Digital Object Identifier (DOI) 10.1007/s10107-005-0672-6

Jiawang Nie · James Demmel · Bernd Sturmfels

Minimizing Polynomials via Sum of Squares over the Gradient Ideal

Received: November 16, 2004 / Accepted: September 15, 2005 Published online: 14 November 2005 – © Springer-Verlag 2005

Abstract. A method is proposed for finding the global minimum of a multivariate polynomial via sum of squares (SOS) relaxation over its gradient variety. That variety consists of all points where the gradient is zero and it need not be finite. A polynomial which is nonnegative on its gradient variety is shown to be SOS modulo its gradient ideal, provided the gradient ideal is radical or the polynomial is strictly positive on the real gradient variety. This opens up the possibility of solving previously intractable polynomial optimization problems. The related problem of constrained minimization is also considered, and numerical examples are discussed. Experiments show that our method using the gradient variety outperforms prior SOS methods.

Key words. Polynomials – Global Optimization – Sum of Squares (SOS) – Semidefinite Programming (SDP) – Radical Ideal– Variety – Gradient Ideal – Algebraic Geometry.

1. Introduction

We consider the global optimization problem

$$f^* = \min_{x \in \mathbb{R}^n} f(x) \tag{1}$$

where $x = (x_1, ..., x_n) \in \mathbb{R}^n$ is a real vector, and f(x) is a multivariate polynomial of degree *d*. As is well-known, the optimization problem (1) is NP-hard even when *d* is fixed to be four [22]. A lower bound can be computed efficiently using the Sum Of Squares (SOS) relaxation

$$f_{sos}^* = \text{maximize } \gamma \quad \text{subject to} \quad f(x) - \gamma \succeq_{sos} 0,$$
 (2)

where the inequality $g \succeq_{sos} 0$ means that the polynomial g is SOS, i.e. a sum of squares of other polynomials. We refer to [19, 23, 26–28] for introductions to SOS techniques and their applications. The SOS relaxation (2) can be reduced to a Semidefinite Program (SDP, see [35] for an introduction). The size of the matrix in the corresponding SDP is $\binom{n+d}{d}$, which is polynomial if either *n* or *d* is fixed. The relationship between (1) and (2) is as follows: $f_{sos}^* \leq f^*$ and the equality holds if and only if $f(x) - f^*$ is SOS.

e-mail: njw@math.berkeley.edu

J. Nie: Dept. of Math., Univ. of California, Berkeley, CA 94720, USA.

J. Demmel: Dept. of Math. and EECS, Univ. of California, Berkeley, CA 94720, USA. e-mail: demmel@cs.berkeley.edu

B. Sturmfels: Dept. of Math., Univ. of California, Berkeley, CA 94720, USA. e-mail: bernd@math.berkeley.edu

Blekherman [4] recently showed that, for fixed even degree $d \ge 4$, the ratio between the volume of all nonnegative polynomials and the volume of all SOS polynomials tends to infinity when *n* goes to infinity. In other words, for large *n*, there are many more nonnegative polynomials than SOS polynomials. Parrilo and Sturmfels [27] have used (2) to solve optimization problems (1) which were drawn at random from a natural distribution. Their test family will be revisited in Section 6.1.

For dealing with the challenging case when f_{sos}^* is strictly less than f^* , Lasserre [19] proposed finding a sequence of lower bounds for f(x) in some large ball $\{x \in \mathbb{R}^n : ||x||_2 \le R\}$. His approach is based on the result [2] that SOS polynomials are dense among polynomials which are nonnegative on some compact set. This sequence converges to f^* when the degrees of the polynomials introduced in the algorithm go to infinity. But it may not converge in finitely many steps, and the degrees of the required auxiliary polynomials can be very large.

In this paper, we introduce a method which can find the global minimum and terminate in finitely many steps, under some mild assumptions. Our point of departure is the observation that all local minima and global minima of (1) occur at points in the *real* gradient variety

$$V_{grad}^{\mathbb{R}}(f) = \{ u \in \mathbb{R}^n : \ (\nabla f)(u) = 0 \}.$$
(3)

The *gradient ideal* of *f* is the ideal in $\mathbb{R}[x_1, \ldots, x_n]$ generated by all partial derivatives of *f*:

$$\mathcal{I}_{grad}(f) = \langle \nabla f(x) \rangle = \langle \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n} \rangle.$$
(4)

There are several recent references on minimizing polynomials by way of the gradients. Hanzon and Jibetean [14] suggest applying perturbations to f to produce a sequence of polynomials f_{λ} (for small λ) with the property that the gradient variety of f_{λ} is finite and the minima f_{λ}^* converge to f^* as λ goes to 0. Laurent [20] and Parrilo [30] discuss the more general problem of minimizing a polynomial subject to polynomial equality constraints (not necessarily partial derivatives). Under the assumption that the variety defined by the equations is finite, the matrix method proposed in [20] has finite convergence even if the ideal generated by the constraints is not radical. Building on [14, 20], Jibetean and Laurent [17] propose to compute f^* by solving a single SDP, provided the gradient variety is finite (radicalness is not necessary).

There are also methods for minimizing polynomials based on exact real computing, such as [1, 13, 27, 32]. If the gradient ideal $\mathcal{I}_{grad}(f)$ is zero-dimensional, i.e., $\nabla f(x) = 0$ has finitely many complex solutions, then we can apply these methods to find all real solutions, and choose those which minimize f(x). If the gradient $\mathcal{I}_{grad}(f)$ is not zero-dimensional, perturbations may be applied to make $\mathcal{I}_{grad}(f)$ zero-dimensional, as in [17, 32]. However, even if $\mathcal{I}_{grad}(f)$ is zero-dimensional, the system $\nabla f(x) = 0$ usually has exponentially many solutions (the Bézout number is $(d-1)^n$), which is not tractable in computations.

The approach of this paper is to find a lower bound f_{grad}^* for (1) by requiring $f - f_{grad}^*$ to be SOS in the quotient ring $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}_{grad}(f)$ instead of in $\mathbb{R}[x_1, \ldots, x_n]$. Under the assumption that the infimum f^* is attained, i.e., there exits

some $x^* \in \mathbb{R}^n$ such that $f(x^*) = f^*$, we can find a monotonically increasing sequence $\{f_{N,grad}^*\}$ such that $\lim_{N\to\infty} f_{N,grad}^* = f^*$, and the equality $f_{N,grad}^* = f^*$ holds (i.e., finite convergence) for some large integer N when the ideal $\mathcal{I}_{grad}(f)$ is radical.

This paper is organized as follows. Section 2 offers a review of fundamental results from (real) algebraic geometry. In Section 3 we prove that a positive polynomial is SOS modulo its gradient ideal. The same holds for non-negative polynomials if the gradient ideal is radical. The resulting algorithms for unconstrained polynomial minimization will be presented in Section 4. Section 5 generalizes our methods to constrained optimization. In Section 6 we discuss numerical experiments. Section 7 draws some conclusions.

2. Tools from Algebraic Geometry

This section will introduce some basic notions from algebraic geometry needed for our discussion. Readers may consult [7, 8, 11] for more details. We write $\mathbb{R}[x] = \mathbb{R}[x_1, \ldots, x_n]$ for the ring of all polynomials in *n* variables $x = (x_1, \ldots, x_n)$ with real coefficients. A subset *I* of $\mathbb{R}[x]$ is an *ideal* if $p \cdot h \in I$ for any $p \in I$ and $h \in \mathbb{R}[x]$. If $g_1, \ldots, g_r \in \mathbb{R}[x]$ then $\langle g_1, \cdots, g_m \rangle$ denotes the smallest ideal containing the g_i . Equivalently, $\langle g_1, \cdots, g_m \rangle$ is the set of all polynomials that are polynomial linear combinations of the g_i . Every ideal arises in this way:

Theorem 1. (Hilbert Basis Theorem, Section 5, Ch. 2, [7]). Every ideal $I \subset \mathbb{R}[x]$ has a finite generating set, i.e., $I = \langle g_1, \dots, g_m \rangle$ for some $g_1, \dots, g_m \in I$.

The *variety* of an ideal *I* is the set of all common **complex** zeros of the polynomials in *I*:

$$V(I) = \{ x \in \mathbb{C}^n : p(x) = 0 \text{ for all } p \in I \}.$$

The subset of all real points in V(I) is the *real variety* of I. It is denoted

$$V^{\mathbb{R}}(I) = \{ x \in \mathbb{R}^n : p(x) = 0 \text{ for all } p \in I \}.$$

If $I = \langle g_1, \ldots, g_m \rangle$ then $V(I) = V(g_1, \ldots, g_m) = \{x \in \mathbb{C}^n : g_1(x) = \cdots = g_m(x) = 0\}$. An ideal $I \subseteq \mathbb{R}[x]$ is *zero-dimensional* if its variety V(I) is a finite set. This condition is much stronger than requiring that the real variety $V^{\mathbb{R}}(I)$ be a finite set. For instance, $I = \langle x_1^2 + x_2^2 \rangle$ is not zero-dimensional: the real variety $V^{\mathbb{R}}(I) = \{(0, 0)\}$ is only one point of the curve V(I).

Theorem 2. (Chapter 5,[7]). *The following conditions are equivalent for an ideal* $I \subset \mathbb{R}[x]$:

- (i) I is zero-dimensional (the variety V(I) is a finite set);
- (ii) the quotient ring $\mathbb{R}[x]/I$ is a finite-dimensional \mathbb{R} -vector space;
- (iii) if G is a Gröbner basis of I, then for each $1 \le i \le n$, there exists an integer $m_i \ge 0$ such that $x_i^{m_i}$ is the leading term of some $g \in G$.

A variety $V \subseteq \mathbb{C}^n$ is *irreducible* if there do not exist two proper subvarieties $V_1, V_2 \subsetneq V$ such that $V = V_1 \cup V_2$. Given a variety $V \subseteq \mathbb{C}^n$, the set of all polynomials that vanish on V is an ideal

$$I(V) = \{ p \in \mathbb{R}[x] : p(u) = 0 \text{ for all } u \in V \}.$$

Given any ideal *I* of $\mathbb{R}[x]$, its *radical* is the ideal

$$\sqrt{I} = \{ q \in \mathbb{R}[x] : q^m \in I \text{ for some } m \in \mathbb{N} \}.$$

Note that $I \subseteq \sqrt{I}$. We say that *I* is a *radical ideal* if $\sqrt{I} = I$. Clearly, the ideal I(V) defined by a variety *V* is a radical ideal. The following theorems offer a converse to this observation:

Theorem 3. (Hilbert's Weak Nullstellensatz). If I is an ideal in $\mathbb{R}[x]$ such that $V(I) = \emptyset$ then $1 \in I$.

Theorem 4. (Hilbert's Strong Nullstellensatz). If I is an ideal in $\mathbb{R}[x]$, then $I(V(I)) = \sqrt{I}$.

In real algebraic geometry, we are also interested in subsets of \mathbb{R}^n of the form

$$S = \{ x \in \mathbb{R}^n : g_1(x) = \dots = g_m(x) = 0, h_1(x) \ge 0, \dots, h_\ell(x) \ge 0 \},\$$

where $g_i, h_j \in \mathbb{R}[x]$. We call *S* a *basic semi-algebraic set*. With the given description of *S*, we associate the following set of polynomials:

$$M(S) = \left\{ \sigma_0(x) + \sum_{i=1}^m \lambda_i(x)g_i(x) + \sum_{j=1}^\ell h_j(x)\sigma_j(x) : \sigma_0, \cdots, \sigma_\ell \text{ are SOS}, \lambda_i(x) \in \mathbb{R}[x] \right\}.$$

Theorem 5. (Putinar, [31]). Assume that the basic semi-algebraic set S is compact and there exists one polynomial $\rho(x) \in M(S)$ such that the set $\{x \in \mathbb{R}^n : \rho(x) \ge 0\}$ is compact. Then every polynomial p(x) which is positive on S belongs to M(S).

Suppose we are given an ideal $I = \langle h_1, \ldots, h_r \rangle$ in $\mathbb{R}[x]$ and a polynomial $f \in \mathbb{R}[x]$. Then we can regard f as an element in the quotient $\mathbb{R}[x]/I$. Even if f is not SOS in $\mathbb{R}[x]$, it is possible for f to be SOS in the quotient ring $\mathbb{R}[x]/I$. For f to be SOS in $\mathbb{R}[x]/I$ means that there exists a $q \in I$ such that f - q is SOS in $\mathbb{R}[x]$, or, more explicitly, that f has a representation

$$f(x) = \sum_{j} q_j^2(x) + \sum_{i} \phi_i(x) h_i(x)$$

for some polynomials $q_j(x)$ and $\phi_i(x)$, Clearly, if f is SOS in $\mathbb{R}[x]/I$ then the function f(x) is non-negative on the real variety $V^{\mathbb{R}}(I)$. The following partial converse holds if V(I) is finite.

Theorem 6. (Parrilo, [30]). Let I be a zero-dimensional radical ideal in $\mathbb{R}[x]$. Then a polynomial $f \in \mathbb{R}[x]$ is nonnegative on the real variety $V^{\mathbb{R}}(I)$ if and only if f(x) is SOS in $\mathbb{R}[x]/I$.

We close with the following theorem, which is a special case of the *real Nullstellensatz*.

Theorem 7. (Real Nullstellensatz, [3, 5, 6]). Let I be an ideal in $\mathbb{R}[x]$ whose real variety $V^{\mathbb{R}}(I)$ is empty. Every polynomial f(x) is SOS in $\mathbb{R}[x]/I$. In particular, -1 is SOS in $\mathbb{R}[x]/I$.

3. Polynomials over their Gradient Varieties

Consider a polynomial $f \in \mathbb{R}[x]$ and its gradient ideal $\mathcal{I}_{grad}(f)$ as in (4). A natural idea in solving (1) is to apply Theorem 6. to the ideal $I = \mathcal{I}_{grad}(f)$, since the minimum of f over \mathbb{R}^n will be attained at a subset of $V^{\mathbb{R}}(I)$ if it is attained at all. However, the hypothesis of Theorem 6. requires that I be zero-dimensional, which means that the complex variety $V_{grad}(f) = V(I)$ of all critical points must be finite. Our results in this section remove this restrictive hypothesis. We shall prove that every nonnegative f is SOS in $\mathbb{R}[x]/I$ as long as the gradient ideal $I = \mathcal{I}_{grad}(f)$ is radical.

Theorem 8. Assume that the gradient ideal $\mathcal{I}_{grad}(f)$ is radical. If the real polynomial f(x) is nonnegative over $V_{grad}^{\mathbb{R}}(f)$, then there exist real polynomials $q_i(x)$ and $\phi_j(x)$ so that

$$f(x) = \sum_{i=1}^{s} q_i(x)^2 + \sum_{j=1}^{n} \phi_j(x) \frac{\partial f}{\partial x_j}.$$
(5)

The proof of this theorem will be based on the following two lemmas. The first is a generalization of the Lagrange Interpolation Theorem from sets of points to disjoint varieties.

Lemma 1. Let V_1, \ldots, V_r be pairwise disjoint varieties in \mathbb{C}^n . Then there exist polynomials $p_1, \ldots, p_r \in \mathbb{R}[x]$ such that $p_i(V_j) = \delta_{ij}$, where δ_{ij} is the Kronecker delta function.

Proof. Our definition of variety requires that each V_j is actually defined by polynomials with real coefficients. If $I_j = I(V_j)$ is the radical ideal of V_j then we have $V_j = V(I_j)$. Fix an index j and let W_j denote the union of the varieties $V_1, \ldots, V_{i-1}, V_{i+1}, \ldots, V_r$. Then

$$I(W_i) = I_1 \cap \cdots \cap I_{i-1} \cap I_{i+1} \cap \cdots \cap I_r.$$

Our hypothesis implies that $V_j \cap W_j = \emptyset$. By Hilbert's Weak Nullstellensatz (Theorem 3.), there exist polynomials $p_j \in I(W_j)$ and $q_j \in I_j$ such that $p_j + q_j = 1$. This identity shows that $p_j(V_j) = 1$ and $p_j(V_k) = 0$ for $k \neq j$. Hence the *r* polynomials p_1, \ldots, p_r have the desired properties.

Now consider the behavior of the polynomial f(x) over its gradient variety $V_{grad}(f)$. We make use of the fact that $V_{grad}(f)$ is a finite union of irreducible subvarieties ([3, §2]).

Lemma 2. Let W be an irreducible subvariety of $V_{grad}(f)$ and suppose that W contains at least one real point. Then f(x) is constant on W.

Proof. If we replace our polynomial ring $\mathbb{R}[x]$ by $\mathbb{C}[x]$ then W either remains irreducible or it becomes a union of two irreducible components $W = W_1 \cup W_2$ which are exchanged under complex conjugation. Let us first consider the case when W is irreducible in the Zariski topology induced from $\mathbb{C}[x]$. Then W is connected in the strong topology on \mathbb{C}^n (see [34]). Any connected algebraic variety in \mathbb{C}^n can be connected by algebraic curves. They may be singular, but they are images of nonsingular curves. So W is smoothly path-connected. Let x, y be two arbitrary points in W. There exists a smooth path $\varphi(t)$ ($0 \le t \le 1$) lying inside W such that $x = \varphi(0)$ and $y = \varphi(1)$. By the Mean Value Theorem of Calculus, it holds that for some $t^* \in (0, 1)$,

$$f(y) - f(x) = \nabla f(\varphi(t^*))^T \varphi'(t^*) = 0,$$

since $\nabla f(x)$ vanishes on W. We conclude that f(x) = f(y), and hence f is constant on W.

Now consider the case when $W = W_1 \cup W_2$ where W_1 and W_2 are exchanged by complex conjugation. We had assumed that W contains a real point p. Since p is fixed under complex conjugation, $p \in W_1 \cap W_2$. By the same argument as above, f(x) = f(p) for all $x \in W$.

Proof of Theorem 8. Consider the irreducible decomposition of $V_{grad}(f)$. We group together all components which have no real point and all components on which f takes the same real value. Hence the gradient variety has a decomposition

$$V_{grad}(f) = W_0 \cup W_1 \cup W_2 \cup \dots \cup W_r, \tag{6}$$

such that W_0 has no real point and f is a real constant on each other variety W_i , say,

$$f(W_1) > f(W_2) > \cdots > f(W_r) \ge 0.$$

The varieties W_i are pairwise disjoint, so by Lemma 1 there exist polynomials $p_i \in \mathbb{R}[x]$ such that $p_i(W_j) = \delta_{ij}$. By Theorem 7., there exists a sum of squares $sos(x) \in \mathbb{R}[x]$ such that f(x) = sos(x) for all $x \in W_0$. Using the non-negative real numbers $\alpha_j := \sqrt{f(W_j)}$, we define

$$q(x) = sos(x) \cdot p_0^2(x) + \sum_{i=1}^r (\alpha_i \cdot p_i(x))^2.$$
 (7)

By construction, f(x) - q(x) vanishes on the gradient variety $V_{grad}(f)$. The gradient ideal $\mathcal{I}_{grad}(f)$ was assumed to be radical. Using Hilbert's Strong Nullstellensatz (Theorem 4.), we conclude that f(x) - q(x) lies in $\mathcal{I}_{grad}(f)$. Hence the desired representation (5) exists.

In Theorem 8, the assumption that $\mathcal{I}_{grad}(f)$ is radical cannot be removed. This is shown by the following counterexample which was suggested to us by Claus Scheiderer.

Example 1. Let n = 3 and consider the polynomial

$$f(x, y, z) = x^{8} + y^{8} + z^{8} + M(x, y, z)$$

where $M(x, y, z) = x^4 y^2 + x^2 y^4 + z^6 - 3x^2 y^2 z^2$ is the Motzkin polynomial, which is is non-negative but not a sum of squares. The residue ring $A = \mathbb{R}[x, y, z]/\mathcal{I}_{grad}(f)$ is a real vector space of dimension $7^3 = 243$ because the three partial derivatives form a Gröbner basis:

$$\frac{\partial f}{\partial x} = \underline{8 \, x^7} + 4 \, x^3 y^2 + 2 \, x y^4 - 6 \, x y^2 z^2,$$

$$\frac{\partial f}{\partial y} = \underline{8 \, y^7} + 2 \, x^4 y + 4 \, x^2 y^3 - 6 \, x^2 y z^2,$$

$$\frac{\partial f}{\partial z} = \underline{8 \, z^7} + 6 \, z^5 - 6 \, x^2 y^2 z.$$

Reduction modulo this Gröbner basis shows that f(x, y, z) is congruent to $\frac{1}{4}M(x, y, z)$ modulo $\mathcal{I}_{grad}(f)$. Hence it suffices to show that M(x, y, z) is not a sum of squares in A. Suppose otherwise. Then there exist polynomials $s_i, \phi_1, \phi_2, \phi_3 \in \mathbb{R}[x, y, z]$ such that

$$M(x, y, z) = \sum_{i} s_{i}^{2} + \frac{\partial f}{\partial x} \phi_{1}(x, y, z) + \frac{\partial f}{\partial y} \phi_{2}(x, y, z) + \frac{\partial f}{\partial z} \phi_{3}(x, y, z).$$
(8)

By inspecting $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial f}{\partial z}$ and M, we see that every monomial in the expansion of $\sum_i s_i^2$ has degree at least six, and the monomials x^6 , y^6 , x^4z^2 , y^4z^2 , x^2z^4 , y^2z^4 cannot occur. This implies

 $s_i(x, y, z) = A_1^{(i)}xy^2 + A_2^{(i)}x^2y + A_3^{(i)}z^3 + A_4^{(i)}xyz +$ higher order terms $\phi_1(x, y, z) = Bx$ + other linear and high order terms $\phi_2(x, y, z) = Cy$ + other linear and high order terms $\phi_3(x, y, z) = Dz$ + other linear and high order terms.

Comparing the terms in M(x, y, z) with the expansion of the right hand side in (8), we get

$$x^{4}y^{2}: \quad 1 = \sum_{i} A_{1}^{(i)^{2}} + 4B + 2C$$
$$x^{2}y^{4}: \quad 1 = \sum_{i} A_{2}^{(i)^{2}} + 2B + 4C$$
$$z^{6}: \quad 1 = \sum_{i} A_{3}^{(i)^{2}} + 6D$$
$$x^{2}y^{2}z^{2}: -3 = \sum_{i} A_{4}^{(i)^{2}} - 6B - 6C - 6D.$$

Summing the above equations together results in

$$0 = \sum_{i} A_{1}^{(i)^{2}} + \sum_{i} A_{2}^{(i)^{2}} + \sum_{i} A_{3}^{(i)^{2}} + \sum_{i} A_{4}^{(i)^{2}}.$$

Thus $A_1^{(i)} = A_2^{(i)^2} = A_3^{(i)^2} = A_4^{(i)} = 0$ and $B = C = D = \frac{1}{6}$. Hence s_i only contains terms of degree ≥ 4 . Let $E^{(i)}$ be the coefficient of z^4 in $s_i(x, y, z)$. Comparing the coefficient of z^8 in (8), we get $0 = \sum_i E^{(i)^2} + 4/3$, which is a contradiction. We conclude that the nonnegative polynomial $f(x, y, z) = M(x, y, z) + x^8 + y^8 + z^8$ is not SOS modulo its gradient ideal.

In cases (like Example 1) when the gradient ideal is not radical, the following still holds.

Theorem 9. Let $f(x) \in \mathbb{R}[x]$ be a polynomial which is strictly positive on its real gradient variety $V_{grad}^{\mathbb{R}}(f)$, Then f(x) is SOS modulo its gradient ideal $\mathcal{I}_{grad}(f)$.

Proof of Theorem 8. We retain the notation from the proof of Theorem 8. Consider the decomposition of the gradient variety in (6). Each W_i is the union of several irreducible components. Consider a primary decomposition of the ideal $\mathcal{I}_{grad}(f)$, and define J_i to be the intersection of all primary ideals in that decomposition whose variety is contained in W_i . Then we have $\mathcal{I}_{grad}(f) = J_0 \cap J_1 \cap \cdots \cap J_r$, where $W_i = V(J_i)$ and, since the W_i are pairwise disjoint, we have $J_i + J_k = \mathbb{R}[x]$ for $i \neq k$. The Chinese Remainder Theorem [11] implies

$$\mathbb{R}[x]/\mathcal{I}_{grad}(f) \simeq \mathbb{R}[x]/J_0 \times \mathbb{R}[x]/J_1 \times \cdots \times \mathbb{R}[x]/J_r.$$
(9)

Here $V^{\mathbb{R}}(J_0) = \emptyset$. Hence, by Theorem 7., there exists a sum of squares $sos(x) \in \mathbb{R}[x]$ such that $f(x) - sos(x) \in J_0$. By assumption, $\alpha_i^2 = f(W_i)$ is strictly positive for all $i \ge 1$. The polynomial $f(x)/\alpha_i^2 - 1$ vanishes on W_i . By Hilbert's Strong Nullstellensatz, there exists an integer m > 0 such that $(f(x)/\alpha_i^2 - 1)^m$ is in the ideal J_i . We construct a square root of $f(x)/\alpha_i^2$ in the residue ring $\mathbb{R}[x]/J_i$ using the familiar Taylor series expansion for the square root function:

$$\left(1 + (f(x)/\alpha_i^2 - 1)\right)^{1/2} = \sum_{k=0}^{m-1} {\binom{1/2}{k}} (f(x)/\alpha_i^2 - 1)^k \mod J_i.$$

Multiplying this polynomial by α_i , we get a polynomial $q_i(x)$ such that $f(x) - q_i^2(x)$ is in the ideal J_i . We have shown that f(x) maps to the vector $(sos(x), q_1(x)^2, q_2(x)^2, \ldots, q_r(x)^2)$ under the isomorphism (9). That vector is clearly a sum of squares in the ring on the right hand side of (9). We conclude that f(x) is a sum of squares in $\mathbb{R}[x]/\mathcal{I}_{grad}(f)$.

Example 2. Let f be the polynomial in Example 1 and let ϵ be any positive constant. Theorem 9 says that $f + \epsilon$ is SOS modulo $\mathcal{I}_{grad}(f)$. Such a representation can be found by symbolic computation as follows. Primary decomposition over $\mathbb{Q}[x, y, z]$ yields

$$\mathcal{I}_{grad}(f) = J_0 \cap J_1,$$

where $V^{\mathbb{R}}(J_0) = \emptyset$ and $\sqrt{J_1} = \langle x, y, z \rangle$. The ideal J_1 has multiplicity 153, and it contains the square f^2 of our given polynomial. The ideal J_0 has multiplicity 190. Its variety $V(J_0)$ consists of 158 distinct points in \mathbb{C}^3 . By elimination, we can reduce to the univariate case. Using the algorithm of [5, 6] for real radicals in $\mathbb{Q}[z]$, we find a sum of squares $sos(z) \in \mathbb{Q}[z]$ such that $f - sos(z) \in J_0$. Running Buchberger's algorithm for $J_0 + J_1 = \langle 1 \rangle$, we get polynomials $p_0 \in J_0$ and $p_1 \in J_1$ such that $p_0 + p_1 = 1$. The following polynomial is a sum of squares,

$$p_1^2 \cdot (sos(z) + \epsilon) + p_0^2 \cdot \epsilon \cdot (1 + \frac{1}{2\epsilon}f)^2, \tag{10}$$

and it is congruent to $f(x, y, z) + \epsilon$ modulo $\mathcal{I}_{grad}(f) = J_0 \cap J_1 = J_0 \cdot J_1$. Note that the coefficients of the right hand polynomial in the SOS representation (10) tend to infinity as ϵ approaches zero. This is consistent with the conclusion of Example 1.

4. Unconstrained Optimization

This section concerns finding the global minimum of a polynomial function f(x) on \mathbb{R}^n . Let $\mathbb{R}[x]_m$ denote the $\binom{n+m}{m}$ -dimensional vector space of polynomials of degree at most *m*. Since the gradient is zero at local or global minimizers, we consider the SOS relaxation

Maximize
$$\gamma$$
 subject to $f(x) - \gamma - \sum_{j=1}^{n} \phi_j(x) \frac{\partial f}{\partial x_j} \succeq_{sos} 0$ and $\phi_j(x) \in \mathbb{R}[x]_{2N-d+1}$.
(11)

Here *d* is the degree of polynomial f(x), and *N* is an integer to be chosen by the user. Let $f_{N,grad}^*$ denote the optimal value γ of the optimization problem (11). This is a lower bound for the global minimum f^* of the polynomial f(x). The lower bound gets better as *N* increases:

$$\cdots \leq f_{N-1,grad}^* \leq f_{N,grad}^* \leq f_{N+1,grad}^* \leq \cdots \leq f^*.$$
(12)

4.1. SOS optimization using the software SOSTOOLS

The problem (11) is a standard SOS program. It can be translated into an SDP as described in [26–28]. The decision variables in (11) are the real number γ and the coefficients the multiplier polynomials $\phi_j(x)$. The resulting SDP is dual to the formulation of Lasserre [19]. The SOS program (11) can be solved using the software package SOSTOOLS. We refer to [29] for the documentation. For instance, if we take N = 4 and f(x, y, z) the trivariate polynomial in Example 1 then (11) translates into an SOSTOOLS program as follows:

```
syms x y z gam;
prog = sosprogram([x;y;z],[gam]);
f = x^8+y^8+z^8+x^4*y^2+x^2*y^4+z^6-3*x^2*y^2*z^2;
```

```
vec = monomials([x;y;z],[0 1]);
[prog,phi_1] = sospolyvar(prog,vec);
[prog,phi_2] = sospolyvar(prog,vec);
[prog,phi_3] = sospolyvar(prog,vec);
G = f-gam-(phi_1*diff(f,x)+phi_2*diff(f,y)+phi_3*diff(f,z));
prog = sosineq(prog,G);
prog = sossetobj(prog,-gam);
[prog, info] = sossolve(prog);
gam = sosgetsol(prog,gam);
```

The system returns the following lower bound $\gamma = f_{4,grad}^*$ for the global minimum $f^* = 0$:

gam = -.12077e - 8

Even if we increase the value of N, the lower bound $f_{N,grad}^*$ always remains negative, since f is not SOS modulo its gradient ideal. However, the sequence $\{f_{N,grad}^*\}_{N\geq 4}$ converges to zero.

4.2. Convergence of the lower bounds

We have the following general result concerning the convergence of the lower bounds.

Theorem 10. Let f(x) be a polynomial in n real variables which attains its infimum f^* over \mathbb{R}^n . Then $\lim_{N\to\infty} f^*_{N,grad} = f^*$. Furthermore, if the gradient ideal $\mathcal{I}_{grad}(f)$ is radical, then f^* is attainable, i.e., there exists an integer N such that $f^*_{N,grad} = f^*_{grad} = f^*$.

Proof. Since f(x) attains its infimum, the global minima of f(x) must occur on the real gradient variety $V_{grad}^{\mathbb{R}}(f)$. It is obvious that any real number γ which satisfies the SOS constraint in (11) is a lower bound of f(x), and we have the sequence of inequalities in (12). Consider an arbitrary small real number $\varepsilon > 0$. The polynomial $f(x) - f^* + \varepsilon$ is strictly positive on its real gradient variety $V_{grad}^{\mathbb{R}}(f)$. By Theorem 9, $f(x) - f^* + \varepsilon$ is SOS modulo $\mathcal{I}_{grad}(f)$. Hence there exists an integer $N(\epsilon)$ such that

 $f_{N \text{ erad}}^* \ge f^* - \varepsilon$ for all $N \ge N(\epsilon)$.

Since the sequence $\{f_{N,grad}^*\}$ is monotonically increasing, it follows that $\lim_{N\to\infty} f_{N,grad}^* = f^*$.

Now suppose $\mathcal{I}_{grad}(f) = \mathcal{I}_{grad}(f - f^*)$ is a radical ideal. The nonnegative polynomial $f(x) - f^*$ is SOS modulo $\mathcal{I}_{grad}(f)$ by Theorem 8. Hence $f^*_{N,grad} = f^*$ for some $N \in \mathbb{Z}_{>0}$.

Remarks: (i) The condition that f(x) attains its infimum cannot be removed. Otherwise the infimum f_{grad}^* of f(x) on $V_{grad}^{\mathbb{R}}(f)$ need not be a lower bound for f(x) on \mathbb{R}^n . A counterexample is $f(x) = x^3$. Obviously f(x) has infimum $f^* = -\infty$ on \mathbb{R}^1 . However, $f_{grad}^* = f_{grad,N}^* = 0$ for all $N \ge 1$ because $f(x) = (\frac{x}{3})f'(x)$ is in the gradient ideal $\mathcal{I}_{grad}(f) = \langle f'(x) \rangle$. (ii) It is also not always the case that $f_{grad}^* = f^*$ when f^* is finite. Consider the bivariate polynomial $f(x, y) = x^2 + (1 - xy)^2$. We can see that $f^* = 0$ is not attained, but $f_{grad}^* = 1 > f^*$.

(iii) If f(x) attains its infimum but $\mathcal{I}_{grad}(f)$ is not radical, we have only that $\lim_{N \to \infty} f_{N,grad}^* = f^*$. But there is typically no integer N with $f_{N,grad}^* = f^*$, as shown in Example 1.

4.3. Duality and an algorithm for finding minimizing points

In this subsection, we describe a dual formulation of the SOS problem (11), and we present an explicit algorithm for finding the global minimizer of a polynomial f(x). First, we introduce some notation. Given any polynomial p(x) in $\mathbb{R}[x]_m$, we write $p(x) = \sum_{|\alpha| \le m} p_\alpha x^\alpha$ where $\alpha \in \mathbb{Z}_{\ge 0}^n$ and $|\alpha| = \sum_{j=1}^n \alpha_j$. In what follows, we denote by $p \in \mathbb{R}^{\binom{n+m}{m}}$ the vector of coefficients p_α of p(x). For any integer N, we write $\mathbb{Z}_N^n = \{\tau \in \mathbb{Z}_{\ge 0}^n : |\tau| \le N\}$ and we denote by $mon_N(x)$ the column vector of monomials of degree up to N, i.e.,

$$mon_N(x) = (1, x_1, \cdots, x_n, x_1^2, x_1x_2, \cdots, x_1^N, \cdots, x_n^N)^T.$$

The dimension of $mon_N(x)$ is the binomial coefficient $\binom{n+N}{N}$. Given any finite or infinite vector $y = (y_\alpha)$, indexed by integer vectors $\alpha \in \mathbb{Z}_{\geq 0}^n$, define $M_N(y)$ to be its moment matrix

$$M_N(y) = (y_{\alpha+\beta})_{\alpha,\beta\in\mathbb{Z}_N^n}$$

The moment matrix represents the linear map $p \mapsto p * y$, where, for any polynomial $p(x) = \sum_{\beta} p_{\beta} x^{\beta}$, the vector p * y has coordinates $(p * y)_{\alpha} = \sum_{\beta} p_{\beta} y_{\alpha+\beta}$. Let f(x) be the polynomial we wish to minimize. Its vector of coefficients is f. Let

Let f(x) be the polynomial we wish to minimize. Its vector of coefficients is f. Let f_i denote the vector of coefficients of the *i*-th partial derivative $\frac{\partial f}{\partial x_i}$. We rewrite (11) as follows:

$$f_{N,grad}^* = \max_{\substack{\gamma \in \mathbb{R}, \sigma \in \mathbb{R}[x]_{2N} \\ \phi_j(x) \in \mathbb{R}[x]_{2N-d+1}}} \gamma \quad \text{subject to} \quad \sigma(x) \succeq_{sos} 0$$

and
$$f(x) - \gamma = \sigma(x) + \sum_{j=1}^{n} \phi_j(x) \frac{\partial f}{\partial x_j}$$
.

We call the formulation above the *dual SDP*, because it is dual to the formulation proposed in [19, 20]. The corresponding *primal SDP* supposes that d is even and it is given by

$$f_{N,mom}^{*} = \min_{y} \quad f^{T} y$$

s.t. $M_{N-d/2}(f_{i} * y) = 0, \ i = 1, \cdots, n$
 $M_{N}(y) \geq 0, \ y_{0} = 1.$

The following theorem relates the primal and dual objective function values $f_{N,mom}^*$ and $f_{N,grad}^*$, and it shows how to extract a point x^* in \mathbb{R}^n at which the minimum of f(x) is attained.

Theorem 11. Assume f(x) attains its infimum f^* over \mathbb{R}^n (hence d is even). Then we have:

- (i) $f_{N,mom}^* = f_{N,grad}^*$ and hence $\lim_{N \to \infty} f_{N,mom}^* = f^*$. This is referred to as strong duality [35].
- (ii) Suppose $f_{N,grad}^* = f^*$ for some N. If $x^* \in \mathbb{R}^n$ minimizes f(x), then $y^* = mon_{2N}(x^*) \in \mathbb{R}^{\binom{n+2N}{2N}}$ solves the primal SDP.
- (iii) If y is a solution to the primal problem with $rank(M_N(y)) = 1$, then factoring $M_N(y)$ as column vector times row vector yields one global minimizer x^* of the polynomial f(x).
- (iv) Suppose that $f_{N,grad}^* = f^*$ and $\sigma(x) = \sum_{j=1}^{\ell} (q_j(x))^2$ solves the dual SDP. Then the set of all global minima of f(x) equals the set of solutions $x \in \mathbb{R}^n$ to the following equations:

$$q_j(x) = 0, \quad j = 1, \dots, \ell$$

 $\frac{\partial f(x)}{\partial x_i} = 0, \quad i = 1, \dots, n.$

Proof. Parts (i) and (ii) are basically a direct application of Theorem 4.2 in [19]. The hypotheses of that theorem can be verified by an "epsilon argument" and applying our Theorem 9. Let us prove part (iii). Since the moment matrix $M_N(y)$ has rank one, there exists a vector $x^* \in \mathbb{R}^n$ such that $y = mon_N(x^*)$. The strong duality result in (i) implies that

$$f(x^*) = f^T y = f^*_{N,mom} = f^*_{N,grad}$$

Since $f_{N,grad}^*$ is a lower bound for f(x), we conclude that this lower bound is attained at the point x^* . Therefore, $f_{N,grad}^* = f^*$ and x^* is a global minimizer. Part (iv) is straightforward.

From Theorem 11 (ii), we can see that there exists one optimal solution y^* to the primal SDP such that $rank(M_N(y^*)) = 1$ if $f_{N,grad}^* = f^*$ for some integer N. However, interior-point solvers for SDP will find a solution with moment matrix of *maximum rank*. So, if there are several global minimizers, the moment matrix $M_N(y^*)$ at relaxation N for which the global minimum is reached, will have rank > 1. Therefore, we need to handle this situation. Fortunately, there is a suitable method in [16] which can detect global optimality and extract optimal solutions. This method has been implemented in the examples in Section 6. We refer to [16] for details. Here we briefly outline the technique.

Suppose for some integer N at optimal solution y^* to the primal SDP, the rank condition

$$\operatorname{rank} M_N(y^*) = \operatorname{rank} M_{N-d/2}(y^*) = r$$
 (13)

holds, which can be verified very accurately by Singular Value Decomposition (SVD). Then as a consequence of Theorem 1.6 in [10], there exist r vectors $x^*(1), \dots, x^*(r) \in \mathbb{R}^n$ such that

$$M_N(y^*) = \sum_{j=1}^r v_j \ mon_N(x^*(j)) \cdot mon_N(x^*(j))^T$$

where $\sum_{j=1}^{r} v_j = 1$ and $v_j > 0$ for all $j = 1, \dots, r$. Henrion and Lasserre [16] proposed a detailed algorithm to find all such vectors $x^*(j)$. The condition (13) can be satisfied for some N when $V_{grad}(f)$ is finite; see [20] for a proof.

Now we discuss how to extract the vectors $x^*(j)$, using the method described in [16]. Since $M_N(y^*) \geq 0$, its (pivoted) Cholesky factorization gives a lower triangular matrix V such that $M_N(y^*) = VV^T$. Reduce V to column echelon form

$$U = \begin{bmatrix} 1 & & \\ * & & \\ 0 & 1 & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \dots & 1 \\ * & * & \dots & * \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ * & * & \dots & * \end{bmatrix}$$

by elementary column operations. Notice that the rows of U are indexed by monomials x^{α} up to degree N. Let β_1, \dots, β_r be the indices corresponding to the ones in the above U. Let $w = [x^{\beta_1} \cdots x^{\beta_r}]^T$. Then $mon_N(x) = Uw$ for all solutions $x = x^*(j), j = 1, \dots, r$. Thus for each variable $x_i, i = 1, \dots, n$, we can extract the *r*-by-*r* submatrix N_i from U such that

$$N_i w = x_i w, \ i = 1, \cdots, n.$$

This means that x_i is an eigenvalue of N_i . Now let $N = \sum_{i=1}^{n} \rho_i N_i$ where $\rho_i \in (0, 1)$ are random numbers such that $\sum_{i=1}^{n} \rho_i = 1$. Then compute the ordered Schur decomposition $N = QTQ^T$ where $Q = [q_1 \cdots q_r]$ is orthogonal and T is real and upper triangular with diagonal entries sorted increasingly. Then

$$x_i^*(j) = q_j^T N_i q_j, \ i = 1, \cdots, n, \ j = 1, \cdots, r.$$

The justification of this process is in [16, 9].

Summarizing our discussion, we get the following algorithm for global minimization of polynomials.

Algorithm 1. Computing the global minimizer(s) (if any) of a polynomial.

Input: A polynomial f(x) of even degree d in n variables $x = (x_1, \ldots, x_n)$. Output: Global minimizers $x^*(1), \cdots, x^*(r) \in \mathbb{R}^n$ of f(x) for some $r \ge 1$. Algorithm: Initialize N = d/2.

Step 1 Solve the pair of primal SDP and dual SDP described above.

Step 2 Check rank condition (13). If it is satisfied, extract r solutions $x^*(1)$, \cdots , $x^*(r)$ by using the above method, where r is the rank of $M_N(y^*)$, and then stop.

Step 3 If (13) is not satisfied, N = N + 1 and then go to Step 1.

As we pointed out after (13) ([20]), this algorithm will terminate if $V_{grad}(f)$ is finite. If $V_{grad}(f)$ is infinite, it is possible to have infinitely many global minimizers and the extraction method in [16] can not be applied generally (it may work sometimes). In such situations we need to solve the equations in (iv) of Theorem 11 to find some global minimizers.

4.4. What if the gradient ideal $\mathcal{I}_{grad}(f)$ is NOT radical ?

The lack of radicalness of the gradient ideal $\mathcal{I}_{grad}(f)$ would be an obstacle for our algorithm. First of all, this does not happen often in practice because $\mathcal{I}_{grad}(f)$ is generically radical. The following result is proved by standard arguments of algebraic geometry. We omit the proof.

Proposition 1. For almost all polynomials f in the finite-dimensional vector space $\mathbb{R}[x]_d$, the gradient ideal $\mathcal{I}_{grad}(f)$ is radical and the gradient variety $V_{grad}(f)$ is a finite subset of \mathbb{C}^n .

Proposition 1 means that, for almost all polynomials f which attain their minimum f^* , Algorithm 1 will compute the minimum in finitely many steps. An a priori bound for a degree N with $f_{N,grad}^* = f^*$ is given in [20].

Let us now consider the unlucky case when $\mathcal{I}_{grad}(f)$ is not radical. This happened for instance, in Example 1. In theory, one can replace the gradient ideal $\mathcal{I}_{grad}(f)$ by its radical $\sqrt{\mathcal{I}_{grad}(f)}$ in our SOS optimization problem. This is justified by the following result.

Corollary 1. If a polynomial f(x) attains its infimum f^* over \mathbb{R}^n then $f(x) - f^*$ is SOS modulo the radical $\sqrt{\mathcal{I}_{grad}(f)}$ of the gradient ideal.

Proof. Consider the decomposition (6) and form the SOS polynomial q(x) in (7). Since f(x) - q(x) vanishes on the gradient variety $V(\mathcal{I}_{grad}(f)) = V(\sqrt{\mathcal{I}_{grad}(f)})$, Hilbert's Strong Nullstellensatz implies that $f(x) - q(x) \in \sqrt{\mathcal{I}_{grad}(f)}$.

Suppose we could compute a set of polynomials $\{h_1, h_2, \ldots, h_r\}$ which generate the radical $\sqrt{\mathcal{I}_{grad}(f)}$ of the Jacobian ideal. Then we can replace the partial derivatives $\partial f/\partial x_i$ by the polynomials h_j in the SOS program (11). The resulting SDP will always have the property that $f_{grad}^* = f^*$ provided this infimum is attainable. While there are known algorithms for computing radicals (see e.g. [12, 18]), and they are implemented in various computer algebra systems, running these algorithms is very time-consuming. We believe that replacing $\mathcal{I}_{grad}(f)$ by its radical $\sqrt{\mathcal{I}_{grad}(f)}$ is not a viable option for efficient optimization algorithms. However, it is conceivable that some polynomials in $\sqrt{\mathcal{I}_{grad}(f)} \setminus \mathcal{I}_{grad}(f)$ are known to the user (for instance, from the geometry of the problem at hand). Including such polynomials in the sum of (11), will surely enhance the speed of convergence of the sequence of lower bounds $f_{grad,N}^*, f_{grad,N+1}^*, \ldots \longrightarrow f^*$.

5. Constrained Optimization

This section discusses how to generalize the method in Section 4 to minimize a polynomial function subject to polynomial constraints. The conditions for optimality are now expressed using the KKT (Karush-Kuhn-Tucker) equations instead of the gradient ideal. We need to reformulate the problem accordingly. Similar results hold as in Section 4. There is another paper ([24]) which studies the representation of nonnegative and positive polynomials over noncompact semialgebraic sets. The authors refer to [24] for its applications in constrained polynomial optimization.

We consider the following constrained optimization problem involving polynomials in $\mathbb{R}[x]$:

$$f^* = \min \quad f(x) \tag{14}$$

s.t.
$$g_i(x) = 0, \quad i = 1, \dots, m.$$
 (15)

One lower bound can be found by SOS relaxation

$$f_{sos}^* = \max_{\substack{\gamma \in \mathbb{R} \\ \phi_i(x) \in \mathbb{R}[x]}} \gamma \quad \text{subject to} \quad f(x) - \gamma - \sum_i g_i(x)\phi_i(x) \succeq_{sos} 0.$$
(16)

There are several recent papers [19, 20, 30] on solving this kind of constrained problem using SOS or moment matrix techniques. The convergence of their methods is based on the assumption that the real variety $V^{\mathbb{R}}(g_1, \ldots, g_m)$ is compact or even finite, which allows the application of Putinar's Theorem 5.. When $V^{\mathbb{R}}(g_1, \ldots, g_m)$ is compact, the methods may not converge within finitely many steps. Laurent [20] established the finite convergence of moment matrix techniques when $V(g_1, \ldots, g_m)$ is finite. However, if $V^{\mathbb{R}}(g_1, \ldots, g_m)$ is not compact, then f_{sos}^* may be smaller than f^* ([19, 31]); or even if $V^{\mathbb{R}}(g_1, \ldots, g_m)$ is compact, we may just get a sequence of bounds that converge to f^* as the degrees of ϕ_i go to infinity [19].

As is well-known in optimization theory, the local or global optimal solutions to problem (14)–(15) satisfy the KKT conditions (under some regularity conditions, see [25])

$$\nabla f(x) + \sum_{i} \lambda_i \nabla g_i(x) = 0 \tag{17}$$

$$g_i(x) = 0. \tag{18}$$

As we can see, the above KKT system is exactly the gradient of the Lagrangian function

$$\mathcal{L}(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x).$$

Therefore the methods in the previous sections can be generalized directly here. Define the *KKT ideal*

$$I_{kkt} = \left\{ p(x,\lambda) \in \mathbb{R}[x,\lambda] : p(x,\lambda) = \sum_{j} \left(\frac{\partial f}{\partial x_{j}} + \sum_{i} \lambda_{i} \frac{\partial g_{i}}{\partial x_{j}}\right) \eta_{j}(x,\lambda) + \sum_{i} g_{i}(x)\phi_{i}(x,\lambda) \right\}.$$

Let $I_{\ell,kkt}$ denote the finite-dimensional \mathbb{R} -linear subspace of the ideal I_{kkt} consisting of all polynomials which have a representation as above where each summand has degree at most ℓ .

A lower bound for (14)–(15) can be found by solving the SOS programming problem

$$f_{N,kkt}^* = \max_{\gamma \in \mathbb{R}, \sigma \in \mathbb{R}[x,\lambda]_{2N}} \gamma \qquad \text{subject to} \quad \sigma(x,\lambda) \succeq_{sos} 0 \tag{19}$$

and $f(x) - \gamma = \sigma(x, \lambda) \mod I_{2N,kkt}$. (20)

Just like in Section 4, we call (19)–(20) the *dual SDP formulation* of our problem. Similarly, we can derive the following results from Theorem 8 and Theorem 9 (see [24]).

Theorem 12. Suppose that either f(x) is positive on V_{kkt} , or f(x) is nonnegative on V_{kkt} and I_{kkt} is a radical ideal. Then f(x) is a sum of squares in the residue ring $\mathbb{R}[x, \lambda]/I_{kkt}$.

Corollary 2. Assume the optimality conditions (17)–(18) hold at some global optima of constrained optimization (14)–(15). Then we have $\lim_{N\to\infty} f_{N,kkt}^* = f^*$. Furthermore, if the ideal I_{kkt} is radical, then f^* is attainable, i.e., there exist SOS polynomial $\sigma(x, \lambda)$

$$f(x) - f^* = \sigma(x, \lambda) \mod I_{2N,kkt}$$

for some large enough integer N.

such that

Corollary 2 does not need assume $V^{\mathbb{R}}(g_1, \dots, g_m)$ to be finite or compact.

6. Numerical Experiments

The examples in this section have been computed using the software GloptiPoly[15] and SOSTOOLS [29]. In Subsection 6.1 we compare our formulation (11) with the formulation (2) by testing the family of polynomials considered in [27]. From the comparison tables listed below, we see that our new formulation (11) is faster by roughly a quarter when compared to (2) on this family of polynomials. In Subsection 6.2, we test our method on examples where the lower bound f_{sos}^* is strictly less than f^* . In all cases our lower bound $f_{N,grad}^*$ equals f^* within rounding errors for suitable N.

6.1. Testing on the Parrilo-Sturmfels family of polynomials

In this subsection we consider the following family of polynomials of even degree d,

$$f(x_1,\cdots,x_n)=x_1^d+\cdots+x_n^d+g(x_1,\cdots,x_n),$$

where $g \in \mathbb{R}[x]$ is a random polynomial of degree $\leq d - 1$ whose coefficients are uniformly distributed between -K and K, for a fixed positive integer K. This family of polynomials was considered in [27] where it was shown experimentally that the SOS formulation (2) almost always yields the global minimum. Without loss of generality, we can set K = 1, because any f(x) in the above form can be scaled to have coefficients between -1 and 1 by taking

$$f_s(x_1, \cdots, x_n) = \alpha^{-d} \cdot f(\alpha x_1, \cdots, \alpha x_n)$$

for some properly chosen α . As observed in [27], this scaling will greatly increase the stability and speed of the numerical computations involved in solving the primal-dual SDP.

We ran a large number of randomly generated examples for various values of d and n. The comparison results are in listed in Table 1 and Table 4. The computations were performed on a Dell Laptop with a Pentium IV 2.0 GHz and 512MB of memory. Table 1 is the comparison of the lower bounds by formulation (2) and (11). Taking N = d/2 in Algorithm 1 appears to be good enough in practice for minimizing the Parrilo-Sturmfels polynomials. Our experiments show that increasing N above d/2 will not increase the lower bound significantly.

From Table 1, we can see that the lower bounds f_{sos}^* and $f_{N,grad}^*$ are close, agreeing to their leading 8 to 10 decimal digits, which confirms the observation made in [27] that almost *all* the polynomials gotten by subtracting their infima are SOS. Tables 2–4 are comparisons of running time in CPU seconds for formulations (2) and (11). The symbol "-" in the tables means that the computation takes more than one hour and we then terminate it. And "*" means we use a different scaling as described below.

"-" means the computation is terminated if it takes more than one hour; "*" means the coefficients of $g(x_1, \dots, x_n)$ are scaled to belong to [-0.1, 0.1].

$d \setminus n$	3	4	5	6	7	8	9	10
4	5	7	9	10	11	13	14	15
6	10	19	38	41	232	_	_	-
8	17	78	186	233	_	_	_	_
10	40	39*	102*	-	-	-	-	_

Table 1. The relative difference $\frac{|f_{N,grad}^{-} - f_{sos}^{*}|}{|f_{soc}^{*}|} \times 10^{10}$, with N = d/2.

$d \setminus n$	3	4	5	6	7	8	9	10
4	0.16	0.24	0.42	0.86	1.86	7.56	25.85	73.69
6	0.32	1.17	8.40	49.04	309.66	-	-	_
8	1.10	12.23	173.98	1618.86	-	_	_	_
10	3.15	64.48*	2144.04*	_	-	-	-	-

Table 2. Running time in CPU seconds via traditional SOS approach (2)

Table 3. Running time in CPU seconds via our approach (11), with N = d/2.

$d \setminus n$	3	4	5	6	7	8	9	10
4	0.12	0.18	0.32	0.68	1.46	5.65	18.85	54.97
6	0.23	0.91	6.39	35.16	241.71	_	_	_
8	0.84	9.54	129.53	1240.23	-	_	_	_
10	2.59	45.14*	1539.80*	-	_	-	-	-

Table 4. The ratio of CPU seconds between (2) and (11), with N = d/2.

$d \setminus n$	3	4	5	6	7	8	9	10
4	0.75	0.75	0.76	0.79	0.78	0.74	0.73	0.75
6	0.72	0.77	0.76	0.72	0.78	_	_	_
8	0.76	0.78	0.74	0.76	_	_	_	_
10	0.82	0.70*	0.71*	-	-	-	-	-

Our formulation (11) uses about three quarters of the running time used by formulation (2). This may be unexpected since the use of gradients introduces many new variables. While we are not sure of the reason, one possible explanation is that adding gradients improves the conditioning and makes the interior-point algorithm for solving the SDP converge faster.

The numerical performance is subtle in this family of test polynomials. In the cases (n, d) = (4, 10) or (n, d) = (5, 10), our formulation (11) has numerical trouble, while (2) does not, and yet (11) is still faster than (2). However, for these two cases, if we scale $f(x_1, \ldots, x_n)$ so that the coefficients of $g(x_1, \ldots, x_n)$ belong to [-0.1, 0.1], both (2) and (11) do not have numerical trouble, and formulation (11) is still faster than (2). In Table 4 we see that the time ratio between (11) and (2) under this scaling is smaller than the time ratio for other values of (n, d). So numerical comparisons in Tables 1–4 for (n, d) = (4, 10) or (n, d) = (5, 10) are implemented under this new scaling, while for other values of (n, d) we still use the old scaling where the coefficients of $g(x_1, \ldots, x_n)$ belong to [-1, 1]. A stability analysis for the scaling and the speed-up caused by adding gradients may be a future research topic.

6.2. Other examples

The following examples demonstrate the effectiveness of our Algorithm 1 for a sample of polynomials that have been discussed in the SOS optimization literature.

Homogeneous Polynomials Let f(x) be a homogeneous polynomial. Regardless of whether f(x) is non-negative, we always have $f_{N,grad}^* = 0$ for any $N \ge d/2$. This comes from the identity $f(x) = \frac{1}{d} \cdot \sum_i x_i \frac{\partial f}{\partial x_i}$, which implies that f(x) lies in its gradient ideal $\mathcal{I}_{grad}(f)$. In order to test global non-negativity of a homogeneous polynomial f(x), we can apply Algorithm 1 to a dehomogenization of f(x), as shown in Examples 4 and 5 below.

Example 3: $f(x, y) = x^2 y^2 (x^2 + y^2 - 1)$. This polynomial is taken from [19]. It has global minimum value $f^* = -1/27 = -0.03703703703703...$ However, $f_{sos}^* = -33.157325$ is considerably smaller than f^* . If we minimize f(x) over its gradient ideal with N = 4, then we get $f_{4,grad}^* = -0.03703703706212$. The difference equals $f^* - f_{4,grad}^* \approx 2.50 \cdot 10^{-11}$. The solutions extracted by GloptiPoly ([15]) are ($\pm 0.5774, \pm 0.5774$).

Example 4: The polynomial $f(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$ is obtained from the Motzkin polynomial by substituting z = 1 as in [28]. We have $f^* = 0 > f_{sos}^* = -\infty$. However, if we minimize f(x, y) over its gradient ideal with N = 4, we get $f_{4,grad}^* = -6.1463 \cdot 10^{-10}$. The solutions extracted by GloptiPoly are $(\pm 1.0000, \pm 1.0000)$.

Example 5: The polynomial $f(x, y) = x^4 + x^2 + z^6 - 3x^2z^2$ is obtained from the Motzkin polynomial by substituting y = 1. Now, $f^* = 0 > f_{sos}^* = -729/4096$. However, if we minimize f(x, z) over its gradient ideal with N = 4, we get $f_{4,grad}^* = -9.5415 \cdot 10^{-12}$. The solutions extracted by GloptiPoly are (0.0000, 0.0000) and ($\pm 1.0000, \pm 1.0000$).

7. Conclusions

This paper proposes a method for minimizing a multivariate polynomial f(x) over its gradient variety. We assume that the infimum f^* is attained. This assumption is non-trivial, and we do not address the (important and difficult) question of how to verify that a given polynomial f(x) has this property. A sufficient condition for f(x) to attain its minimum can be derived from results of Marshall concerning stable compactness [21, Theorem 5.1].

Every polynomial which is strictly positive on its real gradient variety is SOS modulo its gradient ideal, even if the gradient variety is not zero-dimensional or radical. This fact implies that we can find a sequence of lower bounds $\{f_{N,grad}^*\}$ which converges to f^* . Moreover, if the gradient ideal is radical, we showed that every nonnegative polynomial is also SOS modulo its gradient ideal, which implies that $f_{N,grad}^* = f^*$ for some integer N. This finite convergence property holds for random polynomials by Proposition 1. Our method can also be generalized to constrained polynomial optimization. Instead of using gradients, we minimizing the objective polynomial over the variety defined by its KKT system. Similar results hold as in the unconstrained case.

Numerical experiments with SOSTOOLS suggest that our algorithm is effective for unconstrained polynomial optimization. Our method (11) with gradients is faster than the method (2) without gradients on the family of polynomials in Section 6.1. The reason for the speed-up is not clear yet, which might be a future research topic.

The method is also effective for equality constrained optimization, when the number of equality constraints are small compared with the number of decision variables. When there are many equality or inequality constraints, the structure of the KKT system must be exploited for computation efficiency.

Acknowledgement. The authors thank Claus Scheiderer for very helpful comments and for providing the example at the end of Section 2. Jean Lasserre and Pablo Parrilo suggested one example in Section 3. The authors are grateful to Monique Laurent for reviewing an early manuscript version and fruitful suggestions that helped improved this paper. The authors also thank two referees for their valulable comments which improves this paper. Jiawang Nie was supported in part by National Science Foundation (ELA-0122599). Bernd Sturmfels was supported in part by the National Science Foundation (DMS-0200729).

References

- 1. Basu, S., Pollack, R., Roy, M.-F.: Algorithms in Real Algebraic Geometry. Springer-Verlag, 2003
- Berg, C.: The multidimensional moment problem and semi-groups. In: Moments in Mathematics, H.J. Landau (ed.), AMS, Providence, RI, 1980, pp 110–124
- 3. Bochnak, J., Coste, M., Roy, M-F.: Real Algebraic Geometry, Springer, 1998
- 4. Blekherman, G.: There are significantly more nonnegative polynomials than sums of squares. To appear in Israel Journal of Mathematics
- Becker, E., Neuhaus, R.: Computation of real radicals of polynomial ideals. Computational algebraic geometry (Nice, 1992), 1–20, Progress in Mathematics, 109, Birkhäuser, Boston, MA, 1993
- Becker, E., Neuhaus, R.: Computation of real radicals of polynomial ideals. II. J. Pure Appl. Algebra 124, 261–280 (1998)
- Cox, D.A., Little, J.B., D.O'Shea.: Ideals, Varieties and Algorithms: an Introduction to Computational Algebraic Geometry and Commutative Algebra, Second Edition. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1997
- Cox, D.A., Little, J.B., D.O'Shea.: Using Algebraic Geometry, Graduate Texts in Mathematics, Vol. 185. Springer-Verlag, New York, 1998

- Corless, R.M., Gianni, P.M., Trager, B.M.: A reorder Schur factorization method for zero-dimensional polynomial systems with multiple roots. Proc. ACM Int. Symp. Symbolic and Algebraic Computation, Maui, Hawaii, 1997, pp 133–140
- Curto, R.E., Fialkow, L.A.: The truncated complex K-moment problem. Trans. Am. Math. Soc. 352, 2825–2855 (2000)
- Eisenbud, D.: Commutative Algebra with a View Toward Algebraic Geometry, Graduate Texts in Mathematics, Vol. 150. Springer-Verlag, New York, 1995
- Fortuna, E., Gianni, P., Trager, B.: Derivations and radicals of polynomial ideals over fields of arbitrary characteristic. Computer algebra (London, ON, 2001). J. Symbolic Comput. 33 (5), 609–625 (2002)
- Grigoriev, D., Vorobjov, N. N., Jr. Solving systems of polynomial inequalities in subexponential time. J. Symbolic Comput. 5 (1–2), 37–64 (1988)
- Hanzon, B., Jibetean, D.: Global minimization of a multivariate polynomial using matrix methods. J. Global Optimization 27, 1–23 (2003)
- Henrion, D., Lasserre, J. B.: GloptiPoly: Global Optimization over Polynomials with Matlab and SeDuMi. Proceeding of the 41st IEEE Conference on Decision and Control (CDC'2002), Las Vegas, Nevada, December 10–13, 2002, pp 747–752
- Henrion, D., Lasserre, J.: Detecting global optimality and extracting solutions in GloptiPoly. Positive polynomials in control, D. Henrion, A. Garulli Eds., Lecture Notes on Control and Information Sciences, Vol. 312, Springer, Berlin, 2005, pp 293–310
- 17. Jibetean, D., Laurent, M.: Converging SDP bounds for global unconstrained polynomial optimization. Preprint, 2004. Website: www.cwi.nl/~monique
- Krick, T., Logar, A.: An algorithm for the computation of the radical of an ideal in the ring of polynomials. Applied algebra, algebraic algorithms and error-correcting codes (New Orleans, LA, 1991), Lecture Notes in Comput. Sci., 539, Springer, Berlin, 1991, pp 195–205
- Lasserre, J.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. 11 (3), 796–817 (2001)
- 20. Laurent, M.: Semidefinite representations for finite varieties. To appear in Mathematical Programming. Website: www.cwi.nl/~monique
- 21. Marshall, M.: Optimization of polynomial functions. Canad. Math. Bull. 46, 575-587 (2003)
- 22. Nesterov, Y.: Squared functional systems and optimization problems. High Performance Optimization H. Frenk et al. (eds.), Kluwer Academic Publishers, 2000, pp 405–440
- 23. Nie, J., Demmel, J. W.: Minimum ellipsoid bounds for solutions of polynomial systems via sum of squares, to appear in J. Global Optimization, arXiv:math.OC/0411122
- Nie, J., Demmel, J. W., Powers, V.: Representations of positive polynomials on non-compact semialgebraic sets via KKT ideals. Preprint, 2005
- Nocedal, J., Wright, S. J.: Numerical Optimization, Springer Series in Operations Research, Springer-Verlag, New York, 1999
- 26. Parrilo, P.: Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization, Ph.D Thesis, California Institute of Technology, 2000
- Parrilo, P., Sturmfels, B.: Minimizing polynomial functions, Proceedings of the DIMACS Workshop on Algorithmic and Quantitative Aspects of Real Algebraic Geometry in Mathematics and Computer Science (March 2001), S. Basu, L. Gonzalez-Vega (eds.), American Mathematical Society, 2003, pp 83–100
- Parrilo, P.: Semidefinite Programming relaxations for semialgebraic problems. Math. Program. Ser. B 96 (2), 293–320 (2003)
- Prajna, S., Papachristodoulou, A., Parrilo, P.: SOSTOOLS User's Guide. http://control.ee.ethz.ch/~parrilo/SOSTOOLS/
- Parrilo, P.: An explicit construction of distinguished representations of polynomials nonnegative over finite sets, If A Technical Report AUT02-02, March 2002
- 31. Putinar, M.: Positive polynomials on compact semi-algebraic sets. Ind. Univ. Math. J. 42, 203–206 (1993)
- Renegar, J.: On the computational complexity and geometry of the first-order theory of the reals. I-III. J. Symbolic Comput. 13 (3), 255–352 (1992)
- Reznick, B.: Some concrete aspects of Hilbert's 17th problem. In Contemporary Mathematics, volume 253, American Mathematical Society, 2000, pp 251–272
- Shafarevich. Basic algebraic geometry. Die Grundlehren der mathematischen Wissenschaften. Band 213. Springer-Verlag, 1974
- 35. L. Vandenberghe and S. Boyd. Semidefinite Programming. SIAM Review 38, 49-95 (1996)