GLOBAL OPTIMIZATION OF POLYNOMIALS USING GRADIENT TENTACLES AND SUMS OF SQUARES

MARKUS SCHWEIGHOFER

Abstract. We consider the problem of computing the global infimum of a real polynomial f on $\mathbb{R}^n$. Every global minimizer of f lies on its gradient variety, i.e., the algebraic subset of $\mathbb{R}^n$ where the gradient of f vanishes. If f attains a minimum on $\mathbb{R}^n$, it is therefore equivalent to look for the greatest lower bound of f on its gradient variety. Nie, Demmel and Sturmfels proved recently a theorem about the existence of sums of squares certificates for such lower bounds. Based on these certificates, they find arbitrarily tight relaxations of the original problem that can be formulated as semidefinite programs and thus be solved efficiently.

We deal here with the more general case when f is bounded from below but does not necessarily attain a minimum. In this case, the method of Nie, Demmel and Sturmfels might yield completely wrong results. In order to overcome this problem, we replace the gradient variety by larger semialgebraic subsets of $\mathbb{R}^n$ which we call gradient tentacles. It now gets substantially harder to prove the existence of the necessary sums of squares certificates.

1. Introduction

Throughout this article, $\mathbb{N} := \{1, 2, \dots\}$, $\mathbb{R}$ and $\mathbb{C}$ denote the sets of natural, real and complex numbers, respectively. We fix $n \in \mathbb{N}$, and consider real polynomials in n variables $\bar{X} := (X_1, \dots, X_n)$. These polynomials form a commutative ring

$$\mathbb{R}[\bar{X}] := \mathbb{R}[X_1, \dots, X_n].$$

1.1. **The problem.** We consider the problem of computing good approximations for the global infimum

$$f^* := \inf\{f(x) \mid x \in \mathbb{R}^n\} \in \mathbb{R} \cup \{-\infty\}$$

of a polynomial $f \in \mathbb{R}[\bar{X}]$. Since f^* is the greatest lower bound of f, it is equivalent to compute

$$f^* = \sup\{a \in \mathbb{R} \mid f - a \geq 0 \text{ on } \mathbb{R}^n\} \in \mathbb{R} \cup \{-\infty\}. \tag{1}$$

To solve this hard problem, it has become a standard approach to approximate f^* by exchanging in (1) the nonnegativity constraint

$$f - a \geq 0 \text{ on } \mathbb{R}^n \tag{2}$$

Date: May 17, 2006.

1991 *Mathematics Subject Classification.* Primary 13J30, 90C26; Secondary 12Y05, 13P99, 14P10, 90C22.

Key words and phrases. global optimization, polynomial, preorder, sum of squares, semidefinite programming.

Supported by the DFG grant “Barrieren”.

by a computationally more feasible condition and analyze the error caused by this substitution. Typically, the choice of this replacement is related to the interplay between (globally) nonnegative polynomials, sums of squares of polynomials and semidefinite optimization (also called semidefinite programming):

1.2. **Method based on the fact that every sum of squares of polynomials is nonnegative (Shor** [Sho]**, Stetsyuk** [SS]**, Parrilo and Sturmfels** [PS] **et al.)** We start with the most basic ideas concerning these connections which can be found in greater detail in the just cited references. A first try is to replace condition (2) by the constraint

$$f - a \text{ is a sum of squares in the polynomial ring } \mathbb{R}[\bar{X}] \tag{3}$$

since every sum of squares in $\mathbb{R}[\bar{X}]$ is obviously nonnegative on $\mathbb{R}^n$.

The advantage of (3) over (2) is that sums of squares of polynomials can be nicely parametrized. Fix a column vector v whose entries are a basis of the vector space $\mathbb{R}[\bar{X}]_d$ of all real polynomials of degree $\leq d$ in n variables ($d \in \mathbb{N}_0 := \{0\} \cup \mathbb{N}$). This vector has a certain length $k = \dim \mathbb{R}[\bar{X}]_d$. It is easy to see that the map from the vector space $S\mathbb{R}^{k\times k}$ of symmetric $k \times k$-matrices to $\mathbb{R}[\bar{X}]_{2d}$ defined by $M \mapsto v^T M v$ is surjective. Using the spectral theorem for symmetric matrices, it is not hard to prove that a polynomial $f \in \mathbb{R}[\bar{X}]_{2d}$ is a sum of squares in $\mathbb{R}[\bar{X}]$ if and only if $f = v^T M v$ for some positive semidefinite matrix $M \in S\mathbb{R}^{k\times k}$. Use the following remark which is an easy exercise (write the polynomials as sums of their homogeneous parts).

Remark 1. In any representation $f = \sum_i g_i^2$ of a polynomial $f \in \mathbb{R}[\bar{X}]_{2d}$ as a sum of squares $g_i \in \mathbb{R}[\bar{X}]$, we have necessarily $\deg g_i \leq d$.

The described parametrization shows that the modified problem (where we exchange (2) by (3)), i.e., the problem to compute

$$f^{\text{sos}} := \sup\{a \in \mathbb{R} \mid f - a \text{ is a sum of squares in } \mathbb{R}[\bar{X}]\} \in \mathbb{R} \cup \{-\infty\} \tag{4}$$

can be written as a *semidefinite optimization problem* (also called semidefinite program or SDP for short), i.e., as the problem of minimizing (or maximizing) an affine linear function on the intersection of the cone of positive semidefinite matrices with an affine subspace in $S\mathbb{R}^{k\times k}$. For solving SDPs, there exist very good numerical algorithms, perhaps almost as good as for linear optimization problems. Linear optimization can be seen as the restriction of semidefinite optimization to diagonal matrices, i.e., a method to minimize an affine linear function on the intersection of the cone $\mathbb{R}^k_{\geq 0}$ with an affine subspace of $\mathbb{R}^k$. Speaking very vaguely, most concepts from linear optimization carry over to semidefinite optimization because every symmetric matrix can be diagonalized. We refer for example to [Tod] for an introduction to semidefinite programming.

Whereas computing f^* as defined in (1) is a very hard problem, it is relatively easy to compute (numerically to a given precision) f^{sos} defined in (4). Of course, the question arises how f^* and f^{sos} are related. Since (3) implies (2), it is clear that $f^{\text{sos}} \leq f^*$. The converse implication (and thus $f^{\text{sos}} = f^*$) holds in some cases: A globally nonnegative polynomial

- in one variable or
- of degree at most two or
- in two variables of degree at most four

is a sum of squares of polynomials. We refer to [Rez] for an overview of these and related old facts. However, recently Blekherman has shown in [Ble] that for fixed degree $d \geq 4$ and high number of variables n only a very small portion (in some reasonable sense) of the globally nonnegative polynomials of degree at most d in n variables are sums of squares. In particular, f^{sos} will often differ from f^*. For example, the *Motzkin polynomial*

$$M := X^2Y^2(X^2 + Y^2 - 3Z^2) + Z^6 \in \mathbb{R}[X, Y, Z]. \tag{5}$$

is nonnegative but not a sum of squares (see [Rez, PS]). We have $M^* = 0$ but $M^{\text{sos}} = -\infty$. The latter follows from the fact that M is homogeneous and not a sum of squares by the following remark applied to $f := M - a$ for $a \in \mathbb{R}$ (which can again be proved easily by considering homogeneous parts).

Remark 2. If f is a sum of squares in $\mathbb{R}[\bar{X}]$, then so is the highest homogeneous part (the leading form) of f.

We see that the basic problem with this method (computing f^{sos} by solving an SDP and hoping that f^{sos} is close to f^*) is that polynomials positive on $\mathbb{R}^n$ in general do not have a representation as a sum of squares, a fact that Hilbert already knew.

1.3. **The Positivstellensatz.** In the 17th of his famous of 23 problems, Hilbert asked whether every (globally) nonnegative (real) polynomial (in several variables) was a sum of squares of *rational functions*. Artin answered this question affirmatively in 1926 and today there exist numerous refinements of his solution. One of them is the *Positivstellensatz* (in analogy to Hilbert's Nullstellensatz). It is often attributed to Stengle [Ste] who clearly deserves credit for finding it independently and making it widely known. However, Prestel [PD, Section 4.7] recently discovered that Krivine [Kri] knew the result about ten years earlier in 1964. Here we state only the following special case of the Positivstellensatz.

Theorem 3 (Krivine)**.** *For every $f \in \mathbb{R}[\bar{X}]$, the following are equivalent.*

(i) $f > 0$ on $\mathbb{R}^n$
(ii) There are sums of squares s and t in $\mathbb{R}[\bar{X}]$ such that $sf = 1 + t$.

By this theorem, we have of course that f^* is the supremum over all $a \in \mathbb{R}$ such that there are sums of squares $s, t \in \mathbb{R}[\bar{X}]$ with $s(f - a) = 1 + t$. When one tries to write this as an SDP there are two obstacles.

First, each SDP involves matrices of a fixed (finite) size. But with matrices of a fixed size, we can only parametrize sums of squares up to a certain degree. We need therefore to impose a degree restriction on s and t. There are no (at least up to now) *practically relevant* degree bounds that could guarantee that such a restriction would not affect the result. We refer to the tremendous work [Scd] of Schmid on degree bounds. This first obstacle, namely the question of degrees of the sums of squares, will us accompany throughout the article. The answer will always be to model the problem not as a single SDP but as a whole *sequence* of SDPs, each SDP corresponding to a certain degree restriction. As you solve one SDP after the other, the degree restriction gets less restrictive and you hope for fast convergence of the optimal values of the SDPs to f^*. For newcomers in the field, it seems at first glance unsatisfactory having to deal with a whole sequence of SDPs rather than a single SDP. But after all, it is only natural that a very hard problem cannot be modeled by an SDP of a reasonable size so that you have to look for good *relaxations* of the problem which can easier be dealt with and to which the techniques of mathematical optimization can be applied.

The second obstacle is much more severe. It is the fact that the unknown polynomial $s \in \mathbb{R}[\bar{X}]$ is multiplied with the unknown $a \in \mathbb{R}$ on the left hand side of the constraint $s(f-a) = 1+t$. This makes the formulation as an SDP (even after having imposed a restriction on the degree of s and t) impossible (or at least highly non-obvious). Of course, if you *fix* $a \in \mathbb{R}$ and a degree bound $2d$ for s and t, then the question whether there exist sums of squares s and t of degree at most $2d$ such that $s(f-a) = 1+t$ is equivalent to the feasibility of an SDP. But this plays (at least currently) only a role as a criterion that *might help to decide* whether a certain fixed (or guessed) $a \in \mathbb{R}$ is a strict lower bound of f. We refer to [PS] for more details. What one needs are representation theorems for positive polynomials that are better suited for optimization than the Positivstellensatz (even if they are sometimes less aesthetic).

1.4. **"Big ball" method proposed by Lasserre** [L1]**.** In the last 15 years, a lot of progress has been made in proving existence of sums of squares certificates which can be exploited for optimization (although most of the new results were obtained without having in mind the application in optimization which has been established more recently). The first breakthrough was perhaps Schmüdgen's theorem [Sch, Corollary 3] all of whose proofs use the Positivstellensatz. In this article, we will prove a generalization of Schmüdgen's theorem, namely Theorem 9 below. In [L1], Lasserre uses the following special case of Schmüdgen's theorem which has already been proved by Cassier [Cas, Théorème 4] and which can even be derived easily from [Kri, Théorème 12].

Theorem 4 (Cassier)**.** *For $f \in \mathbb{R}[\bar{X}]$ and $R \geq 0$, the following are equivalent.*

(i) $f \geq 0$ on the closed ball centered at the origin of radius R
(ii) For all $\varepsilon > 0$, there are sums of squares s and t in $\mathbb{R}[\bar{X}]$ such that

$$f + \varepsilon = s + t(R^2 - \|\bar{X}\|^2).$$

Here and in the following, we use the notation

$$\|\bar{X}\|^2 := X_1^2 + \cdots + X_n^2 \in \mathbb{R}[\bar{X}].$$

Similar to Subsection 1.2, it can be seen that for any fixed $d \in \mathbb{N}_0$, computing the supremum over all $a \in \mathbb{R}$ such that $f-a = s+t(R^2-\|\bar{X}\|^2)$ for some sums of squares $s, t \in \mathbb{R}[\bar{X}]$ of degree at most $2d$ amounts to solving an SDP. Therefore you get a sequence of SDPs parametrized by $d \in \mathbb{N}_0$. Theorem 4 can now be interpreted as a convergence result, namely the sequence of optimal values of these SDPs converges to the minimum of f on the closed ball around the origin with radius R. If one has a polynomial $f \in \mathbb{R}[\bar{X}]$ attaining a minimum on $\mathbb{R}^n$ and for which one knows moreover a big ball on which this minimum is attained, this method is good for computing f^*. Of course, if you do not know such a big ball in advance you might choose larger and larger R. But at the same time you might have to choose a bigger and bigger degree restriction $d \in \mathbb{N}_0$ and it is not really clear how to get a sequence of SDPs that converges to f^*.

1.5. **Lasserre's high order perturbation method** [L2]**.** Recently, Lasserre used in [L2] a theorem of Nussbaum from operator theory to prove the following result that can be exploited in a similar way for global optimization of polynomials.

Theorem 5 (Lasserre)**.** *For every $f \in \mathbb{R}[\bar{X}]$, the following are equivalent:*

(i) $f \geq 0$ *on* $\mathbb{R}^n$

(ii) For all $\varepsilon > 0$*, there is* $r \in \mathbb{N}_0$ *such that*

$$f + \varepsilon \sum_{i=1}^{n} \sum_{k=0}^{r} \frac{X_i^{2k}}{k!} \text{ is a sum of squares in } \mathbb{R}[\bar{X}].$$

Note that (ii) implies that $f(x) + \varepsilon \sum_{i=1}^{n} \exp(x_i) \geq 0$ for all $x \in \mathbb{R}^n$ and $\varepsilon > 0$ which in turn implies (i). In condition (ii), r depends on ε and f. Using real algebra and model theory, Netzer showed that in fact r depends only on ε, n, the degree of f and a bound on the size of the coefficients of f [Net, LN].

1.6. **"Gradient perturbation" method proposed by Jibetean and Laurent** [JL]**.** The most standard idea for finding the minimum of a function everybody knows from calculus is to compute critical points, i.e., the points where the gradient vanishes. It is a natural question whether the power of classical differential calculus can be combined with the relatively new ideas using sums of squares. Fortunately, it can and the rest of the article will be about how to merge both concepts, sums of squares and differential calculus.

If a polynomial $f \in \mathbb{R}[\bar{X}]$ attains a minimum in $x \in \mathbb{R}^n$, i.e., $f(x) \leq f(y)$ for all $y \in \mathbb{R}^n$, then the gradient ∇f of f vanishes at x, i.e., $\nabla f(x) = 0$. However, there are polynomials that are bounded from below on $\mathbb{R}^n$ and yet do not attain a minimum on $\mathbb{R}^n$. The simplest example is perhaps

$$f := (1 - XY)^2 + Y^2 \in \mathbb{R}[X, Y] \tag{6}$$

for which we have $f > 0$ on $\mathbb{R}^n$ but $f^* = 0$ since $\lim_{x\to\infty} f(x, \frac{1}{x}) = 0$. In the following,

$$(\nabla f) := \left(\frac{\partial f}{\partial X_1}, \ldots, \frac{\partial f}{\partial X_n}\right) \subseteq \mathbb{R}[\bar{X}]$$

denotes the ideal generated by the partial derivatives of f in $\mathbb{R}[\bar{X}]$. We call this ideal the *gradient ideal* of f.

Without going into details, the basic idea of Jibetean and Laurent in [JL] is again to apply a perturbation to f. Instead of adding a truncated exponential like Lasserre, they just add $\varepsilon \sum_{i=1}^{n} X_i^{2(d+1)}$ for small $\varepsilon > 0$ when $\deg f = 2d$. If $f > 0$ on $\mathbb{R}^n$, then the perturbed polynomial $f_\varepsilon := f + \varepsilon \|\bar{X}\|^{2(d+1)}$ is again a sum of squares but this time only modulo its gradient ideal (∇f_ε). In this case, this is quite easy to prove since it turns out that this ideal will be zero-dimensional, i.e., $\mathbb{R}[\bar{X}]/(\nabla f_\varepsilon)$ is a finite-dimensional real algebra. We will later see in Theorems 6 and 46 that this finite-dimensionality is not needed for the sums of squares representation. But the work of Jibetean and Laurent exploits the finite-dimensionality in many ways. We refer to [JL] for details.

1.7. **"Gradient variety" method by Nie, Demmel and Sturmfels** [NDS]**.** The two perturbation methods just sketched rely on introducing very small coefficients in a polynomial. This small coefficients might lead to SDPs which are hard to solve because of numerical instability. It is therefore natural to think of another method which avoids perturbation at all. Nie, Demmels and Sturmfels considered, for a polynomial $f \in \mathbb{R}[\bar{X}]$, its *gradient variety*

$$V(\nabla f) := \{x \in \mathbb{C}^n \mid \nabla f(x) = 0\}.$$

This is the algebraic variety corresponding to the radical of the gradient ideal (∇f). It can be shown that a polynomial $f \in \mathbb{R}[\bar{X}]$ is constant on each irreducible component of the gradient variety (see [NDS] or use an unpublished algebraic argument of Scheiderer based on Kähler differentials). This is the key to show that a polynomial $f \in \mathbb{R}[\bar{X}]$ nonnegative on its gradient variety is a sum of squares modulo its gradient ideal in the case where the ideal is radical. In the general case where the gradient ideal is not necessarily radical, the same thing still holds for polynomials *positive* on their gradient variety. The following is essentially [NDS, Theorem 9] (confer also the recent work [M2]). We will later prove a generalization of this theorem as a byproduct. See Corollary 47 below.

Theorem 6 (Nie, Demmel and Sturmfels)**.** *For every $f \in \mathbb{R}[\bar{X}]$ attaining a minimum on $\mathbb{R}^n$, the following are equivalent.*

(i) $f \geq 0$ *on* $\mathbb{R}^n$

(ii) $f \geq 0$ *on* $V(\nabla f) \cap \mathbb{R}^n$

(iii) For all $\varepsilon > 0$, there exists a sum of squares s in $\mathbb{R}[\bar{X}]$ such that

$$f + \varepsilon \in s + (\nabla f).$$

Moreover, (ii) and (iii) are equivalent for all $f \in \mathbb{R}[\bar{X}]$.

For each degree restriction $d \in \mathbb{N}_0$, the problem of computing the supremum over all $a \in \mathbb{R}$ such that

$$f - a = s + p_1 \frac{\partial f}{\partial X_1} + \cdots + p_n \frac{\partial f}{\partial X_n}$$

for some sum of squares s in $\mathbb{R}[\bar{X}]$ and polynomials $p_1, \ldots, p_n$ of degree at most d, can be expressed as an SDP. Theorem 6 shows that the optimal values of the corresponding sequence of SDPs (indexed by d) tend to f^* provided that f attains a minimum on $\mathbb{R}^n$. However, if f does not attain a minimum on $\mathbb{R}^n$, the computed sequence still tends to the infimum of f on its gradient variety which might however now be very different from f^*. Take for example the polynomial f from (6). It is easy to see that $V(\nabla f) = \{0\}$ and therefore the method computes $f(0) = 1$ instead of $f^* = 0$. In [NDS, Section 7], the authors write:

> "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."

1.8. **Our "gradient tentacle" method.** The reason why the method just described might fail is that the global infimum of a polynomial $f \in \mathbb{R}[\bar{X}]$ is not always a *critical value* of f, i.e., a value that f takes on at least one of its critical points in $\mathbb{R}^n$. Now there is a well-established notion of *generalized critical values* which includes also the *asymptotic critical values* (a kind of critical values at infinity we will introduce in Definition 12 below).

In this article, we will replace the real part $V(\nabla f) \cap \mathbb{R}^n$ of the gradient variety by several larger semialgebraic sets on which the partial derivatives do not necessarily vanish but get very small far away from the origin. These semialgebraic sets often look like tentacles, and that is how we will call them. All tentacles we will consider are defined by a single polynomial inequality that depends only on the polynomial

$$\|\nabla f\|^2 := \left(\frac{\partial f}{\partial X_1}\right)^2 + \cdots + \left(\frac{\partial f}{\partial X_n}\right)^2$$

and expresses that this polynomial gets very small. Given a polynomial f for which you want to compute f^*, the game will consist in finding a tentacle such that two things will hold at the same time:

- There exist suitable sums of squares certificates for nonnegativity on the tentacle.
- The infimum of f on $\mathbb{R}^n$ and on the tentacle coincide.

One can imagine that these two properties are hardly compatible. Taking $\mathbb{R}^n$ as a tentacle, would of course ensure the second condition but we have discussed in Subsection 1.2 that the first one would be badly violated. The other extreme would be to take the empty set as a tentacle. Then the first condition would trivially be satisfied whereas the second would fail badly. How we will roughly be able to find the balancing act between the two requirements is as follows: The second condition will be satisfied by known non-trivial theorems about asymptotic behaviour of polynomials at infinity. The existence of suitable sums of squares certificates will be based on the author's (real) algebraic work [Sr1] on iterated rings of bounded elements (also called real holomorphy rings).

1.9. **Contents of the article.** The article is organized as follows. In Section 2, we prove a general sums of squares representation theorem which generalizes Schmüdgen's theorem we have mentioned in Subsection 1.4. This representation theorem is interesting in itself and will be used in the subsequent sections. In Section 3, we introduce a gradient tentacle (see Definition 17) which is defined by the polynomial inequality

$$\|\nabla f\|^2 \|\bar{X}\|^2 \leq 1.$$

We call this gradient tentacle *principal* since we can prove that it does the job in a large number of cases (see Theorem 25) and there is hope that it works in fact for all polynomials $f \in \mathbb{R}[\bar{X}]$ bounded from below. Indeed, we have not found any counterexamples (see Open Problem 33). In case this hope were disappointed, we present in Section 4 a collection of other gradient tentacles (see Definition 41) defined by the polynomial inequalities

$$\|\nabla f\|^{2N}(1 + \|\bar{X}\|^2)^{N+1} \leq 1 \qquad (N \in \mathbb{N}).$$

Their advantage is that if $f \in \mathbb{R}[\bar{X}]$ is bounded from below and N is large enough for this particular f, then we can prove that the corresponding tentacle does the job (see Theorems 46 and 50). We call these tentacles higher gradient tentacles since the degree of the defining inequality gets unfortunately high when N gets big which has certainly negative consequences for the complexity of solving the SDPs arising from these tentacles. However, if f attains a minimum on $\mathbb{R}^n$, then any choice of $N \in \mathbb{N}$ will be good. Conclusions are drawn in Section 5.

2. The sums of squares representation

In this section, we prove the important sums of squares representation theorem we will need in the following sections. It is a generalization of Schmüdgen's Positivstellensatz (see [PD, Sch]) which is also of independent interest. Schmüdgen's result is not to confuse with the (classical) Positivstellensatz we described in the introduction. The connection between the two is that all known proofs of Schmüdgen's result use the classical Positivstellensatz. Our result, Theorem 9 below, is much harder to prove than Schmüdgen's result. Its proof relies on the theory of iterated *rings of bounded elements* (also called real holomorphy rings) described in [Sr1].

Definition 7. For any polynomial $f \in \mathbb{R}[\bar{X}]$ and subset $S \subseteq \mathbb{R}^n$, the set $R_\infty(f, S)$ of *asymptotic values* of f on S consists of all $y \in \mathbb{R}$ for which there exists a sequence $(x_k)_{k\in\mathbb{N}}$ of points $x_k \in S$ such that

$$\lim_{k\to\infty} \|x_k\| = \infty \qquad \text{and} \qquad \lim_{k\to\infty} f(x_k) = y. \tag{7}$$

We now recall the important notion of a preordering of a commutative ring. Except in the proof of Theorem 9, we need this concept only for the ring $\mathbb{R}[\bar{X}]$.

Definition 8. Let A be a commutative ring (with 1). A subset $T \subseteq A$ is called a *preordering* if it contains all squares f^2 of elements $f \in A$ and is closed under addition and multiplication. The preordering *generated* by $g_1, \dots, g_m \in A$

$$T(g_1, \dots, g_m) = \left\{ \sum_{\delta\in\{0,1\}^m} s_\delta g_1^{\delta_1} \dots g_m^{\delta_m} \mid s_\delta \text{ is a sum of squares in } A \right\} \tag{8}$$

is by definition the smallest preordering containing $g_1, \dots, g_m$.

If $g_1, \dots, g_m \in \mathbb{R}[\bar{X}]$ are polynomials, then the elements of $T(g_1, \dots, g_m)$ have obviously the geometric property that they are nonnegative on the (basic closed semialgebraic) set S they define by (9) below. The next theorem is a partial converse. Namely, if a polynomial satisfies on S some stronger geometric condition, then it lies necessarily in $T(g_1, \dots, g_m)$. In case that S is compact, the conditions (a) and (b) below are empty and the theorem is Schmüdgen's Positivstellensatz (see [PD, Sch]). The more general version we need here is quite hard to prove.

Theorem 9. *Let $f, g_1, \dots, g_m \in \mathbb{R}[\bar{X}]$ and set*

$$S := \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \dots, g_m(x) \geq 0\}. \tag{9}$$

Suppose that

(a) f is bounded on S,

(b) f has only finitely many asymptotic values on S and all of these are positive, i.e., $R_\infty(f, S)$ is a finite subset of $\mathbb{R}_{>0}$, and

(c) $f > 0$ on S.

Then $f \in T(g_1, \dots, g_m)$.

Proof. Write $R_\infty(f, S) = \{y_1, \dots, y_s\} \subseteq \mathbb{R}_{>0}$ and consider the polynomial

$$h := \prod_{i=1}^{s} (f - y_i).$$

This polynomial is "on S small at infinity" by which we mean that for every $\varepsilon > 0$ there exists $k \in \mathbb{N}$ such that for all $x \in S$ with $\|x\| \geq k$, we have $|h(x)| < \varepsilon$.

To show this, assume the contrary. Then there exists $\varepsilon > 0$ and a sequence $(x_k)_{k\in\mathbb{N}}$ of points $x_k \in S$ with $\lim_{k\to\infty} \|x_k\| = \infty$ and

$$|h(x_k)| \geq \varepsilon \qquad \text{for all } k \in \mathbb{N}. \tag{10}$$

Because the sequence $(f(x_k))_{k\in\mathbb{N}}$ is bounded by hypothesis (a), we find an infinite subset $I \subseteq \mathbb{N}$ such that the subsequence $(f(x_k))_{k\in I}$ converges. The limit must be one of the asymptotic values of f on S, i.e., $\lim_{k\in I,k\to\infty} f(x_k) = y_i$ for some $i \in \{1,\dots,s\}$. Using (a), it follows that $\lim_{k\in I,k\to\infty} h(x_k) = 0$ contradicting (10).

Let $A := (\mathbb{R}[\bar{X}], T)$ where $T := T(g_1,\dots,g_m)$. The set

$$H'(A) := \{p \in \mathbb{R}[\bar{X}] \mid N \pm p \in T \text{ for some } N \in \mathbb{N}\}$$

is a subring of A (see, e.g, [Sr1, Definition 1.2]). We endow $H'(A)$ with the preordering $T' := T \cap H'(A)$ and consider it as also as a preordered ring. By [Sr1, Corollary 3.7], the smallness of h at infinity proved above is equivalent to $h \in S_\infty(A)$ in the notation of [Sr1]. By [Sr1, Corollary 4.17], we have $S_\infty(A) \subseteq H'(A)$ and consequently $h \in H'(A)$. The advantage of $H'(A)$ over A is that its preordering is archimedean, i.e., $T' + \mathbb{Z} = H'(A)$. According to an old criterion for an element to be contained in an archimedean preordering (see for example [PD, Proposition 5.2.3 and Lemma 5.2.7] or [Sr1, Theorem 1.3]), our claim $f \in T'$ follows if we can show that $\varphi(f) > 0$ for all ring homomorphisms $\varphi : H'(A) \to \mathbb{R}$ with $\varphi(T') \subseteq \mathbb{R}_{\geq 0}$. For all such homomorphisms possessing an extension $\bar{\varphi} : A \to \mathbb{R}$ with $\bar{\varphi}(T) \subseteq \mathbb{R}_{\geq 0}$, this follows from hypothesis (c) because it is easy to see that such an extension $\bar{\varphi}$ must be evaluation $p \mapsto p(x)$ in the point $x := (\bar{\varphi}(X_1),\dots,\bar{\varphi}(X_n)) \in S$. Using the theory in [Sr1], we will see that the only possibility for such a φ not to have such an extension $\bar{\varphi}$ is that $\varphi(h) = 0$. Then we will be done since $\varphi(h) = 0$ implies $\varphi(f) = y_i > 0$ for some i. We have used here that $f \in H'(A)$ which follows from $h \in H'(A)$ since $H'(A)$ is integrally closed in A (see [Sr1, Theorem 5.3]).

So let us now use [Sr1]. By [Sr1, Corollary 3.7 and Theorem 4.18], the smallness of h at infinity means that

$$A_h = H'(A)_h$$

where we deal on both sides of this equation with the localization of a preordered ring by the element h (see [Sr1, pages 24 and 25]). If $\varphi : H'(A) \to \mathbb{R}$ is a ring homomorphism with $\varphi(T') \subseteq \mathbb{R}_{\geq 0}$ and $\varphi(h) \neq 0$, then φ extends to a ring homomorphism $\tilde{\varphi} : A_h = H'(A)_h \to \mathbb{R}$ with $\tilde{\varphi}(T_h) = \tilde{\varphi}(T'_h) \subseteq \mathbb{R}_{\geq 0}$. Then $\bar{\varphi} := \tilde{\varphi}|_A$ is the desired extension of φ. □

Example 10. Consider the polynomials

$$h_N := 1 - Y^N(1+X)^{N+1} \in \mathbb{R}[X,Y] \qquad (N \in \mathbb{N}) \tag{11}$$

in two variables. We fix $N \in \mathbb{N}$ and apply Theorem 9 with $f = h_{N+1}$, $m = 3$, $g_1 = X$, $g_2 = Y$ und $g_3 = h_N$. The set S defined by the g_i as in (9) is a subset of the first quadrant which is bounded in Y-direction but unbounded in X-direction. Of course, we have $0 \leq h_N \leq 1$ and

$$0 \leq Y(1+X) \leq \frac{1}{\sqrt[N]{1+X}} \qquad \text{on } S$$

showing that 0 is the only asymptotic value of

$$1 - h_{N+1} = (1-h_N)Y(1+X)$$

on S and therefore $R_\infty(h_{N+1}, S) = \{1\}$. It follows also that $0 \leq h_{N+1} \leq 1$ on S. By Theorem 9, we obtain

$$h_{N+1} + \varepsilon \in T(X, Y, h_N) \tag{12}$$

for all $\varepsilon > 0$.

The following lemma shows that (12) holds even for $\varepsilon = 0$, a fact that does not follow from Theorem 9. This lemma will be interesting later to compare the quality of certain SDP relaxations (see Proposition 49). In its proof, we will explicitly construct a representation of h_{N+1} as an element of $T(X, Y, h_N)$. Only part of this explicit representation will be needed in the sequel, namely an explicit polynomial $g \in T(X, Y)$ such that $h_{N+1} \in T(X, Y) + gh_N \subseteq T(X, Y, h_N)$. This explains the formulation of the statement. Theorem 9 will not be used in the proof but gave us good hope before we had the proof. The role of Theorem 9 in this article is above all to prove Theorems 25 and 46 below.

Lemma 11. *For the polynomials h_N defined by (11), we have*

$$h_{N+1} - \left(1 + \frac{1}{N}\right) Y(1+X)h_N \in T(X, Y).$$

Proof. For a new variable Z,

$$\begin{aligned}
(Z-1)^2 \sum_{k=0}^{N-1} (N-k)Z^k &= (Z-1)^2 \left(N \sum_{k=0}^{N-1} Z^k - Z \sum_{k=1}^{N-1} kZ^{k-1} \right) \\
&= (Z-1)^2 \left(N \frac{Z^N - 1}{Z-1} - Z \frac{\partial}{\partial Z} \left(\frac{Z^N - 1}{Z-1} \right) \right) \\
&= N(Z-1)(Z^N - 1) - Z((Z-1)NZ^{N-1} - (Z^N - 1)) \\
&= Z^{N+1} - (N+1)Z + N.
\end{aligned}$$

Specializing Z to $z := Y(1+X)$, we have therefore

$$\begin{aligned}
Nh_{N+1} - (N+1)zh_N &= N(1 - z^{N+1}(1+X)) - (N+1)z(1 - z^N(1+X)) \\
&= z^{N+1}X + (z^{N+1} - (N+1)z + N) \\
&= z^{N+1}X + (z-1)^2 \sum_{k=0}^{N-1} (N-k)z^k \in T(X, Y).
\end{aligned}$$

Dividing by $N = (\sqrt{N})^2$ yields our claim. □

3. The principal gradient tentacle

In this section, we associate to every polynomial $f \in \mathbb{R}[\bar{X}]$ a gradient tentacle which is a subset of $\mathbb{R}^n$ containing the real part of the gradient variety of f and defined by a single polynomial inequality whose degree is not more than twice the degree of f. The infimum of any polynomial $f \in \mathbb{R}[\bar{X}]$ bounded from below on $\mathbb{R}^n$ will coincide with the infimum on its principal gradient tentacle (see Theorem 19). Under some technical assumption (see Definition 20) which is not known to be necessary (see Open Problem 33), we prove a sums of squares certificate for non-negativity of f on its principal gradient tentacle which is suitable for optimization purposes. This representation theorem (Theorem 25) is of independent interest and its proof is mainly based on the nontrivial representation theorem from the previous section and a result of Parusiński on the behaviour of polynomials at infinity ([P1, Theorem 1.4]). In Subsection 3.2, we outline how to get a sequence of SDPs growing in size whose optimal values tend to f^* for any f satisfying the conditions of Theorem 25 (or perhaps for any f with $f^* > -\infty$ if the answer to Open Problem 33 is yes). In the Subsections 3.3 and 3.4, we give a MATLAB code for the sums of squares optimization toolboxes YALMIP [Löf] and SOSTOOLS [PPS] that produces and solves these SDP relaxations. This short and simple code is meant for readers who have little experience with such toolboxes and want nevertheless try our proposed method on their own. In Subsection 3.5, we provide simple examples which have been calculated using the YALMIP code from Subsection 3.3.

We start by recalling the concept of asymptotic critical values developed by Rabier in his 1997 milestone paper [Rab]. For simplicity, we stay in the setting of real polynomials right from the beginning (though part of this theory make sense in a much broader context).

Definition 12. Suppose $f \in \mathbb{R}[\bar{X}]$. The set $K_0(f)$ of *critical values* of f consists of all $y \in \mathbb{R}$ for which there exists $x \in \mathbb{R}^n$ such that $\nabla f(x) = 0$ and $f(x) = y$. The set $K(f)$ of *generalized critical values* of f consists of all $y \in \mathbb{R}$ for which there exists a sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ such that

$$\lim_{k\to\infty} \|\nabla f(x_k)\|(1 + \|x_k\|) = 0 \qquad \text{and} \qquad \lim_{k\to\infty} f(x_k) = y. \tag{13}$$

The set $K_\infty(f)$ of *asymptotic critical values* consists of all $y \in \mathbb{R}$ for which there exists a sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ such that $\lim_{k\to\infty} \|x_k\| = \infty$ and (13) hold.

The following proposition is easy.

Proposition 13. *The set of generalized critical values of a polynomial $f \in \mathbb{R}[\bar{X}]$ is the union of its set of critical and asymptotic critical values, i.e.,*

$$K(f) = K_0(f) \cup K_\infty(f).$$

The following notions go back to Thom [Tho].

Definition 14. Suppose $f \in \mathbb{R}[\bar{X}]$. We say that $y \in \mathbb{R}$ is a *typical value* of f if there is neighbourhood U of y in $\mathbb{R}$ and a smooth (i.e., C^∞) manifold F such that $f|_{f^{-1}(U)} : f^{-1}(U) \to U$ is a (not necessarily surjective) trivial smooth fiber bundle, i.e., there exists a smooth manifold F and a C^∞ diffeomorphism $\Phi : f^{-1}(U) \to F \times U$ such that $f|_{f^{-1}(U)} = \pi_2 \circ \Phi$ where $\pi_2 : F \times U \to U$ is the canonical projection. We call $y \in \mathbb{R}$ an *atypical value* of f if it is not a typical value of f. The set of all atypical values of f is denoted by $B(f)$ and called the *bifurcation set* of f.

Note that a Φ like in the above definition induces a C^∞ diffeomorphism $f^{-1}(y) \to F \times \{y\} \cong F$ for every $y \in U$. In this context, the preimages $f^{-1}(y)$ are called fibers and F is called *the* fiber. We do not require that the fiber bundle $f|_{f^{-1}(U)} : f^{-1}(U) \to U$ is surjective (if it is not then the image is necessarily empty). Hence the fiber F may be empty and a *typical value* is not necessarily a value taken on by f. We make use the following well-known theorem (see, e.g., [KOS, Theorem 3.1]).

Theorem 15. *Suppose $f \in \mathbb{R}[\bar{X}]$. Then $B(f) \subseteq K(f)$ and $K(f)$ is finite.*

The advantage of $K(f)$ over $K_0(f)$ is that $f^* \in K(f)$ even if f does not attain a minimum on $\mathbb{R}^n$. This is an easy consequence of Theorem 15. See Theorem 19 below.

Example 16. Consider again the polynomial $f = (1-XY)^2+Y^2 \in R[X,Y]$ from (6) that does not attain its infimum $f^* = 0$ on $\mathbb{R}^2$. Calculating the partial derivatives, it is easy to see that the origin is the only critical point of f. Because f takes the value 1 at the origin, we have $K_0(f) = \{1\}$ and therefore $f^* = 0 \notin K_0(f)$. Clearly, we have $0 \in B(f)$ since $f^{-1}(-y) = \emptyset \neq f^{-1}(y)$ for small $y \in \mathbb{R}_{>0}$. By Theorem 15, we have therefore $0 \in K_\infty(f) \subseteq K(f)$. To show this directly, a first guess would be that $\|\nabla f(x, \frac{1}{x})\|(1 + \|(x, \frac{1}{x})\|)$ tends to zero when $x \to \infty$ because $\lim_{x\to\infty} f(x, \frac{1}{x}) = 0$. But in fact, this expressions tends to 2 when $x \to \infty$. However, a calculation shows that $\lim_{x\to\infty} \|\nabla f(x, \frac{1}{x})\|(1 + \|(x, \frac{1}{x} - \frac{1}{x^3})\|) = 0$.

Definition 17. For a polynomial $f \in \mathbb{R}[\bar{X}]$, we call

$$S(\nabla f) := \{x \in \mathbb{R}^n \mid \|\nabla f(x)\|\|x\| \leq 1\}$$

the *principal gradient tentacle* of f.

Remark 18. In the definition of $S(\nabla f)$, the inequality $\|\nabla f(x)\|\|x\| \leq 1$ could be exchanged by $\|\nabla f(x)\|\|x\| \leq R$ for some constant $R > 0$. Then all subsequent results will still hold with obvious modifications. Using an R different from 1 might have in certain cases a practical advantage (see Subsection 3.6 below). However, we decided to stay with this definition in order to get not too technical and to keep the paper readable.

As expressed by the notation $S(\nabla f)$, polynomials f with the same gradient ∇f have the same gradient tentacle, in other words

$$S(\nabla(f + a)) = S(\nabla f) \qquad \text{for all } a \in \mathbb{R}.$$

The first important property of $S(\nabla f)$ is stated in the following immediate consequence of Theorem 15.

Theorem 19. *Suppose $f \in \mathbb{R}[\bar{X}]$ is bounded from below. Then $f^* \in K(f)$ and therefore $f^* = \inf\{f(x) \mid x \in S(\nabla f)\}$.*

Proof. By Theorem 15, it suffices to show that $f^* \in B(f)$. Assume that $f^* \notin B(f)$, i.e., f^* is a typical value of f. Then for all y in a neighbourhood of f^*, the fibers $f^{-1}(y)$ are smoothly diffeomorphic to each other. But this is absurd since $f^{-1}(y)$ is empty for $y < f^*$ but certainly not empty in a neighbourhood of f^*. □

Let $\mathbb{P}^{n-1}(\mathbb{C})$ denote the $(n-1)$-dimensional complex projective space over $\mathbb{C}$. For a homogeneous polynomial f and a point $z \in \mathbb{P}^{n-1}(\mathbb{C})$, we simply say $f(z) = 0$ to express that f vanishes on (a non-zero point of) the straight line $z \subseteq \mathbb{C}^n$. Following [P1], we give the following definition.

Definition 20. We say that a polynomial $f \in \mathbb{C}[\bar{X}]$ *has only isolated singularities at infinity* if $f \in \mathbb{C}$ (i.e., f is constant) or $d := \deg f \geq 1$ and there are only finitely many $z \in \mathbb{P}^{n-1}(\mathbb{C})$ such that

$$\frac{\partial f_d}{\partial X_1}(z) = \cdots = \frac{\partial f_d}{\partial X_n}(z) = f_{d-1}(z) = 0 \tag{14}$$

where $f = \sum_i f_i$ and each $f_i \in \mathbb{C}[\bar{X}]$ is zero or homogeneous of degree i.

As shown in [P1, Section 1.1], the geometric interpretation of the above definition is that the projective closure of a generic fiber of f has only isolated singularities.

Remark 21. A generic complex polynomial has only isolated singularities at infinity. In fact, much more is true: A generic polynomial $f \in \mathbb{C}[\bar{X}]$ of degree $d \geq 1$ has *no* isolated singularities at infinity in the sense that there is no $z \in \mathbb{P}^{n-1}(\mathbb{C})$ such that (14) holds. In more precise words, to every $d \geq 2$, there exists a complex polynomial relation that is valid for all coefficient tuples of polynomials $f \in \mathbb{C}[\bar{X}]$ of degree d for which (14) has an infinite number of solutions. This follows from the fact that for a generic homogeneous polynomial $g \in \mathbb{C}[\bar{X}]$ of degree $d \geq 1$, there are only finitely many points $z \in \mathbb{P}^{n-1}(\mathbb{C})$ such that $\frac{\partial f}{\partial X_i}(z) = 0$ for all i. See [Kus, Théorème II] or [Shu, Proposition 1.1.1].

Remark 22. In the two variable case $n = 2$, every polynomial $f \in \mathbb{C}[\bar{X}]$ has only isolated singularities at infinity. This is clear since (14) defines an algebraic subvariety of $\mathbb{P}^1(\mathbb{C})$.

The following theorem follows easily from [P1, Theorem 1.4].

Theorem 23. *Suppose $f \in \mathbb{R}[\bar{X}]$ has only isolated singularities at infinity. Then*

$$R_\infty(f, S(\nabla f)) \subseteq K(f).$$

In particular, $R_\infty(f, S(\nabla f))$ is finite, i.e., f has only finitely many asymptotic values on its principal gradient tentacle.

Proof. Let $(x_k)_{k\in\mathbb{N}}$ be a sequence of points $x_k \in S(\nabla f)$ and $y \in \mathbb{R}$ such that $\lim_{k\to\infty} ||x_k|| = \infty$ and $\lim_{k\to\infty} f(x_k) = y \notin K_0(f)$. We show that $y \in K_\infty(f)$ using implication (i) $\Longrightarrow$ (ii) in [P1, Theorem 1.4]. Because of our sequence $(x_k)_{k\in\mathbb{N}}$, it is impossible that there exists $N \geq 1$ and $\delta > 0$ such that for all $x \in \mathbb{R}^n$ with $\|x\|$ sufficiently large and $f(x)$ sufficiently close to y, we have

$$||x||\|\nabla f(x)\| \geq \delta \sqrt[N]{\|x\|}.$$

This means that condition (ii) in [P1, Theorem 1.4] is violated. The implication (i) $\Longrightarrow$ (ii) in [P1, Theorem 1.4] yields that $y \in B(f)$ (here we use that $y \notin K_0(f)$). But $B(f) \subseteq K(f)$ by Theorem 15. This shows $y \in K(f) \setminus K_0(f) \subseteq K_\infty(f)$ by Proposition 13. □

Lemma 24. *Every $f \in \mathbb{R}[\bar{X}]$ is bounded on $S(\nabla f)$.*

Proof. By the Łojasiewicz inequality at infinity [Spo, Theorem 1], there exist $c_1, c_2 \in \mathbb{N}$ such that for all $x \in \mathbb{C}^n$,

$$|f(x)| \geq c_1 \implies |f(x)| \leq c_2\|\nabla f(x)\|\|x\|.$$

Then $|f| \leq \max\{c_1, c_2\}$ on $S(\nabla f)$. □

3.1. **The principal gradient tentacle and sums of squares.** Here comes one of the main results of this article which is interesting on its own but can later be read as a convergence result for a sequence of optimal values of SDPs (Theorem 30 below).

Theorem 25. *Let $f \in \mathbb{R}[\bar{X}]$ be bounded from below. Furthermore, suppose that f has only isolated singularities at infinity (which is always true in the two variable case $n = 2$) or the principal gradient tentacle $S(\nabla f)$ is compact. Then the following are equivalent.*

(i) $f \geq 0$ *on* $\mathbb{R}^n$
(ii) $f \geq 0$ *on* $S(\nabla f)$
(iii) *For every* $\varepsilon > 0$, *there are sums of squares of polynomials* s *and* t *in* $\mathbb{R}[\bar{X}]$ *such that*

$$f + \varepsilon = s + t(1 - \|\nabla f\|^2 \|\bar{X}\|^2). \tag{15}$$

Proof. First of all, the polynomial $g := 1 - \|\nabla f\|^2 \|\bar{X}\|^2$ is a polynomial describing the principal gradient tentacle

$$S := \{x \in \mathbb{R}^n \mid g(x) \geq 0\} = S(\nabla f).$$

Because sums of squares of polynomials are globally nonnegative on $\mathbb{R}^n$, identity (15) can be viewed as a certificate for $f \geq -\varepsilon$ on S. Hence it is clear that (iii) implies (ii). For the reverse implication, we apply Theorem 9 (with $m = 1$ and $g_1 := g$) to $f + \varepsilon$ instead of f. We only have to check the hypotheses. Condition (a) is clear from Lemma 24. By Theorem 23, we have that $R_\infty(f, S)$ is a finite set if f has only isolated singularities at infinity. If $S(\nabla f)$ is compact, the set $R_\infty(f, S)$ is even empty. Since $f \geq 0$ on S by hypothesis, this set contains clearly only nonnegative numbers. This shows condition (b), i.e., $R_\infty(f + \varepsilon, S) = \varepsilon + R_\infty(f, S)$ is a finite subset of $\mathbb{R}_{>0}$. Finally, the hypothesis $f \geq 0$ on S gives $f + \varepsilon > 0$ on S which is condition (c). Therefore (ii) and (iii) are proved to be equivalent. The equivalence of (i) and (ii) is an immediate consequence of Theorem 19. □

Remark 26. Let $f \in \mathbb{R}[\bar{X}]$ be bounded from below and $S(\nabla f)$ be compact. Then f attains its infimum f^*. To see this, observe that the equivalence of (i) and (ii) in the preceding theorem implies

$$\begin{aligned} f^* &= \sup\{a \in \mathbb{R} \mid f - a \geq 0 \text{ on } \mathbb{R}^n\} \\ &= \sup\{a \in \mathbb{R} \mid f - a \geq 0 \text{ on } S(\nabla f)\} \\ &= \min\{f(x) \mid x \in S(\nabla f)\}. \end{aligned}$$

The following observation is proved in the same way than Remark 2.

Remark 27. If f is a sum of squares in the ring $\mathbb{R}[[\bar{X}]]$ of formal power series, then its lowest (non-vanishing) homogeneous part must be a sum of squares in $\mathbb{R}[\bar{X}]$.

Remark 28. There are polynomials $f \in \mathbb{R}[\bar{X}]$ such that $f \geq 0$ on $\mathbb{R}^n$ but there is no representation (15) for $\varepsilon = 0$. To see this, take a polynomial $f \in \mathbb{R}[\bar{X}]$ such that $f \geq 0$ on $\mathbb{R}^n$ but f is not a sum of squares in the ring $\mathbb{R}[[\bar{X}]]$ of formal power series (the Motzkin polynomial from (5) is such an example by the preceding remark). Then a representation (15) with $\varepsilon = 0$ is impossible since the polynomial $1 - \|\nabla f\|^2 \|\bar{X}\|^2$ has a positive constant term and is therefore a square in $\mathbb{R}[[\bar{X}]]$.

3.2. **Optimization using the gradient tentacle and sums of squares.** Theorem 25 shows that under certain conditions, computation of f^* amounts to computing the supremum over all a such that $f - a = s + t(1 - \|\nabla f\|^2 \|\bar{X}\|^2)$ for some sums of squares s and t in $\mathbb{R}[\bar{X}]$. As sketched in the introduction, sums of squares *of bounded degree* can be nicely parametrized by positive semidefinite matrices. This motivates the following definition.

Definition 29. For all polynomials $f \in \mathbb{R}[\bar{X}]$ and all $k \in \mathbb{N}_0$, we define $f_k^* \in \mathbb{R} \cup \{\pm\infty\}$ as the supremum over all $a \in \mathbb{R}$ such that $f - a$ can be written as a sum

$$f - a = s + t(1 - \|\nabla f\|^2 \|\bar{X}\|^2)) \tag{16}$$

where s and t are sums of squares of polynomials with $\deg t \leq 2k$.

Here and in the following, we use the convention that the degree of the zero polynomial is $-\infty$ so that $t = 0$ is allowed in the above definition. Note that when the degree of t in (16) is restricted then automatically also the degree of s.

Therefore the problem of computing f_k^* can be written as an SDP. How to do this, is already suggested in our introduction. It goes exactly like in the well-known method of Lasserre for optimization of polynomials on compact basic closed semialgebraic sets. We refer to [L1, M1, Sr2] for the details. There are anyway several toolboxes for MATLAB (a software for numerical computation) which can be used to create and solve the corresponding SDPs without knowing these details. The toolboxes we know are YALMIP [Löf] (which is very flexible and good for much more than sums of squares stuff), SOSTOOLS [PPS] (which has a very flexible and nice syntax), GloptiPoly [HL] (very easy to use for simple problems) and SparsePOP [KKW] (specialized for sparse polynomials). Besides MATLAB and such a toolbox one needs also an SDP solver for which the toolbox provides an interface.

A side remark that we want to make here is that to each SDP there is a dual SDP and it is desirable from the theoretical and practical point of view that *strong duality* holds, i.e., the optimal value of the primal and dual SDP coincide. For the SDPs arising from Definition 29, strong duality holds. This follows from the fact that principal gradient tentacles (unlike gradient varieties) always have non-empty interior (they always contain a small neighbourhood of the origin). For a proof confer [L1, Theorem 4.2], [M1, Corollary 3.2] or [Sr2, Corollary 21]. Here we will not define the dual SDP nor discuss its interpretation in terms of the so-called moment problem.

Recalling the definition of f^{sos} in (4), we have obviously

$$f^{\mathrm{sos}} \leq f_0^* \leq f_1^* \leq f_2^* \leq \dots \tag{17}$$

and if f is bounded from below, then all f_k^* are lower bounds (perhaps $-\infty$) of f^* by Theorem 19. Note that the technique from Jibetean and Laurent (see Subsection 1.6 above) gives upper bounds for f^* so that it complements nicely our method. It is easy to see that Theorem 25 can be expressed in terms of the sequence $f_0^*, f_1^*, f_2^*, \dots$ as follows.

Theorem 30. *Let $f \in \mathbb{R}[\bar{X}]$ be bounded from below. Suppose that f has only isolated singularities at infinity (e.g., $n = 2$) or the principle gradient tentacle $S(\nabla f)$ is compact. Then the sequence $(f_k^*)_{k \in \mathbb{N}}$ converges monotonically increasing to f^*.*

The following example shows that it is unfortunately in general not true that $f_k^* = f^*$ for big $k \in \mathbb{N}$.

Example 31. Let f be the Motzkin polynomial from (5). By Theorem 30, we have $\lim_{k \to \infty} f_k = 0$. But it is not true that $f_k = 0$ for some $k \in \mathbb{N}$. By Definition 29, this would imply that for all $\varepsilon > 0$, there is an identity (15) with sums of squares s and t such that $\deg s \leq k$. Because $S(\nabla f)$ has non-empty interior (note that $\nabla f(1, 1, 1) = 0$ since $f(1, 1, 1) = 0$), we can use [PS, Proposition 2.6(b)] (see [Sr2, Theorem 4.5] for a more elementary exposition) to see that such an identity would then also have to exist for $\varepsilon = 0$. But this is impossible as we have seen in Remark 28.

Unfortunately, the assumption that f is bounded from below is necessary in Theorem 30 as shown by the following trivial example.

Example 32. Consider $f := X \in \mathbb{R}[X]$ (i.e., let $n = 1$ and write X instead of X_1). Then $K(f) = \emptyset$, $S(\nabla f) = [-1, 1]$ and $(f_k^*)_{k\in\mathbb{N}}$ converges monotonically increasing to $\inf\{f(x) \mid -1 \leq x \leq 1\} = -1 \neq -\infty = f^*$.

Open Problem 33. *Do Theorems 25 and 30 hold without the hypothesis that f has only isolated singularities at infinity or $S(\nabla f)$ is compact?*

By the above arguments, it is easy to see that this question could be answered in the affirmative if $R_\infty(f, S(\nabla f))$ were finite for all polynomials $f \in \mathbb{R}[\bar{X}]$ bounded from below on $\mathbb{R}^n$. But this is not true as the following counterexample shows. We are grateful to Zbigniew Jelonek for pointing out to us this adaption of an example of Parusiński [P2, Example 1.11].

Example 34. Consider the polynomial $h := X + X^2Y + X^4YZ \in \mathbb{R}[X, Y, Z]$, set $f := h^2$ and define for fixed $a > 0$ the curve

$$\gamma : \mathbb{R}_{>0} \to \mathbb{R}^3 : s \mapsto \left(s, \frac{2a}{s^2}, -\frac{\left(1 + \frac{s}{4a}\right)}{2s^2}\right).$$

Observe that

$$h(\gamma(s)) = \frac{3}{4}s + a \qquad \text{and} \qquad \frac{\partial h}{\partial X}(\gamma(s)) = 0$$

and therefore $f(\gamma(s)) = (\frac{3}{4}s + a)^2$ and

$$\|\nabla f\|^2(\gamma(s)) = 4f\|\nabla h\|^2(\gamma(s)) = 4s^4\left(\frac{3}{4}s + a\right)^2\left(\left(\frac{1}{2} - \frac{s}{8a}\right)^2 + (2a)^2\right).$$

It follows that $\|\nabla f\|^2(\gamma(s))\|\gamma(s)\|^2$ equals

$$\left(4s^6 + 16a^2 + \left(1 + \frac{s}{4a}\right)^2\right)\left(\frac{3}{4}s + a\right)^2\left(\left(\frac{1}{2} - \frac{s}{8a}\right)^2 + (2a)^2\right)$$

which tends to $(16a^2+1)a^2(1/4+4a^2)$ for $s \to 0$. We now see that for $s \to 0$, $\|\gamma(s)\|$ tends to infinity, $f(\gamma(s))$ tends to a^2 and, when a is a sufficiently small positive number, $\|\nabla f\|^2(\gamma(s))\|\gamma(s)\|^2$ tends to a real number smaller than 1. This shows that $a^2 \in R_\infty(f, S(\nabla f))$ for every sufficiently small positive number a. Hence f is an example of a polynomial bounded from below such that $R_\infty(f, S(\nabla f))$ is infinite.

3.3. **Implementation in YALMIP.** We show here how to encode computation of f_k^* (as well as of $f_{-1}^* := f^{\text{sos}}$) for any $k \in \mathbb{N}$ with YALMIP. First you have to declare the variables appearing in the polynomial f (here x and y) as well as the variable a to maximize.

```
sdpvar x y a
```

Now you specify the polynomial f and the degree bound k (-1 for computing f^{sos}). Here we take the dehomogenization $f := M(X, Y, 1)$ where M is the Motzkin polynomial introduced in (5).

```
f = x^4 * y^2 + x^2 * y^4 - 3 * x^2 * y^2 + 1, k = 0
```

Now compute the partial derivatives with respect to the variables (here x and y) and specify the polynomial g defining the gradient tentacle.

```
df = jacobian(f, [x y]), g = 1 - (df(1)^2 + df(2)^2) * (x^2 + y^2)
```

Define a polynomial variable t of degree $\leq 2k$ and impose the constraints that t and $f - a - tg$ are sums of squares (for some reason the current version of YALMIP does here not accept a degree zero polynomial t so that this has to be modeled as a scalar variable).

```
if k > 0
  v = monolist([x; y], 2*k), coeffVec = sdpvar(length(v), 1)
  t = coeffVec' * v
  constraints = set(sos(f - a - t * g)) + set(sos(t))
elseif k == 0
  coeffVec = sdpvar(1, 1), t = coeffVec
  constraints = set(sos(f - a - t * g)) + set(t > 0)
else
  coeffVec = []
  constraints = set(sos(f - a))
end
```

Now solve the SDP and output the result for a.

```
solvesos(constraints, -a, [], [a; coeffVec]), double(a)
```

3.4. **Implementation in SOSTOOLS.** Below we give an SOSTOOLS code which even slightly easier to read but essentially analogous to the YALMIP code. In contrast to the YALMIP code above, the MATLAB Symbolic Math Toolbox is required to execute the code below.

```
syms x y a t
f = x^4 * y^2 + x^2 * y^4 - 3 * x^2 * y^2 + 1, k = 0
df = jacobian(f, [x y]), g = 1 - (df(1)^2 + df(2)^2) * (x^2 + y^2)
prog = sosprogram([x; y], a)

if k > 0
  v = monomials([x; y], [0 : k]), [prog, t] = sossosvar(prog, v)
  prog = sosineq(prog, f - a - t * g)
elseif k == 0
  prog = sosdecvar(prog, t), prog = sosineq(prog, t)
  prog = sosineq(prog, f - a - t * g)
else
  prog = sosineq(prog, f - a)
end

prog = sossetobj(prog, -a), prog = sossolve(prog)
sosgetsol(prog, a)
```

3.5. **Numerical results.** The following examples have been computed on an ordinary PC with MATLAB 7, YALMIP 3 and the SDP solver SeDuMi 1.1. Most of the computations took a few seconds, some of them a few minutes. The first example corresponds exactly to the code in Subsection 3.3. To compute the others, the variables, the polynomial f and the degree bound k has to be changed in that code.

Example 35. Let $f := M(X, Y, 1)$ be the dehomogenization of the Motzkin polynomial M from (5), i.e., $f := M(X, Y, 1) = X^4Y^2 + X^2Y^4 - 3X^2Y^2 + 1 \in \mathbb{R}[X, Y]$.

We have $f^* = 0$ but $f^{\text{sos}} = -\infty$ (the latter is an easy exercise). If we execute the program from Subsection 3.3 with $k = -1$ instead of $k = 0$, the computer answers that the SDP is infeasible which means indeed that $f^{\text{sos}} = -\infty$. Executing the same program for $k = 0, 1, 2$ yields $f_0^* \approx -0.0017$, $f_1^* \approx -0.0013$ and $f_2^* \approx 0.000066$ which is already very close to $f^* = 0$. By Theorem 30, the sequence $f_0, f_1, f_2, \dots$ converges monotonically to $f^* = 0$. But the computed value $f_2^* \approx 0.000066$ is positive so that there are obviously numerical problems. Confer [PS, Example 2].

Example 36. Define $f := M(X, 1, Z) \in \mathbb{R}[X, Z]$ where M is the Motzkin polynomial from (5), i.e., $f = X^4 + X^2 + Z^6 - 3X^2Z^2 \in \mathbb{R}[X, Z]$. Computation yields $f^{\text{sos}} \approx -0.1780$, $f_0^* \approx -5.1749 \cdot 10^{-5}$, $f_1^* \approx -1.2520 \cdot 10^{-7}$ and $f_2^* = 8.7662 \cdot 10^{-10}$ which "equals numerically" $f^* = 0$. This is in accordance with Theorem 25 which guarantees convergence to f^* since we are in the two variable case. Confer [PS, Example 3].

Example 37. Consider the Berg polynomial $f := X^2Y^2(X^2 + Y^2 - 1) \in \mathbb{R}[X, Y]$ with global minimum $f^* = -1/27$ attained in $(\pm 1/\sqrt{3}, \pm 1/\sqrt{3})$. We have $f^{\text{sos}} = -\infty$ and running the corresponding program gives indeed an output saying that the corresponding SDP is infeasible. The computed optimal values of the first principal tentacle relaxations are $f_0^* \approx -0.0564$, $f_1^* \approx -0.0555$, $f_2^* \approx -0.0371$ and $f_3^* \approx -0.0370 \approx -1/27 = f^*$. Confer [L1, Example 3], [NDS, Example 3] and [JL, Example 4].

Example 38. Being a polynomial in two variables of degree at most four, we have that for $f := (X^2 + 1)^2 + (Y^2 + 1)^2 - 2(X + Y + 1)^2 \in \mathbb{R}[X, Y]$, $f - f^*$ must be a sum of squares (see introduction) whence $f^* = f^{\text{sos}}$. By computation, we obtain for all values $f^{\text{sos}}, f_0^*, f_1^*, f_2^*$ approximately -11.4581. That all these computed values are the same can be expected by $f^* = f^{\text{sos}}$ and the monotonicity (17). Confer [L1, Example 2] and [JL, Example 3].

Example 39. In [LL], it is shown that

$$f := \sum_{i=1}^{5} \prod_{j \neq i} (X_i - X_j) \in \mathbb{R}[X_1, X_2, X_3, X_4, X_5]$$

is nonnegative on $\mathbb{R}^5$ but not a sum of squares of polynomials. Therefore $f^{\text{sos}} = -\infty$ by Remark 2 since f is homogeneous. The SDP solver detects indeed infeasibility of the corresponding SDP. We have computed $f_0^* \approx -0.2367$, $f_1^* \approx -0.0999$ and $f_2^* \approx -0.0224$. Solving the SDP relaxation computing f_2^* took already the time of a coffee break. As in [JL, Example 6], we observe therefore that minimizing f is after the change of variables $X_i \mapsto X_1 - Y_i$ $(i = 2, 3, 4, 5)$ equivalent to minimizing

$$h := Y_2Y_3Y_4Y_5 + \sum_{i=2}^{5} (-Y_i) \prod_{j \neq i} (Y_j - Y_i) \in \mathbb{R}[Y_2, Y_3, Y_4, Y_5].$$

Computing h^{sos} results in infeasibility. The numerical results using the principle gradient tentacle are $h_0^* \approx -0.2380$, $h_1^* \approx -0.0351$, $h_2^* \approx -0.0072$, $h_3^* \approx -0.0019$ and $h_4^* \approx -0.00086285$ which is already very close to $h^* = 0$. The condition in Theorem 30 is satisfied neither for f nor for h and yet it seems that we have convergence to h^*. This is a typical observation that might give hope that Open Problem 33 has a positive answer.

Example 40. Consider once more the polynomial $f = (1-XY)^2 + Y^2$ from (6) and Example 16 that does not attain its infimum $f^* = 0$ on $\mathbb{R}^2$. Since this polynomial is by definition a sum of squares, we have $f^{\mathrm{sos}} = 0 = f^*$ and therefore $f_k^* = 0$ for all $k \in \mathbb{N}$ by (17). By computation, we get $f^{\mathrm{sos}} \approx 1.5142 \cdot 10^{-12}$ which is almost zero but also $f_0^* \approx 0.0016$, $f_1^* \approx 0.0727$ and $f_2^* \approx 0.1317$ which shows that there are big numerical problems. We have verified that the corresponding SDPs have nevertheless been solved quite accurately. The problem is that small numerical errors in the coefficients of a polynomial can perturb its infimum quite a lot whenever the infimum is not attained (or attained very far from the origin). It should be subject to further research how to fight this problem. Anyway, the gradient tentacle method still performs in this example much better than the gradient variety method which yields the wrong answer 1 (as described in Subsection 1.7 above). The method of Jibetean and Laurent gives the best results in this case [JL, Example 5].

3.6. **Numerical stability.** If the coefficients of f and $\|\nabla f\|\|\bar{X}\|$ have an order of magnitude very different from 1, then the defining polynomial $g = 1 - \|\nabla f\|^2\|\bar{X}\|^2$ for the gradient tentacle should be better exchanged by $R - \|\nabla f\|^2\|\bar{X}\|^2$ where R is a real number of that order of magnitude. This is justified by Remark 18 above.

Example 40 and other experiments that we did with polynomials bounded from below that do not attain a minimum are a bit disappointing and show that for this "hard" class of polynomials (exactly the class we were attacking), a lot of work remains to be done, at least on the numerical side. The corresponding semidefinite programs tend to be numerically unstable.

For polynomials attaining their minimum, the method in [NDS] is often much more efficient, e.g., for Example 39.

4. Higher gradient tentacles

In this section, we associate to every polynomial $f \in \mathbb{R}[\bar{X}]$ a *sequence* of gradient tentacles. Each of these is defined by a polynomial inequality just as the principal tentacle from Section 3 was. But the degree of this polynomial inequality for the N-th tentacle in this sequence will be roughly $2N$ times the degree of f. This has the disadvantage that the corresponding SDP relaxations get very big for large N. Also, we have to deal for *each* N with a sequence of SDPs. All in all, we have therefore a double sequence of SDPs. The advantage is however that we can prove a sums of squares representation theorem (Theorem 46) applicable for *all* $f \in \mathbb{R}[\bar{X}]$ bounded from below independently of what is the answer to Open Problem 33. Again, we think that this theorem is also of theoretical interest. Implementation of the higher gradient tentacle method is analogous to Subsections 3.3 and 3.4. This time we do not give numerical examples because of Open Problem 33, Remark 21 and numerical problems for big N.

Definition 41. For $f \in \mathbb{R}[\bar{X}]$ and $N \in \mathbb{N}$, we call

$$S(\nabla f, N) := \{x \in \mathbb{R}^n \mid \|\nabla f(x)\|^{2N}(1 + \|x\|^2)^{N+1} \leq 1\}$$

the N-th *gradient tentacle* of f.

A trivial fact that one should keep in mind is that $\|\nabla f(x)\|^2(1 + \|x\|^2) \leq 1$ and in particular $\|\nabla f(x)\|\|x\| \leq 1$ for all $x \in S(\nabla f, N)$. This shows that

$$V(\nabla f) \cap \mathbb{R}^n \subseteq S(\nabla f, 1) \subseteq S(\nabla f, 2) \subseteq S(\nabla f, 3) \subseteq \ldots \subseteq S(\nabla f).$$

The definition of $S(\nabla f, N)$ is motivated by the following definition which is taken from [KOS, page 79].

Definition 42. Suppose $f \in \mathbb{R}[\bar{X}]$ and $N \in \mathbb{N}$. The set $K_\infty^N(f)$ consists of all $y \in \mathbb{R}$ for which there exists a sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ such that

$$\lim_{k\to\infty} ||x_k|| = \infty, \qquad \lim_{k\to\infty} \|\nabla f(x_k)\| \|x_k\|^{1+\frac{1}{N}} = 0 \quad \text{and} \qquad \lim_{k\to\infty} f(x_k) = y. \tag{18}$$

Clearly, we have

$$K_\infty^1(f) \subseteq K_\infty^2(f) \subseteq K_\infty^3(f) \subseteq \cdots \subseteq K_\infty(f).$$

The next lemma says that this chain actually gets stationary and reaches $K_\infty(f)$. For the proof, we refer to [KOS, Lemma 3.1].

Lemma 43 (Kurdyka, Orro and Simon)**.** *For all $f \in \mathbb{R}[\bar{X}]$, there exists $N \in \mathbb{N}$ such that*

$$K_\infty(f) = K_\infty^N(f).$$

Now we prove for sufficiently large gradient tentacles what was Corollary 19 for the principal gradient tentacle (which contains all higher gradient tentacles).

Theorem 44. *Suppose $f \in \mathbb{R}[\bar{X}]$ is bounded from below. Then $f^* \in K(f)$ and there is $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$,*

$$f^* = \inf\{f(x) \mid x \in S(\nabla f, N)\}. \tag{19}$$

Proof. We know already from Theorem 19 that $f^* \in K(f)$. By Proposition 13, at least one of the following two cases therefore must occur. The first case is that $f^* \in K_0(f)$. Then f^* is attained by f on its gradient variety and therefore on the N-th gradient tentacle for actually all $N \in \mathbb{N}$. Hence we can set $N_0 := 1$. In the second case $f^* \in K_\infty(f)$, we can choose some $N_0 \in \mathbb{N}$ such that $f^* \in K_\infty^N(f)$ by the previous Lemma. Then $f^* \in K_\infty^N(f)$ for any $N \geq N_0$. This means that there exists a sequence $(x_k)_{k\in\mathbb{N}}$ satisfying (18). Therefore $\|\nabla f(x)\| \|x_k\|^{1+1/N} \leq \frac{1}{2}$ and consequently

$$\|\nabla f(x_k)\|^{2N}(1 + \|x_k\|^2)^{N+1} \leq \|\nabla f(x_k)\|^{2N}(2\|x_k\|^2)^{N+1} \leq 1$$

for all large k since $\|x_k\| \geq 1$ and $2^{N+1} \leq 2^{2N}$. This shows that $x_k \in S(\nabla f, N)$ for all large k which implies our claim. $\square$

The great advantage of the higher gradient tentacles over the principal one is that they are *always* small enough to admit only finitely many asymptotic values, i.e., there is no counterpart to Example 34.

Theorem 45. *For every $f \in \mathbb{R}[\bar{X}]$, $R_\infty(f, S(\nabla f)) \subseteq K_\infty(f)$. In particular, every $f \in \mathbb{R}[\bar{X}]$ has only finitely many asymptotic values on each of its higher gradient tentacles, i.e., the set $R_\infty(f, S(\nabla f, N))$ is finite for all $N \in \mathbb{N}$.*

Proof. Let $y \in \mathbb{R}$ be such that (7) holds for some sequence $(x_k)_{k\in\mathbb{N}}$ of points $x_k \in S(\nabla f, N)$. By Definition 41,

$$\|\nabla f(x_k)\|^N \|x_k\|^N \leq \frac{1}{\|x_k\|} \to 0 \qquad \text{for } k \to \infty$$

implying (13). This shows $y \in K_\infty(f)$. $\square$

4.1. **Higher gradient tentacles and sums of squares.** We are now able to prove the third important sums of squares representation theorem of this article besides Theorems 9 and 25.

Theorem 46. *For all $f \in \mathbb{R}[\bar{X}]$ bounded from below, there is $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$, the following are equivalent.*

(i) $f \geq 0$ *on* $\mathbb{R}^n$

(ii) $f \geq 0$ *on* $S(\nabla f, N)$

(iii) For every $\varepsilon > 0$, there are sums of squares of polynomials s and t in $\mathbb{R}[\bar{X}]$ such that

$$f + \varepsilon = s + t(1 - \|\nabla f\|^{2N}(1 + \|\bar{X}\|^2)^{N+1}). \tag{20}$$

Moreover, these conditions are equivalent for all f attaining a minimum on $\mathbb{R}^n$ and all $N \in \mathbb{N}$. Finally, (ii) and (iii) are equivalent for all $f \in \mathbb{R}[\bar{X}]$ and $N \in \mathbb{N}$.

Proof. We first show that (ii) and (iii) are always equivalent. To see this, observe that $g_1 := 1 - \|\nabla f\|^{2N}\|\bar{X}\|^{2N+2}$ is a polynomial that defines the set $S := \{x \in \mathbb{R}^n \mid g_1 \geq 0\} = S(\nabla f, N)$. Because sums of squares of polynomials are globally nonnegative on $\mathbb{R}^n$, identity (20) can be viewed as a certificate for $f \geq -\varepsilon$ on S. Hence it is clear that (iii) implies (ii). For the reverse implication, we apply Theorem 9 to $f + \varepsilon$ instead of f. We only have to check the hypotheses. Condition (a) is clear from Lemma 24. By Corollary 45, we have that $R_\infty(f, S)$ is a finite set. Since $f \geq 0$ on S by hypothesis, this set contains clearly only nonnegative numbers. This shows condition (b), i.e., $R_\infty(f + \varepsilon, S) = \varepsilon + R_\infty(f, S)$ is a finite subset of $\mathbb{R}_{>0}$. Finally, the hypothesis $f \geq 0$ on S gives $f + \varepsilon > 0$ on S which is condition (c).

Now suppose that $f \in \mathbb{R}[\bar{X}]$ attains a minimum $f(x^*) = f^*$ in a point $x^* \in \mathbb{R}^n$. Then $\nabla f(x^*) = 0$ and therefore $x^* \in S(\nabla f, N)$ for all $N \in \mathbb{N}$. This shows that (i) and (ii) are in this case equivalent for *all* $N \in \mathbb{N}$.

By what has already been proved, it remains only to show that (i) and (ii) are equivalent for large $N \in \mathbb{N}$ when $f \in \mathbb{R}[\bar{X}]$ is bounded from below but does not attain a minimum. But in this case, (19) holds by Theorem 44 yielding the equivalence of the first two conditions. □

Without needing it for our application, we draw the following immediate corollary. Taking $N = 1$ in the second part of this corollary yields Theorem 6 above of Nie, Demmel and Sturmfels.

Corollary 47. *Suppose $f \in \mathbb{R}[\bar{X}]$ and $f \geq 0$ on $V(\nabla f) \cap \mathbb{R}^n$. Then $f + \varepsilon$ is for all $\varepsilon > 0$ a sum of squares modulo any principal ideal generated by a power of the polynomial $\|\nabla f\|^2(1 + \|\bar{X}\|^2)$, i.e., for every $\varepsilon > 0$ and $N \in \mathbb{N}$, there is a sum of squares s in $\mathbb{R}[\bar{X}]$ and a polynomial $p \in \mathbb{R}[\bar{X}]$ such that*

$$f = s + p(\|\nabla f\|^2(1 + \|\bar{X}\|^2))^N.$$

In particular, $f + \varepsilon$ is for all $\varepsilon > 0$ a sum of squares modulo each power of its gradient ideal, i.e., for every $\varepsilon > 0$ and $N \in \mathbb{N}$, there is a sum of squares s in $\mathbb{R}[\bar{X}]$ such that

$$f \in s + (\nabla f)^N.$$

Proof. The second claim follows from the first one. The first claim follows immediately from implication (i) $\Longrightarrow$ (iii) in Theorem 46 which always holds for all $N \in \mathbb{N}$. □

4.2. **Optimization using higher gradient tentacles and sums of squares.** The following definition can be motivated in the same way than Definition 29 in Section 3.

Definition 48. For all polynomials $f \in \mathbb{R}[\bar{X}]$, all $N \in \mathbb{N}$ and all $k \in \mathbb{N}_0$, we define $f^*_{N,k} \in \mathbb{R} \cup \{\pm\infty\}$ as the supremum over all $a \in \mathbb{R}$ such that $f - a$ can be written as a sum

$$f - a = s + t(1 - \|\nabla f\|^{2N}(1 + \|\bar{X}\|^2)^{N+1}) \tag{21}$$

where s and t are sums of squares of polynomials with $\deg t \leq 2k$.

Again, like in Section 3 outlined, computation of $f_{N,k}$ amounts to solving an SDP for each fixed $N \in \mathbb{N}$ and $k \in \mathbb{N}_0$. Recalling the definition of f^{sos} in (4), we have for each fixed $N \in \mathbb{N}$,

$$f^{\text{sos}} \leq f^*_{N,0} \leq f^*_{N,1} \leq f^*_{N,2} \leq \dots$$

and if f is bounded from below, then all $f^*_{N,k}$ are lower bounds of f^* by Theorem 44. It would be desirable to have also information how the $f_{N,k}$ are related to each other when not only k but also N varies. All we know about that is the following proposition.

Proposition 49. *For all $f \in \mathbb{R}[\bar{X}]$, $N \in \mathbb{N}$ and $k \in \mathbb{N}_0$,*

$$f^*_{N+1,k} \leq f^*_{N,k+d}.$$

Proof. Let us define the polynomials h_N like in (11) and substitute in the identity proved in Lemma 11, the polynomials $\|\nabla f\|^2$ for Y and $\|\bar{X}\|^2$ for $\bar{X}$. Then we get

$$1 - \|\nabla f\|^{2(N+1)}(1 + \|\bar{X}\|^2)^{N+2} = p + q(1 - \|\nabla f\|^{2N}(1 + \|\bar{X}\|^2)^{N+1}). \tag{22}$$

where p and

$$q := \left(1 + \frac{1}{N}\right) \|\nabla f\|^2(1 + \|\bar{X}\|^2)$$

are sums of squares of polynomials. The degree of q is no higher than $2(d-1)+2 = 2d$. Now if for $a \in \mathbb{R}$ we have an identity

$$f - a = s + t(1 - \|\nabla f\|^{2(N+1)}(1 + \|\bar{X}\|^2)^{N+2})$$

for sums of squares s and t with $\deg t \leq 2k$, then for the same a

$$f - a = (s + tp) + tq(1 - \|\nabla f\|^{2N}(1 + \|\bar{X}\|^2)^{N+1})$$

and $\deg(tq) \leq 2(k + d)$. □

We conclude by interpreting Theorem 46 as a convergence result concerning the optimal values $f^*_{N,k}$ of the proposed relaxations. This is the counterpart to Theorem 30 from Section 2.

Theorem 50. *For all $f \in \mathbb{R}[\bar{X}]$ bounded from below, $(f^*_{N,k})_{k\in\mathbb{N}}$ converges monotonically increasing to f^* provided that $N \in \mathbb{N}$ is sufficiently large (depending on f). If f attains a minimum on $\mathbb{R}^n$, $(f^*_{N,k})_{k\in\mathbb{N}}$ converges monotonically increasing to f^* no matter what $N \in \mathbb{N}$ is.*

5. Conclusions

We have proposed a method for computing numerically the infimum of a real polynomial in n variables which is bounded from below on $\mathbb{R}^n$. Like in [JL] and [NDS], the approach is to find semidefinite relaxations relying on sums of squares certificates and critical point theory. As one could expect, polynomials that do not attain a minimum on $\mathbb{R}^n$ (that are either unbounded from below or have a finite infimum that is not attained) are particularly hard to handle. In [JL], this problem (among others) was solved by perturbing the coefficients of the polynomial to guarantee a minimum (in particular, boundedness from below). Though the results in [JL] are quite good, we are convinced that one should also look for other methods that avoid perturbations and the danger of numerical ill-conditioning coming along with them. Proving sums of squares representations for polynomials positive on their gradient variety, it was shown by Nie, Demmel and Sturmfels [NDS] that an approach without perturbation is possible. The computational performance of their method is extremely good. However, for polynomials that do not attain a minimum, their method yields wrong answers. Combining considerable machinery from differential geometry and real algebraic geometry, we have shown that part of this limitation can be removed. By using our gradient tentacles instead of the gradient variety, polynomials that do not attain a minimum but are bounded from below can also be handled. Our method has three major problems. First, we do not address the important question of how to check efficiently if a polynomial is bounded from below. For such polynomials, our method still gives a wrong answer (see Example 32). Second, it turns out that solving semidefinite programs that arise from a polynomial that does not attain a minimum takes sometimes surprisingly long time. And third, small numerical inaccuracies might lead to big changes in the infimum of a polynomial if the infimum is not attained. All three problems should be subject to further research. Polynomials not attaining a minimum remain hard to handle in practice. On the theoretical side, we have combined the theory of generalized critical values with the the theory of real holomorphy rings and have obtained new interesting characterizations of nonnegative polynomials.

Acknowledgments

We are most grateful to Zbigniew Jelonek for the discussions in Passau where he showed us Example 34 and Parusinski's Theorem 23. Our thanks go also to Mohab Safey El Din for shifting our attention to Theorem 19, to Richard Leroy for helping to prove Lemma 11 and to Krzysztof Kurdyka for interesting discussions in Paris.

References

[Ble] G. Blekherman. There are Significantly More Nonnegative Polynomials than Sums of Squares. Preprint.
`http://arxiv.org/abs/math.AG/0309130`

[Cas] G. Cassier. Problème des moments sur un compact de $\mathbb{R}^n$ et décomposition de polynômes à plusieurs variables. J. Funct. Anal. **58**, 254–266 (1984)

[HL] D. Henrion and J. Lasserre. GloptiPoly: Global optimization over polynomials with Matlab and SeDuMi. ACM Trans. Math. Softw. **29**, No. 2, 165–194 (2003)

[JL] D. Jibetean and M. Laurent. Semidefinite approximations for global unconstrained polynomial optimization. SIAM Journal on Optimization **16**, No. 2, 490–514 (2005)

[KKW] M. Kojima, S. Kim and H. Waki. Sparsity in sums of squares of polynomials. Math. Program. **103**, No. 1 (A), 45–62 (2005)

[KOS] K. Kurdyka, P. Orro and S. Simon: Semialgebraic Sard theorem for generalized critical values. J. Differ. Geom. **56**, No.1, 67-92 (2000)

[Kri] J. Krivine. Anneaux preordonnes. J. Anal. Math. **12**, 307–326 (1964)

[Kus] A. Kushnirenko. Polyèdres de Newton et nombres de Milnor. Invent. Math. **32**, 1–31 (1976)

[L1] J. Lasserre. Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**, No. 3, 796–817 (2001)

[L2] J. Lasserre. A sum of squares approximation of nonnegative polynomials. SIAM J. Optim. **16**, No. 3, 751–765 (2006)

[LL] A. Lax and P. Lax. On sums of squares. Linear Algebra Appl. **20**, 71–75 (1978)

[LN] J. Lasserre and T. Netzer. SOS approximations of nonnegative polynomials via simple high degree perturbations. To appear in Math. Z.
`http://arxiv.org/abs/math.math.AG/0510456`

[Löf] J. Löfberg. YALMIP: A MATLAB toolbox for rapid prototyping of optimization problems
`http://control.ee.ethz.ch/~joloef/yalmip.php`

[M1] M. Marshall. Optimization of polynomial functions. Can. Math. Bull. **46**, No. 4, 575–587 (2003)

[M2] M. Marshall. Representations of non-negative polynomials, degree bounds and applications to optimization. Preprint.
`http://math.usask.ca/~marshall/`

[NDS] J. Nie, J. Demmel, and B. Sturmfels. Minimizing polynomials via sum of squares over the gradient ideal. Math. Prog., Ser. A, **106**, No. 3, 587–606 (2006)

[Net] T. Netzer. High degree perturbations of nonnegative polynomials. Diplomarbeit, Universität Konstanz (2005)
`http://www.math.uni-konstanz.de/~netzer/`

[P1] A. Parusiński. On the bifurcation set of a complex polynomial with isolated singularities at infinity. Compos. Math. **97**, No. 3, 369–384 (1995)

[P2] A. Parusiński. A note on singularities at infinity of complex polynomials. Proceedings of the Banach Center symposium on differential geometry and mathematical physics in Spring 1995. Banach Cent. Publ. **39**, 131–141 (1997)

[PD] A. Prestel, C. Delzell. Positive polynomials. Monographs in Mathematics. Springer, Berlin (2001)

[PPS] S. Prajna, A. Papachristodoulou, P. Seiler and P. Parrilo. SOSTOOLS and its control applications. Lecture Notes in Control and Information Sciences **312**, 273–292 (2005)

[PS] P. Parrilo and B. Sturmfels. Minimizing polynomial functions. Ser. Discrete Math. Theor. Comput. Sci. **60**, 83-99 (2003)

[Rab] P. Rabier. Ehresmann fibrations and Palais-Smale conditions for morphisms of Finsler manifolds. Ann. Math. (2) **146**, No. 3, 647–691 (1997)

[Rez] B. Reznick. Some concrete aspects of Hilbert's 17th problem. Contemp. Math. **253**, 251–272 (2000)

[Sch] K. Schmüdgen. The K-moment problem for compact semi-algebraic sets. Math. Ann. **289**, No. 2, 203–206 (1991)

[Scd] J. Schmid. On the degree complexity of Hilbert's 17th problem and the Real Nullstellensatz. Habilitationsschrift, Universität Dortmund (1998)

[Sho] N. Shor. Class of global minimum bounds of polynomial functions. Cybernetics **23**, No. 6, 731–734 (1987)

[Shu] E. Shustin. Critical points of real polynomials, subdivisions of Newton polyhedra and topology of real algebraic hypersurfaces. Transl., Ser. 2, Am. Math. Soc. **173**, 203–223 (1996)

[Spo] S. Spodzieja. Lojasiewicz inequalities at infinity for the gradient of a polynomial. Bull. Pol. Acad. Sci., Math. **50**, No. 3, 273–281 (2002)

[Sr1] M. Schweighofer. Iterated rings of bounded elements and generalizations of Schmüdgen's Positivstellensatz. J. Reine Angew. Math. **554**, 19–45 (2003). Erratum available at
`http://arxiv.org/abs/math.AC/0510675`

[Sr2] M. Schweighofer. Optimization of polynomials on compact semialgebraic sets. SIAM Journal on Optimization **15**, No. 3, 805–825 (2005)

[SS] N. Shor and P. Stetsyuk. Modified r-algorithm to find the global minimum of polynomial functions. Cybern. Syst. Anal. **33**, No. 4, 482–497 (1997)

[Ste] G. Stengle. A Nullstellensatz and a Positivstellensatz in semialgebraic geometry. Math. Ann. **207**, 87–97 (1974)

[Tho] R. Thom. Ensembles et morphismes stratifies. Bull. Am. Math. Soc. **75**, 240–284 (1969)

[Tod] M. Todd. Semidefinite Optimization. Acta Numerica **10**, 515-560 (2001)

Universität Konstanz, Fachbereich Mathematik und Statistik, 78457 Konstanz, Allemagne

E-mail address: Markus.Schweighofer@uni-konstanz.de