Representations of positive polynomials on noncompact semialgebraic sets via KKT ideals

James Demmel\textsuperscript{a}, Jiawang Nie\textsuperscript{b}, Victoria Powers\textsuperscript{c,*}

\textsuperscript{a} Department of Mathematics and EECS, University of California, Berkeley, CA 94720, United States
\textsuperscript{b} Department of Mathematics, University of California, Berkeley, CA 94720, United States
\textsuperscript{c} Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, United States

Received 2 December 2005; received in revised form 25 April 2006
Available online 18 July 2006
Communicated by R. Parimala

Abstract

This paper studies the representation of a positive polynomial \( f(x) \) on a noncompact semialgebraic set \( S = \{ x \in \mathbb{R}^n : g_1(x) \geq 0, \ldots, g_s(x) \geq 0 \} \) modulo its KKT (Karush–Kuhn–Tucker) ideal. Under the assumption that the minimum value of \( f(x) \) on \( S \) is attained at some KKT point, we show that \( f(x) \) can be represented as sum of squares (SOS) of polynomials modulo the KKT ideal if \( f(x) > 0 \) on \( S \); furthermore, when the KKT ideal is radical, we argue that \( f(x) \) can be represented as a sum of squares (SOS) of polynomials modulo the KKT ideal if \( f(x) \geq 0 \) on \( S \). This is a generalization of results in [J. Nie, J. Demmel, B. Sturmfels, Minimizing polynomials via sum of squares over the gradient ideal, Mathematical Programming (in press)], which discusses the SOS representations of nonnegative polynomials over gradient ideals.

\( \text{MSC: 13J30; 11E25; 14P10; 90C22} \)

1. Introduction

There has been much recent interest in developing algorithms for optimizing polynomial functions on semialgebraic sets using representation theorems from real algebraic geometry for positive polynomials. The idea is to turn a problem of this type into a question about the existence of a representation involving sums of squares (SOS) polynomials and the polynomials defining the semialgebraic set — an SOS representation for short. This can then be implemented as a semidefinite program (SDP), and solved numerically \([18,22]\). In the global case, i.e., when the semialgebraic set is the whole space \( \mathbb{R}^n \), an SOS representation gives a convex relaxation of the original problem and hence a lower bound for the minimum. In the case of compact semialgebraic sets, using results on SOS representations, Lasserre \([12]\) gave a procedure for finding natural sequences of computationally feasible SDP relaxations of the original problem, whose solutions converge to a solution of the original problem.

However, these methods do not always work well. In the global case, the resulting SDP might not have a solution even if the polynomial attains a minimum. This can also occur in the case of a semialgebraic set which is not compact.

* Corresponding author.
E-mail addresses: demmel@cs.berkeley.edu (J. Demmel), njw@math.berkeley.edu (J. Nie), vicki@mathcs.emory.edu (V. Powers).
In the compact case, the procedure proposed by Lasserre in [12] can generate a sequence of lower bounds which converge to the minimum under a certain constraint qualification condition. Recently, Nie and Schweighofer [16] gave results on the convergence rate of these lower bounds. However, Lasserre’s procedure is based on SOS representations of positive polynomials on compact semialgebraic sets and the lower bounds generated usually have only asymptotic convergence, i.e., the finite convergence is usually not guaranteed, as shown in an example due to Stengle [27].

As is well known, most numerical optimization methods targeting local (including global) minimizers are often based on the optimality conditions: the Karush–Kuhn–Tucker (KKT) system. In the unconstrained global case, the KKT system reduces to the zero gradient condition. Thus an approach with great potential in global optimization is based on the optimality conditions.

Theorem 1.1 (Schm"udgen [24]). If \( S(G) \) is compact, then \( f > 0 \) on \( S(G) \) implies \( f \in P(G) \).

In general, even with the assumption that \( S(G) \) is compact, this does not hold if we replace \( P(G) \) by \( M(G) \), or if we assume only that \( f \geq 0 \) on \( S(G) \). See [20] for details.

A quadratic module \( M \) is archimedean if there exists \( p(x) \in M \) such that the set \( \{ x \in \mathbb{R}^n : p(x) \geq 0 \} \) is compact, equivalently, if there exists \( N \in \mathbb{N} \) such that \( N - \sum_{i=1}^{m} x_i^2 \in M \), see [3, 5.3.8]. Note that if \( M(S) \) or \( P(S) \) is archimedean, then \( S \) is compact.

Theorem 1.2 (Putinar [23]). Suppose \( M(G) \) is archimedean, then for any \( f \in \mathbb{R}[X] \), \( f > 0 \) on \( S(G) \) implies \( f \in M(G) \).
Remarks 1.1. (i) There are examples of compact S(G) for which the corresponding quadratic module M(G) is not archimedean and the conclusion of Putinar’s Theorem does not hold, see Example 6.3.1 in [3]. In the case of the preorder P(G), it is a deep theorem of Schmüdgen [24] that if S(G) is compact then P(G) is archimedean.

(ii) The Putinar and Schmüdgen theorems say that if the conditions are satisfied, then there always exists an SOS representation of f positive on S(G). Thus, in this case, there is trivially an SOS representation modulo the gradient ideal. On the other hand, all of the assumptions of the theorem are necessary.

Given \( f \in \mathbb{R}[X] \), let \( f^* \) denote the minimum of \( f \) on \( S \), i.e., the solution to the optimization problem

\[
 f^* := \min_{x \in \mathbb{R}^n} f(x)
\]

s.t. \( g_i(x) \geq 0, \quad i = 1, \ldots, s \). (1.1)

The KKT system associated to this optimization problem is

\[
 \nabla f - \sum_{j=1}^{s} \lambda_j \nabla g_j = 0 \]

\[
 g_j \geq 0, \quad \lambda_j g_j = 0, \quad j = 1, \ldots, s
\]

where the variables \( \lambda := [\lambda_1 \cdots \lambda_s]^T \) are called Lagrange multipliers and \( \nabla f \) denotes the gradient of \( f \), i.e., the vector of partial derivatives. Under certain regularity conditions, for example, if the gradients of the \( g_j \)'s are linearly independent (see [17]), the local (including global) minimizers of \( f(x) \) on \( S \) satisfy the KKT system above. A point is said to be a KKT point if the KKT system holds at that point. We note that we do not include the condition that the Lagrange multipliers \( \lambda_j \) are nonnegative, as is usual. It turns out that we do not need the nonnegativeness of \( \lambda_j \) to obtain our representation theorems, as we shall see. Since taking the sign of \( \lambda_j \) into account adds an unnecessary complication to the representation, we omit it.

It is possible that the KKT system sometimes fails at some minimizers; thus the assumption that the KKT system holds may be very restrictive in some situations. However, in most practical applications, the minimizers often satisfy the KKT system. Hence, for this reason most optimization theory and methods are based on KKT systems. Thus using the KKT system is a natural way to proceed from the point of view of practical techniques for optimization, although it might be restrictive sometimes. Most numerical algorithms targeting local (including global) minimizers generate a sequence of points \( \{x^{(k)}, \lambda^{(k)}\} \) whose limit or accumulation points satisfy the KKT system (1.3) and (1.4). We refer to [17] and the references therein for general numerical methods in optimization.

We work in the polynomial rings \( \mathbb{C}[X, \lambda] := \mathbb{C}[x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_s] \) and \( \mathbb{R}[X, \lambda] \). Let \( F_i = \frac{\partial f}{\partial x_i} - \sum_{j=1}^{s} \lambda_j \frac{\partial g_j}{\partial x_i} \) and define the KKT ideal \( I_{KKT} \) and the varieties associated with KKT system (1.3) and (1.4) as follows:

\[
 I_{KKT} = (F_1, \ldots, F_n, \lambda_1 g_1, \ldots, \lambda_s g_s),
\]

\[
 V_{KKT} = \{(x, \lambda) \in \mathbb{C}^n \times \mathbb{C}^s : p(x, \lambda) = 0, \forall p \in I_{KKT}\}.
\]

\[
 V_{KKT}^{\mathbb{R}} = \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^s : p(x, \lambda) = 0, \forall p \in I_{KKT}\}.
\]

Keeping in mind that we are now working in the larger polynomial ring, we use \( P(G) \), resp. \( M(G) \), to denote the preorder, resp. quadratic module, in \( \mathbb{R}[X, \lambda] \) generated by \( G \). The associated KKT preorder \( P_{KKT} \) and KKT quadratic module \( M_{KKT} \) in \( \mathbb{R}[X, \lambda] \) are defined as

\[
 P_{KKT} = P(G) + I_{KKT}
\]

\[
 M_{KKT} = M(G) + I_{KKT}.
\]

Finally, let \( \mathcal{H} \) be the set satisfying constraints (1.2):

\[
 \mathcal{H} = \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^s : g_j(x) \geq 0, \quad j = 1, \ldots, s\}.
\]

The main results of this paper are the following: assume \( f^* \) is attained at some KKT point. If \( I_{KKT} \) is radical and \( f \geq 0 \) on \( V_{KKT}^{\mathbb{R}} \cap \mathcal{H} \), then \( f \in P_{KKT} \); if \( I_{KKT} \) is not radical but \( f > 0 \) on \( V_{KKT}^{\mathbb{R}} \cap \mathcal{H} \), then \( f \in P_{KKT} \).

This paper is organized as follows. Section 2 contains some background in algebraic geometry and real algebra. In Section 3 we study the SOS representations of polynomials modulo KKT ideals. Section 4 shows the applications
of this kind of SOS representations in optimization on noncompact semialgebraic sets. We draw some conclusions in Section 5.

2. Preliminaries

In this section we present some notions and results from algebraic geometry and real algebra needed for our discussion. Readers may consult [1,2,26] for more details.

Throughout this section, denote by \( \mathbb{R}[Z] \) the ring of polynomials in \( Z = (z_1, \ldots, z_m) \) with real coefficients. Given an ideal \( I \subseteq \mathbb{R}[Z] \), define its variety to be the set

\[
V(I) = \{ z \in \mathbb{C}^m : p(z) = 0 \text{ for all } p \in I \},
\]

and its real variety to be

\[
V^R(I) = \{ z \in \mathbb{R}^m : p(z) = 0 \text{ for all } p \in I \}.
\]

An ideal \( I \subseteq \mathbb{R}[X] \) is said to be zero-dimensional if its variety \( V(I) \) is a finite set. This condition is much stronger than requiring that the real variety \( V^R(I) \) be a finite set. For example, \( I = (Z_1^2 + Z_2^2) \) is not zero-dimensional, however the real variety \( V^R(I) = \{(0, 0)\} \) consists of one point of the curve \( V(I) \).

A nonempty variety \( V = V(I) \subseteq \mathbb{C}^m \) is irreducible if there do not exist two proper subvarieties \( V_1, V_2 \subseteq V \) such that \( V = V_1 \cup V_2 \). The reader should note that in this paper “irreducible” means that the set of complex zeros cannot be written as a proper union of subvarieties defined by real polynomials.

Given any ideal \( I \) of \( \mathbb{R}[Z] \), its radical ideal \( \sqrt{I} \) is defined to be the following ideal

\[
\sqrt{I} = \left\{ q \in \mathbb{R}[Z] : q^\ell \in I \text{ for some } \ell \in \mathbb{N} \right\}.
\]

Clearly, \( I \subseteq \sqrt{I} \); \( I \) is a radical ideal if \( \sqrt{I} = I \). As usual, for a variety \( V \subseteq \mathbb{C}^m \), \( I(V) \) denotes the ideal in \( \mathbb{C}[Z] \) of polynomials vanishing on \( V \). We will write \( I^R(V) \) for the ideal \( \mathbb{R}[Z] \cap I(V) \).

We need versions of the Nullstellensatz for varieties defined by polynomials in \( \mathbb{R}[Z] \). The following are normally stated for ideals in \( \mathbb{C}[Z] \); however, keeping in mind that \( V(I) \) lies in \( \mathbb{C}^m \), they hold as stated.

**Theorem 2.1 ([2]).** If \( I \) is an ideal in \( \mathbb{R}[Z] \) such that \( V(I) = \emptyset \) then \( 1 \in I \).

**Theorem 2.2 ([2]).** If \( I \) is an ideal in \( \mathbb{R}[Z] \) then \( I^R(V(I)) = \sqrt{I} \).

Finally, we need the following real algebra version of Theorem 2.1 see e.g. [3, 4, 2, 13].

**Theorem 2.3.** Suppose \( S(G) \) and \( P(G) \) are defined as above, then \( S(G) = \emptyset \) if and only if \( -1 \in P(G) \).

We will also need the following lemma which is the “variety version” of Lagrangian interpolation:

**Lemma 2.4 (Lemma 1 [15]).** Let \( V_1, \ldots, V_r \) be pairwise disjoint varieties of \( \mathbb{C}^m \). Then there exist polynomials \( p_1, \ldots, p_r \in \mathbb{C}[X] \) such that \( p_i(V_j) = \delta_{ij} \), where \( \delta_{ij} \) is the Kronecker delta function.

**Remark 2.5.** If each \( V_\ell \) is conjugate symmetric, i.e. a point \( z \in \mathbb{C}^m \) belongs to \( V_\ell \) if and only if its complex conjugate \( \bar{z} \in V_\ell \), then the polynomials \( p_\ell \) can be chosen such that \( p_\ell \in \mathbb{R}[Z] \), since we can replace \( p_i(Z) \) by \( (p_i(Z) + \bar{p}_i(Z))/2 \), where \( \bar{p}_i(Z) \) is obtained from \( p_i(Z) \) by conjugating its coefficients.

3. Sums of squares modulo KKT ideals

In this section, we discuss the SOS representation of nonnegative and positive polynomials on a noncompact basic closed semialgebraic set \( S \) modulo the corresponding KKT ideals.

When \( S = \mathbb{R}^n \), the problem is reduced to the SOS representation of nonnegative or positive polynomials modulo gradient ideals, as discussed in [15]. Nie et al. [15] showed that if a polynomial \( f \in \mathbb{R}[X] \) is nonnegative on its real gradient variety and its gradient ideal is radical, then \( f \) has a representation as a sum of squares modulo the gradient ideal; if the gradient ideal of \( f(x) \) is not radical but \( f(x) \) is positive on its real gradient variety, then \( f(x) \) also has
a representation as a sum of squares modulo its gradient ideal. When \( f(x) \) is just nonnegative on its real gradient variety and its gradient ideal is not radical, the polynomial \( f(x) \) might not have such an SOS representation modulo its gradient ideal, as shown in Example 1 in [15].

In this section we generalize this result to real polynomials which are nonnegative on a basic closed semialgebraic set. The real gradient variety and real gradient ideal are replaced by a variety and an ideal defined by the KKT system corresponding to the optimization (1.1) and (1.2).

Fix \( G = \{ g_1, \ldots, g_s \} \subseteq \mathbb{R}[X] \) and let \( S = S(G) \). Given \( f \in \mathbb{R}[X] \), define the ideal \( I_{\text{KKT}} \), varieties \( V_{\text{KKT}} \), \( V_{\text{KKT}}^{\mathbb{R}} \), preorder \( P_{\text{KKT}} \) and quadratic module \( M_{\text{KKT}} \) associated to the KKT system (1.3) and (1.4) defined in Section 1.

As is well-known, if an ideal \( f \) in a polynomial ring is zero-dimensional, then every PSD polynomial \( f \) on \( V(I) \) is SOS modulo \( \sqrt{I} \). This follows easily from the Chinese Remainder Theorem, for a proof see, e.g., [22]. From this fact, we immediately obtain the following representation theorem:

**Theorem 3.1.** Assume \( I_{\text{KKT}} \) is zero-dimensional and radical. If \( f(x) \) is nonnegative on \( V_{\text{KKT}}^{\mathbb{R}} \cap H \), then \( f(x) \) belongs to \( M_{\text{KKT}} \).

Using a proof similar to that of Theorem 8 in [15], we can remove the restrictive hypothesis that \( I_{\text{KKT}} \) is zero-dimensional; however to obtain the most general result we must replace the quadratic module \( M_{\text{KKT}} \) by the preorder \( P_{\text{KKT}} \).

**Theorem 3.2.** Assume \( I_{\text{KKT}} \) is radical. If \( f(x) \) is nonnegative on \( V_{\text{KKT}}^{\mathbb{R}} \cap H \), then \( f(x) \) belongs to \( P_{\text{KKT}} \).

To prove the above theorem, we need the following lemma.

**Lemma 3.3.** Let \( W \) be an irreducible component of \( V_{\text{KKT}} \) and assume \( W^{\mathbb{R}} \neq \emptyset \). Then \( f(x) \) is constant on \( W \).

**Proof.** Since \( W \) is irreducible and contains a real point, it remains irreducible if we replace \( \mathbb{R}[X, \lambda] \) by \( \mathbb{C}[X, \lambda] \). Thus \( W \) is connected in the strong topology on \( \mathbb{C}^{n+2} \) and hence is path-connected (see e.g. [28, 4.1.3]).

The Lagrangian function

\[
\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^{s} \lambda_i g_i(x)
\]

is equal to \( f(x) \) on \( V_{\text{KKT}} \), which contains \( W \). Choose two arbitrary points \( (x^{(1)}, \lambda^{(1)}), (x^{(2)}, \lambda^{(2)}) \) in \( W \). We claim that \( f(x^{(1)}) = f(x^{(2)}) \).

First, assume both \( (x^{(1)}, \lambda^{(1)}) \) and \( (x^{(2)}, \lambda^{(2)}) \) are nonsingular points. Since the set of nonsingular points is a manifold and \( W \) is path-connected, there exists a piecewise-smooth path \( \varphi(\tau) = (x(\tau), \lambda(\tau)) \) \( (0 \leq \tau \leq 1) \) lying inside \( W \) such that \( \varphi(0) = (x^{(1)}, \lambda^{(1)}) \) and \( \varphi(1) = (x^{(2)}, \lambda^{(2)}) \). Let \( \mu_j(\tau) \) be the principle square root of \( \lambda_j(\tau) \), \( 1 \leq j \leq s \) (for a complex number \( z = |z| \exp(\sqrt{-1}\theta) \) with \( 0 \leq \theta < 2\pi \), its principle square root is defined to be \( \sqrt{|z|} \exp(\frac{1}{2}\sqrt{-1}\theta) \)). From the KKT system (1.3) and (1.4), we can see that the function

\[
f(x) + \sum_{i=1}^{s} \mu_i^2 g_i(x)
\]

has zero gradient on the path \( \varphi(\tau) \) \( (0 \leq \tau \leq 1) \). By the Mean Value Theorem, it follows that \( f(x^{(1)}) = f(x^{(2)}) \).

Now suppose that at least one of \( (x^{(1)}, \lambda^{(1)}) \) and \( (x^{(2)}, \lambda^{(2)}) \) is singular. Since the set of nonsingular points of \( W \) is dense and open in \( W \) ([28, Chap. 4]), we can choose arbitrarily close nonsingular points to approximate \( (x^{(1)}, \lambda^{(1)}) \) and \( (x^{(2)}, \lambda^{(2)}) \). By continuity of \( f(x) \), we immediately have \( f(x^{(1)}) = f(x^{(2)}) \) and hence that \( f \) is constant on \( W \).

**Proof of Theorem 3.2.** Decompose \( V_{\text{KKT}} \) into its irreducible components and let \( W_0 \) be the union of all the components whose intersection with \( H \) is empty. We note that this includes all components \( W \) with \( W^{\mathbb{R}} = \emptyset \). Thus, by Lemma 3.3, \( f \) is constant on each of the remaining components. We group together all components for which \( f \) takes the same value, then we have disjoint components \( W_1, \ldots, W_r \) such that \( f \) is constant on each \( W_i \). Further, since each contains a real point and \( f \) is nonnegative on \( V_{\text{KKT}} \), the value of \( f \) on each \( W_i \) is real and non-negative.
Suppose $f = \alpha_i \geq 0$ on $W_i$ for $1 \leq i \leq r$. We have $V_{KKT} = W_0 \cup W_1 \cup \cdots \cup W_r$, and the $W_i$ are pairwise disjoint. Note that by our definition of irreducibility, each $W_i$ is conjugate symmetric. By Lemma 2.4, there exist polynomials $p_0, p_1, \ldots, p_r \in \mathbb{R}[x, \lambda]$ such that $p_i(W_j) = \delta_{ij}$, where $\delta_{ij}$ is the Kronecker delta function.

By assumption, $W_0 \cap \mathcal{H} = \emptyset$ and so, by Theorem 2.3, there is some $v_0 \in P := P(g_1, \ldots, g_s)$ such that $-1 \equiv v_0 \mod I(W_0)$. We have $f = s_1 - s_2$ for the SOS polynomials $s_1 = (f + \frac{1}{2})^2$ and $s_2 = (f^2 + (\frac{1}{2})^2)$. Hence

$$f \equiv s_1 + v_0 \cdot s_2 \mod I(W_0).$$

Let $q_0 = s_1 + v_0 \cdot s_2 \in P$. Recall that $f(x) = \alpha_i$, a nonnegative real constant, on each $W_i (1 \leq i \leq r)$. Set $q_i(x) = \sqrt{\alpha_i}$, then $f(x) = q_i(x)^2$ on $I(W_i)$.

Now let $q = q_0(p_0)^2 + (\sum_{i=1}^r q_i p_i)^2$. Then $f - q$ vanishes on $V_{KKT}$ and hence $f - q \in I_{KKT}$ since $I_{KKT}$ is radical. It follows that $f \in P_{KKT}$.

**Remark 3.4.** The assumption that $I_{KKT}$ is radical is needed in Theorem 3.2, as shown by Example 3.4 in [15]. However, when $I_{KKT}$ is not radical, the conclusion also holds if $f(x)$ is strictly positive on $V_{KKT}^\mathbb{R}$.

**Theorem 3.5.** If $f > 0$ on $V_{KKT}^\mathbb{R} \cap \mathcal{H}$, then $f$ belongs to $P_{KKT}$.

**Proof.** As in the proof of Theorem 3.2, we decompose $V_{KKT}$ into subvarieties $W_0, W_1, \ldots, W_r$ such that $W_0 \cap \mathcal{H} = \emptyset$, the $W_i$’s are pairwise disjoint, and for $i = 1, \ldots, r$, $f(x) = \alpha_i$, a real constant, on $W_i$. Since $f > 0$ on $V_{KKT}^\mathbb{R}$, each $\alpha_i > 0$.

Consider the primary decomposition $I_{KKT} = \bigcap_{i=0}^r J_i$ corresponding to our decomposition of $V_{KKT}$, i.e., $V(J_i) = W_i$ for $i = 0, 1, \ldots, r$. Since $W_i \cap W_j = \emptyset$, we have $J_1 + J_2 = \mathbb{R}[x, \lambda]$ by Theorem 2.1. The Chinese Remainder Theorem, see e.g. [4, 2.13], implies that there is an isomorphism

$$\rho : \mathbb{R}[x, \lambda]/I_{KKT} \to \mathbb{R}[x, \lambda]/J_0 \times \mathbb{R}[x, \lambda]/J_1 \times \cdots \times \mathbb{R}[x, \lambda]/J_r.$$

For any $p \in \mathbb{R}[x, \lambda]$, let $[p]$ and $\rho([p])$, denote the equivalence classes of $p$ in $\mathbb{R}[x, \lambda]/I_{KKT}$ and $\mathbb{R}[x, \lambda]/J_i$ respectively.

Recall that $V(J_0) \cap \mathcal{H} = \emptyset$, hence by Theorem 2.3 there exist SOS polynomials $u_{\theta} (\theta \in \{0, 1\}^s)$ such that

$$-1 \equiv \sum_{\theta \in \{0, 1\}^s} u_{\theta} \rho([g^\theta_1]) \cdots \rho([g^\theta_s]) \equiv u_0 \mod J_0.$$

As in the proof of Theorem 3.2, we write $f = f_1 - f_2$ for SOS polynomials $f_1, f_2$ and then we have

$$f \equiv f_1 + u_0 f_2 \equiv \sum_{\theta \in \{0, 1\}^s} v_{\theta} \rho([g^\theta_1]) \cdots \rho([g^\theta_s]) \equiv q_0 \mod J_0$$

for some SOS polynomials $v_{\theta} (\theta \in \{0, 1\}^s)$. Thus the preimage $\rho^{-1}((q_0, 0, \ldots, 0)) \in P_{KKT}$.

Now on each $W_i$, $1 \leq i \leq r$, $f = \alpha_i > 0$, and hence $(f/\alpha_i) - 1$ vanishes on $W_i$. Then by Theorem 2.2 there exists some integer $\ell \in \mathbb{N}$ such that $(f/\alpha_i - 1)^\ell \in J_i$. From the binomial theorem, it follows that

$$\left(1 + \frac{f}{\alpha_i} - 1\right)^{\ell/2} = \sum_{k=1}^{\ell-1} \binom{\ell/2}{k} (f/\alpha_i - 1)^k \equiv q_i/\sqrt{\alpha_i} \mod J_i.$$

Thus $\rho((f))_i = q_i^2$ is SOS modulo $J_i$, and hence $\rho^{-1}(q_i^2 e_{i+1})$ is SOS modulo $I_{KKT}$, where $e_{i+1}$ is the $(i + 1)$-st standard unit vector in $\mathbb{R}^{r+1}$.

Finally, we see that $\rho((f)) = (q_0, q_1^2, \ldots, q_s^2)$. The preimage of the latter is

$$\rho^{-1}((q_0, q_1^2, \ldots, q_s^2)) = \rho^{-1}(q_0 e_1) + \sum_{i=1}^r \rho^{-1}(q_i^2 e_{i+1}),$$

which implies that $f \in P_{KKT}$.

**Remark.** The conclusions in Theorems 3.2 and 3.5 cannot be strengthened to show that $f(x) \in M_{KKT}$, as seen in the following example:
Example 3.6. Let \( g_1 = 1 - x_1, g_2 = x_2, \) and \( g_3 = x_3 - x_2 - 1 \) and set \( G = \{ g_1, g_2, g_3 \}. \) Let \( f = (x_3 - x_1^2 x_2)^2 - 1 + \epsilon, \) where \( 0 < \epsilon < 1. \) It is easy to see that the minimum of \( f^* \) on \( S := S(G) \) is \( f^* = \epsilon. \) In particular, \( f > 0 \) on \( S. \) The corresponding KKT ideal

\[
I_{KKT} = \left\{ 2x_1 x_2 (x_3 - x_1^2 x_2) - \lambda_1 x_1, 2 x_1^2 (x_3 - x_1^2 x_2) + \lambda_2 - \lambda_3, 2 (x_3 - x_1^2 x_2) - \lambda_3, \lambda_1 (1 - x_1^2), \lambda_2 x_2, \lambda_3 (x_3 - x_2 - 1) \right\}
\]

is radical (verified in Macaulay 2 [5]). However, \( f \not\in M_{KKT}. \) Suppose to the contrary that \( f \in M_{KKT}, \) then there exist SOS polynomials \( \sigma_0, \sigma_1, \sigma_2, \sigma_3 \) and general polynomials \( \phi_1, \phi_2, \phi_3 \) such that

\[
f(x) = \sigma_0 + \sigma_1 g_1 + \sigma_2 g_2 + \sigma_3 g_3 + \phi_1 \left( \frac{\partial f}{\partial x_1} - \lambda_1 x_2 \right) + \phi_2 \left( \frac{\partial f}{\partial x_2} - \lambda_2 + \lambda_3 \right) + \phi_3 \left( \frac{\partial f}{\partial x_3} - \lambda_3 \right).
\]

Plugging \( \lambda = (0, 0, 0) \) into the above identity yields

\[
0 = 1 - \epsilon + \sigma_0 + \sigma_1 (1 - x_1^2) + \sigma_2 x_2 + \sigma_3 (x_3 - x_2 - 1) + \phi(x_3 - x_1^2 x_2)
\]

where \( \phi = -4 x_1 \phi_1 - x_1^2 \phi_2 + 2 \phi_3 - (x_3 - x_1^2 x_2). \) Now substitute \( x_3 = x_1^2 x_2 \) in the above, yielding

\[
\sigma_3 ((1 - x_1^2) x_2 + 1) = 1 - \epsilon + \sigma_0 + \sigma_1 (1 - x_1^2) + \sigma_2 x_2.
\]

Here \( \sigma_0, \sigma_1, \sigma_2, \sigma_3 \) are now considered SOS polynomials in \( (x_1, x_2). \) Since \( 1 - \epsilon > 0, \) \( \sigma_3 \) cannot be the zero polynomial. If \( \sigma_3 = \sigma_2 (x_1) \) is independent of \( x_2, \) we can derive a contradiction using an argument identical to the argument in the proof of [20, Thm. 2]. Thus \( 2m = \deg_x \sigma_3 (x_1, x_2) \geq 2 \) and \( 2d = \deg_x \sigma_3 (x_1, x_2) \geq 0. \) On the left hand side, the leading term is of the form \( A \cdot 2d^2 + 2 \cdot x_2^{2m+1} \) with coefficient \( A < 0. \) Since the degree in \( x_2 \) on the left hand side is odd, the leading term on the right hand side must come from \( \sigma_2 (x_1, x_2) x_2, \) and is of the form \( B \cdot 2d^2 + 2 \cdot x_2^{2m+1} \) with \( B > 0. \) This is a contradiction. Therefore we can conclude that \( f \not\in M_{KKT}. \)

4. Applications in optimization

Given \( f, g_1, \ldots, g_s \in \mathbb{R}[X], \) recall the optimization problem from the introduction

\[
f^* := \min_{x \in \mathbb{R}^n} f(x), \quad \text{s.t.} \quad g_i(x) \geq 0, \quad i = 1, \ldots, s
\]

and suppose we are interested in computing numerically the optimal value \( f^*. \) In other words, we wish to compute the minimum of \( f \) on the basic closed semialgebraic set \( S(G), \) where \( G = \{ g_1, \ldots, g_s \}. \)

Finding the global optimal solutions to (4.1) and (4.2) is an NP-hard problem, even if \( f \) is quadratic and the \( g_i \) are linear. For instance, the Maximum-Cut problem for graphs is of this form, and it is NP-hard [6]. Recently, techniques using SOS relaxations and moment matrix methods have made it possible to approximate the global optimal solutions to (4.1) and (4.2) by approximating nonnegative polynomials with SOS polynomials, which allows the problem to be implemented as an SDP which can then be solved numerically. For details about these methods and their applications, see [11–15,18,19,25].

In the case where \( S \) is compact, the SOS methods are based on representation theorems for positive polynomials on compact semialgebraic sets, i.e., the theorems of Schm"udgen and Putinar. However, these theorems do not hold in the case where \( S \) is not compact. As discussed in the introduction, a more traditional approach in numerical optimization methods is to use the first order optimality conditions (the Karush–Kuhn–Tucker (KKT) system in the constrained case). Using Theorems 3.2 and 3.5, we combine these two methods to give a procedure for approximating \( f^* \) in the case where the semialgebraic set is not necessarily compact.

Recall the KKT system corresponding to (4.1) and (4.2):

\[
\nabla f(x) - \sum_{j=1}^s \lambda_j \nabla g_j(x) = 0
\]

\[
g_j(x) \geq 0, \quad \lambda_j g_j(x) = 0, \quad j = 1, \ldots, s.
\]
Let $f^*_{KKT}$ be the global minimum of $f(x)$ over the KKT system defined by (4.3) and (4.4). Assume the KKT system holds at least one global optimizer. Then we claim that $f^* = f^*_{KKT}$. First, $f^* \leq f^*_{KKT}$ follows immediately from the fact that all solutions to the KKT system are feasible. Now let $x^*$ be a global minimizer such that $f(x^*) = f^*$, then by assumption, there exist Lagrange multipliers $\lambda^*$ such that $(x^*, \lambda^*)$ satisfies the above KKT system. Thus $f^* \geq f^*_{KKT}$ and hence they are equal.

In order to implement membership in $P_{KKT}$ as an SDP, we need a bound on the degrees of the sums of squares involved. Thus, for $N \in \mathbb{N}$, we define the truncated KKT ideal

$$I_{N,KKT} = \left\{ \sum_{k=1}^n \phi_k F_k + \sum_{j=1}^s \psi_j \lambda_j g_j \mid \deg(\phi_k F_k), \deg(\psi_j \lambda_j g_j) \leq 2N \right\}$$

and the truncated preorder

$$P_{N,KKT} = \left\{ \sum_{\theta \in [0,1]^s} \sigma_0 g_1^{\theta_1} g_2^{\theta_2} \cdots g_s^{\theta_s} \mid \deg(\sigma_0 g_1^{\theta_1} \cdots g_s^{\theta_s}) \leq 2N \right\} + I_{N,KKT}.$$

Then we define a sequence $\{f^*_N\}$ of SOS relaxations of the optimization problem (4.1) and (4.2) as follows:

$$f^*_N = \max_{\gamma \in \mathbb{R}} \gamma$$

subject to $f(x) - \gamma \in P_{N,KKT}.$

Obviously each $\gamma$ feasible in (4.6) is a lower bound of $f^*$. So $f^*_N \leq f^*$. When we increase $N$, the feasible region defined by (4.6) is increasing, and hence the sequence of lower bounds $\{f^*_N\}$ is also monotonically increasing. Thus we have

$$f_1^* \leq f_2^* \leq f_3^* \leq \cdots \leq f^*.$$

It can be shown that the sequence of lower bounds $\{f^*_N\}$ obtained from (4.5) and (4.6) converges to $f^*$ in (1.1) and (1.2), provided that $f^*$ is attained at one KKT point. We summarize in the following theorem:

**Theorem 4.1.** Assume $f(x)$ has a minimum $f^* := f(x^*)$ at one KKT point $x^*$ of (1.1) and (1.2). Then $\lim_{N \to \infty} f^*_N = f^*$. Furthermore, if $I_{KKT}$ is radical, then there exists some $N \in \mathbb{N}$ such that $f^*_N = f^*$, i.e., the SOS relaxations (4.5) and (4.6) converge in finitely many steps.

**Proof.** The sequence $\{f^*_N\}$ is monotonically increasing, and $f^*_N \leq f^*$ for all $N \in \mathbb{N}$, since $f^*$ is attained by $f(x)$ in the KKT system (1.3) and (1.4) by assumption and the constraint (4.6) implies that $\gamma \leq f^*$. Now for arbitrary $\epsilon > 0$, let $\gamma_\epsilon = f^* - \epsilon$ and replace $f(x)$ by $f(x) - \gamma_\epsilon$ in (1.1) and (1.2). The KKT system remains unchanged, and $f(x) - \gamma_\epsilon$ is strictly positive on $V_{KKT}^R$. By Theorem 3.5, $f(x) - \gamma_\epsilon \in P_{KKT}$. Since $f(x) - \gamma_\epsilon$ is fixed, there must exist some integer $N_1$ such that $f(x) - \gamma_\epsilon \in P_{N_1,KKT}$. Hence $f^* - \epsilon \leq f^*_N \leq f^*$. Therefore we have that $\lim_{N \to \infty} f^*_N = f^*$.

Now assume that $I_{KKT}$ is radical. Replace $f(x)$ by $f(x) - f^*$ in (1.1) and (1.2). The KKT system still remains the same, and $f(x) - f^*$ is now nonnegative on $V_{KKT}^R$. By Theorem 3.2, $f(x) - f^* \in P_{KKT}$. So there exists some integer $N_2$ such that $f(x) - f^* \in P_{N_2,KKT}$, and hence $f^*_N \geq f^*$. Then $f^*_N \leq f^*$ for all $N$ implies that $f^*_N = f^*$.

**Remarks.** The assumption in Theorem 4.1 that $f$ has a minimum at a KKT point is nontrivial and cannot be removed, as the following example shows.

**Example 4.2.** Consider the optimization: $\min x$ s.t. $x^3 \geq 0$. Obviously $f^* = 0$ and the global minimizer $x^* = 0$. However, the KKT system

$$1 - \lambda \cdot 3x^2 = 0, \quad \lambda \cdot x^3 = 0, \quad x^3 \geq 0$$

is not satisfied, since $V_{KKT} = \emptyset$. Actually we can see that the lower bounds $\{f^*_N\}$ given by (4.5) and (4.6) tend to infinity. By Theorem 2.3, $V_{KKT} = \emptyset$ implies that $1 \in P_{KKT}$, i.e.,

$$(1 + 3\lambda x^2)(1 - 3\lambda x^2) + 9\lambda^2 x \cdot \lambda x^3 = 1.$$
In the SOS relaxation (4.5) and (4.6), for arbitrarily large \( \gamma, x - \gamma \in P_{\text{KKT}} \), since
\[
x - \gamma = (x - \gamma)(1 + 3\lambda x^2)(1 - 3\lambda x^2) + 9\nu^2 x(x - \gamma) \cdot \lambda x^3 \in P_{\text{KKT}}.
\]
Thus \( f^*_g = \infty \).

The SOS relaxation (4.5) and (4.6) is essentially a semidefinite program [18, 19, 29] and can be solved numerically. The dual problem of (4.5) and (4.6) is to minimize a linear functional over some linear moment matrix inequalities. It can also be obtained by applying moment matrix methods [12] to minimize \( f \) over the semialgebraic set defined by KKT system (1.3) and (1.4). Using software like Gloptipoly [8] and SOSTOOLS [21], the SOS program (4.5) and (4.6) or its dual problem can be solved, and in many cases, the global minimizer \( x^* \) and the Lagrange multiplier \( \lambda^* \) can be extracted. For more details about extracting minimizers from SOS relaxations or moment matrix methods, we refer to [9].

**Example 4.3** (Exercise 2.18, [10]). Consider the global optimization problem:
\[
\begin{align*}
\min \quad & (-4x_1^2 + x_2^2)(3x_1 + 4x_2 - 12) \\
\text{s.t.} \quad & 3x_1 - 4x_2 \leq 12, \quad 2x_1 - x_2 \leq 0, \quad -2x_1 - x_2 \geq 0.
\end{align*}
\]
The semialgebraic set \( S \) defined by the constraints is non-compact. The global minimum \( f^* = -\frac{1024}{55} \approx -18.6182 \) and the minimizer \( x^* = (-24/55, 128/55) \approx (-0.4364, 2.3273) \). The lower bound obtained from (4.5) and (4.6) is \( f^*_g \approx -18.6182 \). The extracted minimizer \( \hat{x} = (-0.4364, 2.3273) \).

**Example 4.4.** Consider the Quadratically Constrained Quadratic Program (QCQP):
\[
\begin{align*}
\min \quad & -\frac{4}{3}x_1^2 + \frac{2}{3}x_2^2 - 2x_1x_2 \\
\text{s.t.} \quad & x_2^2 - x_1^2 \geq 0, \quad -x_1x_2 \geq 0.
\end{align*}
\]
The global minimum \( f^* = 0 \) and minimizer \( x^* = (0, 0) \). The semialgebraic set \( S \) defined by the constraints is non-compact. The lower bound returned by (4.5) and (4.6) is \( f^*_g = -2.6 \times 10^{-15} \) (Note: this computation was done in double precision floating point, with round off error bounded by \( 2^{-53} \sim 10^{-16} \)). The extracted minimizer is \( \hat{x} = (6.1 \times 10^{-16}, -9.0 \times 10^{-17}) \).

We conclude with another application of our theorem, to a nonconvex QCQP problem which was posed by Zhi-Quan Luo and communicated to us by Paul Tseng.

**Example 4.5.** Consider the following nonconvex quadratic optimization
\[
\begin{align*}
\min_{x \in \mathbb{R}^2} \quad & f(x) := x_1^2 + x_2^2 \\
\text{s.t.} \quad & g_1(x) := x_2^2 - 1 \geq 0 \\
& g_2(x) := x_1^2 - Mx_1x_2 - 1 \geq 0 \\
& g_3(x) := x_1^2 + Mx_1x_2 - 1 \geq 0
\end{align*}
\]
over a noncompact semialgebraic set, where \( M \) is a positive constant. Simple calculation shows that the global minimum is
\[
f^* = \frac{1}{2} \left( M^2 + M\sqrt{M^2 + 4} \right) + 2
\]
and the global minimizers are
\[
\left( \pm \frac{1}{2}(M + \sqrt{M^2 + 4}), 1 \right), \quad \left( \pm \frac{1}{2}(M + \sqrt{M^2 + 4}), -1 \right).
\]
Let \( P := P(g_1, g_2, g_3) \), the preorder in \( \mathbb{R}[x_1, x_2] \) generated by \( \{g_1, g_2, g_3\} \). Suppose we apply the standard SOS method, i.e., we find the maximum \( \gamma \) so that \( f - \gamma \in P \). Note that \( f - 2 = g_2 + g_3 \in P \); we claim that the maximum \( \gamma \) is 2.
Suppose we have a representation
\[
 f(x) - \gamma = \sigma_0 + g_1 \sigma_1 + g_2 \sigma_2 + g_3 \sigma_3 + g_1 g_2 \sigma_{12} + g_1 g_3 \sigma_{13} + g_2 g_3 \sigma_{23} + g_1 g_2 g_3 \sigma_{123}
\]  
(4.11)
where \( \sigma_0, \sigma_1, \sigma_2, \sigma_3, \sigma_{12}, \sigma_{13}, \sigma_{23}, \sigma_{123} \) are all SOS polynomials in \( \mathbb{R}[x_1, x_2] \). Since the highest (total) degree monomial of each \( g_i \) has coefficient 1 and the leading coefficients of sums of squares are positive, the coefficients of the monomials of highest degree for each term on the right are positive. It follows that there is no leading term cancellation on the right and hence every term on the right has degree 2 or less. Thus we must have that \( \sigma_{12} = \sigma_{13} = \sigma_{23} = \sigma_{123} = 0 \) and \( \sigma_1, \sigma_2, \) and \( \sigma_3 \) are all constant. It is then easy to see that (4.11) is impossible if \( \gamma > 2 \). This proves our claim.

Since the maximum \( \gamma \) such that \( f - \gamma \in \mathcal{P} \) is 2, no matter how large we take the degree for the SOS relaxation using the Lasserre method, we can get only the lower bound 2. Thus the ratio of this lower bound to the true global minimum tends to zero when \( M \) goes to infinity. This shows that the SOS lower bounds can be arbitrarily bad.

Of course, the reason that the SOS method using the preorder fails in this case is that the feasible set (4.8)–(4.10) is noncompact, hence Schmüdgen’s Theorem does not apply. To overcome this problem, we might consider using the standard SOS method to solve problem (4.7)–(4.10) by adding a redundant condition like
\[
 R - x_1^2 - x_2^2 \geq 0
\]
as introduced in [12]. Here \( R \) is a sufficiently large positive number. We implemented this approach using SOSTOOLS and found that the lower bounds obtained this way are still very bad. The bigger the \( M \) is, the worse the bound we obtain.

Let us apply instead the optimization method described at the beginning of this section, using our representation theorems based on the KKT system. The KKT system for problem (4.7)–(4.10) is
\[
 \begin{align*}
 2(1 - \lambda_2 - \lambda_3)x_1 + (\lambda_2 - \lambda_3)Mx_2 &= 0 \\
 2(1 - \lambda_1)x_1 + (\lambda_2 - \lambda_3)Mx_1 &= 0 \\
 (x_2^2 - 1)\lambda_1 &= 0 \\
 (x_1^2 - Mx_1x_2 - 1)\lambda_2 &= 0 \\
 (x_1^2 + Mx_1x_2 - 1)\lambda_3 &= 0.
\end{align*}
\]
Using Macaulay 2 [5], we check that the KKT ideal \( I_{KKT} \) in this case is radical. Now let
\[
 q(x) = \rho_1 \left( x_1^2 - \frac{1}{4} \left( M + \sqrt{M^2 + 4} \right)^2 \right)^2 \lambda_1^2 \left( (x_1^2 + Mx_1x_2 - 1)\lambda_2^2 + (x_1^2 - Mx_1x_2 - 1)\lambda_3^2 \right)
 + \rho_2 \lambda_1^2 \left( \left( \frac{2\lambda_1}{2 + M^2} - 1 \right)^2 - \left( \frac{M^2}{M^2 + 4} \right)^2 \right)^2 \lambda_2 \left( (x_1^2 + Mx_1x_2 - 1) \right)
 + \rho_3 (4\lambda_1 \lambda_2)^2 (x_2^2 - 1) + \rho_4 (x_1^2 + Mx_1x_2 - 1)^2 (x_1^2 - Mx_1x_2 - 1)^2 (x_2^2 - 1)
 + (\lambda_2 (1 - 2\lambda_3))^2 \left( \rho_5 \left( \sqrt{M^2 + 1x_1^2} + 1 \right) (x_2^2 - 1) + \rho_6 \left( \sqrt{M^2 + 1x_1^2} - 1 \right)^2 (x_2^2 - 1) \right)
 + (\lambda_3 (1 - 2\lambda_3))^2 \left( \rho_5 \left( \sqrt{M^2 + 1x_1^2} + 1 \right) (x_2^2 - 1) + \rho_6 \left( \sqrt{M^2 + 1x_1^2} - 1 \right)^2 (x_2^2 - 1) \right).
\]
Here the constants are defined as
\[
 \rho_1 = \frac{4(M^2 + 4)^3/2}{(M^2 + 4)^{5/2} M^2 (2 + M^2)^2}, \quad \rho_3 = 1 + \frac{1}{2} M \left( M + \sqrt{M^2 + 4} \right),
\]
\[
 \rho_2 = \frac{1 + \frac{1}{2} M \left( M + \sqrt{M^2 + 4} \right)}{\left( \frac{M^4}{M^2 + 2} - \frac{M^2}{M^2 + 4} \right)^2}, \quad \rho_4 = 2 + \frac{1}{2} M \left( M + \sqrt{M^2 + 4} \right)
\]
\[
\rho_5 = \frac{M^{10} \sqrt{M^2 + 1} \left( \frac{1}{2} M \left( M + \sqrt{M^2 + 4} \right) + 2 - 2 \frac{\sqrt{M^2 + 1 - 1}}{M^2} \right)}{8(M^2 + 2) \left( \sqrt{M^2 + 1} - 1 \right)^3 \left( M^2 + 4 - 4\sqrt{M^2 + 1} \right)^2}.
\]
\[
\rho_6 = \frac{M^{10} \sqrt{M^2 + 1} \left( \frac{1}{\sqrt{M^2 + 1}} + \frac{1}{2} M \left( M + \sqrt{M^2 + 4} \right) + 2 + \frac{\left( \frac{\sqrt{M^2 + 1}}{M^2} \right)^2}{\sqrt{M^2 + 1}} \right)}{16(M^2 + 2) \left( \sqrt{M^2 + 1} + 1 \right)^3 \left( M^2 + 4 + 4\sqrt{M^2 + 1} \right)^2}.
\]

Then \( q(x) \) is visibly in \( M_{\text{KKT}} \) and hence in \( P_{\text{KKT}} \). It can be shown, e.g. using Macaulay 2, that
\[
 f(x) - f^* \equiv q(x) \mod I_{5, \text{KKT}}.
\]
This implies that \( f^*_5 = f^* \), hence we converge to the exact solution for \( N = 5 \). So for problem (4.7)–(4.10), our method returns the global minimum exactly. Thus the KKT system plays a crucial role in this example.

5. Conclusions

This paper studies representations of positive polynomials on noncompact semialgebraic sets via the KKT ideal. We give a representation theorem for polynomials positive on a basic closed semialgebraic set, even in the case where the semialgebraic set is not compact. This theorem can be used to numerically solve an optimization problem of the form (1.1) and (1.2) in the case where the feasible region is not compact. However, we must make the assumption that one of the global minimizers satisfies the KKT system. As discussed in [15], this assumption is sometimes very restrictive. Also, in general, the SOS relaxations (4.5) and (4.6) are very hard to solve when there are many constraints, since this introduces many Lagrange multipliers. The structure of (4.5) and (4.6) should be exploited to improve the efficiency of the method.

Acknowledgments

The authors would like to thank Zhi-Quan Luo and Paul Tseng for communicating to us the nonconvex quadratic optimization problem in Example 4.5. We also thank Bernd Sturmfels and the referee for many helpful comments.

The first and second authors were supported in part by the National Science Foundation (ELA-0122599). The third author was supported in part by the National Security Agency (H98230-05-1-00).

References