes if there exists a complex vector $(x_0, x_1, \ldots, x_n, \lambda_1) \neq 0$ satisfying

$$(4.14) \qquad \begin{cases} x_0^{a+1} \cdot \nabla_x \widetilde{p_0} - \lambda_1 \nabla_x \widetilde{p_1} = 0, \quad \widetilde{p_1}(x_0, \ldots, x_n) = 0, \\ \det \begin{bmatrix} x_0^{a+1} \cdot \nabla_x^2 \widetilde{p_0} - \lambda_1 \nabla_x^2 \widetilde{p_1} & \nabla_x \widetilde{p_1} \\ \nabla_x \widetilde{p_1}^T & 0 \end{bmatrix} = 0. \end{cases}$$

Let $(u_1, \ldots, u_n) \neq 0$ be a complex zero of $\widetilde{p_1}(0, x_1, \ldots, x_n)$. Then, $(0, u_1, \ldots, u_n, 0)$ is a nonzero solution of (4.14). So, for any p_0, p_1, (4.14) always has a nonzero complex solution like $(0, u_1, \ldots, u_n, 0)$. This means that the discriminant in (4.13) identically vanishes. On the other hand, the polynomial $\mathscr{D}$ in (4.11) does not vanish identically in $(p_0, \ldots, p_k) \in \mathbb{R}[x]_{d_0} \times \cdots \times \mathbb{R}[x]_{d_k}$, for any given positive degrees $d_0, \ldots, d_k$. A proof for this fact is given in the Appendix.

Typically, the polynomials $\mathscr{R}$ in (4.6) and $\mathscr{D}$ in (4.11) are very difficult to compute explicitly. They are mostly for theoretical interests.

4.2. **Zariski openness of optimality conditions.** This section is to prove that the constraint qualification, strict complementarity and second order sufficiency conditions all hold at every local minimizer of (1.1) if a finite set of polynomials, which are in the coefficients of polynomials $f, h_i \, (i \in [m_1]), g_j \, (j \in [m_2])$, do not not vanish at the input polynomials. (That is, they hold in a Zariski open set in the space of input polynomials.) These polynomials are listed as follows.

Condition 4.3. *The polynomials $f_0 \in \mathbb{R}[x]_{d_0}$, $h_i \in \mathbb{R}[x]_{d_i}$ $(i \in [m_1])$, and $g_j \in \mathbb{R}[x]_{d_j'}$ $(j \in [m_2])$ with $m_1 \le n$ satisfy (Res, Δ are from §2.4, $\mathscr{R}$ from (4.6) and $\mathscr{D}$ from (4.11)):*

(a) *If $m_1 + m_2 \ge n+1$, for all $1 \le j_1 < \cdots < j_{n-m_1+1} \le m_2$,*

$$Res(h_1, \ldots, h_{m_1}, g_{j_1}, \ldots, g_{j_{n-m_1+1}}) \neq 0.$$

(b) *For all $1 \le j_1 < \cdots < j_r \le m_2$ with $0 \le r \le n - m_1$,*

$$\Delta(h_1, \ldots, h_{m_1}, g_{j_1}, \ldots, g_{j_r}) \neq 0.$$

(c) *For all $1 \le j_1 < \cdots < j_r \le m_2$ with $0 \le r \le n - m_1$,*

$$\mathscr{R}(f, p_1, \ldots, p_k; p_{k+1}) \neq 0,$$

where $(p_1, \ldots, p_k, p_{k+1})$ is a re-ordering of $(h_1, \ldots, h_{m_1}, g_{j_1}, \ldots, g_{j_r})$.

(d) *For all $1 \le j_1 < \cdots < j_r \le m_2$ with $0 \le r \le n - m_1$,*

$$\mathscr{D}(f, h_1, \ldots, h_{m_1}, g_{j_1}, \ldots, g_{j_r}) \neq 0.$$

First, we study the relationship between Condition 4.3 and properties of critical points. Let $u \in K$ be a critical point of (1.1) (i.e., (1.3) and (1.4) are satisfied for some λ_i, μ_j, excluding the sign conditions $\mu_j \ge 0$). Let $J(u) := \{j_1, \ldots, j_r\}$ be the index set of active inequality constraints. Denote

$$L(x) := f(x) - \sum_{i=1}^{m_1} \lambda_i h_i(x) - \sum_{j \in J(u)} \mu_j g_j(x),$$

$$G(x) := \begin{bmatrix} \nabla h_1(x) & \cdots & \nabla h_{m_1}(x) & \nabla g_{j_1}(x) & \cdots & \nabla g_{j_r}(x) \end{bmatrix}^T,$$

$$H(x) := \begin{bmatrix} \nabla_x^2 L(x) & G(x)^T \\ G(x) & 0 \end{bmatrix}.$$

Proposition 4.4. *Let $u \in K$ and λ_i, μ_j satisfy (1.3)-(1.4) (excluding the sign conditions $\mu_j \ge 0$), and $L(x), G(x), H(x)$ be as above. Condition 4.3 has the following properties:*

i) *Item (a) implies that at most $n - m_1$ of g_j's are active at every point of K.*

ii) *Item (b) implies that the constraint qualification condition holds at every point of K.*

iii) *Item (c) implies that $\lambda_i \neq 0, \mu_j \neq 0$ for all $i \in [m_1]$ and $j \in J(u)$.*

iv) *Item (d) implies that $H(u)$ is nonsingular, i.e., $\det H(u) \neq 0$.*

Proof. i) If more than $n - m_1$ of g_j's vanish at a point $u \in K$, say, $g_{j_1}, \ldots, g_{j_{n-m_1+1}}$, then there are $n+1$ polynomials vanishing at u, including $h_1, \ldots, h_{m_1}$. This implies the resultant

$$Res(h_1, \ldots, h_{m_1}, g_{j_1}, \ldots, g_{j_{n-m_1+1}}) = 0,$$

which violates item (a) of Condition 4.3. So, the item i) is true.

ii) By item (b) of Condition 4.3, $\Delta(h_1, \ldots, h_{m_1}, g_{j_1}, \ldots, g_{j_r}) \neq 0$. By the definition of Δ (cf. §2.4), the gradients of $h_1, \ldots, h_{m_1}, g_{j_1}, \ldots, g_{j_r}$ at u are linearly independent, i.e., the constraint qualification condition holds at u.

iii) Suppose otherwise one of $\lambda_i (i \in [m_1])$ or $\mu_j (j \in J(u))$ is zero, say, $\mu_{j_r} = 0$, then u is also a critical point of the optimization problem

$$\min \quad f(x) \quad s.t. \quad h_i(x) = 0 \, (i \in [m_1]), \ g_j(x) = 0 \, (j \in J(u)/\{j_r\}).$$

Note that $g_{j_r}(u) = 0$. By definition of $\mathscr{R}$ in (4.6) and Proposition 4.1, we get

$$\mathscr{R}(f, h_1, \ldots, h_{m_1}, g_{j_1}, \ldots, g_{j_{r-1}}; g_{j_r}) = 0,$$

which contradicts item (c) of Condition 4.3. So, the item iii) must be true.

iv) This is implied by definition of $\mathscr{D}$ in (4.11) and Proposition 4.2. $\square$

Second, we study the relationship between Condition 4.3 and the optimality conditions. This is summarized as follows.

Proposition 4.5. *If Condition 4.3 holds, then the constraint qualification, strict complementarity and second order sufficiency conditions all hold at every local minimizer of (1.1). This is implied by the following properties:*

1) *Item (a) of Condition 4.3 implies that at most $n - m_1$ of g_j's are active at every local minimizer of (1.1).*
2) *Item (b) of Condition 4.3 implies that the constraint qualification condition holds at every local minimizer of (1.1).*
3) *Items (b) and (c) of Condition 4.3 imply that the strict complementarity condition holds at every local minimizer of (1.1).*
4) *Items (b) and (d) of Condition 4.3 imply that the second order sufficiency condition holds at every local minimizer of (1.1).*

Proof. Let u be a local minimizer of (1.1).

1) and 2) are implied by i), ii) of Proposition 4.4 respectively.

3) By item 2), the constraint qualification condition holds at u. So, there exist λ_i, μ_j satisfying (1.3)-(1.4) with all $\mu_j \geq 0$. If $j \notin J(u)$, then $g_j(u) > 0$ and $\mu_j + g_j(u) > 0$; if $j \in J(u)$, then $\mu_j \neq 0$ by item iii) of Proposition 4.4, and hence $\mu_j > 0$ and $\mu_j + g_j(u) > 0$. This means that the strict complementarity condition holds at u.

4) By item 2), the constraint qualification condition holds at u. So, (1.3) and (1.4) are satisfied. The second order sufficiency condition is then implied by item iv) of Proposition 4.4 and Lemma 4.6 in the below. $\square$

Lemma 4.6. *Let u be a local minimizer of (1.1), λ_i, μ_j satisfy (1.3)-(1.4), and $L(x), G(x), H(x)$ be as defined preceding Proposition 4.4. If $G(u)$ has full rank, then (1.7) holds at u if and only if $\det H(u) \neq 0$.*

Proof. First, assume (1.7) holds. Then, for $\eta > 0$ big enough,

$$V := \nabla_x^2 L(u) + \eta G(u)^T G(u) \succ 0.$$

By the matrix equation

$$\begin{bmatrix} I_n & \frac{1}{2}\eta G(u)^T \\ 0 & I_{m_1+r} \end{bmatrix} \begin{bmatrix} \nabla_x^2 L(u) & G(u)^T \\ G(u) & 0 \end{bmatrix} \begin{bmatrix} I_n & 0 \\ \frac{1}{2}\eta G(u) & I_{m_1+r} \end{bmatrix} =$$
$$\begin{bmatrix} \nabla_x^2 L(u) + \eta G(u)^T G(u) & G(u)^T \\ G(u) & 0 \end{bmatrix},$$

one can see that

$$\det H(u) = \det(V) \cdot \det\left(-G(u)V^{-1}G(u)^T\right) \neq 0,$$

because of the positive definiteness of V and nonsingularity of $G(u)$.

Second, assume $\det H(u) \neq 0$. Suppose otherwise (1.7) fails. Then there exists $0 \neq v \in G(u)^\perp$ such that $v^T \nabla_x^2 L(0) v \leq 0$. Since $G(u)$ has full rank, the constraint

qualification condition holds at u. So, the second order necessity condition (1.6) is satisfied at u. It implies that v is a minimizer of the problem

$$\min_{z \in \mathbb{R}^n} \quad z^T \left(\nabla_x^2 L(u) \right) z \quad s.t. \quad G(u)z = 0.$$

By the first order optimality condition for the above, there exists ν such that $\nabla_{xx}^2 L(u)v = G(u)^T \nu$, which then implies

$$\begin{bmatrix} \nabla_{xx}^2 L(u) & G(u)^T \\ G(u) & 0 \end{bmatrix} \begin{bmatrix} v \\ -\nu \end{bmatrix} = 0.$$

This contradicts $\det H(u) \neq 0$, because $v \neq 0$. So, (1.7) must hold at u.   □

We conclude this section with the proof of Theorem 1.2.

Proof of Theorem 1.2. Let $\varphi_1, \ldots, \varphi_L$ be the finite set of polynomials given in Condition 4.3. Theorem 1.2 is then implied by Proposition 4.5.   □

5. Some discussions

Our main conclusions are Theorems 1.1 and 1.2. Lasserre's hierarchy has finite convergence when the constraint qualification, strict complementarity and second order sufficiency conditions hold at every global minimizer, under the archimedean condition. These optimality conditions are all satisfied at every local minimizer if the vector of coefficients of input polynomials lies in a Zariski open set. This gives a connection between the classical nonlinear programming theory and Lasserre's hierarchy of semidefinite relaxations in polynomial optimization. These results give an interpretation for the phenomenon that Lasserre's hierarchy often has finite convergence in solving polynomial optimization problems.

Under the assumptions that Condition 2.2 holds at every $u \in K$ and K is irreducible and bounded, Marshall [17, Corollary 4.5] proved that, for each $d \geq 2$, the set

$$\left\{ f \in \mathbb{R}[x]_d : f \text{ satisfies BHC at each global minimizer on } K \right\}$$

is open and dense in $\mathbb{R}[x]_d$. This interesting result can also be implied by Theorems 1.2 and 3.1. Indeed, they can imply the following stronger conclusions:

- the boundary hessian condition is satisfied in a Zariski open set in the space of input data (not every open dense set is Zariski open, e.g., $\mathbb{R}^n \backslash \mathbb{Z}^n$);
- Condition 2.2 also holds in a Zariski open set;
- for the case $d = 1$, the boundary hessian condition also holds in a Zariski open set;
- the defining polynomials for K are also allowed to be generic; the set K is not required to be irreducible or bounded.

We would like to remark that Putinar's Positivstellensatz (cf. Theorem 2.1) also holds generically for polynomials that are nonnegative on K. Assume $\langle h \rangle + Q(g)$ is archimedean and the ideal $\langle h \rangle$ is real. Let $P_d(K)$ be the cone of polynomials in $\mathbb{R}[x]_d$ that are nonnegative on K, and $\partial P_d(K)$ be the boundary of $P_d(K)$. Theorems 1.2, 2.4 and 3.1 imply that if p lies generically on $\partial P_d(K)$ then $p \in \langle h \rangle + Q(g)$. In [11], Lasserre interpreted Putinar's Positivstellensatz as a generalized KKT condition for global optimality. Therefore, the classical KKT conditions for local optimality and the generalized KKT condition (i.e., Putinar's Positivstellensatz, under archimedeanness) for global optimality, both hold generically.

A theoretically interesting question is whether there is a uniform bound on the number of steps to achieve finite convergence for Lasserre's hierarchy in the generic case. That is, whether there exists an integer N, which only depends on the degree of f and a set of defining polynomials for K, such that $f_k = f_{min}$ for all generic f of a given degree and $k \geq N$? Unfortunately, such a bound N typically does not exist. This could be implied by a result of Scheiderer [25] on the non-existence of degree bounds for weighted SOS representations. For instance, when K is the 3-dimensional unit ball, such a bound does not exist (cf. [21, Section 5]).

The archimedean condition cannot be removed in Theorems 1.1, 2.4. For instance, consider the unconstrained optimization

$$\min \quad x_1^2 x_2^2 (x_1^2 + x_2^2 - 3x_3^2) + x_3^6 + x_1^2 + x_2^2 + x_3^2.$$

The origin 0 is the unique minimizer. The archimedean condition failed, because the feasible set is the entire space $\mathbb{R}^n$ and is not compact. The objective f is the sum of the Motzkin polynomial and the positive definite quadratic form $x^T x$. The second order sufficiency condition hold at 0. However, for all scalar γ, $f - \gamma$ is not SOS. In this case, Lasserre's hierarchy does not converge.

The archimedean condition is not generically satisfied. To see this fact, consider the simple case that $m_1 = 0$ and $m_2 = 1$. For a given d, let $\mathscr{A}(d)$ be the set of polynomials $g \in \mathbb{R}[x]_d$ such that $Q(g)$ is archimedean. The set $\mathscr{A}(d)$ is not dense in $\mathbb{R}[x]_d$. For instance, when $d = 2$, both $\mathscr{A}(d)$ and its complement $\mathbb{R}[x]_d \backslash \mathscr{A}(d)$ have nonempty interior:

- Let $b_1 = x^T x - 1$. Clearly, $b_1 \notin \mathbb{R}[x]_2 \backslash \mathscr{A}(2)$. For all $q \in \mathbb{R}[x]_2$ with sufficiently small coefficients, we have $b_1 + q \notin \mathscr{A}(2)$.
- Let $b_2 = 1 - x^T x$. Clearly, $b_2 \in \mathscr{A}(2)$. For all $q \in \mathbb{R}[x]_2$ with small sufficiently coefficients, we have $b_2 + q \in \mathscr{A}(2)$.

There exist polynomial optimization problems that Lasserre's hierarchy fails to have finite convergence, e.g., minimizing the Motzkin polynomial over the unit ball. Such problems always exist when the feasible set has dimension three or higher, as shown by Scheiderer [28]. So, we are also interested in methods that have finite convergence for optimizing *all* polynomials over a given set K. The Jacobian SDP relaxation is a method that has this property (cf. [20]).

Theorems 1.1 does not tell how to check when finite convergence happens. This can be done by using *flat truncation*, which is a rank condition on the dual optimizers of (1.2). Flat truncation is a sufficient condition for Lasserre's hierarchy to have finite convergence. In the generic case, flat truncation is also a necessary condition for Lasserre's hierarchy to have finite convergence (cf. [21]).

No matter Lasserre's hierarchy has finite convergence or not, if there are finitely many global minimizers and the archimedean condition holds, then the flat truncation condition is always asymptotically satisfied (cf. [21, §3]). So, in numerical experiments, we might also observe that Lasserre's hierarchy has finite convergence even if it does not have in exact mathematical computations. However, if there are infinitely many global minimizers, the flat truncation condition is typically not satisfied (cf. Laurent [14, §6.6]). For instance, consider the problem

$$\begin{cases} \min & x_1^2 x_2^2 (x_1^2 + x_2^2) + (x_3 - 1)^6 - 3x_1^2 x_2^2 (x_3 - 1)^2 \\ s.t. & x_1^2 + x_2^2 + x_3^2 - \frac{1}{2} \geq 0, 2 - x_1^2 - x_2^2 - x_3^2 \geq 0. \end{cases}$$

The objective f is shifted from the Motzkin polynomial (i.e., $f(x_1, x_2, x_3 + 1)$ is the Motzkin polynomial). It has infinitely many global minimizers and $f_{min} = 0$.

Lasserre's hierarchy for this problem does not have finite convergence. This can be implied by the proof of Prop. 6.1 of Scheiderer [27], because $(0, 0, 1)$ is a zero f lying in the interior of the feasible set and f is a nonnegative but non-SOS form in $(x_1, x_2, x_3 - 1)$. The flat truncation condition is typically not satisfied for dual optimizers of (1.2). When `GloptiPoly 3` is applied to solve this problem numerically, the convergence did not occur for the orders $k = 3, 4, \ldots, 12$.

Appendix A. Non-identically Vanishing of $\mathscr{R}$ and $\mathscr{D}$

Given any positive degrees $d_0, d_1, \ldots, d_k, d_{k+1}$, we show that the polynomial $\mathscr{R}(p_0, \ldots, p_k; q)$ defined in (4.6) and the polynomial $\mathscr{D}(p_0, p_1, \ldots, p_k)$ defined in (4.11) do not vanish identically in $p_i \in \mathbb{R}[x]_{d_i}$ $(i = 0, \ldots, k)$ and $q \in \mathbb{R}[x]_{d_{k+1}}$. Without loss of generality, we can assume all $d_1, \ldots, d_k > 1$ because linear constraints in (4.2) can be removed by eliminating variables.

First, we prove that the polynomial $\mathscr{R}$ defined in (4.6) does not vanish identically in the space $\mathbb{R}[x]_{d_0} \times \cdots \times \mathbb{R}[x]_{d_k} \times \mathbb{R}[x]_{d_{k+1}}$. We only consider the case $k < n$, because if $k = n$ then $\widetilde{p_1}(\tilde{x}) = \cdots = \widetilde{p_k}(\tilde{x}) = \widetilde{q}(\tilde{x})$ has no nonzero complex solution in the generic case. By Proposition 4.1, it is enough to show that the homogeneous polynomial system (4.5) does not have a complex solution $\tilde{x} \neq 0$ for generic $p_0, p_1, \ldots, p_k, q$. We prove this in two cases:

- ($\mathbf{x_0} \neq \mathbf{0}$) We can scale as $x_0 = 1$, and the system (4.5) is then reduced to (4.4). When $p_0, p_1, \ldots, p_k$ are generic, the set $\mathcal{K}(p)$ defined in (4.3) is finite (cf. [18, Prop. 2.1]). Thus, when q is also generic, (4.4) does not have a solution in $\mathbb{C}^n$.
- ($\mathbf{x_0} = \mathbf{0}$) The system (4.5) is then reduced to

(A.1)
$$\begin{cases} \operatorname{rank} \begin{bmatrix} \nabla_x p_0^h(x) & \nabla_x p_1^h(x) & \cdots & \nabla_x p_k^h(x) \end{bmatrix} \leq k, \\ \quad p_1^h(x) = \cdots = p_k^h(x) = q^h(x) = 0. \end{cases}$$

  (Here, f^h denotes the homogeneous part of the highest degree for a polynomial f.) When $p_1, \ldots, p_k$ are generic, we have $\Delta(p_1^h, \ldots, p_k^h) \neq 0$. By definition of Δ (cf. §2.4), if $p_1^h(x) = \cdots = p_k^h(x) = 0$ and $x \neq 0$, then

$$\operatorname{rank} \begin{bmatrix} \nabla_x p_1^h(x) & \cdots & \nabla_x p_k^h(x) \end{bmatrix} = k.$$

  So, if x satisfies (A.1), there must exist scalars $c_1, \ldots, c_k$ such that

$$\nabla_x p_0^h(x) = c_1 \nabla_x p_1^h(x) + \cdots + c_k \nabla_x p_k^h(x).$$

  Since each p_i^h is a form, by Euler's formula for homogeneous polynomials (cf. [19, §2]), we can get

$$d_0 p_0^h(x) = x^T \nabla_x p_0^h(x) = \sum_{i=1}^{k} c_i x^T \nabla_x p_k^h(x) = \sum_{i=1}^{k} c_i d_k p_k^h(x) = 0.$$

  This means that (A.1) implies

$$p_0^h(x) = p_1^h(x) = \cdots = p_k^h(x) = 0,$$
$$\operatorname{rank} \begin{bmatrix} \nabla_x p_0^h(x) & \nabla_x p_1^h(x) & \cdots & \nabla_x p_k^h(x) \end{bmatrix} = k.$$

Any x satisfying the above must be zero if $\Delta(p_0^h, p_1^h, \ldots, p_k^h) \neq 0$.

Combining the above two cases, we know the polynomial system (4.5) has no complex solution $\tilde{x} \neq 0$ when $p_0, p_1, \ldots, p_k, q$ are generic.

Second, we show that the polynomial $\mathscr{D}(p_0, p_1, \ldots, p_k)$ defined in (4.11) does not identically vanish in the space $\mathbb{R}[x]_{d_0} \times \mathbb{R}[x]_{d_1} \times \cdots \times \mathbb{R}[x]_{d_k}$. By Proposition 4.2, it is enough to prove that there exist $p_i \in \mathbb{R}[x]_{d_i}$ $(i = 0, \ldots, k)$ such that (4.10) has no complex solution $(\tilde{x}, y)$ with $\tilde{x} \neq 0, y \neq 0$. We prove this in two cases.

- ($\mathbf{x_0 \neq 0}$) We scale as $x_0 = 1$, and (4.10) is then reduced to (4.9). Choose polynomials $\hat{p}_i$ as follows:

$$\hat{p}_0 := f_0 \in \mathbb{R}[x_{k+1}, \ldots, x_n]_{d_0}, \ \hat{p}_1 := x_1^{d_1} - 1, \ldots, \hat{p}_k := x_k^{d_k} - 1.$$

Clearly, on the variety $V(\hat{p}_1, \ldots, \hat{p}_k)$, the gradients $\nabla_x \hat{p}_1, \ldots, \nabla_x \hat{p}_k$ are linearly independent, and so are

$$\begin{bmatrix} \nabla_x \hat{p}_i \\ (\nabla_x^2 \hat{p}_i) y \end{bmatrix}, \begin{bmatrix} 0 \\ \nabla_x \hat{p}_i \end{bmatrix} \quad (i = 1, \ldots, k).$$

Thus, (4.9) is equivalent to (4.8). If (x, λ) is a critical pair, then $\lambda_1 = \cdots = \lambda_k = 0$ and $D := \operatorname{diag}(d_1 x_1^{d_1-1}, \ldots, d_k x_k^{d_k-1})$ is invertible. Denote $x_I := (x_1, \ldots, x_k,)$ and $x_{II} := (x_{k+1}, \ldots, x_n)$. Note that $(\hat{p} := (\hat{p}_0, \ldots, \hat{p}_k))$

$$H_{\hat{p}}(x, 0) = \left[\begin{array}{c|c|c} 0 & 0 & D \\ \hline 0 & \nabla_{x_{II}}^2 f_0 & 0 \\ \hline D & 0 & 0 \end{array}\right].$$

(In the above, the 0's denote zero matrices of proper dimensions.) The matrix $H_{\hat{p}}(x, 0)$ is nonsingular if and only if $\nabla_{x_{II}}^2 f_0$ is nonsingular. Therefore, (4.8) has a solution if and only if there exists $u \in \mathbb{C}^{n-k}$ satisfying

$$\nabla_{x_{II}} f_0(u) = 0, \quad \det \nabla_{x_{II}}^2 f_0(u) = 0.$$

However, the above is possible only if

$$\Delta\left(\frac{\partial f_0}{\partial x_{k+1}}, \ldots, \frac{\partial f_0}{\partial x_n}\right) = 0.$$

So, if f_0 is generic, then $H_{\hat{p}}(x, 0)$ is nonsingular for all (x, λ) satisfying (4.2) corresponding to $\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_k$.

By continuity of roots of polynomials, $H_p(x, \lambda)$ is nonsingular for every pair (x, λ) satisfying (4.2), if each p_i is generic and close enough to $\hat{p}_i$.

- ($\mathbf{x_0 = 0}$) The polynomial system (4.10) is then reduced to

(A.2)
$$\begin{cases} \operatorname{rank} Q(x, y) \leq 2k, \quad p_1^h(x) = \cdots = p_k^h(x) = 0, \\ \left(\nabla_x p_1^h(x)\right)^T y = \cdots = \left(\nabla_x p_1^h(x)\right)^T y = 0. \end{cases}$$

In the above, $Q(x, y)$ denotes the matrix

$$\begin{bmatrix} \nabla_x p_0^h & \cdots & \nabla_x p_k^h \\ (\nabla_x^2 p_0^h) y & \cdots & (\nabla_x^2 p_k^h) y & \nabla_x p_1^h & \cdots & \nabla_x p_k^h \end{bmatrix}.$$

We show that if $p_0, p_1, \ldots, p_k$ are generic, then (A.2) has no complex solution (x, y) with $x \neq 0, y \neq 0$. When all p_i are generic, for every $x \neq 0$ satisfying

$$p_1^h(x) = \cdots = p_k^h(x) = 0,$$

the gradients $\nabla_x p_1^h, \ldots, \nabla_x p_k^h$ are linearly independent. When rank $Q(x, y) \le 2k$, there exist scalars $c_1, \ldots, c_k$ such that

(A.3)
$$\nabla_x p_0^h(x) - \sum_{i=1}^{k} c_i \nabla_x p_i^h(x) = 0.$$

By Euler's formula for homogeneous polynomials, the above implies

$$d_0 p_0^h(x) = x^T \nabla_x p_0^h(x) = \sum_{i=1}^{k} \lambda_i x^T \nabla_x p_i^h(x) = \sum_{i=1}^{k} \lambda_i d_i p_i^h(x) = 0.$$

This means that if some $x \ne 0$ satisfies (A.2) then the polynomial system

$$p_0^h(x) = p_1^h(x) = \cdots = p_k^h(x) = 0$$

is singular. But this is impossible unless $\Delta(p_0^h, p_1^h, \ldots, p_k^h) = 0$.

Combining the above two cases, we know that there exist polynomials $p_i \in \mathbb{R}[x]_{d_i}$ ($i = 0, \ldots, k$) such that there are no complex $\tilde{x} \ne 0, y \ne 0$ satisfying (4.10). This shows that $\mathscr{D}(p_0, \ldots, p_k)$ does not identically vanish.

References

[1] D. Bertsekas. *Nonlinear Programming*, second edition. Athena Scientific, 1995.

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

[3] D. Cox, J. Little and D. O'Shea. *Ideals, varieties, and algorithms. An introduction to computational algebraic geometry and commutative algebra.* Third edition. Undergraduate Texts in Mathematics. Springer, New York, 1997.

[4] D. Cox, J. Little and D. O'Shea. *Using algebraic geometry.* Graduate Texts in Mathematics, 185. Springer-Verlag, New York, 1998.

[5] I. Gelfand, M. Kapranov, and A. Zelevinsky. *Discriminants, resultants, and multidimensional determinants.* Mathematics: Theory & Applications, Birkhäuser, 1994.

[6] J. Harris. *Algebraic Geometry, A First Course.* Springer Verlag, 1992.

[7] R. Hartshorne. *Algebraic geometry.* Graduate Texts in Mathematics, No. 52. Springer-Verlag, New York, 1977.

[8] D. Henrion, J. Lasserre and J. Loefberg. *GloptiPoly 3: moments, optimization and semidefinite programming.* `http://homepages.laas.fr/henrion/software/gloptipoly3/`

[9] D. Henrion and J.B. Lasserre. GloptiPoly : Global Optimization over Polynomials with Matlab and SeDuMi. *ACM Trans. Math. Soft.* 29, pp. 165–194, 2003.

[10] D. Henrion and J.B. Lasserre. Detecting global optimality and extracting solutions in GloptiPoly. *Positive polynomials in control*, 293310, Lecture Notes in Control and Inform. Sci., 312, Springer, Berlin, 2005.

[11] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3): 796-817, 2001.

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

[13] M. Laurent. Semidefinite representations for finite varieties. *Mathematical Programming*, Vol. 109, pp. 1–26, 2007.

[14] 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* (Eds. M. Putinar and S. Sullivant), Springer, pages 157-270, 2009.

[15] M. Marshall. Representation of non-negative polynomials with finitely many zeros. *Annales de la Faculte des Sciences Toulouse*, 15 (2006), pp. 599–609.

[16] M. Marshall. *Positive Polynomials and Sums of Squares.* Mathematical Surveys and Monographs, 146. American Mathematical Society, Providence, RI, 2008.

[17] M. Marshall. Representation of non-negative polynomials, degree bounds and applications to optimization. *Canad. J. Math.*, 61 (2009), pp. 205–221.

[18] J. Nie and K. Ranestad. Algebraic Degree of Polynomial Optimization. *SIAM Journal on Optimization*, Vol. 20, No. 1, pp. 485–502, 2009.

[19] J. Nie. Discriminants and Nonnegative Polynomials. *Journal of Symbolic Computation*, Vol. 47, No. 2, pp. 167–191, 2012.

[20] J. Nie. An exact Jacobian SDP relaxation for polynomial optimization. *Mathematical Programming*, Ser. A, Vol. 137, pp. 225–255, 2013.

[21] J. Nie. Certifying convergence of Lasserre's hierarchy via flat truncation. *Mathematical Programming*, to appear.

[22] J. Nie. Polynomial optimization with real varieties. *Preprint*, 2012.

[23] M. Putinar. Positive polynomials on compact semi-algebraic sets, *Ind. Univ. Math. J.* 42 (1993), 969–984.

[24] B. Reznick. Some concrete aspects of Hilbert's 17th problem. In *Contemp. Math.*, Vol. 253, pp. 251-272. American Mathematical Society, 2000.

[25] C. Scheiderer. Non-existence of degree bounds for weighted sums of squares representations. *Journal of Complexity*, 21, 823-844 (2005).

[26] C. Scheiderer. Distinguished representations of non-negative polynomials. *Journal of Algebra*, 289, 558-573 (2005).

[27] C. Scheiderer. Sums of squares of regular functions on real algebraic varieties. *Trans. Am. Math. Soc.*, 352, 1039-1069 (1999).

[28] C. Scheiderer. Positivity and sums of squares: A guide to recent results. *Emerging Applications of Algebraic Geometry (M. Putinar, S. Sullivant, eds.)*, IMA Volumes Math. Appl. 149, Springer, 2009, pp. 271–324.

[29] B. Sturmfels. *Solving systems of polynomial equations*. CBMS Regional Conference Series in Mathematics, 97. American Mathematical Society, Providence, RI, 2002.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, SAN DIEGO
E-mail address: njw@math.ucsd.edu