Convergence of the Lasserre Hierarchy of SDP Relaxations for Convex Polynomial Programs without Compactness*

V. Jeyakumar[†], T. S. Phạm [‡] and G. Li[§]

Revised Version: September 18, 2013

Abstract

The Lasserre hierarchy of semidefinite programming (SDP) relaxations is a powerful scheme for solving polynomial optimization problems with *compact* semi-algebraic sets. In this paper, we show that, for convex polynomial optimization, the Lasserre hierarchy with a slightly extended quadratic module *always converges asymptotically* even in the case of non-compact semi-algebraic feasible sets. We do this by exploiting a coercivity property of convex polynomials that are bounded below. We further establish that the positive definiteness of the Hessian of the associated Lagrangian at a saddle-point (rather than the objective function at each minimizer) guarantees *finite convergence* of the hierarchy. We obtain finite convergence by first establishing a new sum-of-squares polynomial representation of convex polynomials over convex semi-algebraic sets under a saddle-point condition.

Keywords: Convex polynomial optimization, sums-of-squares of polynomials, positivstellensatz, representations, semidefinite programming

AMS subject class: 90C60, 90C56, 90C26

---

*The authors are grateful to the referee and the editors for their constructive comments and helpful suggestions which have contributed to the final preparation of the paper. Research was partially supported by a grant from the Australian Research Council

[†]Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia. E-mail: v.jeyakumar@unsw.edu.au

[‡]Center of Research and Development, Duy Tan University, K7/25, Quang Trung, Danang, Vietnam, and Department of Mathematics, University of Dalat, 1, Phu Dong Thien Vuong, Dalat, Vietnam. E-mail: sonpt@dlu.edu.vn. This work was carried out while the author was a visitor to the School of Mathematics and Statistics, University of New South Wales, Sydney, Australia. This work is also partially funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED), Grant 101.04–2013.07.

[§]Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia. E-mail: g.li@unsw.edu.au

1 Introduction

Consider the polynomial optimization problem:

$$\min_{x\in\mathbb{R}^n}\{f(x) \mid g_i(x) \le 0,\ i = 1, 2, \dots, m\}, \tag{1.1}$$

where $f, g_1, \dots, g_m$ are polynomials on $\mathbb{R}^n$ and

$$K := \{x \in \mathbb{R}^n \mid g_i(x) \le 0,\ i = 1, 2, \dots, m\}$$

is a non-empty basic closed semi-algebraic set. A powerful approach to solving problem (1.1) globally is the Lasserre's hierarchy of semidefinite programming (SDP) relaxations [16, 17]. Under the standard archimedean assumption (see Definition 2.2), the sequence of optimal values of the Lasserre hierarchy of SDP relaxations converges to the optimal value of the original problem (1.1) and we can also get a sequence of points in $\mathbb{R}^n$ converging to a minimizer of the problem (1.1) (see [15, 17, 22, 28]). The archimedean assumption guarantees that the feasible set K is compact. The proof of convergence relies on the powerful sum-of-squares polynomial representation of positive polynomials over compact semi-algebraic sets from real algebraic geometry [24, 27]. Moreover, it has recently been shown that finite convergence occurs for Lasserre's hierarchy generically (see [21]).

On the other hand, it is known that, under the archimedean assumption, Lasserre's hierarchy has finite convergence whenever $f, g_1, \dots, g_m$ are convex polynomials on $\mathbb{R}^n$ and the Hessian of the objective function f is positive definite at each minimizer [14, 16, 17], requiring strict convexity of f (see Lemma 2.1 in Section 2). However, to the best knowledge of the authors, not much is known about convergence of Lasserre's hierarchy for problem (1.1) in the case where K is *not compact* and $f, g_1, \dots, g_m$ are convex polynomials on $\mathbb{R}^n$.

In this paper, without assuming compactness of K, we show that, if $f, g_1, \dots, g_m$ are convex polynomials on $\mathbb{R}^n$ then the Lasserre hierarchy of SDP relaxations with an extended *quadratic module*, generated in terms of both the convex polynomial objective function f and the convex polynomials g_i, $i = 1, 2, \dots, m$, *converges asymptotically.* We prove this by exploiting a coercivity property of convex polynomials that are bounded below. In addition, if the Hessian of the associated Lagrangian function of (1.1) is positive definite at a saddle-point of the Lagrangian, then we show that our hierarchy has finite convergence.

We derive the finite convergence result by first proving that a convex polynomial with *positive definite Hessian at a single point* is strictly convex and coercive, and then establishing that the positive definiteness of the Hessian of the Lagrangian at a saddle-point guarantees a sum-of-squares representation of a convex polynomial over a convex (not necessarily compact) semi-algebraic set. The significance of our sum-of-squares polynomial representation is that it allows us to construct a hierarchy of SDP approximations in terms of *quadratic modules rather than pre-orderings* [3, 4, 5] even in the case of convex polynomial programs with non-compact feasible sets. Also, our representation extends the corresponding known representations of convex polynomials over compact feasible sets, given in [14, 16].

A necessary condition for finite convergence of the Lasserre hierarchy has recently been given in [21, Proposition 3.4], where it was shown that if the Lasserre hierarchy has finite convergence then the well-known Kurash-Kuhn-Tucker (KKT) first-order optimality condition holds at every minimizer of the given non-convex polynomial optimization problem whenever the hierachy achieves its optimal value. Indeed [21, Proposition 3.4] holds for polynomial optimization problems, where the feasible sets are not necessarily compact sets. As a consequence of [21, Proposition 3.4], we obtain that the existence of a saddle-point of the Lagrangian function of problem (1.1) at each minimizer is necessary for finite convergence of our hierarchy of SDP relaxations whenever our hierarchy achieves its optimal value.

2 Asymptotic Convergence without Compactness

We begin by fixing notation, definitions and preliminaries. Throughout this paper $\mathbb{R}^n$ denotes the Euclidean space with dimension n. The inner product in $\mathbb{R}^n$ is defined by $\langle x, y\rangle := x^T y$ for all $x, y \in \mathbb{R}^n$. The non-negative orthant of $\mathbb{R}^n$ is denoted by $\mathbb{R}^n_+$ and is defined by $\mathbb{R}^n_+ := \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_i \geq 0\}$. Denote by $\mathbb{R}[\underline{x}]$ the ring of polynomials in $x := (x_1, x_2, \ldots, x_n)$ with real coefficients. For a polynomial f with real coefficients, we use $\deg f$ to denote the degree of f.

A symmetric $n \times n$ matrix A is said to be *positive definite*, denoted by $A \succ 0$, whenever $x^T Ax > 0$ for all $x \in \mathbb{R}^n, x \neq 0$. The gradient and the Hessian of a real polynomial $f \in \mathbb{R}[\underline{x}]$ at a point x^* are denoted by $\nabla f(x^*)$ and $\nabla^2 f(x^*)$ respectively. Moreover, for a function $L\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, we use $\nabla^2_{xx} L(x^*, \lambda^*)$ to denote the Hessian of L at $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$ with respect to the variable x.

We say that a real polynomial $f \in \mathbb{R}[\underline{x}]$ is *sum-of-squares* (SOS) if there exist real polynomials $f_j, j = 1, \ldots, r$, such that $f = \sum_{j=1}^r f_j^2$. The set of all sum-of-squares real polynomials is denoted by Σ^2.

Definition 2.1. (Quadratic module) *A* quadratic module generated by polynomials $-g_1, \ldots, -g_m \in \mathbb{R}[\underline{x}]$ *is defined as*

$$\mathbf{M}(-g_1, \ldots, -g_m) := \{\sigma_0 - \sigma_1 g_1 - \cdots - \sigma_m g_m \mid \sigma_i \in \Sigma^2, i = 0, 1, \ldots, m\}.$$

A quadratic module is a subset of polynomials that are non-negative on the set $\{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \ldots, m\}$ and it possess a very nice certificate for this property.

The following archimedean condition (see [17, 19]) has played a key role in the study of polynomial optimization.

Definition 2.2. (Archimedean) *The quadratic module* $\mathbf{M}(-g_1, \ldots, -g_m)$ *is called* Archimedean *if there exists* $p \in \mathbf{M}(-g_1, \ldots, -g_m)$ *such that* $\{x : p(x) \geq 0\}$ *is compact.*

When the quadratic module $\mathbf{M}(-g_1, \dots, -g_m)$ is Archimedean, we have the following important characterization of positivity of a polynomial over a semialgebraic set.

Lemma 2.1. (Putinar positivstellensatz) [24] *Let $f, g_i,\ i = 1, \dots, m,$ be real polynomials with $K := \{x : g_i(x) \le 0, i = 1, \dots, m\} \neq \emptyset$. Suppose that $M(-g_1, \dots, -g_m)$ is Archimedean. If $f(x) > 0$ for all $x \in K$, then $f \in M(-g_1, \dots, -g_m)$.*

Now, consider the convex polynomial programming problem, discussed in the Introduction:

$$f^* \ := \ \min_{x \in \mathbb{R}^n} \{f(x) \mid g_i(x) \le 0,\ i = 1, 2, \dots, m\}, \tag{2.2}$$

where $f, g_1, \dots, g_m$ are convex polynomials on $\mathbb{R}^n$ and

$$K := \{x \in \mathbb{R}^n \mid g_1(x) \le 0, \dots, g_m(x) \le 0\} \neq \emptyset.$$

Let $c \in \mathbb{R}$ be such that $c > f(x^0)$ for some $x^0 \in K$. For each integer k, we define the truncated quadratic module $\mathbf{M}_k$ generated by the polynomials $c - f$ and $-g_1, \dots, -g_m$ as

$$\begin{aligned}\mathbf{M}_k \ := \ \{\sigma_0 - \sum_{i=1}^m \sigma_i g_i + \sigma(c - f) \ \mid \ & \sigma, \sigma_0, \sigma_1, \dots, \sigma_m \in \Sigma^2 \subset \mathbb{R}[\underline{x}], \\ & \deg\ \sigma_0 \le 2k,\ \deg\ \sigma_i g_i \le 2k \text{ and } \deg\ \sigma(c - f) \le 2k\}.\end{aligned}$$

Consider the following SDP relaxation problem of (2.2)

$$(P_k) \qquad f_k^* := \sup_{\mu \in \mathbb{R}} \{\mu \mid f - \mu \in \mathbf{M}_k\}. \tag{2.3}$$

It is well known that computing the supremum f_k^* can be equivalently reformulated as a semidefinite programming problem (see [15], [17], [19]). Moreover, it can be easily verified that

$$f_k^* \le f_{k+1}^* \le \cdots \le f^*.$$

Note that the relaxation problem (2.3) is the Lasserre hierarchy of SDP relaxations of the convex program

$$\min_{x \in \mathbb{R}^n} \{f(x) : g_i(x) \le 0,\ i = 1, \dots, m,\ f(x) - c \le 0\}.$$

Definition 2.3. (Coercivity) *A polynomial $f : \mathbb{R}^n \to \mathbb{R}$ is coercive whenever the lower level set $\{x \in \mathbb{R}^n : f(x) \le \alpha\}$ is a (possibly empty) compact set, for all $\alpha \in \mathbb{R}$.*

The following useful coercivity property of a convex polynomial, that is bounded below, allows us to establish that the Lasserre hierarchy of SDP relaxations of Problem (2.2) *has an asymptotic convergence* in the sense that $f_k^* \uparrow f^*$ as $k \to \infty$.

Lemma 2.2 (**Coercivity and Convex Polynomials**). *Let $h \in \mathbb{R}[\underline{x}]$ be a convex polynomial which is bounded below on $\mathbb{R}^n$. Then there exist an orthogonal $n \times n$ matrix A and a coercive polynomial $g\colon \mathbb{R}^l \to \mathbb{R}$, $1 \le l \le n$, such that*

$$h(Ax) = h(A(x_1, \dots, x_l, \dots, x_n)) = g(x_1, \dots, x_l),\ \forall\ x = (x_1, \dots, x_l, \dots, x_n)^T \in \mathbb{R}^n.$$

In particular, h attains its infimum on $\mathbb{R}^n$.

Proof. The proof is given in the Appendix. □

Recall that for a function $f : \mathbb{R}^n \to \mathbb{R}$ and a set $C \subset \mathbb{R}^n$, $\mathrm{argmin}_{x \in C} f(x)$ is defined by

$$\mathrm{argmin}_{x\in C} f(x) = \{a \in C : f(a) = \inf_{x \in C} f(x)\}.$$

The following known existence result of a convex polynomial program will also be useful for the proof of the asymptotic convergence.

Lemma 2.3. *[1] Let $f_0, f_1, \dots, f_m \in \mathbb{R}[\underline{x}]$ be convex polynomials. Let $C := \{x \in \mathbb{R}^n : f_i(x) \le 0, i = 1, \dots, m\}$. Suppose that $\inf_{x \in C} f_0(x) > -\infty$. Then, $\mathrm{argmin}_{x\in C} f_0(x) \neq \emptyset$.*

Theorem 2.1 (**Asymptotic Convergence**). *Let f and $g_1, \dots, g_m \in \mathbb{R}[\underline{x}]$ be convex polynomials. Let x^* be a minimizer of the convex polynomial optimization problem* (2.2)*. Then, $\lim_{k\to\infty} f_k^* = f^*$.*

Proof. [**Positivity of Approximate Lagrangian by Convex Programming Duality**]. Let $\epsilon > 0$. We first prove that there exists $\lambda \in \mathbb{R}^m_+$ such that

$$f(x) - f(x^*) + \sum_{i=1}^m \lambda_i g_i(x) + \epsilon > 0, \qquad \forall x \in \mathbb{R}^n. \tag{2.4}$$

To see this, note that $f - f(x^*) \ge 0$ on K, where $K := \{x \in \mathbb{R}^n \mid g_1(x) \le 0, \dots, g_m(x) \le 0\}$. Then, $f + \epsilon - f(x^*) > 0$ on K. So, there exists $\delta > 0$ such that $f + \epsilon - f(x^*) > 0$ on K_δ, where $K_\delta := \{x \in \mathbb{R}^n \mid g_1(x) \le \delta, \dots, g_m(x) \le \delta\}$. [Otherwise, we can find a sequence $\{\delta_k\} \subset \mathbb{R}_+$, $\delta_k \to 0$ and $\{x_k\} \subset \mathbb{R}^n$ such that $g_i(x_k) \le \delta_k$, $i = 1, 2, \dots, m$ and $f(x_k) + \epsilon - f(x^*) \le 0$. Then,

$$\begin{aligned} 0 &\le \inf_{x, z_1, \dots, z_m} \{\sum_{i=1}^m z_i^2 \mid f(x) + \epsilon - f(x^*) \le 0,\ g_i(x) - z_i \le 0,\ i = 1, \dots, m\} \\ &\le \sum_{i=1}^m \delta_k^2 = m\delta_k^2 \to 0, \quad \text{as} \quad k \to \infty. \end{aligned}$$

So, from Lemma 2.3 that there exist $y^* \in \mathbb{R}^n$ and $z^* = (z_1^*, \dots, z_m^*) \in \mathbb{R}^m$ such that $f(y^*) + \epsilon - f(x^*) \le 0$, $g_i(y^*) - z_i^* \le 0, i = 1, \dots, m$, and $\sum_{i=1}^m z_i^{*2} = 0$. Thus, $f(y^*) + \epsilon - f(x^*) \le 0$ and $g_i(y^*) \le 0$, $i = 1, 2, \dots, m$. This contradicts the property that $f + \epsilon - f(x^*) > 0$ on K].

Now, by Lemma 2.3, f attains its minimizer at $w^* \in K_\delta$ with $f(w^*)+\epsilon-f(x^*) > 0$. As $g_i(x^*) \leq 0 < \delta, i = 1, 2, \ldots, m$, the Slater condition holds for the constraints, $g_1(x) \leq \delta, \ldots, g_m(x) \leq \delta$. So, by the convex programming duality [8, 9, 10], there exist $\lambda_i \geq 0, i = 1, 2, \ldots, m$ such that, for all $x \in \mathbb{R}^n$, $f(x) + \sum_{i=1}^m \lambda_i(g_i(x) - \delta) \geq f(w^*)$. This gives us that, for all $x \in \mathbb{R}^n$, $f(x)+\sum_{i=1}^m \lambda_i g_i(x) \geq f(w^*)+\sum_{i=1}^m \lambda_i\delta \geq f(w^*) > f(x^*) - \epsilon$. Thus, our claim (2.4) holds.

[**Asymptotic Representation by Putinar Positivstellensatz**]. Define $h : \mathbb{R}^n \to \mathbb{R}$ by

$$h(x) := f(x) - f(x^*) + \sum_{i=1}^m \lambda_i g_i(x) + \epsilon, \ \forall \ x \in \mathbb{R}^n.$$

Clearly, h is a convex polynomial which is positive over $\mathbb{R}^n$. Lemma 2.2 shows that there exist an orthogonal $n \times n$ matrix A and a coercive polynomial $g\colon \mathbb{R}^l \to \mathbb{R}$ such that

$$h(A(x_1, \ldots, x_l, \ldots, x_n)) = g(x_1, \ldots, x_l), \ \forall \ x = (x_1, \ldots, x_l, \ldots, x_n) \in \mathbb{R}^n. \quad (2.5)$$

Let $T = \{x \in \mathbb{R}^n : h(x) \leq c - f(x^*) + \epsilon\}$. Clearly, T is nonempty as $x^0 \in T$. As g is coercive on $\mathbb{R}^l$, it follows from (2.5) that

$$S := \{x \in \mathbb{R}^l \mid g(x_1, \ldots, x_l) \leq c - f(x^*) + \epsilon\}$$

is a nonempty and compact set. Since h is positive, it follows from (2.5) that $g > 0$ over $\mathbb{R}^l$, and in particular $g > 0$ over S.

Let $p(x) := g(x) - c + f(x^*) - \epsilon$, for $x \in \mathbb{R}^l$. Then, the quadratic module $\mathbf{M}(-p)$ is Archimedean as $-p \in \mathbf{M}(-p)$ and $\{x : -p(x) \geq 0\} = S$ is compact. Now, from the Putinar Positivstellensatz (Lemma 2.1), we get that there exist sum-of-squares polynomials σ_0, σ_1 over $\mathbb{R}^l$ such that

$$g = \sigma_0 + \sigma_1(c - f(x^*) - g + \epsilon).$$

From (2.5), we see that, for each $x = (x_1, \ldots, x_l, x_{l+1}, \ldots, x_n) \in \mathbb{R}^n$, $h(Ax) = g(x_1, \ldots, x_l)$. So, for each $x = (x_1, \ldots, x_l, x_{l+1}, \ldots, x_n) \in \mathbb{R}^n$,

$$h(Ax) = \sigma_0(x_1, \ldots, x_l) + \sigma_1(x_1, \ldots, x_l)(c - f(x^*) - h(Ax) + \epsilon).$$

Then, for each $z \in \mathbb{R}^n$,

$$h(z) = \sigma_0\big((A^{-1}z)_1, \ldots, (A^{-1}z)_l\big) + \sigma_1\big((A^{-1}z)_1, \ldots, (A^{-1}z)_l\big)(c - f(x^*) - h(z) + \epsilon).$$

Using the definition of h, we obtain that, for each $z \in \mathbb{R}^n$,

$$\begin{aligned} & f(z) - f(x^*) + \sum_{i=1}^m \lambda_i g_i(z) + \epsilon \\ = \ & \sigma_0\big((A^{-1}z)_1, \ldots, (A^{-1}z)_l\big) + \sigma_1\big((A^{-1}z)_1, \ldots, (A^{-1}z)_l\big)(c - f(z) - \sum_{i=1}^m \lambda_i g_i(z)). \end{aligned}$$

Thus, for each $z \in \mathbb{R}^n$,

$$
\begin{aligned}
& f(z) - f(x^*) + \epsilon \qquad (2.6)\\
= \; & \sigma_0\big((A^{-1}z)_1, \dots, (A^{-1}z)_l\big) + \sigma_1\big((A^{-1}z)_1, \dots, (A^{-1}z)_l\big)(c - f(z)) \\
& - \sum_{i=1}^{m} \left(\sigma_1\big((A^{-1}z)_1, \dots, (A^{-1}z)_l\big)\lambda_i + \lambda_i\right) g_i(z) \ ,
\end{aligned}
$$

where $z \mapsto \sigma_i\big((A^{-1}z)_1, \dots, (A^{-1}z)_l\big)$, $i = 0, 1$, are sum-of-squares polynomials and $\lambda_i \geq 0$, for $i = 1, 2, \dots, m$.

[**Convergence from Asymptotic Representation**]. Equation (2.6) shows that, for each $\epsilon > 0$, $f - f^* + \epsilon \in \mathbf{M}(-g_1, \dots, -g_m, c - f)$. So, for each $\epsilon > 0$, there exists $k \in \mathbb{N}$ such that $f^* - \epsilon \leq f_k^*$. This together with the fact that $f_k^* \leq f_{k+1}^* \leq \dots \leq f^*$ gives us that $\lim_{k \to \infty} f_k^* = f^*$. □

3 Representations and Finite Convergence

In this section, we present new representation results for non-negativity of convex polynomials over convex semi-algebraic sets. For related results, see [6, 7, 20, 23, 28, 29] and other references therein.

The following Lemma on strict convexity and coercivity of convex polynomials plays a key role in proving the desired representation of convex polynomials and also the finite convergence of our hierarchy.

Lemma 3.1 (**Hessian Condition for Coercivity and Strict Convexity).** *Let $f \in \mathbb{R}[\underline{x}]$ be a convex polynomial. If $\nabla^2 f(x_0) \succ 0$ at some point $x_0 \in \mathbb{R}^n$, then f is coercive and strictly convex on $\mathbb{R}^n$.*

Proof. A simple proof is given in the Appendix. □

Let f and $g_1, \dots, g_m \in \mathbb{R}[\underline{x}]$ be convex polynomials with $K := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \dots, m\} \neq \emptyset$. Suppose that $\text{argmin}_{x \in K} f(x) \neq \emptyset$ and that there exists $x^* \in \text{argmin}_{x \in K} f(x)$. Then, convex programming duality [8, 9, 10, 11] shows that if there exists $x^0 \in \mathbb{R}^n$ such that $g_i(x^0) < 0$ for $i = 1, \dots, m$, then there exists $\lambda^* \in \mathbb{R}^m_+$ such that (x^*, λ^*) is *a saddle-point of the Lagrangian function* $L(x, \lambda) := f(x) + \sum_{i=1}^m \lambda_i g_i(x)$ in the sense that,

$$
(\forall x \in \mathbb{R}^n) \ (\forall \lambda \in \mathbb{R}^m_+) \ \ L(x, \lambda^*) \geq L(x^*, \lambda^*) \geq L(x^*, \lambda). \qquad (3.7)
$$

Theorem 3.1 (**Representation of Convex Polynomials).** *Let f and $g_1, \dots, g_m \in \mathbb{R}[\underline{x}]$ be convex polynomials with $K := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \dots, m\} \neq \emptyset$. Let $L\colon \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}$ be the Lagrangian function defined by $L(x, \lambda) := f(x) + \sum_{i=1}^m \lambda_i g_i(x)$. If the Lagrangian function L has a saddle-point $(x^*, \lambda^*) \in K \times \mathbb{R}^m_+$ with $\nabla^2_{xx} L(x^*, \lambda^*) \succ 0$, then, for any $c \in \mathbb{R}$ with $c > f(x^*)$, we have*

$$
f - f(x^*) \in \mathbf{M}(-g_1, \dots, -g_m, c - f).
$$

Proof. Since (x^*, λ^*) is a saddle-point of the Lagrangian function L and $x^* \in K$, it follows that, for each $x \in \mathbb{R}^n$, $L(x, \lambda^*) \geq L(x^*, \lambda^*) = f(x^*)$ and x^* is a minimizer of f over K. Let

$$h(x) := L(x, \lambda^*) - f(x^*) = f(x) - f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x), \qquad \forall x \in \mathbb{R}^n.$$

Clearly h is a convex polynomial and $h(x) \geq 0$, for all $x \in \mathbb{R}^n$. Moreover, it is easy to check that $h(x^*) = 0 = \inf_{x \in \mathbb{R}^n} h(x)$; in particular, $\nabla h(x^*) = 0$. By a direct calculation, we see that the Hessian $\nabla^2 h$ of h at x^* is positive definite. Now, Lemma 3.1 shows us that the polynomial h is strictly convex and coercive, which implies that x^* is the unique minimizer of h on $\mathbb{R}^n$ and that

$$S := \{x \in \mathbb{R}^n \mid h(x) \leq c - f(x^*)\}$$

is a nonempty compact set. Since

$$-h(x) + c - f(x^*) = c - f(x) - \sum_{i=1}^{m} \lambda_i^* g_i(x) \in \mathbf{M}(c - f, -g_1, \ldots, -g_m)$$

and $S = \{x \in \mathbb{R}^n \mid -h(x) + c - f(x^*) \geq 0\}$ is compact, it follows that the quadratic module $\mathbf{M}(c - f, -g_1, \ldots, -g_m)$ is Archimedean.

We apply [26, Corollary 3.6] (see also [25, Example 3.18]) to conclude that there exist sum-of-squares polynomials $\sigma_0, \sigma_1 \in \Sigma^2$ such that, for each $x \in \mathbb{R}^n$,

$$h(x) \quad = \quad \sigma_0(x) + \sigma_1(x)(-h(x) + c - f(x^*)).$$

So, for each $x \in \mathbb{R}^n$,

$$f(x) - f(x^*) \quad = \quad \sigma_0 - \sum_{i=1}^{m} (\lambda_i^* + \lambda_i^* \sigma_1) g_i(x) + \sigma_1(c - f(x)).$$

Hence, the conclusion follows. □

Example 3.1 (Importance of positive definite Hessian of L at a saddle–point for representation). Let $p \in \mathbb{R}[x]$ be a convex form (i.e., homogeneous polynomial) on $\mathbb{R}^n$ with degree at least 4 which is not a sum-of-squares polynomial. See [2] for the existence of such polynomials.

Let f, g be convex polynomials on $\mathbb{R}^n \times \mathbb{R}$ defined by $f(x, y) := p(x)$ and $g(x, y) := y^2 - 1$. Then, f is not strictly convex. Let $f^* := \min_{x \in K} f(x, y)$, where $K := \{(x, y) \in \mathbb{R}^n \times \mathbb{R} \mid g(x, y) \leq 0\} = \mathbb{R}^n \times [-1, 1]$.

Then $f^* = 0$ because $f(0, 1) = 0$, $\nabla f(0, 1) = 0$ and f is convex. Consider the corresponding Lagrangian $L\colon \mathbb{R}^{n+1} \times \mathbb{R}_+ \to \mathbb{R}$ defined by $L(x, y, \lambda) := f(x, y) + \lambda g(x, y)$. Clearly, $(x^*, y^*, \lambda^*) := (0, 1, 0)$ is a saddle point of L as $L(x^*, y^*, \lambda) = L(x^*, y^*, \lambda^*) = 0 \leq f(x, y) = L(x, y, \lambda^*)$ for all $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}_+$. Moreover, as

$\nabla^2 p(x^*) = \nabla^2 p(0) = 0$, the Hessian of the Lagrangian function L is not positive definite at the point (x^*, y^*, λ^*).

We now show that the representation of Theorem 3.1 fails. To see this, note that the quadratic module $\mathbf{M}(1 - \|x\|^2) \subset \mathbb{R}[x]$ is Archimedean. So, there exists $c > f(x^*, y^*) = 0$ such that $c - p \in \mathbf{M}(1 - \|x\|^2)$ (for example, see [19, Corollary 5.2.4]).

On the contrary, suppose that the representation of Theorem 3.1 holds. Then,

$$f(x,y) = \sigma_0(x,y) - \sigma_1(x,y)g(x,y) + \sigma(x,y)(c - f(x,y)), \text{ for all } (x,y) \in \mathbb{R}^n \times \mathbb{R},$$

for some sum-of-squares polynomials $\sigma, \sigma_0, \sigma_1$ in the ring $\mathbb{R}[x,y]$. Letting $y = 1$ and noting that $g(x,1) = 0$, we see that, for all $x \in \mathbb{R}^n$

$$\begin{aligned} p(x) = f(x,1) &= \sigma_0(x,1) + \sigma(x,1)(c - f(x,1)) \\ &= \sigma_0(x,1) + \sigma(x,1)(c - p(x)). \end{aligned}$$

So, $p \in \mathbf{M}(c - p) \subset \mathbb{R}[x]$. Then we have $p \in \mathbf{M}(1 - \|x\|^2)$. By Proposition 4 in De Klerk, Laurent, and Parrilo [13], a form belongs to the quadratic module $\mathbf{M}(1-\|x\|^2)$ if and only if it is a sum-of-squares polynomial. This contradicts our assumption that the polynomial p is not a sum-of-squares. Thus, the representation fails in this case.

As an easy application of Theorem 3.1, we obtain the following representation under the Archimedean assumption. For related results, see [16, Theorem 3.4] and [14, Corollary 3.3].

Corollary 3.1 (Representation with Archimedean Condition). *Let $f, g_1, \ldots, g_m$ be convex polynomials, and let $K := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \ldots, m\} \neq \emptyset$. Suppose that the following assumptions hold:*

(i) *There exists $x^0 \in \mathbb{R}^n$ such that $g_i(x^0) < 0$ for $i = 1, \ldots, m$;*

(ii) *$\nabla^2_{xx} L(x^*, \lambda^*) \succ 0$ at a saddle-point $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m_+$ of the Lagrangian function L;*

(iii) *The quadratic module $\mathbf{M}(-g_1, \ldots, -g_m)$ is Archimedean.*

Then, $f - f(x^*) \in \mathbf{M}(-g_1, \ldots, -g_m)$.

Proof. The assumption (iii) implies that the set K is compact, and so $\operatorname{argmin}_{x \in K} f(x) \neq \emptyset$. The assumption (i) guarantees that there exists $\lambda^* \in \mathbb{R}^m_+$ such that (x^*, λ^*) is a saddle-point of the Lagrangian function L. Let $c \in \mathbb{N}$ be an arbitrary natural number satisfying $c > f(x^*)$. It follows from Theorem 3.1 that $f - f(x^*) \in \mathbf{M}(-g_1, \ldots, -g_m, c - f)$.

On the other hand, by taking c large enough, if necessary, from the assumption (iii) we may assume that $c - f \in \mathbf{M}(-g_1, \ldots, -g_m)$. Therefore $f - f(x^*) \in \mathbf{M}(-g_1, \ldots, -g_m, c - f) = \mathbf{M}(-g_1, \ldots, -g_m)$, which completes the proof. □

Remark 3.1 (Comparisons with known recent results). In the special case where the Hessian $\nabla^2 f$ of the objective function f is positive definite at a minimizer $x^* \in \text{argmin}_{x\in K} f(x)$, then the Slater condition ensures that there exists $\lambda^* \in \mathbb{R}^m_+$ such that (x^*, λ^*) is a saddle-point of the Lagrangian function L, and so, the Hessian $\nabla^2_{xx} L$ of L is positive definite at (x^*, λ^*). Hence, it is easy to see that the above corollary extends the representation results for convex polynomial optimization established in [16, Theorem 3.4] and [14, Corollary 3.3].

The following simple one dimensional example illustrates that our representation result can be applied to the case where the Hessian $\nabla^2 f$ of the objective function f is not positive definite at a minimizer.

Example 3.2. (Verifying representation: Non-positive definiteness case of the Hessian $\nabla^2 f$) Let $f(x) = x$ and $g(x) = x^2 - 1$. Then, $K := \{x \in \mathbb{R} \mid g(x) \leq 0\} = [-1, 1]$. Clearly, $\text{argmin}_{x\in} K f(x) = \{-1\}$ and f is not positive definite at the unique minimizer $x^* := -1$. On the other hand, direct verification shows that $(x^*, \lambda^*) := (-1, \frac{1}{2})$ is a saddle point of the Lagrangian function $L(x, \lambda) := f(x) + \lambda g(x) = x + \lambda(x^2 - 1)$, and $\nabla^2_{xx} L(x^*, \lambda^*) \succ 0$. Moreover, Slater condition is satisfied and the quadratic module $\mathbf{M}(-g)$ is Archimedean. So, it follows from the previous corollary that $f - f(x^*) = f + 1 \in \mathbf{M}(-g)$. Indeed, $f - f(x^*) = x + 1 = \frac{1}{2}(x + 1)^2 + \frac{1}{2}(1 - x^2) \in \mathbf{M}(-g)$.

As we see in the following theorem, under the Slater condition and the positive definiteness of the Hessian of f at a minimizer, we obtain a sharper representation than the one in Theorem 3.1.

Theorem 3.2 (Sharp Representation with positive definite $\nabla^2 f(x^*)$). *Let f and $g_1, \ldots, g_m \in \mathbb{R}[\underline{x}]$ be convex polynomials with $K := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \ldots, m\} \neq \emptyset$. Let $\text{argmin}_{x\in K} f(x) \neq \emptyset$ and $x^* \in \text{argmin}_{x\in K} f(x)$. If there exists $x^0 \in \mathbb{R}^n$ such that $g_i(x^0) < 0$, for $i = 1, \ldots, m$ and if $\nabla^2 f(x^*) \succ 0$ then, for any $c > f(x^*)$, there exist sum-of-squares polynomials $\sigma_0, \sigma_1 \in \Sigma^2$ and Lagrange multipliers $\lambda_i^* \geq 0$, $i = 1, 2, \ldots, m$ such that*

$$f - f(x^*) = \sigma_0 - \sum_{i=1}^{m} \lambda_i^* g_i + \sigma_1(c - f).$$

Proof. The Slater condition and convex programming duality guarantee that there exists $\lambda^* \in \mathbb{R}^m_+$ such that (x^*, λ^*) is a saddle-point of the Lagrangian function $L(x, \lambda) := f(x) + \sum_{i=1}^m \lambda_i g_i(x)$. So, for each $x \in \mathbb{R}^n$, $L(x, \lambda^*) \geq L(x^*, \lambda^*) = f(x^*)$. Then,

$$h(x) := L(x, \lambda^*) - f(x^*) = f(x) - f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x) \geq 0, \qquad \forall x \in \mathbb{R}^n.$$

Now, as $\nabla^2 f(x^*) \succ 0$, Lemma 3.1 shows that f is a strictly convex and coercive polynomial. Then, the convex set $\bar{S} := \{x \in \mathbb{R}^n \mid f(x) \leq c\}$ is nonempty and compact. On the other hand, as $c - f \in \mathbf{M}(c-f)$ and $\{x \in \mathbb{R}^n \mid c - f(x) \geq 0\} = \bar{S}$ is compact, it follows that the quadratic module $\mathbf{M}(c-f)$ is Archimedean. Now, as $h \geq 0$ on $\bar{S}$, [26, Corollary 3.6] (see also [25, Example 3.18]) guarantees that there exist sum-of-squares polynomials $\sigma_0, \sigma_1 \in \Sigma^2$ such that, for each $x \in \mathbb{R}^n$,

$$h(x) \quad = \quad \sigma_0(x) + \sigma_1(x)(c - f(x)).$$

So, for each $x \in \mathbb{R}^n$, $f(x) - f(x^*) = \sigma_0 - \sum_{i=1}^m \lambda_i^* g_i(x) + \sigma_1(c - f(x))$. Hence, the conclusion follows. □

Remark 3.2 (Constraint qualifications). In Corollary 3.1 and Theorem 3.2, we have used the Slater condition for guaranteeing the existence of a saddle-point. For other general constraint qualifications ensuring the existence of a saddle point of the Lagrangian function, see [10, 11, 12].

Definition 3.1. (Finite convergence) *We say that the hierarchy of SDP relaxations (P_k) of problem (2.2) has* finite convergence *whenever $f_k^* = f^*$ for some integer k.*

We now show that our hierarchy of SDP relaxations of problem (2.2) has *finite convergence* under suitable conditions.

Theorem 3.3 (Finite convergence of our hierachy). *For the problem* (2.2), *let f and $g_1, \ldots, g_m \in \mathbb{R}[\underline{x}]$ be convex polynomials. Let $L\colon \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}$ be the Lagrangian function defined by $L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i g_i(x)$. Assume that the Lagrangian function L has a saddle-point $(x^*, \lambda^*) \in K \times \mathbb{R}^m_+$ with $\nabla^2_{xx} L(x^*, \lambda^*) \succ 0$. Then there exists an integer k such that $f_k^* = f^*$ and the problem (P_k) achieves its optimal value.*

Proof. We know that $f_k^* \leq f^*$ for all $k \geq 1$. On the other hand, it follows from Theorem 3.1 that there exist sum-of-squares polynomials $\sigma, \sigma_0, \sigma_1, \ldots, \sigma_m \in \Sigma^2$ such that

$$f - f^* \quad = \quad \sigma_0 - \sigma_1 g_1 - \cdots - \sigma_m g_m + \sigma(c - f).$$

Hence $f_k^* = f^*$ for some $k \in \mathbb{N}$. As $(x^*, \lambda^*) \in K \times \mathbb{R}^m_+$ is a saddle-point of L, x^* is minimizer of Problem (2.2) and $f^* = f(x^*)$ which is also a solution of (P_k). □

Remark 3.3. It is worth noting that in the case where $\nabla^2 f(x^*)$ is positive definite at a minimizer x^* of Problem (2.2), using Theorem 3.2, one can establish finite convergence of a sharper form of approximation problem (2.3), where σ_i, $i = 1, 2, \ldots, m$, in the truncated quadratic module $\mathbf{M}_k$ are replaced by the Lagrange multipliers, λ_i^*, $i = 1, 2, \ldots, m$, associated with the minimizer x^*.

The following example shows that the finite convergence in the preceding theorem may fail if the saddle-point condition does not hold at a minimizer.

Example 3.3 (Importance of Saddle-point Condition for Finite Convergence). Consider the minimization problem

$$\min\{f(x,y) \mid g(x,y) \le 0 \}, \tag{3.8}$$

where $f(x,y) = x^2+y^2+x+y$, $g(x,y) = x^2+y^2$ and $K := \{(x,y) \in \mathbb{R}^2 \mid g(x,y) \le 0\}$.

Clearly, the unique minimizer of (3.8) is $(x^*,y^*) = (0,0)$, $f^* := f(x^*,y^*) = 0$ and $\nabla^2 f(x^*,y^*) = \mathrm{diag}(2,2) \succ 0$. It is easy to check that the saddle-point condition is not satisfied at $(x^*,y^*) = (0,0)$.

Now, let c be a real number such that $c > f^* = 0$. For each k, the kth-order relaxation problem of (3.8) is

$$\sup_{\mu\in\mathbb{R}}\{\mu \mid f-\mu \in \mathbf{M}_k\},$$

where $\mathbf{M}_k := \{\sigma_0 - \sigma_1 g + \sigma(c-f) \mid \sigma,\sigma_0,\sigma_1 \in \Sigma^2, \deg\sigma_0 \le 2k, \deg\sigma_1 g \le 2k, \deg\sigma(c-f) \le 2k\}$. We now show that the finite convergence fails. We establish this by the method of contradiction. Suppose that problem (3.8) has finite convergence. Then, there exists $k_0 \in \mathbb{N}$, $\sigma,\sigma_0,\sigma_1 \in \Sigma^2$ with $\deg\sigma_0 \le 2k_0, \deg\sigma_1 g \le 2k_0$ and $\deg\sigma(c-f) \le 2k_0$ such that $f = f - f^* = \sigma_0 - \sigma_1 g + \sigma(c-f)$. This gives us that, for each $(x,y) \in \mathbb{R}^2$,

$$\big(1+\sigma(x,y)+\sigma_1(x,y)\big)(x^2+y^2)+(1+\sigma(x,y))(x+y) = \sigma_0(x,y)+c\sigma(x,y) \ge 0. \tag{3.9}$$

Letting $(x,y) = (-\frac{1}{k},-\frac{1}{k})$ in (3.9), where $k \in \mathbb{N}$, yields

$$\big(1+\sigma(-\frac{1}{k},-\frac{1}{k})+\sigma_1(-\frac{1}{k},-\frac{1}{k})\big)\frac{2}{k^2} - (1+\sigma_1(-\frac{1}{k},-\frac{1}{k}))\frac{2}{k} \ge 0.$$

Then,

$$1+\sigma(-\frac{1}{k},-\frac{1}{k})+\sigma_1(-\frac{1}{k},-\frac{1}{k}) \ge k(1+\sigma_1(-\frac{1}{k},-\frac{1}{k})) \ge k,$$

which is impossible as the left hand side converges to $1+\sigma(0,0)+\sigma_1(0,0)$.

As a consequence of [21, Proposition 3.4], we obtain that the existence of a saddle-point of the Lagrangian function of problem (2.2) at each minimizer is necessary for finite convergence of our hierachy of SDP relaxations whenever the SDP relaxation (P_k) of problem (2.2) achieves its optimal value f_k^*, for all k sufficiently large.

Proposition 3.1 (Necessity of Saddle-point for Finite Convergence). *For problem (2.2), let f and $g_1,\ldots,g_m \in \mathbb{R}[\underline{x}]$ be convex polynomials. Suppose that the SDP relaxation (P_k) of problem (2.2) achieves its optimal value f_k^*, for all k sufficiently large. If the saddle-point condition (3.7) fails at a minimizer of problem (2.2), then the hierarchy of SDP relaxations (P_k) of problem (2.2) does not have finite convergence.*

Proof. By contrary, assume that the hierarchy of SDP relaxations (P_k) of problem (2.2) has finite convergence. Let $x^* \in K$ with $f^* := f(x^*) = \min_{x\in K} f(x)$. Then, x^* is a minimizer of the following convex program:

$$\min_{x\in\mathbb{R}^n}\{f(x) : g_i(x) \le 0,\, i = 1,\dots,m,\;\; f(x) - c \le 0\} \tag{3.10}$$

and (P_k) is the Lasserre hierarchy of SDP relaxations of (3.10), where c is a real number such that $c > f(x^0)$ for some $x^0 \in K = \{x : g_i(x) \le 0, i = 1,\dots,m\}$. Now, by our assumption, $f^*_{k_0} = f^*$ for some integer k_0. Since $f_k \le f^*$ and the sequence $\{f_k\}$ is monotonically increasing, we get that $f^*_k = f^*$ for all $k \ge k_0$. By increasing k_0 if necessary, we can assume without loss of generality that problem (P_{k_0}) achieves its optimal value $f^*_{k_0}$. So, it follows from [21, Proposition 3.4] that the first-order necessary optimality conditions hold for problem (3.10) at x^*. Thus, there exist $\lambda^* \in \mathbb{R}^m$ and $\gamma^* \in \mathbb{R}$ such that

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \gamma^* \nabla f(x^*) = 0,\;\; \sum_{i=1}^m \lambda_i^* g_i(x^*) = 0,\;\; \gamma^*(f(x^*) - c) = 0.$$

As $c > f(x_0) \ge f(x^*)$, $\gamma^* = 0$. So, $\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) = 0$. Hence, by convexity of $f + \sum_{i=1}^m \lambda_i^* g_i$, we get that, for each $x \in \mathbb{R}^n$,

$$f(x) + \sum_{i=1}^m \lambda_i^* g_i(x) \ge f(x^*) + \sum_{i=1}^m \lambda_i^* g_i(x^*).$$

It is now easy to check that (x^*, λ^*) is a saddle-point of the Lagrangian function of problem (2.2). □

4 Appendix:Proofs of Lemma 2.2 and Lemma 3.1

Proof of Lemma 2.2. Let

$$E_h := \{d \in \mathbb{R}^n \mid h(x + td) = h(x),\, \forall\, t \in \mathbb{R} \text{ and } \forall\, x \in \mathbb{R}^n\}.$$

Then, it is easy to verify directly that E_h is a subspace of $\mathbb{R}^n$. Let $l := n - \dim E_h$, and let $e_1,\dots,e_n \in \mathbb{R}^n$ be an orthonormal basis such that $\operatorname{span}\{e_{l+1},\dots,e_n\} = E_h$ and $\operatorname{span}\{e_1,\dots,e_l\} = E_h^\perp$, where $E_h^\perp$ is the orthogonal complement of E_h. Let $A := [e_1,\dots,e_n]$. Then, A is an orthogonal matrix. Define $g\colon \mathbb{R}^l \to \mathbb{R}$ by $g(x_1,\dots,x_l) := h\left(\sum_{i=1}^l x_i e_i\right)$. Then, g is a convex polynomial which is bounded below on $\mathbb{R}^l$. Moreover, for all $x \in \mathbb{R}^n$, we have

$$h(Ax) = h\left(\sum_{i=1}^n x_i e_i\right) = h\left(\sum_{i=1}^l x_i e_i + \sum_{i=l+1}^n x_i e_i\right) = h\left(\sum_{i=1}^l x_i e_i\right) = g(x_1,\dots,x_l),$$

where the third equality follows by the fact that $\sum_{i=l+1}^n x_i e_i \in E_h$.

To verify that g is indeed coercive, we assume, on the contrary, that $S := \{x : g(x) \leq \alpha\}$ is unbounded for some $\alpha \in \mathbb{R}$. Let $\{a_k\} \subseteq S$ such that $\|a_k\| \to +\infty$ as $k \to \infty$. Let $a \in \mathbb{R}^l$. Then, by passing to subsequence if necessary, we may assume that $\frac{a_k - a}{\|a_k - a\|} \to v \neq 0$. Let $t \geq 0$. For sufficiently large k, we have $0 < \frac{t}{\|a_k - a\|} < 1$, and so

$$\begin{aligned} g\left(a + t\frac{a_k - a}{\|a_k - a\|}\right) &= g\left(\left(1 - \frac{t}{\|a_k - a\|}\right)a + \frac{t}{\|a_k - a\|}a_k\right) \\ &\leq \left(1 - \frac{t}{\|a_k - a\|}\right)g(a) + \frac{t}{\|a_k - a\|}g(a^k) \\ &\leq \max\{g(a), \alpha\}. \end{aligned}$$

Letting $k \to \infty$, we get that $g(a+tv) \leq \max\{g(a), \alpha\}$ for all $t \geq 0$. By assumption, g is bounded below. So, $t \mapsto g(a+tv)$ is either a constant or a polynomial with even degree ≥ 2. It then follows that g takes a constant value on $\{a + tv : t \geq 0\}$ for all $a \in \mathbb{R}^l$. Then, for all $t \geq 0$ and for any $a \in \mathbb{R}^l$, $g(a - tv) = g(a - tv + tv) = g(a)$. Thus,

$$g(a) = g(a + tv) \text{ for all } a \in \mathbb{R}^l \text{ and } t \in \mathbb{R}. \tag{4.11}$$

Let $\tilde{v} := (v^T, 0, \ldots, 0)^T \in \mathbb{R}^n$ and $d := A\tilde{v} = \sum_{i=1}^l v_i e_i \in E_h^\perp$. Since $v \neq 0$, $d \neq 0$. Moreover, for all $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$,

$$h(x + td) = h(A(A^{-1}x + t\tilde{v})) = g(z + tv) = g(z) = h(x),$$

where $z = \big((A^{-1}x)_1, \ldots, (A^{-1}x)_l\big) \in \mathbb{R}^l$. So, by definition, $d \in E_h$. Consequently, we obtain that $d \in (E_h \cap E_h^\perp)\backslash\{0\}$, which is impossible. Hence, g is coercive.

Since the polynomial g is coercive, there exists $z^* := (z_1^*, \ldots, z_l^*) \in \mathbb{R}^l$ such that $g(z^*) = \inf_{z \in \mathbb{R}^l} g(z)$. Let $x^* := A\binom{z^*}{0} = z_1^* e_1 + \cdots + z_l^* e_l \in E_h^\perp \subset \mathbb{R}^n$. Then, $h(x^*) = \inf_{x \in \mathbb{R}^n} h(x) = g(z^*)$.

Proof of Lemma 3.1. (**Coercivity**) Let c be any real number such that $c \geq f(x_0)$. To prove coercivity of f on $\mathbb{R}^n$, we show that the set

$$S := \{x \in \mathbb{R}^n \mid f(x) \leq c\}$$

is compact. On the contrary, suppose that there exists a sequence $\{a_k\}_{k \geq 0} \subset S$ such that $\|a_k\| \to \infty$ as $k \to \infty$. Without lost of generality, we may assume that there exists $v \neq 0$ such that

$$v := \lim_{k \to \infty} \frac{a_k - x_0}{\|a_k - x_0\|}.$$

Let $t \geq 0$. For sufficiently large k, we have $0 < \frac{t}{\|a_k - x_0\|} < 1$, and so

$$\begin{aligned} f\left(x_0 + t\frac{a_k - x_0}{\|a_k - x_0\|}\right) &= f\left(\left(1 - \frac{t}{\|a_k - x_0\|}\right)x_0 + \frac{t}{\|a_k - x_0\|}a_k\right) \\ &\leq \left(1 - \frac{t}{\|a_k - x_0\|}\right)f(x_0) + \frac{t}{\|a_k - x_0\|}f(a_k) \\ &\leq c. \end{aligned}$$

Letting $k \to \infty$, we get

$$f(x_0 + tv) \le c, \quad \text{for all} \quad t \ge 0.$$

On the other hand, as the Hessian $\nabla^2 f(x_0)$ is positive definite, $\langle \nabla^2 f(x_0)v, v\rangle > 0$ and so, for each $t \in \mathbb{R}$,

$$f(x_0 + tv) \;=\; f(x_0) + \langle \nabla f(x_0), v\rangle t + \frac{1}{2}\langle \nabla^2 f(x_0)v, v\rangle t^2 + \text{ higher order terms in } t.$$

Hence, the one dimensional convex polynomial $t \mapsto f(x_0+tv)$ is of even degree ≥ 2. This is a contradiction since $f(x_0 + tv) \le c \quad$ for all $\quad t \ge 0$.
(**Strict Convexity**) We establish strict convexity of f by the method of contradiction and suppose that f is not strictly convex. Then, there exist $x, y \in \mathbb{R}^n, x \ne y$, and $t_0 \in (0,1)$ such that

$$f((1-t_0)x + t_0 y) = (1-t_0)f(x) + t_0 f(y).$$

Define $h : [0,1] \to \mathbb{R}$ by $h(t) = f((1-t)x + ty) - (1-t)f(x) - tf(y)$. Then, h is a convex polynomial, $h(t) \le 0$, for each $t \in [0,1]$ and $h(t_0) = 0 = \max_{t\in[0,1]} h(t)$. As h is a convex function on $[0,1]$, it attains its maximum on the extreme points of $[0,1]$, and so,

$$f((1-t)x + ty) = (1-t)f(x) + tf(y), \quad \forall t \in [0,1].$$

Now, define a polynomial φ on $\mathbb{R}$ by $\varphi(\lambda) := f\big(x + \lambda(y-x)\big)$, $\lambda \in \mathbb{R}$. Clearly, φ is affine on $[0,1]$, and moreover, it is coercive on $\mathbb{R}$ because f is coercive on $\mathbb{R}^n$, shown above. We show that φ is indeed affine over $\mathbb{R}$. Let the degree of the one-dimensional polynomial φ be d. Then, for each $\lambda \in \mathbb{R}$,

$$\varphi(\lambda) = \varphi(0) + \varphi'(0)\lambda + \frac{\varphi''(0)}{2}\lambda^2 + \cdots + \frac{\varphi^{(d)}(0)}{d!}\lambda^d.$$

As φ is affine over $[0,1]$, $\varphi^{(i)}(0) = 0$ for $i = 2, \ldots, d$, and so, $\varphi(\lambda) = \varphi(0) + \varphi'(0)\lambda$. Hence, φ is affine over $\mathbb{R}$. This contradicts the fact that φ is coercive on $\mathbb{R}$.

Remark 4.1. The conclusion of Lemma 3.1 may also be derived from error bound results of convex polynomials (see e.g [18, 30] and other references therein). However, for the sake of simplicity and self-containment, we have given an elementary direct proof for Lemma 3.1.

References

[1] E. G. Belousov, and D. Klatte, *A Frank-Wolfe type theorem for convex polynomial programs,* Comp. Optim. & Appl., 22 (2002), 37-48.

[2] G. Blekherman, *Nonnegative polynomials and sums of squares,* Amer. Math. Soc. 25 (2012), 617-635.

[3] J. Demmel, J. W. Nie, and V. Powers, *Representations of positive polynomials on noncompact semi-algebraic sets via KKT ideals,* J. Pure Appl. Algebra 209(1) (2007), 189-200.

[4] H. V. Hà, and T. S. Phạm, *Global optimization of polynomials using the truncated tangency variety and sums of squares,* SIAM J. Optim., 19(2) (2008), 941-951.

[5] H. V. Hà, and T. S. Phạm, *Solving polynomial optimization problems via the truncated tangency variety and sums of squares,* J. Pure Appl. Algebra, 213 (2009), 2167-2176.

[6] H. V. Hà, and T. S. Phạm, *Representations of positive polynomials and optimization on noncompact semi-algebraic sets,* SIAM J. Optim., 20 (2010), 3082-3103.

[7] J. W. Helton, and J. W. Nie, *Semidefinite representation of convex sets,* Math. Program., 122(1) (2010), Ser. A, 21-64.

[8] J. B. Hiriart-Urruty, and C. Lemaréchal, *Convex analysis and minimization algorithms I*, Grundlehren der mathematischen Wissenschaften. Springer, 1993.

[9] V. Jeyakumar, and D. T. Luc, *Nonsmooth vector functions and continuous optimization,* Springer Optimization and Its Applications, 10. Springer, New York, 2008.

[10] V. Jeyakumar, *Constraint qualifications characterizing Lagrangian duality in convex optimization,* J. Optim. Theor. Appl. (136) (2008), 31-41.

[11] V. Jeyakumar, G. M. Lee, and N. Dinh, *New sequential Lagrange multiplier conditions characterizing optimality without constraint qualification for convex programs,* SIAM J. Optim. 14 (2003), 534-547.

[12] V. Jeyakumar, and G. Li, *Exact SDP relaxations for classes of nonlinear semidefinite programming problems,* Oper. Res. Lett., 40(6) (2012), 529–536.

[13] E. De Klerk, M. Laurent and P. Parrilo *On the equivalence of algebraic approaches to the minimization of forms on the simplex,* Positive Polynomials in Control, D. Henrion and A. Garulli, eds., Lecture Notes on Control and Information Sciences, Springer Verlag, Berlin, 312, (2005) 121-133.

[14] E. De Klerk, and M. Laurent, *On the Lasserre hierarchy of semidefinite programming relaxations of convex polynomial optimization problems,* SIAM J. Optim., 21 (2011), 824-832.

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

[16] J. B. Lasserre, *Convexity in semi-algebraic geometry and polynomial optimization,* SIAM J. Optim. 19(4) (2009), 1995–2014.

[17] J. B. Lasserre, *Moments, positive polynomials and their applications*, Imperial College Press, 2009.

[18] G. Li, *On the asymptotic well behaved functions and global error bound for convex polynomials*, SIAM J. Optim.,, 20(2010), no. 4, 1923–1943,

[19] M. Marshall, *Positive polynomials and sums of squares,* Mathematical Surveys and Monographs 146. Providence, RI: American Mathematical Society, (2008).

[20] M. Marshall, *Representations of non-negative polynomials, degree bounds and applications to optimization,* Canad. J. Math., 61(1) (2009), 205-221.

[21] J. Nie, *Optimality conditions and finite convergence of Lasserre's hierarchy,* Math. Prog., DOI: 10.1007/s10107-013-0680-x, (2013).

[22] J. Nie, *Certifying Convergence of Lasserre's Hierarchy via Flat Truncation,* Math. Prog., DOI: 10.1007/s10107-012-0589-9, (2013).

[23] J. Nie, J. Demmel, and B. Sturmfels, *Minimizing polynomials via sum of squares over the gradient ideal,* Math. Prog., Ser. A, 106(3) (2006), 587-606.

[24] M. Putinar, Positive polynomials on compact semi-algebraic sets, *Ind. Uni. Math. J.* 41 (1993), 49-95.

[25] C. Scheiderer, *Sums of squares on real algebraic curves,* Math. Z., 245 (2003), pp. 725-760.

[26] C. Scheiderer, *Distinguished representations of non-negative polynomials,* J. Algebra 289 (2005), no. 2, 558-573.

[27] K. Schmüdgen, *The K-moment problem for compact semi-algebraic sets*, Math. Ann., 289 (2) (1991), 203–206.

[28] M. Schweighofer, *Optimization of polynomials on compact semialgebraic sets,* SIAM J. Optim., 15 (3) (2005), 805-825.

[29] M. Schweighofer, *Global optimization of polynomials using gradient tentacles and sums of squares,* SIAM J. Optim., 17 (3) (2006), 920-942.

[30] W. H. Yang, *Error bounds for convex polynomials,* SIAM J. Optim., 19 (2008), 1633-1647.