SEMIDEFINITE APPROXIMATIONS FOR GLOBAL UNCONSTRAINED POLYNOMIAL OPTIMIZATION

DORINA JIBETEAN AND MONIQUE LAURENT

ABSTRACT. We consider the problem of minimizing a polynomial function on $\mathbb{R}^n$, known to be hard even for degree 4 polynomials. Therefore approximation algorithms are of interest. Lasserre [15] and Parrilo [23] have proposed approximating the minimum of the original problem using a hierarchy of lower bounds obtained via semidefinite programming relaxations. We propose here a method for computing tight upper bounds based on perturbing the original polynomial and using semidefinite programming. The method is applied to several examples.

January 27, 2005

1. INTRODUCTION

We consider the problem:

$$(1) \qquad\qquad p^* := \inf_{x \in \mathbb{R}^n} \ p(x)$$

of minimizing a polynomial p in n indeterminates over $\mathbb{R}^n$. We may assume that p has an even degree $2m$, since otherwise $p^* = -\infty$. There are three possibilities: Either p has an infinite infimum (i.e., $p^* = -\infty$), or p has a finite infimum (e.g., for the polynomial $p(x_1, x_2) = x_1^2 + (x_1 x_2 - 1)^2$), or p has a minimum. Computing the infimum of a polynomial is a hard problem, already for degree 4 polynomials. Indeed, it contains the problem of deciding whether a matrix is copositive, known to be co-NP-hard [21]; an $n \times n$ matrix P being copositive if $p(x) := \sum_{i,j=1}^n P_{ij} x_i^2 x_j^2 \geq 0$ for all $x \in \mathbb{R}^n$, i.e., if $p^* = 0$. Alternatively, problem (1) contains the problem of deciding whether an integer sequence $a_1, \dots, a_n$ can be partitioned, known to be NP-complete [7]; $a_1, \dots, a_n$ being partitionable if there exists $x \in \{\pm 1\}^n$ such that $a^T x = 0$, i.e., if the infimum of the polynomial $p(x) := (a^T x)^2 + \sum_{i=1}^n (x_i^2 - 1)^2$ is equal to 0.

1.1. **Some known approaches to polynomial unconstrained minimization.** An approach followed by some authors (e.g., by Hägglöf et al. [8]) is to look at the first order conditions $\partial p / \partial x_i = 0$ $(i = 1, \dots, n)$. Various algebraic techniques can be used for determining the real solutions to this system of polynomial equations; e.g., using Groebner bases and the eigenvalue method, using resultants and discriminants, or homotopy methods (see, e.g., [3]; see [25] for a discussion and comparison). However, there are several difficulties with such an approach. It is computationally expensive (e.g., computing a Gröbner basis may be computationally very demanding), the number of critical points

can be infinite and, moreover, this approach applies *only* if the polynomial p attains its minimum. We will come back to this type of approach later in this section.

Hanzon and Jibetean [9] (see also Jibetean [12]) propose to go around these difficulties by considering a perturbation:

$$(2) \qquad p_\lambda(x) := p(x) + \lambda \left(\sum_{i=1}^{n} x_i^{2m+2} \right)$$

of the original polynomial p for small $\lambda > 0$. Set

$$p_\lambda^* := \inf_{x \in \mathbb{R}^n} p_\lambda(x).$$

Thus, $p^* \le p_\lambda^* \le p^* + \lambda \|x^*\|^{2m+2}$ if x^* is a global minimizer of p. The perturbed polynomial has the following properties: p_λ attains its minimum, the set of critical points of p_λ is finite, and the limit of the minima p_λ^* as $\lambda \to 0$ is equal to the infimum p^* of p. Moreover, if p has a global minimum, then the limit set as $\lambda \downarrow 0$ of the set of global minimizers of p_λ is contained in the set of global minimizers of p, and each connected component of the set of global minimizers of p contains a point which is the limit of a branch of local minimizers of p_λ. Exploiting these facts, Hanzon and Jibetean propose an exact algorithm for computing the limit p^* of the minima p_λ^* as well as a global minimizer of p (if some exists). Their algorithm uses algebraic techniques, some of them closely related to the algebraic machinery developed by Basu, Pollack and Roy [1]. Hanzon and Jibetean's method applies to any polynomial p, i.e., no assumption is made on the existence of a minimum. However, its computational cost is very high and the algorithm can be applied in practice only to small instances.

Another type of approach consists of solving a convex (in fact, semidefinite) relaxation of the original problem; see, e.g., Lasserre [15], Parrilo [22, 23], Shor [28]. The approach applies more generally to the problem:

$$(3) \qquad p^* := \inf_{x \in K} p(x), \text{ where } K := \{x \in \mathbb{R}^n \mid h_1(x) \ge 0, \dots, h_\ell(x) \ge 0\}$$

of minimizing p over a set defined by polynomial inequalities and equations (treating an equation $h(x) = 0$ as two opposite inequalities: $h(x) \ge 0$, $-h(x) \ge 0$). Following Lasserre [15], set $d_i := \lceil \deg(h_i)/2 \rceil$ and, for any integer $k \ge \max(\lceil \deg(p)/2 \rceil, d_1, \dots, d_\ell)$, consider the semidefinite program:

$$(4) \qquad p_{L,k}^* := \inf \ p^T y \text{ s.t. } M_k(y) \succeq 0, \ M_{k-d_i}(h_i y) \succeq 0 \ (i = 1, \dots, \ell), \ y_0 = 1$$

(the *moment relaxation of order k* of (3)) and its dual:

$$(5) \qquad \begin{aligned} \rho_k^* := \sup \ \rho \text{ s.t.} \quad & p(x) - \rho = u_0 + \textstyle\sum_{i=1}^{\ell} u_i h_i, \text{ where} \\ & u_0, u_1, \dots, u_\ell \text{ are sum of squares of polynomials} \\ & \text{and } \deg(u_0), \deg(u_1 h_1), \dots, \deg(u_\ell h_\ell) \le 2k \end{aligned}$$

(the *s.o.s. relaxation of order k* of (3)). Program (4) uses the variables $y = (y_\alpha)_{\alpha \in S_{2k}}$, $M_k(y) := (y_{\alpha+\alpha'})_{\alpha,\alpha' \in S_k}$ is the moment matrix of order k, $M_{k-d_i}(h_i y)$ are localizing matrices, and for an integer k, we set $S_k := \{\alpha \in \mathbb{Z}_+^n \mid |\alpha| := \sum_{i=1}^{n} \alpha_i \le k\}$. Then,

$\rho_k^* \le p_{L,k}^* \le p^*$, $\rho_k^* \le \rho_{k+1}^*$, and $p_{L,k}^* \le p_{L,k+1}^*$. Under some assumption on K, there is asymptotic convergence of the parameters ρ_k^*, μ_k^* to p^*. The following cases are of particular interest for our purpose:

(I) $K = \{x \in \mathbb{R}^n \mid \sum_{i=1}^n x_i^2 \le R^2\}$. Then there is asymptotic convergence of ρ_k^* and $p_{L,k}^*$ to p^* ([15]).

(II) $K = \{x \in \mathbb{R}^n \mid h_1(x) = 0, \ldots, h_\ell(x) = 0\}$ and the polynomials $h_1, \ldots, h_\ell$ generate a zero-dimensional ideal I (i.e., they have finitely many common complex zeros). Then there is finite convergence of $p_{L,k}^*$ to p^*, and of ρ_k^* when $h_1, \ldots, h_\ell$ form a Groebner basis of I ([18]) or when I is radical ([24]).

(III) $K = \{x \in \mathbb{R}^n \mid \frac{\partial p}{\partial x_i}(x) = 0 \ (i = 1, \ldots, n)\}$. Then there is asymptotic convergence of ρ_k^* and p_k^* to p^*, and finite convergence when the ideal I_{grad} generated by the polynomials $\frac{\partial p}{\partial x_i}$ $(i = 1, \ldots, n)$ is radical ([6]). (By Case (II) there is finite convergence of p_k^* to p^* when I_{grad} is zero-dimensional.)

Henrion and Lasserre [11] give the following stopping criterion: If the optimum solution y to (4) satisfies the rank condition:

(6) $$\operatorname{rank} M_k(y) = \operatorname{rank} M_{k-d}(y), \text{ where } d := \max(d_1, \ldots, d_\ell),$$

then $p_k^* = p^*$. See Section 2.2 for details.

For our original unconstrained minimization problem (1) (then $\ell = 0$ and $K = \mathbb{R}^n$), we have: $p_{L,k}^* = p_{L,m}^* \le p^*$ for all $k \ge m$, with equality: $p_{L,m}^* = p^*$ if and only if $p - p^*$ is a sum of squares. One possible option to better approximate p^* is to transform the unconstrained problem (1) into a constrained problem of the form (3). This is possible if p attains its minimum as p^* can then be formulated as

(7) $$p^* = p_{grad}^* := \inf \ p(x) \text{ s.t. } \partial p(x)/\partial x_i = 0 \ (i = 1, \ldots, n).$$

The equality $p^* = p_{grad}^*$ does not hold in general if p does not attain its minimum; for instance, $p^* = 0$ and $p_{grad}^* = 1$ for $p(x_1, x_2) = x_1^2 + (x_1 x_2 - 1)^2$; $p^* = -\infty$ and $p_{grad}^* = 0$ for $p(x) = x^3$. If p has a minimum and if some upper bound R is known a priori on the norm of a global minimizer, then p^* can also be expressed as

(8) $$p^* = \min \ p(x) \text{ s.t. } \sum_{i=1}^n x_i^2 \le R^2.$$

A major drawback of approaches based on formulations like (7) or (8) is that it is not clear how to test whether a polynomial has a minimum and, for (8), how to find a ball containing a global minimizer. We will however present in Section 2.1 a result of Marshall [19] concerning a class of polynomials for which such a ball can be determined before hand.

1.2. **Our approach.** In this paper we propose the following strategy for going around these difficulties. Following Hanzon and Jibetean [9], we consider the perturbed polynomial p_λ from (2). As computing the exact limit p^* of the minima p_λ^* is not a realistic option for large problems, we set up to the less ambitious goal of computing a good upper approximation p_λ^* of p^* for some small value of λ. As mentioned earlier, the polynomial

p_λ enjoys several properties (that p may not have in general). Namely, p_λ attains its minimum, which can thus be formulated as

$$(9) \quad p_\lambda^* = \min_{x \in V_\lambda \cap \mathbb{R}^n} p_\lambda(x), \text{ where } V_\lambda := \{x \in \mathbb{C}^n \mid h_{\lambda,i}(x) := \frac{\partial p_\lambda}{\partial x_i}(x) = 0 \ (i = 1, \ldots, n)\}$$

and the set V_λ is finite ($|V_\lambda| \leq (2m+1)^n$). Moreover, one can give an explicit radius:

$$(10) \qquad\qquad\qquad\qquad R_\lambda = \frac{n^m}{\lambda} \sum_{\alpha \neq 0} |p_\alpha|$$

for a ball containing the global minima of p_λ (see Corollary 3); thus

$$(11) \qquad p_\lambda^* = \min_{x \in B_\lambda} p_\lambda(x), \text{ where } B_\lambda := \{x \in \mathbb{R}^n \mid h_{\lambda,0}(x) := R_\lambda^2 - \sum_{i=1}^n x_i^2 \geq 0\}.$$

By minimizing $p(x)$ over the algebraic set $V_\lambda \cap \mathbb{R}^n$ or over the ball B_λ, one obtains even better bounds μ_λ^* and β_λ^*, respectively; that is,

$$p^* \leq \mu_\lambda^* := \min_{x \in V_\lambda \cap \mathbb{R}^n} p(x) \leq p_\lambda^*, \ p^* \leq \beta_\lambda^* := \min_{x \in B_\lambda} p(x) \leq p_\lambda^*.$$

As the parameters μ_λ^* and β_λ^* are expressed via constrained polynomial programs of the form (3), a first option is to apply Lasserre's approach for computing them. Namely, for any integer $k \geq m + 1$, consider the programs:

$$(12) \qquad \mu_{L,k,\lambda}^* := \inf \ p^T y \ \text{ s.t. } \ y_0 = 1, \ M_k(y) \succeq 0, \ M_{k-m-1}(h_{\lambda,i} y) = 0 \ (i = 1, \ldots, n),$$

$$(13) \qquad\qquad \beta_{L,k,\lambda}^* := \inf \ p^T y \ \text{ s.t. } \ y_0 = 1, \ M_k(y) \succeq 0, \ M_{k-1}(h_{\lambda,0} y) \succeq 0.$$

Then,

$$\mu_{L,k,\lambda}^* \leq \mu_{L,k+1,\lambda}^* \leq \mu_\lambda^*, \ \beta_{L,k,\lambda}^* \leq \beta_{L,k+1,\lambda}^* \leq \beta_\lambda^* \ \text{ for } k \geq m + 1.$$

As k goes to infinity, there is asymptotic convergence of $\beta_{L,k,\lambda}^*$ to β_λ^* (recall Case (I)) and finite convergence of the parameters $\mu_{L,k,\lambda}^*$ to μ_λ^* (recall Case (II)).

As the set V_λ is finite, another option for computing the bound μ_λ^* is to apply the semidefinite representation result for finite varieties of Laurent [18]. Namely, μ_λ^* can be expressed as the optimum of the semidefinite program:

$$(14) \qquad\qquad \mu_\lambda^* = \min p^T y \ \text{ such that } M_{\mathcal{B}}(y) \succeq 0, \ y_0 = 1,$$

involving a combinatorial moment matrix $M_{\mathcal{B}}(y)$. Here, $y = (y_\beta)_{\beta \in \mathcal{B}} \in \mathbb{R}^{\mathcal{B}}$, where

$$\mathcal{B} := \{\beta \in \mathbb{Z}^n \mid 0 \leq \beta_i \leq 2m \ (i = 1, \ldots, n)\}$$

has the property that the set of monomials $\{x^\beta \mid \beta \in \mathcal{B}\}$ forms a basis of $\mathbb{R}[x_1, \ldots, x_n]/I_\lambda$, and I_λ is the ideal generated by $h_{\lambda,i} = \partial p_\lambda / \partial x_i$ ($i = 1, \ldots, n$). The matrix $M_{\mathcal{B}}(y)$ is obtained from a classical moment matrix by 'factoring' through I_λ which, roughly speaking, means that the equations $h_{\lambda,i}(x) = 0$ are used for expressing any y_α ($\alpha \in \mathbb{Z}_+^n$) in terms of y_β ($\beta \in \mathcal{B}$). As a by-product, this implies the finite convergence of the bounds $\mu_{L,k,\lambda}^*$ from (12) to μ_λ^*; more precisely, $\mu_{L,k,\lambda}^* = \mu_\lambda^*$ for $k \geq 2nm$ (by Theorem 23 in [18]).

The semidefinite program (14) is more compact than (12) (for any k ensuring finite convergence). Indeed, the program (14) involves only one linear matrix inequality (LMI) and $|\mathcal{B}| = (2m+1)^n$ variables whereas (12) involves $n + 1$ LMI's and $\binom{n+2k}{2k}$ variables.

Moreover the size of the matrix $M_{\mathcal{B}}(y)$ is $|\mathcal{B}| = (2m+1)^n$, which is smaller than the size $\binom{n+k}{k}$ of the matrix $M_k(y)$ for any $k \geq 2nm$. Solving the semidefinite program (14) is, however, still out of reach for large n or m. Moreover, the entries of $M_{\mathcal{B}}(y)$ are polynomial in $1/\lambda$ (and linear in y) and thus, for λ close to 0, they may be ill-conditioned. These difficulties can be addressed in the following way. Given an integer k, $m \leq k \leq 2nm$, consider the truncated semidefinite program obtained by considering the principal submatrix of $M_{\mathcal{B}}(y)$, denoted $M_{\mathcal{B}_k}(y)$, indexed by the subset $\mathcal{B}_k := \mathcal{B} \cap S_k$ and set

(15) $$\mu_{k,\lambda}^* := \inf p^T y \text{ such that } M_{\mathcal{B}_k}(y) \succeq 0, y_0 = 1.$$

Thus,

$$\mu_{k,\lambda}^* \leq \mu_{k+1,\lambda}^* \leq \mu_{2nm,\lambda}^* = \mu_\lambda^*.$$

When the optimum solution $M_{\mathcal{B}_k}(y)$ satisfies the following rank condition:

(16) $$\text{rank} M_{\mathcal{B}_h}(y) = \text{rank} M_{\mathcal{B}_{h-1}}(y)$$

for some $m \leq h \leq k$, one can conclude that the optimum value of the truncated problem (15) is an upper bound for the infimum p^*; that is, $p^* \leq \mu_{k,\lambda}^* \leq \mu_\lambda^*$. Moreover, one can extract a point x for which $p^* \leq p(x) \leq \mu_{k,\lambda}^*$, thus giving a certificate for the claimed upper bound $\mu_{k,\lambda}^*$ on p^* (see Corollary 19). In this way, one is (often) able to compute a very good upper approximation of p^* by solving a much smaller semidefinite program. Moreover the degree in $1/\lambda$ of the entries of $M_{\mathcal{B}_k}(y)$ is at most $k - m$ (see Theorem 18) and thus remains small for small values of k. Several examples illustrating this procedure will be given in Section 3.2. In most cases one is able to conclude that the parameter $\mu_{k,\lambda}^*$ from the program (15) is an upper bound for p^* already for $k = m+1$ or $m+2$, in which case the entries of $M_{\mathcal{B}_k}(y)$ are at most quadratic in $1/\lambda$ and we are thus able to carry out the computations for a small perturbation parameter $\lambda \sim 10^{-4}$ and sometimes even smaller. By the results of [9], for such small λ, the extracted minimizer x_λ is very close to a global minimizer of p (if some exists); this will be verified on the examples.

Given an integer $k \geq m$, the program (15) can be seen as a 'compact' analogue of the program (12). We can prove the following interlacing property for their optimal values (see Theorem 17):

(17) $$\mu_{k,\lambda}^* \leq \mu_{L,k+1,\lambda}^* \leq \mu_{k+1,\lambda}^*$$

for $m \leq k \leq 2nm$, with equality $\mu_{2nm,\lambda}^* = \mu_{L,2nm,\lambda}^* = \mu_\lambda^*$; see Examples 4, 5, 6 for a numerical comparison. Program (12) involves matrices of size $|S_k| = \binom{n+k}{k}$ and $|S_{2k}| = \binom{n+2k}{2k}$ variables, whereas its compact analogue (15) involves matrices of size $|\mathcal{B}_k| = |S_k \cap \mathcal{B}|$ and $|\mathcal{B}_{2k}| = |S_{2k} \cap \mathcal{B}|$ variables. For $k \leq 2m$, $\mathcal{B}_k = S_k$, but $\mathcal{B}_{2k}$ is then already significantly smaller than S_{2k}. This is illustrated in the next table, which displays some values of $|S_{2k} \setminus \mathcal{B}_{2k}| = |S_{2k} \setminus \mathcal{B}|$ for $k = m+1, m+2$.

	$|S_{2m+2} \setminus \mathcal{B}|$	$|S_{2m+4} \setminus \mathcal{B}|$ for $m \geq 2$	$|S_{2m+4} \setminus \mathcal{B}|$ for $m = 1$
n	$n(n+1)$	$4n + 12\binom{n}{2} + 12\binom{n}{3} + 4\binom{n}{4}$	$4n + 11\binom{n}{2} + 12\binom{n}{3} + 4\binom{n}{4}$
$n = 2$	6	20	19
$n = 3$	12	60	57
$n = 4$	20	140	136
$n = 5$	30	280	275
$n = 10$	110	2860	2850

Gain in number of variables when using program (15) instead of program (12)

1.3. **Contents of the paper.** The paper is organized as follows. Section 2 contains preliminaries about polynomials and about classical and combinatorial moment matrices and their application to polynomial optimization. In Section 3, we present our method for computing the upper approximations μ_λ^* for the infimum p^* of a polynomial p over $\mathbb{R}^n$ and, in Section 3.2, several examples on which our method has been tested.

2. PRELIMINARIES

2.1. **Polynomials.** We begin with some preliminaries on ideals of polynomials. Throughout the paper, $\mathbb{R}[x_1, \ldots, x_n]$ denotes the ring of real polynomials in n indeterminates. For an integer $k \geq 0$, S_k denotes the set of $\alpha \in \mathbb{Z}_+^n$ with $|\alpha| := \sum_{i=1}^n \alpha_i \leq k$. Write a polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ with (total) degree at most k as $p(x) = \sum_{\alpha \in S_k} p_\alpha x^\alpha$, where x^α denotes the monomial $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. As usual, we identify a polynomial p of degree at most k with the sequence of its coefficients $p = (p_\alpha)_{\alpha \in S_k}$.

Let I be an ideal in $\mathbb{R}[x_1, \ldots, x_n]$. The set $V = V(I) := \{x \in \mathbb{C}^n \mid f(x) = 0 \ \forall f \in I\}$ is its associated (complex) variety. The ideal I is said to be *zero-dimensional* if $|V| < \infty$. The sets $I(V) := \{f \in \mathbb{R}[x_1, \ldots, x_n] \mid f(v) = 0 \ \forall v \in V\}$, and $\sqrt{I} := \{f \in \mathbb{R}[x_1, \ldots, x_n] \mid f^k \in I$ for some integer $k \geq 1\}$ are again ideals in $\mathbb{R}[x_1, \ldots, x_n]$, which obviously contain the ideal I. The Nullstellensatz asserts that these two ideals coincide; namely, $\sqrt{I} = I(V)$. The ideal I is said to be *radical* when $I = \sqrt{I}$. Hence, by the Nullstellensatz,

$$(18) \qquad I \text{ is radical} \iff \begin{array}{l} \text{the polynomials vanishing at all points of } V \\ \text{are precisely the polynomials in } I. \end{array}$$

The following result, relating the dimension of the quotient vector space $\mathbb{R}[x_1, \ldots, x_n]/I$ and the cardinality of V, can be found, e.g., in [2, §5.3].

$$(19) \qquad \begin{array}{c} |V| < \infty \iff \dim \mathbb{R}[x_1, \ldots, x_n]/I < \infty, \\ |V| \leq \dim \ \mathbb{R}[x_1, \ldots, x_n]/I, \text{ with equality if and only if } I \text{ is radical.} \end{array}$$

We now recall a result of Marshall [19] giving a sufficient condition for a polynomial to have a minimum. Given a nonzero polynomial p, let $\tilde{p}$ be its *highest degree homogeneous component*, defined as the sum of the terms of p having maximum degree, and set

$$\tilde{p}_S := \min_{x \in S} \tilde{p}(x) \ \text{ where } S := \{x \in \mathbb{R}^n \mid \sum_{I=1}^n x_i^2 = 1\}.$$

If $\tilde{p}_S < 0$ then p has obviously an infinite infimum, i.e., $p^* = -\infty$. If $\tilde{p}_S > 0$ then, following Marshall [19], p is said to be *stably bounded from below* and, as the next result shows, p

attains its minimum. On the other hand, no conclusion can be drawn when $\tilde{p}_S = 0$; indeed, p may have an infinite infimum (e.g., for $p(x_1, x_2) = x_1^2 + x_2$), or a finite infimum (e.g., for $p(x_1, x_2) = x_1^2 + (x_1 x_2 - 1)^2$), or a minimum (e.g., for $p(x_1, x_2) = x_1^2 x_2^2$).

LEMMA 1. [19] *Assume p is stably bounded from below. Given $x \in \mathbb{R}^n$,*

$$(20) \qquad p(x) \leq 0 \Longrightarrow \|x\| \leq max(\frac{1}{\tilde{p}_S} \sum_{\alpha : |\alpha| \leq \deg(p)-1} |p_\alpha|, 1).$$

In particular, any global minimum of p belongs to the ball centered at the origin with radius
$R_p := \max(1, \frac{1}{\tilde{p}_S} \sum_{\alpha : 1 \leq |\alpha| \leq \deg(p)-1} |p_\alpha|).$

Proof. Say p has degree d and write $p = \tilde{p} + g$, where all terms of $\tilde{p}$ have degree d and all terms of g have degree $\leq d - 1$. Let $x \in \mathbb{R}^n \setminus \{0\}$ such that $p(x) \leq 0$. Thus, $\tilde{p}(x) \leq -g(x) \leq \sum_{\alpha : |\alpha| \leq d-1} |p_\alpha| |x^\alpha|$. By assumption, $\tilde{p}(x) = \|x\|^d \tilde{p}(\frac{x}{\|x\|}) \geq \|x\|^d \tilde{p}_S > 0$. On the other hand, if $\|x\| \geq 1$ and $|\alpha| \leq d - 1$, then $|x^\alpha| \leq \|x\|^{|\alpha|} \leq \|x\|^{d-1}$. Combining these two facts, we find the relation (20). If x is a global minimum of p, then $p(x) \leq p(0)$ and thus $\|x\| \leq R_p$ follows from (20) applied to the polynomial $p - p(0)$. ∎

In general, the polynomial p may not be stably bounded from below and it may not even have a minimum. However, for any positive λ, the perturbed polynomial p_λ is stably bounded from below. Indeed, if p has degree $2m$, then the highest degree homogeneous component of p_λ is equal to $\lambda \sum_{i=1}^n x_i^{2m+2}$, whose minimum value over the unit sphere is equal to $\frac{\lambda}{n^m}$ as the next lemma shows.

LEMMA 2. *Given an integer $m \geq 2$, the minimum value taken by $\sum_{i=1}^n x_i^{2m}$ over the unit sphere is equal to $\frac{1}{n^{m-1}}$.*

Proof. By evaluating $f(x) := \sum_i x_i^{2m}$ at the point $x := \frac{1}{\sqrt{n}}(1, \ldots, 1)$, we find that the minimum value f_S of f over the unit sphere is at most $\frac{1}{n^{m-1}}$. To show the reverse inequality, note that f_S is equal to the minimum value of $g(x) := \sum_{i=1}^n x_i^m$ over $x \in \mathbb{R}_+^n$ with $\sum_{i=1}^n x_i = 1$. Let x be a minimizer to this program. Applying the Karush-Kuhn-Tucker conditions, there exist $\lambda \in \mathbb{R}$, $z \in \mathbb{R}_+^n$ such that $\nabla g(x) - \lambda e - z = 0$ and $x^T z = 0$. As $x, z \geq 0$, $x_i z_i = 0$ for all i and $\frac{\partial g}{\partial x_i}(x) = \lambda$ if $z_i = 0$. Say $z_1 = \ldots = z_p = 0$, $z_{p+1}, \ldots, z_n > 0$ for some $p \leq n$; thus $x_{p+1} = \ldots = x_n = 0$. For $i = 1, \ldots, p$, $\frac{\partial g}{\partial x_i}(x) = \lambda = m x_i^{m-1}$. From this follows that $x_1 = \ldots = x_p = \frac{1}{p}$. Now, $g(x) = \frac{1}{p^{m-1}} \geq \frac{1}{n^{m-1}}$ as $p \leq n$. ∎

COROLLARY 3. *Given a polynomial p of degree $2m$, the global minima of the perturbed polynomial $p_\lambda(x) = p(x) + \lambda(\sum_{i=1}^n x_i^{2m+2})$ are located in the ball B_λ with radius $R_\lambda := \frac{n^m}{\lambda} \sum_{\alpha \neq 0} |p_\alpha|$.* ∎

7

2.2. **Moment matrices.** We recall here some results about moment matrices, that we need in the paper. Given a probability measure μ on $\mathbb{R}^n$, the quantity $y_\alpha := \int x^\alpha \mu(dx)$ is called its *moment of order* α. A probability measure with finite support is of the form: $\mu = \sum_{i=1}^r \lambda_i \delta_{x_i}$, where $\lambda_i > 0$, $\sum_{i=1}^r \lambda_i = 1$, $x_i \in \mathbb{R}^n$ (the *atoms* of the measure); then μ is said to be *r-atomic*. Here, δ_x is the Dirac measure at $x \in \mathbb{R}^n$, having mass 1 at x and mass 0 elsewhere.

The moment problem concerns the characterization of the sequences $y \in \mathbb{R}^{S_{2k}}$ ($k \geq 1$) that are the sequences of moments of some probability measure μ; in that case one also says that μ is a representing measure for y. Given $y \in \mathbb{R}^{S_{2k}}$, its *moment matrix of order* k is the matrix $M_k(y)$ indexed by S_k with (α, β)-th entry $y_{\alpha+\beta}$, for $\alpha, \beta \in S_k$. Given a polynomial $h(x)$ of degree $2d$ or $2d - 1$, define the vector hy with entries $(hy)_\alpha := \sum_\gamma h_\gamma y_{\alpha+\gamma}$ for $\alpha \in S_{2k-2d}$; $M_{k-d}(hy)$ is known as a *localizing moment matrix*. A well known necessary condition for the existence of a representing measure for y is the positive semidefiniteness of its moment matrix.

LEMMA 4. *If $y \in \mathbb{R}^{S_{2k}}$ has a representing measure μ, then $M_k(y) \succeq 0$. Moreover, if the support of μ is contained in the set $\{x \mid h(x) \geq 0\}$, where $h(x)$ is a polynomial of degree $2d$ or $2d - 1$, then $M_{k-d}(hy) \succeq 0$.*

Proof. For $p \in \mathbb{R}^{S_k}$, we have:

$$p^T M_k(y) p = \sum_{\alpha, \beta \in S_k} p_\alpha p_\beta y_{\alpha+\beta} = \sum_{\alpha, \beta \in S_k} p_\alpha p_\beta \int x^{\alpha+\beta} d\mu(x) = \int p(x)^2 d\mu(x) \geq 0,$$

which shows that $M_k(y) \succeq 0$. If the support of μ is contained in $\{x \mid h(x) \geq 0\}$, one can verify that $p^T M_{k-d}(hy) p = \int p(x)^2 h(x) d\mu(x) \geq 0$ for all $p \in \mathbb{R}^{S_{k-d}}$, which shows that $M_{k-d}(hy) \succeq 0$. ∎

Curto and Fialkow [4, 5] prove some results showing that, under some rank condition, the necessary conditions from the above lemma are also sufficient for the existence of a representing measure. A key notion is that of 'flat extension'. Let X be a symmetric matrix and let A be a principal submatrix of X. One says that X is a *flat extension* of A if rank X = rank A. Then, $X \succeq 0 \iff A \succeq 0$.

THEOREM 5. [4] *Let $y \in \mathbb{R}^{S_{2k}}$. If $M_k(y) \succeq 0$ and $M_k(y)$ is a flat extension of $M_{k-1}(y)$, then y has a representing measure which is $(\text{rank } M_k(y))$-atomic.*

The proof uses the following property of the kernel of $M_k(y)$, which also permits to derive Corollary 7 below.

LEMMA 6. [4] *Assume that $M_k(y) \succeq 0$ and let $f, g \in \mathbb{R}[x_1, \ldots, x_n]$ whose product $h := fg$ has degree $\deg(h) \leq k - 1$. Then, $M_k(y)f = 0$ implies $M_k(y)h = 0$.*

COROLLARY 7. *If $M_k(y) \succeq 0$ and rank $M_h(y)$ = rank $M_{h-1}(y)$ for some $1 \leq h \leq k - 1$, then rank $M_{k-1}(y)$ = rank $M_{k-2}(y)$.*

THEOREM 8. *([5], see [17] for a short proof) Let $y \in \mathbb{R}^{S_{2k}}$, $h_1, \ldots, h_\ell \in \mathbb{R}[x_1, \ldots, x_n]$, $d_i := \lceil deg(h_i)/2 \rceil$, and $d := \max(d_1, \ldots, d_\ell)$. Assume that $M_k(y) \succeq 0$, $M_{k-d_i}(h_i y) \succeq 0$*

(for $i = 1, \ldots, \ell$), and rank $M_k(y) =$ rank $M_{k-d}(y)$. *Then y has a representing measure μ supported by the set $\{x \mid h_1(x) \geq 0, \ldots, h_\ell(x) \geq 0\}$; moreover μ is* (rank $M_k(y)$)*-atomic.*

The above results underlie the semidefinite relaxations (4) and (5) of problem (3). In particular, as an application of Theorem 8, one finds the stopping criterion of Henrion and Lasserre [11]: If $M_k(y)$ is an optimum solution to (4) satisfying the rank condition (6), then $p^*_{L,k} = p^*$. This is a very useful fact, as it permits very often in practice to conclude that the relaxation (4) of a given order k solves the original problem (3) at optimality for small values of k. The following two results imply the asymptotic (or finite) convergence of the parameters ρ^*_k and $p^*_{L,k}$ to the optimum p^*, in the cases (I) and (III) mentioned in Section 1.1.

THEOREM 9. [26] *Let $K = \{x \in \mathbb{R}^n \mid h_1(x) \geq 0, \ldots, h_\ell(x) \geq 0\}$ and $M := \{u_0 + \sum_{i=1}^{\ell} u_i h_i \mid u_0, u_1, \ldots, u_\ell$ are sums of squares of polynomials\}. Assume that K is compact and that there exists a polynomial $u \in M$ for which the set $\{x \in \mathbb{R}^n \mid u(x) \geq 0\}$ is compact. Then every positive polynomial on K belongs to M.*

THEOREM 10. [6] *Given a polynomial p, define $K := \{x \in \mathbb{R}^n \mid \frac{\partial p}{\partial x_i}(x) = 0 \ (i = 1, \ldots, n)\}$ and let I_{grad} be the ideal generated by $\partial p / \partial x_i \ (i = 1, \ldots, n)$. If p is positive on K then p is a sum of squares of polynomials modulo I_{grad}. When I_{grad} is radical, the same conclusion holds if p is nonnegative on K.*

2.3. **Combinatorial moment matrices.** Let I be a zero-dimensional ideal in $\mathbb{R}[x_1, \ldots, x_n]$ with $V = V(I)$ as associated complex variety. With respect to a given monomial ordering, let G be a Gröbner basis of I and let $\mathcal{S}$ be the associated set of standard monomials, consisting of the monomials that are not divisible by the leading term of any polynomial in G. Let $\mathcal{B}$ be the set of exponents of the standard monomials; that is, $\mathcal{S} = \{x^\beta \mid \beta \in \mathcal{B}\}$. The set $\mathcal{S}$ is a basis of $\mathbb{R}[x_1, \ldots, x_n]/I$; that is, for every polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$, there exists a unique polynomial $r(x) = \sum_{\beta \in \mathcal{B}} r_\beta x^\beta$ for which $f - r \in I$; r is called the *residue* of f modulo I.

Given $y = (y_\beta)_{\beta \in \mathcal{B}} \in \mathbb{R}^{\mathcal{B}}$, let $M_{\mathcal{B}}(y)$ be the $\mathcal{B} \times \mathcal{B}$ matrix whose (α, β)-th entry is equal to $\sum_{\gamma \in \mathcal{B}} r_\gamma y_\gamma$, for $\alpha, \beta \in \mathcal{B}$, where $\sum_{\gamma \in \mathcal{B}} r_\gamma x^\gamma$ is the residue of $x^\alpha x^\beta$ modulo I; $M_{\mathcal{B}}(y)$ is called the *combinatorial moment matrix* of y. In words, $M_{\mathcal{B}}(y)$ is obtained from a classical moment matrix by expressing all entries of y in terms of those indexed by the standard monomials using the equations defining I. For $v \in \mathbb{R}^n$, define the vector $\zeta_v := (v^\beta)_{\beta \in \mathcal{B}} \in \mathbb{R}^{\mathcal{B}}$. It is not difficult to check that, if $v \in V \cap \mathbb{R}^n$, then $M_{\mathcal{B}}(\zeta_v) = \zeta_v \zeta_v^T$ is positive semidefinite. Hence, $M_{\mathcal{B}}(y) \succeq 0$ if y belongs to the cone generated by the vectors $\zeta_v \ (v \in V \cap \mathbb{R}^n)$. Laurent [18] shows that equivalence holds.

THEOREM 11. [18] *Let I be a zero-dimensional ideal in $\mathbb{R}[x_1, \ldots, x_n]$, let V be the associated variety, and let $\{x^\beta \mid \beta \in \mathcal{B}\}$ be the set of standard monomials with respect to some monomial ordering. Let $y \in \mathbb{R}^{\mathcal{B}}$ and $M_{\mathcal{B}}(y)$ its associated combinatorial moment matrix. Then, $M_{\mathcal{B}}(y) \succeq 0$ if and only if y belongs to the cone generated by $\zeta_v \ (v \in V \cap \mathbb{R}^n)$; that is, y is the sequence of moments (of order $\alpha \in \mathcal{B}$) of a nonnegative atomic measure μ whose support is contained in $V \cap \mathbb{R}^n$.*

2.4. Truncated combinatorial moment matrices. We assume in this section and the next one that the ideal I is generated by n polynomials of the form:

$$(21) \qquad h_i(x) := x_i^{2m+1} - \tilde{h}_i(x), \quad \text{for } i = 1, \ldots, n$$

where $\deg(\tilde{h}_i) \le 2m$ and $m \ge 1$ is a given integer. In that case, we can prove some results about flat extensions of truncated combinatorial moment matrices, which will be useful for our application to optimization.

The polynomials $h_1, \ldots, h_n$ form a Gröbner basis of the ideal I (w.r.t. a total degree monomial ordering) (apply [2, §2.6]). Therefore, the set of standard monomials is $\mathcal{S} = \{x^\beta \mid \beta \in \mathcal{B}\}$, where

$$(22) \qquad \mathcal{B} := \{\beta \in \mathbb{Z}^n \mid 0 \le \beta_i \le 2m \; \forall i = 1, \ldots, n\}.$$

It follows from (19) that the ideal I is zero-dimensional. Given an integer $1 \le k \le 2nm$, define

$$(23) \qquad \mathcal{B}_k := \mathcal{B} \cap S_k = \{\beta \in \mathcal{B} \mid |\beta| \le k\}.$$

LEMMA 12. *Given $f \in \mathbb{R}[x_1, \ldots, x_n]$, let r be its residue modulo I. Then, $\deg(r) \le \deg(f)$.*

Proof. Fix a total degree monomial ordering. Then the division algorithm applied for dividing f by $h_1, \ldots, h_n$ yields a decomposition $f = \sum_{i=1}^n u_i h_i + r$, where $r(x) = \sum_{\beta \in \mathcal{B}} r_\beta x^\beta$ is the residue of f, and $\deg(u_i h_i) \le \deg(f)$ whenever $u_i \ne 0$ (see [2, §2.3]). Therefore, $\deg(r) \le \deg(f)$. ∎

For a monomial x^α, let $r^{(\alpha)}(x)$ denote its residue modulo I; by Lemma 12, $r^{(\alpha)}(x)$ is of the form $r^{(\alpha)}(x) = \sum_{\beta \in \mathcal{B}_k} r_\beta^{(\alpha)} x^\beta$ if $|\alpha| \le k$. Therefore, given a truncated sequence $y \in \mathbb{R}^{\mathcal{B}_{2k}}$, one can define its truncated combinatorial moment matrix $M_{\mathcal{B}_k}(y)$ as the matrix indexed by $\mathcal{B}_k$ whose (α, β)-th entry is $y^T r^{(\alpha+\beta)}$, for $\alpha, \beta \in \mathcal{B}_k$. We now indicate how to extend a combinatorial moment matrix to a classical moment matrix.

DEFINITION 13. *Given $y \in \mathbb{R}^{\mathcal{B}_{2k}}$, extend y to $\tilde{y} \in \mathbb{R}^{S_{2k}}$ by setting*

$$(24) \qquad \tilde{y}_\gamma := y^T r^{(\gamma)} \text{ for } \gamma \in S_{2k}$$

where $r^{(\gamma)}(x)$ is the residue of x^γ modulo I.

LEMMA 14. *Let $y \in \mathbb{R}^{\mathcal{B}_{2k}}$ and $\tilde{y} \in \mathbb{R}^{S_{2k}}$ its extension from (24). Let f be a polynomial of degree at most $2k$ and r its residue modulo I. Then $f^T \tilde{y} = r^T y$.*

Proof. Using (24), we find that $f^T \tilde{y} = \sum_\delta f_\delta \tilde{y}_\delta = \sum_\delta f_\delta y^T r^{(\delta)} = \sum_{\beta, \delta} f_\delta r_\beta^{(\delta)} y_\beta$, while $y^T r = \sum_\beta r_\beta y_\beta$. Hence it suffices to show that the two polynomials $r(x)$ and $s(x) := \sum_{\beta, \delta} f_\delta r_\beta^{(\delta)} x^\beta$ are identical. For this, note that $s(x) = \sum_\delta f_\delta r^{(\delta)}(x) \equiv \sum_\delta f_\delta x^\delta = f(x) \equiv r(x)$ modulo I. Hence, $r = s$, since both r and s are polynomials using only standard monomials. ∎

LEMMA 15. *Let $y \in \mathbb{R}^{\mathcal{B}_{2k}}$ and $\tilde{y} \in \mathbb{R}^{S_{2k}}$ its extension from (24). Then $M_k(\tilde{y})$ is a flat extension of $M_{\mathcal{B}_k}(y)$.*

Proof. By the definition of $\tilde{y}$, the principal submatrix of $M_k(\tilde{y})$ indexed by $\mathcal{B}_k$ coincides with $M_{\mathcal{B}_k}(y)$. Consider a column C_γ of $M_k(\tilde{y})$ indexed by $\gamma \in S_k \setminus \mathcal{B}_k$. We verify that $C_\gamma = \sum_{\beta \in \mathcal{B}_k} r_\beta^{(\gamma)} C_\beta$; that is,

$$\tilde{y}_{\alpha+\gamma} = \sum_{\beta \in \mathcal{B}_k} r_\beta^{(\gamma)} \tilde{y}_{\alpha+\beta} \ \forall \alpha \in S_k.$$

For this consider the polynomial $f(x) := x^{\alpha+\gamma} - \sum_{\beta \in \mathcal{B}_k} r_\beta^{(\gamma)} x^{\alpha+\beta}$. As f has degree at most $2k$ and $f \in I$, it follows from Lemma 14 that $f^T \tilde{y} = 0$, which gives the desired relation. ∎

COROLLARY 16. *Let $y \in \mathbb{R}^{\mathcal{B}_{2k}}$ and $\tilde{y} \in \mathbb{R}^{S_{2k}}$ its extension from (24). Assume that $M_{\mathcal{B}_h}(y)$ is a flat extension of $M_{\mathcal{B}_{h-1}}(y)$ for some $1 \le h \le k$. [Then this holds for $h = k$ or $k-1$ by Corollary 7.] Then $(\tilde{y}_\alpha)_{\alpha \in S_{2h}}$ (and thus $(y_\alpha)_{\alpha \in \mathcal{B}_{2h}}$) is the sequence of moments of an r-atomic measure μ, where $r := \mathrm{rank}\, M_{\mathcal{B}_h}(y)$. Moreover, if $h \ge 2m+1$, then the support of μ is contained in V.*

Proof. By Lemma 15, $M_h(\tilde{y})$ is a flat extension of $M_{h-1}(\tilde{y})$. Hence, by Theorem 5, $(y_\alpha)_{\alpha \in S_{2h}}$ has a r-atomic representing measure μ, where $r = \mathrm{rank}\, M_h(\tilde{y}) = \mathrm{rank}\, M_{\mathcal{B}_h}(y)$. If $h \ge 2m+1$, then the polynomials $h_i(x)$ $(i = 1, \ldots, n)$ generating the ideal I belong to the kernel of $M_h(\tilde{y})$ (by the construction of $\tilde{y}$). Hence, the support of μ is contained in the set of common zeros of the h_i's, i.e., in the variety V. ∎

2.5. **Optimization and extraction of solutions.** Given a polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$, consider the problem:

$$p^* := \min \ p(x) \ \text{s.t.} \ h_1(x) = 0, \ldots, h_n(x) = 0$$

where $h_1, \ldots, h_n$ are as in (21). We can assume that p has degree at most $2m$, else replace p by its residue modulo the ideal I. We first compare the following two hierarchies of lower bounds for p^*, defined for $k \ge m$:

(25) $$\mu_k^* := \inf p^T y \text{ such that } M_{\mathcal{B}_k}(y) \succeq 0, \ y_0 = 1,$$

(26) $$\mu_{L,k}^* := \inf p^T y \text{ such that } M_k(y) \succeq 0, \ M_{k-m-1}(h_i y) = 0 \ (i = 1, \ldots, n), \ y_0 = 1$$

where we omit the condition $M_{k-m-1}(h_i y) = 0$ when $k = m$. If $k = m$, the two programs (25) and (26) are identical and thus $\mu_m^* = \mu_{L,m}^*$. Moreover, by Theorem 11 (and Theorem 23 in [18]), $\mu_{2nm}^* = \mu_{L,2nm}^* = p^*$. One can show the following interlacing property for the parameters μ_k^* and $\mu_{L,k}^*$, which implies the interlacing property (17) for the two hierarchies of bounds from (12) and (15).

THEOREM 17. $\mu_{k-1}^* \le \mu_{L,k}^* \le \mu_k^*$ *for all $k \ge m+1$.*

Proof. Let z be a feasible solution to (26), i.e., $M_k(z) \succeq 0$, $M_{k-m-1}(h_i z) = 0$, $z_0 = 1$. We observe first that $f^T z = 0$ for every polynomial $f \in I$ with degree at most $2k-1$. Indeed, as

$f \in I$, $f = \sum_{i=1}^n u_i h_i$ where $\deg(u_i h_i) \le \deg(f) \le 2k-1$, i.e., $\deg(u_i) \le 2k-1-(2m+1) = 2k-2m-2$ whenever $u_i \ne 0$. Moreover, $f(x) = \sum_{i=1}^n \sum_{\gamma,\delta} (u_i)_\gamma (h_i)_\delta x^{\gamma+\delta}$. Hence,

$$f^T z = \sum_\beta f_\beta z_\beta = \sum_\beta z_\beta \sum_{i=1}^n \sum_{\gamma,\delta|\gamma+\delta=\beta} (u_i)_\gamma (h_i)_\delta = \sum_{i=1}^n \sum_\gamma (u_i)_\gamma \sum_\delta (h_i)_\delta z_{\gamma+\delta}.$$

Now, $\sum_\delta (h_i)_\delta z_{\gamma+\delta} = (h_i z)_\gamma = 0$ since $|\gamma| \le \deg(u_i) \le 2k-2m-2$ and $M_{k-m-1}(h_i z) = 0$. Therefore, we find that $f^T z = 0$. Hence, if we denote by y the restriction of z to $\mathbb{R}^{\mathcal{B}_{2k}}$, then $z_\gamma = y^T r^{(\gamma)}$ for $|\gamma| \le 2k-1$. Hence $M_{\mathcal{B}_{k-1}}(y)$ coincides with the principal submatrix of $M_k(z)$ indexed by $\mathcal{B}_{k-1}$ and thus $M_{\mathcal{B}_{k-1}}(y) \succeq 0$. This implies that $p^T z = p^T y \ge \mu_{k-1}^*$ and thus $\mu_{L,k}^* \ge \mu_{k-1}^*$.

Consider now a feasible solution $y \in \mathbb{R}^{\mathcal{B}_{2k}}$ to (25). Let $\tilde{y}$ be its extension to $\mathbb{R}^{S_{2k}}$ from (24). Then, $M_k(\tilde{y}) \succeq 0$ by Lemma 15. Remains to verify that $M_{k-m-1}(h_i \tilde{y}) = 0$, i.e., that $(h_i \tilde{y})_\alpha = \sum_\gamma (h_i)_\gamma \tilde{y}_{\alpha+\gamma}$ is equal to 0 for $|\alpha| \le 2k-2m-2$. As the polynomial $f(x) := h_i(x) x^\alpha$ belongs to I and its degree is at most $2k$, it follows from Lemma 14 that $f^T \tilde{y} = 0$, which gives the desired relation. Hence, $\tilde{y}$ is feasible for (26), which implies that $p^T y = p^T \tilde{y} \ge \mu_{L,k}^*$ and thus $\mu_k^* \ge \mu_{L,k}^*$. ∎

Let y be an optimum solution to (25). Assume that $\mathrm{rank} M_{\mathcal{B}_h}(y) = \mathrm{rank} M_{\mathcal{B}_{h-1}}(y) =: r$ for some $1 \le h \le k$. By Corollary 16, $(y_\beta)_{\beta \in \mathcal{B}_{2h}}$ is the sequence of moments of a measure $\mu = \sum_{i=1}^r \lambda_i \delta_{v_i}$ ($\lambda_i > 0$, $\sum_i \lambda_i = 1$, $v_i \in \mathbb{R}^n$). If $h \ge m$, then $p^* \ge \mu_k^* = p^T y = \sum_i \lambda_i p(v_i) \ge \min_i p(v_i)$; moreover, $v_1, \ldots, v_r$ belong to $V(I)$ and thus are global minimizers of p over the set $\{x \in \mathbb{R}^n \mid h_1(x) = \ldots = h_n(x) = 0\}$ when $h \ge 2m+1$. We now indicate how to extract the points $v_1, \ldots, v_r$ from the matrix $M_{\mathcal{B}_h}(y)$; this is analogous to the extraction procedure in [11] (for the program (26)).

As $\mathrm{rank} M_{\mathcal{B}_h}(y) = \mathrm{rank} M_{\mathcal{B}_{h-1}}(y) = r$, one can find a subset $\mathcal{A}$ of $\mathcal{B}_{h-1}$, $|\mathcal{A}| = r$, indexing a positive definite principal submatrix A of $M_{\mathcal{B}_h}(y)$. If $h \le 2m$, let J denote the ideal generated by the kernel of $M_{\mathcal{B}_h}(y)$ and, if $h \ge 2m+1$, let J be the ideal generated by I and the kernel of $M_{\mathcal{B}_h}(y)$. Obviously, $\{v_1, \ldots, v_r\} \subseteq V(J)$. On the other hand, $\mathcal{A}$ is a basis of $\mathbb{R}[x_1, \ldots, x_n]/J$ (easy to verify) and thus $\dim \mathbb{R}[x_1, \ldots, x_n]/J = r$, which implies that $|V(J)| \le r$ (by (19)). Therefore, $V(J) = \{v_1, \ldots, v_r\}$ and J is a zero-dimensional radical ideal. Thus, determining $v_1, \ldots, v_r$ amounts to finding the commom zeros to the polynomials in J, which can be done with the eigenvalue method, briefly described below (see, e.g., chap. 2 §4 in [3]).

For a polynomial f, the *multiplication matrix* M_f is the $|\mathcal{A}| \times |\mathcal{A}|$ matrix whose α-th column (for $\alpha \in \mathcal{A}$) contains the coefficients in the base $\mathcal{A}$ of the residue modulo J of the polynomial $x^\alpha f(x)$. If f is chosen in such a way that the values $f(v)$ are distinct for $v \in V(J)$, then the right eigenspaces of M_f are 1-dimensional and spanned by the vectors $(v^\alpha)_{\alpha \in \mathcal{A}}$ (for $v \in V(J)$) (Proposition 4.7 in [3]). Hence, the points $v_1, \ldots, v_r$ of $V(J)$ can be determined from the right eigenvectors of M_f.

In our extraction procedure, we construct the base $\mathcal{A}$ in a 'greedy manner'; starting from the constant monomial 1, we insert in $\mathcal{A}$ as many low degree monomials as possible. Then, given an eigenvector $(v^\alpha)_{\alpha \in \mathcal{A}}$ (or a scalar multiple of it), it is easy to recover

the components of v (in fact, immediate, if $\mathcal{A}$ contains the monomials $x_1, \ldots, x_n$). We determine the multiplication matrices M_{x_i} (for $f = x_i$, $i = 1, \ldots, n$) in the following way. As before let A be the principal submatrix of $M_{\mathcal{B}_h}(y)$ indexed by $\mathcal{A}$ and let U_i be the submatrix of $M_{\mathcal{B}_h}(y)$ with row indices $\mathcal{A}$ and column indices the set $x_i \mathcal{A} := \{x_i x^\alpha \mid \alpha \in \mathcal{A}\}$. When $h \leq 2m$ (which is the case considered for practical applications), $M_{x_i} = A^{-1} U_i$. (Indeed, given $\beta \in \mathcal{A}$, let v be the column of $M_{\mathcal{B}_h}(y)$ indexed by $x_i x^\beta$, $u := (v_\alpha)_{\alpha \in \mathcal{A}}$ the corresponding column of U_i, and $c = (c_\alpha)_{\alpha \in \mathcal{A}}$ the unique scalars permitting to express v as $v = \sum_{\alpha \in \mathcal{A}} c_\alpha C_\alpha$; C_α being the column of $M_{\mathcal{B}_h}(y)$ indexed by x^α. Then $c = A^{-1} u$ and the polynomial $x_i x^\beta - \sum_{\alpha \in \mathcal{A}} c_\alpha x^\alpha$ belongs to the ideal J generated by the kernel of $M_{\mathcal{B}_h}(y)$. Thus $\sum_{\alpha \in \mathcal{A}} c_\alpha x^\alpha$ is the residue of $x_i x^\beta$ modulo J, i.e., c is the corresponding column of M_{x_i}.) Then, for an arbitrary polynomial f, its multiplication matrix M_f is given by $M_f = f(M_{x_1}, \ldots, M_{x_n})$, whose eigenvectors can be used for extracting the global optimizers.

Let us make a comment at this point. For solving our original problem of minimizing p over the set of real points in $V(I)$, one could follow the following strategy: Determine all points in $V(I)$ (using the eigenvalue method) and evaluate p at the real points. This is however computationally expensive, as this involves computing the eigenvalues of a multiplication matrix whose size is $|\mathcal{B}| = (2m + 1)^n$, thus exponential in the number of variables. Instead, we propose to solve the relaxed convex program (25) for small values of k. Typically it has an optimum solution of small rank r and, when the rank condition holds, one can extract a solution by computing the eigenvalues of a much smaller matrix of size r.

3. Application to unconstrained polynomial minimization

3.1. **Our method.** Let us return to the problem (1) of computing the infimum p^* of a polynomial p over $\mathbb{R}^n$. As before, we assume that p has degree $2m$ and, for $\lambda > 0$, we consider the perturbed polynomial p_λ as in (2) and set $p_\lambda^* := \inf_{x \in \mathbb{R}^n} p_\lambda(x)$. For $i = 1, \ldots, n$, let

$$(27) \qquad h_{\lambda,i}(x) := \partial p_\lambda(x)/\partial x_i = \partial p(x)/\partial x_i + \lambda(2m + 2)x_i^{2m+1}$$

denote the partial derivatives of $p_\lambda(x)$. Let I_λ be the ideal generated by $h_{\lambda,1}, \ldots, h_{\lambda,n}$ and let $V_\lambda := V(I_\lambda)$ be its associated variety. Up to a constant factor, each $h_{\lambda,i}(x)$ is of the form $x_i^{2m+1} + \tilde{h}_i(x)$, where $\tilde{h}_i(x)$ has degree at most $2m - 1$, and thus we are in the situation of Section 2.4. Therefore, for $\lambda \neq 0$, the set $\{x^\beta \mid \beta \in \mathcal{B}\}$, where $\mathcal{B}$ is as in (22), is the set of standard monomials, forming a basis of of $\mathbb{R}[x_1, \ldots, x_n]/I_\lambda$, and I_λ is a zero-dimensional ideal.

As p_λ attains its minimum, it follows that it attains its minimum at a critical point. That is, $\inf_{x \in \mathbb{R}^n} p_\lambda(x) = \min_{x \in V_\lambda \cap \mathbb{R}^n} p_\lambda(x)$. If x^* is a global minimizer of p, then $p^* \leq p_\lambda^* \leq p_\lambda(x^*) \leq p^* + \lambda \|x^*\|^{2m+2}$. As $p(x) \leq p_\lambda(x)$ for all x, we have:

$$p^* \leq \mu_\lambda^* := \min_{x \in V_\lambda \cap \mathbb{R}^n} p(x) \leq \min_{x \in V_\lambda \cap \mathbb{R}^n} p_\lambda(x).$$

As I_λ is a zero-dimensional ideal, we can apply Theorem 11 and compute the bound μ_λ^* via the following semidefinite program:

(28) $$\mu_\lambda^* = \min p^T y \text{ subject to } M_{\mathcal{B}}(y) \succeq 0, \ y_0 = 1,$$

where $\mathcal{B}$ is defined in (22). Given an integer $m \le k \le 2nm$, one can consider the following semidefinite program, involving truncated combinatorial moment matrices:

(29) $$\mu_{k,\lambda}^* := \inf p^T y \text{ subject to } M_{\mathcal{B}_k}(y) \succeq 0, \ y_0 = 1,$$

where $\mathcal{B}_k$ is as in (23). These parameters define a hierarchy of lower bounds for μ_λ^*:

(30) $$\mu_{m,\lambda}^* \le \ldots \le \mu_{k,\lambda}^* \le \ldots \le \mu_{2nm,\lambda}^* = \mu_\lambda^*,$$

where the last equality holds since $\mathcal{B}_{2nm} = \mathcal{B}$.

Let us give some information about the structure of the matrix $M_{\mathcal{B}_k}(y)$. For $\alpha, \beta \in \mathcal{B}_k$, the (α, β)-th entry of $M_{\mathcal{B}_k}(y)$ is equal to $y^T r^{(\alpha+\beta)}$, where $r^{(\alpha+\beta)}(x)$ is the residue of $x^{\alpha+\beta}$ modulo the ideal I_λ. This residue is obtained by dividing the monomial $x^{\alpha+\beta}$ by the polynomials $h_{\lambda,i}$ from (27), forming a Gröbner basis of I_λ. Hence, the entries of $M_{\mathcal{B}_k}(y)$ are polynomial in $1/\lambda$ (and linear in y). The next result gives an estimate on the degree in $1/\lambda$ of the entries of $M_{\mathcal{B}_k}(y)$.

THEOREM 18. *For $k = m, \ldots, 2nm$, the matrices $M_{\mathcal{B}_k}(y)$ are polynomial matrices in $1/\lambda$; the maximal degree in $1/\lambda$ of the entries of $M_{\mathcal{B}_k}(y)$ is at most $k - m$.*

Proof. Consider a monomial x^γ where $\gamma \in \mathbb{Z}_+^n$ with $|\gamma| \ge 2m$. We show by induction on $|\gamma|$ that the coefficients of the residue of x^γ modulo the ideal I_λ are polynomial in $1/\lambda$ with degree at most $\lceil (|\gamma| - 2m)/2 \rceil$. If $\gamma_i \le 2m$ for all $i = 1, \ldots, n$, then x^γ is a standard monomial; that is, its residue is x^γ whose degree in $1/\lambda$ is 0. Suppose, e.g., that $\gamma_1 \ge 2m + 1$. Then, $x^\gamma = x_1^{2m+1} x^{\tilde{\gamma}}$, where $\tilde{\gamma}_1 = \gamma_1 - 2m - 1$ and $\tilde{\gamma}_i = \gamma_i$ for $i \ge 2$. Thus, $|\tilde{\gamma}| = |\gamma| - 2m - 1$ and $x^\gamma \equiv -\frac{1}{2m+2} \frac{1}{\lambda} \frac{\partial p(x)}{\partial x_1} x^{\tilde{\gamma}}$ modulo I_λ. As the degree of $x^{\tilde{\gamma}} \partial p(x)/\partial x_1$ is at most $2m - 1 + |\tilde{\gamma}| = |\gamma| - 2$, we know by induction that the degree in $1/\lambda$ of its residue is at most $\lceil (|\gamma| - 2 - 2m)/2 \rceil = \lceil (|\gamma| - 2m)/2 \rceil - 1$. Therefore, the degree in $1/\lambda$ of the residue of x^γ is at most $\lceil (|\gamma| - 2m)/2 \rceil$. The theorem now follows since each entry of $M_{\mathcal{B}_k}(y)$ is the residue of a monomial of degree at most $2k$. ∎

As $M_{\mathcal{B}_m}(y)$ does not depend on λ, the matrix $M_{\mathcal{B}_m}(y)$ coincides with the classical matrix $M_m(y)$. Hence, the first member $\mu_{m,\lambda}^*$ in the hierarchy (30) does not depend on λ and is equal to $p_{L,m}^*$, the Lasserre lower bound for p^* from (4); thus,

$$\mu_{m,\lambda}^* = p_{L,m}^* \le p^*.$$

It is not clear a priori on which side of p^* the parameter $\mu_{k,\lambda}^*$ is located when $m < k < 2nm$. In some cases, one can derive this information with the help of the following result.

COROLLARY 19. *Let $M_{\mathcal{B}_k}(y)$ be an optimum solution to the program (29) defining $\mu_{k,\lambda}^*$. Assume that $\operatorname{rank} M_{\mathcal{B}_h}(y) = \operatorname{rank} M_{\mathcal{B}_{h-1}}(y)$ for some $m \le h \le k$. Then, $p^* \le \mu_{k,\lambda}^* \le \mu_\lambda^*$ and one can extract a point $x \in \mathbb{R}^n$ for which $p^* \le p(x) \le \mu_{k,\lambda}^*$. Moreover, $\mu_{k,\lambda}^* = \mu_\lambda^*$ if $h \ge 2m + 1$.*

14

Proof. By Corollary 16, $(y_\alpha)_{\alpha \in \mathcal{B}_{2h}}$ is the sequence of moments of a probability measure $\mu = \sum_{i=1}^r \lambda_i \delta_{v_i}$. Hence, $\mu_{k,\lambda}^* = p^T y = \sum_{i=1}^r \lambda_i p(v_i) \geq \min_i p(v_i) \geq p^*$. If $h \geq 2m+1$, then $v_1, \ldots, v_r \in V_\lambda$ and thus $\mu_{k,\lambda}^* = \mu_\lambda^*$. ∎

Let us point out that, for the problem of computing the minimum p^* of a polynomial of the form $p = \sum_{i=1}^n c_i x_i^{2m+2} + p_0$ where $\deg p_0 \leq 2m+1$ ($c_1, \ldots, c_n \in \mathbb{R} \setminus \{0\}$), our method can be applied directly to p, without any perturbation. Namely, let r be the residue of p modulo the ideal generated by $\partial p / \partial x_i$ ($i = 1, \ldots, n$), then

$$p^* = \min \ r^T y \ \text{ s.t. } M_{\mathcal{B}}(y) \succeq 0, \ y_0 = 1,$$

the parameters μ_k^* from (25) are lower bounds for p^*, with equality $\mu_k^* = p^*$ if $\text{rank} M_{\mathcal{B}_h}(y) = \text{rank} M_{\mathcal{B}_{h-1}}(y)$ for some $2m+1 \leq h \leq k$. See Examples 9 and 10 in the next section for an illustration.

We now illustrate our method on two small examples; both will be revisited in the next section.

EXAMPLE 1. Consider the polynomial $p(x_1, x_2) = x_1^2 + x_2$ and its perturbation $p_\lambda(x_1, x_2) = p(x_1, x_2) + \lambda(x_1^4 + x_2^4)$. Then, $p^* = -\infty$. One can compute explicitely the set V_λ of solutions to the system:

$$\frac{\partial p_\lambda}{\partial x_1} = 2x_1(2\lambda x_1^2 + 1) = 0, \quad \frac{\partial p_\lambda}{\partial x_2} = 4\lambda x_2^3 + 1 = 0.$$

Namely, V_λ consists of the nine points (x_1, x_2) with $x_1 = 0, \pm i \sqrt{\frac{1}{2\lambda}}$, and $x_2 = -\sqrt[3]{\frac{1}{4\lambda}}$, $-j\sqrt[3]{\frac{1}{4\lambda}}$, $-j^2\sqrt[3]{\frac{1}{4\lambda}}$ (where $i, j \in \mathbb{C}$, $i^2 = -1$, $j^3 = 1$). Hence, $(0, -\sqrt[3]{\frac{1}{4\lambda}})$ is the only real point in V_λ and thus the unique minimizer of p over V_λ. This implies that $\mu_\lambda^* = -\sqrt[3]{\frac{1}{4\lambda}}$.

EXAMPLE 2. Consider the polynomial $p(x_1, x_2) = (x_1^2 + x_2^2 - 1)^2$ whose minimum is $p^* = 0$ attained at all points on the unit circle. One can verify that the set V_λ contains 25 points, among them 9 real points, namely $(0, 0)$ and

(i) $(x_1, x_2) = \pm(0, a), \pm(a, 0)$, where $a := \sqrt{\frac{-1 + \sqrt{6\lambda + 1}}{3\lambda}}$

(ii) $(x_1, x_2) = (\pm b, \pm b)$ where $b := \sqrt{\frac{-2 + \sqrt{6\lambda + 4}}{3\lambda}}$.

The minimum of p over V_λ is $\mu_\lambda^* = (2b^2 - 1)^2$, which is attained at the points $(\pm b, \pm b)$ in (ii). As $a = 1 + o(1)$ and $b = 1/\sqrt{2} + o(1)$, the limit as $\lambda \downarrow 0$ of the real points in V_λ are the points $(0, \pm 1)$, $(\pm 1, 0)$, $(\pm 1/\sqrt{2}, \pm 1/\sqrt{2})$ on the unit circle together with the origin.

3.2. **Examples.** We present here several examples on which our method has been tested. Let p be the polynomial whose infimum p^* is to be found and let $2m$ be its degree. We compute the approximations $\mu_{k,\lambda}^*$ of p^* provided by the program (29). The computation is carried out for several values of λ, ranging typically from 10^{-1} to 10^{-4} (sometimes much smaller). We solve the program (29) for increasing values of k starting from $k = m$. Let $M_{\mathcal{B}_k}(y^*)$ be the returned optimum solution and $\mu_{k,\lambda}^*$ the returned optimum value. At $k = m$, we find the Lasserre lower bound $p_{L,m}^*$ for p^*.

At each step k, we check whether the rank condition (16) holds; if not, we go to the next step $k + 1$. More precisely,

If $k = m$, then $\mu^*_{m,\lambda} = p^*_{L,m} \leq p^*$. Moreover, if rank $M_{\mathcal{B}_m}(y^*) = $ rank $M_{\mathcal{B}_{m-1}}(y^*)$, then $\mu^*_{m,\lambda} = p^*_{L,m} = p^*$, i.e., the infimum p^* has been found.

If $k \geq m + 1$, and rank $M_{\mathcal{B}_h}(y^*) = $ rank $M_{\mathcal{B}_{h-1}}(y^*) =: r$ for $h = k$ or $h = k - 1$, then $\mu^*_{k,\lambda} \geq p^*$; moreover, one can extract r points $x \in \mathbb{R}^n$ and evaluating p at any such point x gives a certified upper bound on p^*.

There are two phases in the resolution of the program (29): (1) Compute the entries of the matrix variable $M_{\mathcal{B}_k}(y)$ in (29); that is, compute the residue of $x^{\alpha+\beta}$ modulo I_λ with respect to the basis $\mathcal{B}$, for each $\alpha, \beta \in \mathcal{B}_k$; and (2) Solve the semidefinite program (29). The first phase is carried out using Mathematica 4.2 and the SDP problem is solved with SeDuMi 1.05 (used with accuracy parameter pars.eps=0). When evaluating the rank of a matrix we consider the eigenvalues with a precision of 10^{-3}; that is, we ignore all decimals starting with the 5th one.

In the tables below, at a given order k, (r_k, r_{k-1}, r_{k-2}) is the triple consisting of the ranks of the matrices $M_{\mathcal{B}_k}(y^*)$, $M_{\mathcal{B}_{k-1}}(y^*)$, $M_{\mathcal{B}_{k-2}}(y^*)$, where $M_{\mathcal{B}_k}(y^*)$ is the optimum solution to (29) returned by the algorithm.

In some examples, we also compute the upper approximations $\mu^*_{L,k,\lambda}$ on p^* obtained from the program (12), and some other approximations obtained by minimizing p over a ball. Then, $(r_k, r_{k-1}, ..)$ contains the ranks of the matrices $M_k(y^*)$, $M_{k-1}(y^*)$,.., where $M_k(y^*)$ is the optimum solution to (12) (or (4) when optimizing over a ball).

EXAMPLE 1. (REVISITED) Consider again the polynomial $p(x_1, x_2) = x_1^2 + x_2$ with infimum $p^* = -\infty$. Then, $n = 2$, $m = 1$, $|\mathcal{B}_1| = 3$, $|\mathcal{B}_2| = 6$, $|\mathcal{B}| = 9$. When computing the Lasserre lower bound $p^*_{L,1}$, GloptiPoly returns as expected that the 'Sedumi dual may be unbounded'. As can be seen in Table 1, our algorithm retrieves a very accurate estimate of the minimizer $(0, -\sqrt[3]{\frac{1}{4\lambda}})$.

λ	order k	(r_k, r_{k-1}, r_{k-2})	$\mu^*_{k,\lambda}$	extracted solutions
10^{-3}	2	(1,1,1)	-6.2996	(0,-6.2996)
10^{-6}	2	(1,1,1)	-62.9961	(0,-62.9961)
10^{-9}	2	(1,1,1)	-629.9606	(0,-629.9606)

TABLE 1. Bounds $\mu^*_{k,\lambda}$ for Example 1

EXAMPLE 2. (REVISITED) Consider again the polynomial $p(x_1, x_2) = (x_1^2 + x_2^2 - 1)^2$ with infimum $p^* = 0$ attained at the points of the unit circle. Then, $n = 2$, $m = 2$, $|\mathcal{B}_2| = 6$, $|\mathcal{B}_3| = 10$, $|\mathcal{B}_4| = 15$, $|\mathcal{B}| = 25$. The Lasserre lower bound is $p^*_{L,2} = 2.82 \; 10^{-11} \leq p^*$ (with $r_2 = 5$, $r_1 = 3$).

Again, one can see in Table 2 that the algorithm retrieves very accurate estimates of the four minimizers $(\pm b, \pm b)$ of p over V_λ. Moreover, $\mu^*_{4,10^{-3}} \geq p^*$ and $\mu^*_{4,10^{-3}} \sim 10^{-7}$ is an accurate estimate of $p^* = 0$.

λ	order k	(r_k, r_{k-1}, r_{k-2})	$\mu_{k,\lambda}^*$	extracted solutions
10^{-2}	3	(4,4,4)	$1.3854 \ 10^{-5}$	$(\pm 0.7058, \pm 0.7058)$
10^{-3}	3	(9,5,3)	$1.3320 \ 10^{-7}$	none
10^{-3}	4	(4,4,4)	$1.4043 \ 10^{-7}$	$(\pm 0.7070, \pm 0.7070)$

TABLE 2. Bounds $\mu_{k,\lambda}^*$ for Example 2

EXAMPLE 3. Consider the polynomial $p(x_1, x_2) = (x_1^2 + 1)^2 + (x_2^2 + 1)^2 - 2(x_1 + x_2 + 1)^2$. Then, $n = 2$, $m = 2$, $|\mathcal{B}_3| = 10$, $|\mathcal{B}_4| = 15$, $|\mathcal{B}| = 25$. It is known (see [15]) that $p^* = -11.4581$ is attained at the point $(1.3247, 1.3247)$, and that the polynomial $p(x) - p^*$ is a sum of squares. Indeed, $p_{L,2}^* = -11.4581$ and, as $r_2 = r_1 = 1$, GloptiPoly extracts the minimizer $(1.3247, 1.3247)$. Nevertheless Table 3 shows the behaviour of our method on this example.

We have also computed the bound $\beta_{L,k,\lambda}^*$ from (13), computing the order k moment relaxation for the minimum of p over the ball with radius R_λ as in (10). Here, $R_\lambda = \frac{56}{\lambda}$. For $\lambda = 10^{-1}$, $k = 2$, $R_\lambda = 560$ and GloptiPoly returns the value $\beta_{L,2,1/10}^* = -11.4581$ and extracts the solution $(1.3247, 1.3247)$.

λ	order k	(r_k, r_{k-1}, r_{k-2})	$\mu_{k,\lambda}^*$	extracted solutions
10^{-2}	3	(1,1,1)	-11.4548	$(1.3109, 1.3109)$
10^{-3}	3	(1,1,1)	-11.4580	$(1.3233, 1.3233)$
10^{-4}	3	(1,1,1)	-11.4581	$(1.3246, 1.3246)$
10^{-5}	3	(1,1,1)	-11.4581	$(1.3247, 1.3247)$

TABLE 3. Bounds $\mu_{k,\lambda}^*$ for Example 3

EXAMPLE 4. Consider the polynomial $p(x_1, x_2) = 1/27 + x_1^2 x_2^2 (x_1^2 + x_2^2 - 1)$, a dehomogenized version of the Motzkin polynomial, considered in [11]. Then $n = 2$, $m = 3$, $|\mathcal{B}_3| = 10$, $|\mathcal{B}_4| = 15$, $|\mathcal{B}_5| = 21$, $|\mathcal{B}| = 49$. It is known that p has minimum $p^* = 0$, attained at $(\pm 1/\sqrt{3}, \pm 1/\sqrt{3})$, and p is not a sum of squares. As Table 4a shows, our algorithm finds a very accurate estimate of p^* and of its minimizers at the relaxation of order 5 when using the perturbation $\lambda = 10^{-4}$.

λ	order k	(r_k, r_{k-1}, r_{k-2})	$\mu_{k,\lambda}^*$	extracted solutions
10^{-2}	4	(11,8,6)	$-8.8740 \ 10^{-5}$	none
10^{-2}	5	(4,4,4)	$2.1500 \ 10^{-6}$	$(\pm 0.5761, \pm 0.5761)$
10^{-3}	4	(11,8,6)	-0.0060	none
10^{-3}	5	(4,4,4)	$2.1897 \ 10^{-8}$	$(\pm 0.5772, \pm 0.5772)$
10^{-4}	4	(11,8,6)	-0.0336	none
10^{-4}	5	(4,4,4)	$1.9042 \ 10^{-10}$	$(\pm 0.5773, \pm 0.5773)$

TABLE 4a. Bounds $\mu_{k,\lambda}^*$ for Example 4

λ	order k	$(r_k, r_{k-1}, \ldots, r_1)$	$\mu^*_{L,k,\lambda}$	extracted solutions
10^{-2}	4	(15,10,6,3)	-1.0815	none
10^{-2}	5	(21,15,8,6,3)	-0.0060	none
10^{-2}	6	(11,4,4,4,4,3)	$2.1904\ 10^{-8}$	none*
10^{-2}	7	(-,4,4,4,4,3)	$2.1904\ 10^{-8}$	$(\pm 0.5772, \pm 0.5772)$
10^{-3}	4	(15,10,6,3)	-1.3072	none
10^{-3}	5	(21,15,8,6,3)	-0.0332	none
10^{-3}	6	(11,4,4,4,4,3)	$2.1993\ 10^{-10}$	none*
10^{-3}	7	(-,4,4,4,4,3)	$2.3084\ 10^{-10}$	$(\pm 0.5773, \pm 0.5773)$
10^{-4}	4	(15,10,6,3)	-1.1225	none
10^{-4}	5	(21,15,10,6,3)	-0.0909	none
10^{-4}	6	(11,4,4,4,4,3)	$1.0209\ 10^{-11}$	none*
10^{-4}	7	(-,4,4,4,4,3)	$1.498\ 10^{-11}$	$(\pm 0.5773, \pm 0.5773)$

TABLE 4b. Bounds $\mu^*_{L,k,\lambda}$ for Example 4

We have also computed the parameters $\mu^*_{L,k,\lambda}$ from (12) using GloptiPoly. Table 4b shows the results. We have: $|S_4| = 15$, $|S_5| = 21$, $|S_6| = 28$, $|S_7| = 36$. (At the relaxation of order 7, GloptiPoly does not return the value of the rank of $M_7(y)$ which is indicated by '-' in the table.) At the relaxation of order 6, GloptiPoly does not yet extract a solution since the stronger rank condition (6) does not hold. However, this stronger condition is needed only to be able to claim that the extracted solution does satisfy the constraints $\partial p_\lambda / \partial x_i = 0$ $(i = 1, \ldots, n)$. As $\text{rank} M_{k-1}(y) = \text{rank} M_{k-2}(y)$ one could already extract a solution at order 6, which permits to claim that $\mu^*_{L,6,\lambda} \geq p^*$. Note, however, that our algorithm based on combinatorial moment matrices is able to find an upper bound for p^* at order $k = 5$ already. Moreover, at a given order k, the parameter $\mu^*_{k,\lambda}$ is a more accurate approximation of p^* than the parameter $\mu^*_{L,k,\lambda}$.

Finally we have computed the bounds $\beta^*_{L,k,\lambda}$ from (13). Here, the radius is $R_\lambda = \frac{24}{\lambda}$. For $\lambda = 1/10$ and $k = 3, 4, 5$, Sedumi reports that the 'dual may be unbounded'. For $\lambda = 1$, one finds $\beta^*_{L,3,\lambda} = -3.9722$ (with $(r_3, r_2, r_1) = (8, 6, 3)$, thus no solution extracted) and Sedumi reports that the 'dual may be unbounded' for $k \geq 4$.

When using the radius $R = 20$ (instead of R_λ), the smallest order k for which the rank condition holds for the moment relaxation is $k = 6$, where we find the upper bound $8.5345\ 10^{-12}$ for p^* and GloptiPoly extracts the solution $(\pm 0.5774, \pm 0.5774)$.

If we use a smaller radius $R = 2$, then the rank condition holds already at the moment relaxation of order $k = 3$, where we find the upper bound $1.2561\ 10^{-13}$ for p^* and GloptiPoly extracts the solutions $(\pm 0.5774, \pm 0.5774)$.

Therefore, the approach via optimization on a ball seems to work well only if one knows a priori a small ball containing a global minimizer.

EXAMPLE 5. Consider the polynomial $p(x_1, x_2) = x_2^2 + (x_1 x_2 - 1)^2$. This is a classical example of a polynomial having a finite infimum, which is not attained; $p^* = 0$ as $\lim_{\epsilon \downarrow 0} p(1/\epsilon, \epsilon) = 0$. Here, $n = 2$, $m = 2$, $|\mathcal{B}_2| = 6$, $|\mathcal{B}_3| = 10$, $|\mathcal{B}| = 25$. The Lasserre lower bound is $p^*_{L,2} = 5.4776\ 10^{-5}$.

λ	order k	(r_k, r_{k-1}, r_{k-2})	$\mu^*_{k,\lambda}$	extracted solutions
10^{-2}	3	(2,2,2)	0.3385	$\pm(1.3981, 0.4729)$
10^{-3}	3	(2,2,2)	0.2082	$\pm(1.9499, 0.4060)$
10^{-4}	3	(2,2,2)	0.1232	$\pm(2.6674, 0.3287)$
10^{-5}	3	(2,2,2)	0.0713	$\pm(3.6085, 0.2574)$
10^{-6}	3	(2,2,2)	0.0408	$\pm(4.8511, 0.1977)$
10^{-7}	3	(3,2,2)	0.0231	$\pm(6.4986, 0.1503)$
10^{-8}	3	(3,2,2)	0.0131	$\pm(8.6882, 0.1136)$
10^{-9}	3	(8,4,2)	0.0074	none
10^{-10}	3	(7,4,2)	0.0041	none

TABLE 5a. Bounds $\mu^*_{k,\lambda}$ for Example 5

We have also computed the bounds $\mu^*_{L,k,\lambda}$, shown in Table 5b. When the order k is marked with a star (like 6^*), this means that we have rescaled the problem for Sedumi (setting pars.scaling=[1 10]). (This is advised when the expected solutions have large entries; see the manual [10] for GloptiPoly. Without rescaling the solution returned by GloptiPoly is approximatively 1, which is the value of p at the point $(0,0)$ of V_λ, thus not the true minimum.) Recall that $|S_3| = 10$, $|S_4| = 15$, $|S_5| = 21$, $|S_6| = 28$.

λ	order k	$(r_k, \ldots, r_1)$	$\mu^*_{L,k,\lambda}$	extracted solutions
10^{-2}	3	(10,6,2)	0.0096	none
10^{-2}	4	(7,2,2,2)	0.3385	none
10^{-2}	5	(-,2,2,2,2)	0.3385	$\pm(1.3981, 0.4729)$
10^{-3}	3	(10,6,2)	0.0105	none
10^{-3}	4	(7,2,2,2)	0.2082	none
10^{-3}	5	(-,2,2,2,2)	0.2082	$\pm(1.9499, 0.4060)$
10^{-4}	3	(10,6,2)	0.0095	none
10^{-4}	4	(7,2,2,2)	0.1232	none
10^{-4}	5	(-,2,2,2,2)	0.1233	$\pm(2.6674, 0.3287)$
10^{-5}	5	(-,2,2,2,2)	0.0718	$\pm(3.6085, 0.2574)$
10^{-6}	6^*	(-,-,2,2,2,2)	0.0408	$\pm(4.8511, 0.1977)$
10^{-7}	6^*	(-,-,2,2,2,2)	0.0231	$\pm(6.4986, 0.1503)$
10^{-8}	6^*	(-,-,2,2,2,2)	0.0131	$\pm(8.6882, 0.1136)$
10^{-9}	6^*	(-,-,2,2,2,2)	0.0074	$\pm(11.6026, 0.0856)$
10^{-10}	6^*	(-,-,2,2,2,2)	0.0042	$\pm(15.4849, 0.0643)$

TABLE 5b. Bounds $\mu^*_{L,k,\lambda}$ for Example 5

One can make the following observations regarding the results from Tables 5a,5b. While our algorithm extracts the correct solutions at order $k = 3$, when using the moment relaxation to the program (7) GloptiPoly needs to go to higher orders to be able to extract solutions. We have computed (with Mathematica) the points in the gradient variety V_λ; it turns out that there are three real points which are $(0,0)$ and the two points extracted by the algorithms for the given values of λ in Tables 5a, 5b.

EXAMPLE 6. Our next example is the polynomial $q(z_1, z_2, z_3, z_4, z_5) = \sum_{i=1}^5 \prod_{j \neq i}(z_i - z_j)$, which is again an instance of a nonnegative polynomial which is not a sum of squares, due

19

to Lax-Lax and Schmüdgen. More such examples can be found e.g. in [27]. Introducing new variables $x_i := z_1 - z_{i+1}$ $(i = 1, \ldots, 4)$, minimizing $q(z)$ is equivalent to minimizing a polynomial p in the four variables $x_1, \ldots, x_4$. After performing this substitution, we have: $n = 4$, $m = 2$, $|\mathcal{B}_2| = 15$, $|\mathcal{B}_3| = 35$, $|\mathcal{B}_4| = 70$, $|\mathcal{B}| = 625$. When computing the lower bound $p^*_{L,2}$, Sedumi reports that the 'primal problem is infeasible' and the 'dual problem may be unbounded'.

λ	order k	(r_k, r_{k-1}, r_{k-2})	$\mu^*_{k,\lambda}$	extracted solutions
10^{-1}	3	$(20,10,5)$	-0.0575	none
10^{-1}	4	$(5,5,5)$	$-8.9342 \ 10^{-8}$	$\pm(0.0407, 0.0445, 0.0482, 0.0520)$
				approx. $(0,0,0,0)$ three times

TABLE 6a. Bounds $\mu^*_{k,\lambda}$ for Example 6

Table 6b gives some values of the parameter $\mu^*_{L,k,\lambda}$. At order $k = 3$, for $\lambda = 10^{-1}, 10^{-2}$, Sedumi reports that the 'dual problem may be unbounded'. On this example the parameter $\mu^*_{L,k,\lambda}$ appears to be a more accurate approximation of p^* than $\mu^*_{k,\lambda}$.

λ	order k	(r_3, r_2, r_1)	$\mu^*_{L,k,\lambda}$	extracted solutions
10^{-1}	4	$(1,1,1)$	$6.0249 \ 10^{-15}$	$10^{-8}(-0.6138, -0.7014, 0.5825, 0.9606)$
10^{-2}	4	$(1,1,1)$	$3.9252 \ 10^{-14}$	$10^{-8}(0.0602, 0.4502, -0.0416, -0.2084)$

TABLE 6b. Bounds $\mu^*_{L,k,\lambda}$ for Example 6

As the polynomial p is homogeneous, i.e., $p(tx) = t^{2m}p(x)$ for all x ($m = 2$ here), there are in fact two possibilities for its infimum: Either, $p^* = 0$ if p is nonnegative, or $p^* = -\infty$ otherwise. The parameters $\mu^*_{k,\lambda}$ and $\mu^*_{L,k,\lambda}$ are upper bounds for p^*. Hence, if for some small λ, they are close to 0, it is then quite likely that $p^* = 0$ (since μ^*_λ converges to p^* as $\lambda \downarrow 0$) but this cannot be claimed with certitude. On the other hand, such upper bounds will be useful for proving that $p^* = -\infty$. Indeed, if we find a negative upper bound for p^*, then we can conclude that $p^* = -\infty$; moreover, any extracted solution gives a certificate for this. See Example 8 for an illustration.

When p is homogeneous, one can also test its nonnegativity by computing its minimum p^*_B over the unit ball B. Indeed, either $p^*_B = 0$ if p is nonnegative, or $p^*_B < 0$ otherwise. However, if p is nonnegative but not a sum of squares, then the moment relaxation (4) of any order k is never exact, i.e., the inequality $p^*_{L,k} \leq p^*_B$ is always strict (and thus the optimum matrix does not satisfy the rank condition). (Indeed, suppose that $p^*_{L,k} = p^*_B = 0$. For k large enough, there is no duality gap between (4) and (5), and (5) attains its supremum (see [15]). Hence, $\rho^*_k = p^*_{L,k} = p^*_B = 0$, implying that p can be written as $p = u + (1 - \sum_i x_i^2)v$ where u, v are sums of squares. As p is homogeneous, this implies easily that p must be a sum of squares (see [14]), yielding a contradiction.) Let us illustrate this on our current example. Table 6c shows the values $p^*_{L,k}$ obtained for the moment relaxations (4) for the minimum p^*_B of p over the unit ball. Recall that $|S_5| = 126$, $|S_6| = 210$.

order k	$(r_k, r_{k-1}, \ldots, r_1)$	$p^*_{L,k}$	extracted solutions
2	(10,5)	-0.0375	none
3	(25,15,5)	-0.0035	none
4	(39,29,15,5)	$-7.7935 \; 10^{-4}$	none
5	(55,44,29,15,5)	$-2.7268 \; 10^{-4}$	none
6	(210,126,70,29,15,5)	$-1.1936 \; 10^{-4}$	none

TABLE 6c. Bounds from optimizing over a ball for Example 6

EXAMPLE 7. Consider the matrix:

$$P = \begin{pmatrix} 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 \end{pmatrix}$$

and the associated homogeneous polynomial $q(x) = \sum_{i,j=1}^{5} x_i^2 x_j^2 P_{ij}$ (Example 5.4 in [22]). The matrix P is said to be *copositive* when q is nonnegative. Testing matrix copositivity is a co-NP-complete problem [21]. Although some necessary and sufficient conditions for the copositivity of a matrix are known (see e.g. [13]), their algorithmic application is computationally too expensive. An alternative consists therefore of using numerical algorithms for testing (non)copositivity. Parrilo [22, 23] introduced the following criterion, useful for proving copositivity. Namely, if the polynomial $(\sum_{i=1}^{n} x_i^2)^r q(x)$ is a sum of squares for some integer $r \geq 0$, then q is nonnegative and thus P is copositive. For the matrix P considered in the present example, it is known that this criterion is satisfied for $r = 1$.

Let us nevertheless see the behaviour of our method on this example. Due to symmetry, the polynomial q is nonnegative if and only if the (dehomogenized) polynomial $p(x) := q(x_1, x_2, x_3, x_4, 1)$ is nonnegative. Then, $n = 4$, $m = 2$, $|\mathcal{B}_2| = 15$, $|\mathcal{B}_3| = 35$, $|\mathcal{B}_4| = 70$, $|\mathcal{B}_5| = 122$, $|\mathcal{B}| = 625$. The Lasserre lower bound is $p^*_{L,2} = -1.4955 \; 10^6$ with $(r_1, r_2) = (5, 15)$.

λ	order k	(r_k, r_{k-1}, r_{k-2})	$\mu^*_{k,\lambda}$	extracted solutions
10^{-2}	3	(18,6,3)	-1.5407	none
10^{-2}	4	(4,4,4)	$1.3854 \; 10^{-5}$	$(\pm 0.7058, 0, 0, \pm 0.7058)$
10^{-3}	4	(4,4,4)	$1.3854 \; 10^{-5}$	$(\pm 0.7058, 0, 0, \pm 0.7058)$
10^{-4}	4	(9,7,5)	$1.5544 \; 10^{-7}$	none

TABLE 7. Bounds $\mu^*_{k,\lambda}$ for Example 7

EXAMPLE 8. Let $G = (V, E)$ be a graph with node set $V = \{1, \ldots, n\}$ and let A_G be its adjacency matrix, with $(A_G)_{ij} = 1$ if $ij \in E$ and $(A_G)_{ij} = 0$ otherwise, for $i, j \in V$. Consider the matrix:

$$P := t(I + A_G) - J,$$

where $t \in \mathbb{R}$, I is the identity matrix and J is the all-ones matrix, and the associated homogeneous polynomial $p(x) := \sum_{i,j=1}^{n} x_i^2 x_j^2 P_{ij}$. By Motzkin-Straus theorem [20], p is nonnegative (i.e., $p^* = 0$) (equivalently, P is a copositive matrix) if and only if $t \geq \alpha(G)$,

where $\alpha(G)$ is the stability number of G, i.e., the largest cardinality of a stable set in G. In Example 7, G is the circuit $(1,4,2,5,3)$ on 5 nodes with $\alpha(G) = 2$ and $P = 2(I + A_G) - J$ which is therefore copositive. Consider now the case when G is the path $(1,4,2,5,3)$ on 5 nodes and $t = 2$, giving the matrix

$$P = \begin{pmatrix} 1 & -1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 \end{pmatrix}.$$

Then P is not copositive, as $t < \alpha(G) = 3$ (note also $p(1,1,1,0,0) = -3$). This is confirmed by the results about p^* from Table 8a, where we have: $n = 5$, $m = 2$, $|\mathcal{B}_1| = 6$, $|\mathcal{B}_2| = 21$, $|\mathcal{B}_3| = 56$, $|\mathcal{B}_4| = 126$, $|\mathcal{B}| = 3125$.

λ	order k	(r_k, r_{k-1}, r_{k-2})	$\mu^*_{k,\lambda}$	extracted solutions
1	3	(8,7,4)	-1.3333	none
1	4	(8,8,7)	-1.3333	two of the extracted solutions: $\pm(0.8165, 0.8165, 0.8165, 0, 0)$ $(0.8165 \sim \sqrt{2/3})$
10^{-1}	3	(8,7,4)	-133.3333	none
10^{-1}	4	(8,8,7)	-133.3333	two of the extracted solutions: $\pm(2.5820, 2.5820, 2.5820, 0, 0)$ $(2.5820 \sim \sqrt{20/3})$
10^{-2}	3	(8,7,4)	$-1.3333 \ 10^4$	none

TABLE 8a. Bounds $\mu^*_{k,\lambda}$ for Example 8, when G is the path on 5 nodes and $t = 2$

Consider now the case when G is the circuit $(1,2,3,4,5,6)$ on 6 nodes and $t = 2$. Again the corresponding matrix P is not copositive, since $t < \alpha(G) = 3$. This is confirmed by the results about p^* from Table 8b. Because of symmmetry, we made the computations for the polynomial $p(x_1, x_2, x_3, x_4, x_5, 1)$. Then, $n = 5$, $m = 2$, $|B_2| = 21$, $|\mathcal{B}_3| = 56$, $|\mathcal{B}_4| = 126$.

λ	order k	(r_k, r_{k-1}, r_{k-2})	$\mu^*_{k,\lambda}$	extracted solutions
1	3	(4,4,3)	-2.2660	$(0, \pm0.9036, 0, \pm0.9036, 0)$
10^{-1}	3	(8,7,4)	-106.6640	none
10^{-1}	4	(8,8,7)	-106.6640	-
10^{-2}	3	(8,7,4)	$-1.3067 \ 10^4$	none

TABLE 8b. Bounds $\mu^*_{k,\lambda}$ for Example 8, when G is the circuit on 6 nodes and $t = 2$

In both instances we find a point x with $p(x) < 0$ (which certifies that P is not copositive) at the relaxation of order 3 or 4, already for the perturbation $\lambda = 1$. The bounds $\mu^*_{3,\lambda}$ decrease rapidly as λ goes to 0.

As last instance, consider the case when G is the circuit $(1,2,3,4,5,6,7)$ on 7 nodes and $t = 2$. Again, P is not copositive since $t < \alpha(G) = 3$. Due to symmetry it suffices to consider the polynomial p where we set $x_7 = 1$. Then, $n = 6$, $m = 2$, $|\mathcal{B}_2| = 28$, $|\mathcal{B}_3| = 84$ and $|\mathcal{B}_4| = 210$. Table 8c shows some parameters $\mu^*_{3,\lambda}$ which again decrease rapidly as λ becomes small.

λ	order k	(r_k, r_{k-1}, r_{k-2})	$\mu^*_{k,\lambda}$	extracted solutions
1	3	(62,24,7)	-4.1114	none
10^{-1}	3	(62,24,7)	-304.7340	none
10^{-2}	3	(66,24,7)	$-3.0745\ 10^4$	none
10^{-3}	3	(83,27,7)	$-3.0808\ 10^6$	none

TABLE 8c. Bounds $\mu^*_{k,\lambda}$ for Example 8, when G is the circuit on 7 nodes and $t = 2$

EXAMPLE 9. Consider the polynomial $p(x) = \sum_{i=1,2,3} x_i^8 + p_0(x)$, where $p_0(x)$ is the Motzkin polynomial $x_1^2 x_2^2(x_1^2 + x_2^2 - 3x_3^2) + x_3^6$. It is known that $p^* = 0$ and that p is not a sum of squares (in fact, p is not a sum of squares modulo its gradient ideal [6]). In view of the form of p, we can apply directly our method for computing p^*, without perturbing p. Table 9 shows values of the parameter μ^*_k from (25); as $\mu^*_k \le p^* \le 0$, we can conclude that $p^* \sim 0$ already at the relaxation of order $k = 4$. Here $n = 3$, $m = 3$, $|\mathcal{B}_3| = 20$, $|\mathcal{B}_4| = 35$, $|\mathcal{B}_5| = 56$, $|\mathcal{B}_6| = 84$.

order k	(r_k, r_{k-1}, r_{k-2})	μ^*_k	extracted solutions
3	(12,2,1)	-1	none
4	(4,4,4)	$-1.1990\ 10^{-9}$	$\pm(0.0220, 0.0440, 0.0263)$ and approx. $(0,0,0)$ twice
5	(4,4,4)	$-1.9880\ 10^{-10}$	$\pm(0.0160, 0.0319, 0.0274)$ and approx. $(0,0,0)$ twice
6	(4,4,4)	$-8.8465\ 10^{-11}$	$\pm(0.0143, 0.0285, 0.0256)$ and approx. $(0,0,0)$ twice

TABLE 9. Bounds $\mu^*_{k,\lambda}$ for Example 9

EXAMPLE 10. Consider the polynomial $p(x) = (a^T x)^2 + \sum_{i=1}^n (x_i^2 - 1)^2$, where $a_1, \ldots, a_n$ are given positive integers. As mentioned in the Introduction, the sequence $a = (a_1, \ldots, a_n)$ can be partitioned if and only if $p^* = 0$, in which case a global minimizer is ± 1-valued and thus provides a partition of the sequence. Deciding whether an integer sequence can be partitioned is an NP-complete problem and, more generally, computing the parameter $\gamma := \min_{z \in \{\pm 1\}^n} |a^T z|$ (the *minimum gap* of the sequence $a_1, \ldots, a_n$) is NP-hard.

It is interesting to note[1] that $\gamma = 0$ (resp., $\gamma = 1$) if $p^* \le \frac{1}{s^2}$ and $s := \sum_{i=1}^n a_i$ is even (resp., odd); moreover, a partition realizing the minimum gap can be obtained from a real point x with $p(x) \le \frac{1}{s^2}$ by letting $z := sign(x)$ (with $z_i = 1$ if $x_i > 0$ and $z_i = -1$ otherwise). More generally, a similar argument permits to show that a partition realizing the minimum gap γ can be derived from a global minimizer x to the polynomial $p_C(x) := (a^T x)^2 + C^2 \sum_{i=1}^n (x_i^2 - 1)^2$, by letting $z := sign(x)$, $C := \frac{1}{2}(\max_i a_i)(\sum_i a_i)$.

Again we can apply directly our method (without perturbation) for computing the minimum p^* of the polynomial p. If we find a positive lower bound μ^*_k, then we can conclude that the sequence cannot be partitioned. Although this approach can be used

---

[1] Indeed, let $x \in \mathbb{R}^n$ such that $p(x) \le \frac{1}{s^2}$; thus $|a^T x|, |x_i^2 - 1| \le \frac{1}{s}$. Define $z := sgn(x)$, i.e., $z_i := 1$ if $x_i > 0$ and $z_i = -1$ otherwise. Then, $|a^T z| \le |a^T(x - z)| + |a^T x| \le 1 + \frac{1}{s} < 2$; indeed, $|a^T(x - z)| \le \sum_i a_i |x_i - z_i| \le \sum_i a_i |x_i - z_i||x_i + z_i| = \sum_i a_i |1 - x_i^2| \le \frac{1}{s} \sum_i a_i = 1$. As $|a^T z|$ has the same parity as $s = \sum_i a_i$, $a^T z = 0$ if s is even, and $a^T z = \pm 1$ otherwise, which shows that the ± 1-vector z provides a partition of the sequence $a_1, \ldots, a_n$ realizing the minimum gap.

only for sequences of small length n (where the minimum gap could in fact easily be found directly), we consider below some sequences of length $n = 5, 6, 7, 10, 11$ to see the behaviour of the method. We have: $m = 1$, $(|\mathcal{B}_1|, |\mathcal{B}_2|, |\mathcal{B}_3|) = (6, 21, 51)$ (resp., $(7, 28, 78)$, $(8, 36, 113)$, $(11, 66, 276)$, $(12, 78, 353)$) if $n = 5$ (resp., $n = 6$, $n = 7$, $n = 10$, $n = 11$) and $|\mathcal{B}| = 3^n$.

order k	(r_k, r_{k-1}, r_{k-2})	μ_k^*	extracted solutions
2	(10,5,1)	$2.3994\ 10^{-9}$	none
3	(2,2,2)	$2.5072\ 10^{-10}$	$\pm(1, 1, 1, -1, -1)$

TABLE 10a. The sequence $a = (2, 2, 2, 3, 3)$ is partitionable
with $a^T x = 0$ at the returned solutions

order k	(r_k, r_{k-1}, r_{k-2})	μ_k^*	extracted solutions
2	(12,5,1)	0.0639	none
3	(2,2,2)	0.0657	$\pm(1.0157, 1.0308, -0.9477, 1.0590, -0.9069)$

TABLE 10b. The sequence $a = (1, 2, 3, 4, 5)$ is not partitionable
as $p^* \geq \mu_3^* \geq \mu_2^* > 0$; its minimum gap is 1, realized at $\pm(1, 1, -1, 1, -1)$,
obtained by rounding the extracted solutions

order k	(r_k, r_{k-1}, r_{k-2})	μ_k^*	extracted solutions
2	(2,2,1)	$2.2649\ 10^{-12}$	$\pm(1, 1, -1, 1, -1)$

TABLE 10c. The sequence $a = (2, 2, 3, 4, 5)$ is partitionable
with $a^T x = 0$ at the returned solutions

order k	(r_k, r_{k-1}, r_{k-2})	μ_k^*	extracted solutions
2	(15,6,1)	$-1.5649\ 10^{-8}$	none
3	(4,4,3)	$1.3816\ 10^{-8}$	$\pm(1, -1, -1, 1, 1, -1)$, $\pm(1, -1, 1, -1, -1, 1)$

TABLE 10d. The sequence $a = (3, 3, 4, 5, 6, 7)$ is partitionable
with $a^T x = 0$ at the returned solutions

order k	(r_k, r_{k-1}, r_{k-2})	μ_k^*	extracted solutions
2	(2,2,1)	0.0188	$\pm(1.0045, 1.0045, 1.0090, 1.0090, 1.0135, 1.0135, -0.9342)$

TABLE 10e. The sequence $a = (1, 1, 2, 2, 3, 3, 13)$ is not partitionable
as $p^* \geq \mu_2^* > 0$; its minimum gap is 1, realized at $\pm(1, 1, 1, 1, 1, 1, -1)$,
obtained by rounding the extracted solutions

order k	(r_k, r_{k-1}, r_{k-2})	μ_k^*	extracted solutions
2	(2,2,1)	0.0628	$\pm(1.0073, 1.0073, 1.0145, 1.0145, 1.0215, 1.0215, -0.8736)$

TABLE 10f. The sequence $a = (1, 1, 2, 2, 3, 3, 14)$ is not partitionable,
as $p^* \geq \mu_2^* > 0$; its minimum gap is 2, realized at $\pm(1, 1, 1, 1, 1, 1, -1)$,
obtained by rounding the extracted solutions

order k	(r_k, r_{k-1}, r_{k-2})	μ_k^*	extracted solutions
2	(2,2,1)	0.0758	$\pm(1.0015, 1.0029, 1.0044, 1.0282, 1.0073,$ $1.0087, 1.0101, 1.0144, 1.0158, -0.8580)$

TABLE 10g. The sequence $a = (1, 2, 3, 20, 5, 6, 7, 10, 11, 77)$ is not partitionable as $p^* \geq \mu_2^* > 0$; its minimum gap is 12, realized at $\pm(1, 1, 1, 1, 1, 1, 1, 1, 1, -1)$, obtained by rounding the extracted solutions

order k	(r_k, r_{k-1}, r_{k-2})	μ_k^*	extracted solutions
2	(2,2,1)	0.0441	$\pm(1.0012, 1.0023, 1.0035, 1.0225, 1.0058,$ $1.0069, 1.0080, 1.0114, 1.0126, -0.8943, 1.0035)$

TABLE 10h. The sequence $a = (1, 2, 3, 20, 5, 6, 7, 10, 11, 77, 3)$ is not partitionable as $p^* \geq \mu_2^* > 0$; its minimum gap is 9, realized at $\pm(1, 1, 1, 1, 1, 1, 1, 1, 1, -1, 1)$, obtained by rounding the extracted solutions

4. CONCLUSIONS

We consider the problem of computing the global infimum p^* of a multivariate polynomial p of degree $2m$. We propose a method for determining upper approximations μ_λ^* (or $\mu_{k,\lambda}^*$ for some integer $k \geq m$) for the infimum, that converge to p^* as λ goes to 0. In the examples on which our method was tested, a tight upper bound $\mu_{k,\lambda}^*$ for p^* is very often found for k small ($k = m + 1$ or $m + 2$) by solving a semidefinite program of reasonable size, together with a real point x whose evaluation $p(x)$ gives a certificate for the upper bound. For small λ, $p(x)$ is in fact very close to the infimum p^* and x is close to a global minimizer (if some exists), which has been confirmed on the examples.

Our method applies to any polynomial; in particular, no assumption about existence of a minimum is needed. In fact, it works with a perturbation p_λ of p, which has the property of having a minimum as well as a finite set V_λ of critical points. Moreover, the minima μ_λ^* of p over the set V_λ converge to p^* as λ goes to 0. One has two options for computing the minimum μ_λ^*: Either, apply the moment relaxations of Lasserre [15], or apply the more compact relaxations via combinatorial moment matrices of Laurent [18] as proposed here. A feature of this second approach is that one has to solve smaller semidefinite programs and, moreover, one can often extract a solution (giving a certified upper bound for p^*) at an earlier stage than in the approach based on the classical moment relaxation. In fact, our method can be applied directly to polynomials of the form $p = \sum_i c_i x_i^{2m} + p_0$, where $c_i \neq 0$ and $\deg(p_0) \leq 2m - 1$, without perturbing p; then it gives a monotonically nondecreasing hierarchy of lower bounds μ_k^* on the infimum. A limitation for our method is the size of the matrix variable $M_{\mathcal{B}_k}(y)$ which has to be generated and then processed by the semidefinite solver. Thus it applies only to medium size problems.

Previous methods of Lasserre [15] and Parrilo [23] approximate the infimum of p by giving a hierarchy of lower bounds for p^*. Thus in a sense the various methods complement each other.

Parrilo's method computes for an integer $k \geq 0$ the parameter $\gamma_k^* := \sup \gamma$ s.t. $(\sum_i x_i^2)^k (p(x) - \gamma(\sum_i x_i^2)^m)$ is a sum of squares of polynomials. It is useful for proving that a homogeneous polynomial p is nonnegative, i.e., $p^* = 0$; indeed, if $\gamma_k^* \geq 0$ for

some k, then p is nonnegative. On the other hand, our method is useful for proving that a homogeneous polynomial is not nonnegative (e.g., for proving that a matrix is not coposiitive). Indeed, if one finds an upper bound $\mu_{k,\lambda}^* < 0$ for p^*, then p is not nonnegative; in the examples such certified negative upper bounds on the infimum p^* are (often) found for a small order $k = m + 1$ or $m + 2$.

When applied to the unconstrained minimization of p, Lasserre's approach gives a lower bound $p_{L,m}^*$ for p^*, with equality $p^* = p_{L,m}^*$ if and only if $p - p^*$ is a sum of squares. One can construct a hierarchy of lower bounds converging to p^* by considering the constrained problem of minimizing p over its gradient variety (when p has a minimum) or over a ball (when a ball is known a priori containing a global minimum).

Let us finally mention another method based on perturbations recently introduced by Lasserre [16]. Given $\epsilon > 0$ and an integer $k \geq 0$, define the perturbed polynomial $p_{k,\epsilon} := p + \epsilon \sum_{r=0}^{k} \sum_{i=1}^{n} \frac{x_i^{2r}}{r!}$. Lasserre [16] defines the parameter:

$$\ell_{k,\epsilon}^* := \inf \ p_{k,\epsilon}^T y \ \text{ s.t. } M_k(y) \succeq 0, \ y_0 = 1$$

and shows that, given $\epsilon > 0$, $p^* \leq \ell_{k,\epsilon}^*$ for k large enough, and $\ell_{k,\epsilon}^* \leq p^* + \epsilon \sum_{i=1}^{n} e^{x_i^2}$ if x is a global minimum of p. From the numerical results given in [16], it appears that the bound $\ell_{k,\epsilon}^*$ is sensitive to the parameter ϵ (e.g., $\ell_{k,\epsilon}^*$ does not approximate p^* very well for some values of k and small ϵ) and $\ell_{k,\epsilon}^*$ provides less good approximations of p^* than when solving a constrained program with the first order conditions (which is however allowed only when p has a minimum).

REFERENCES

[1] S. Basu, R. Pollack, M.-F. Roy. *Algorithms in Real Algebraic Geometry*. Springer, 2003.

[2] D. Cox, J. Little, and D. O'Shea. *Ideals, varieties and algorithms*. Springer, New York, 2nd edition, 1997.

[3] D. Cox, J. Little, and D. O'Shea. *Using Algebraic Geometry*. Graduate Textys in Mathematics, Springer, New York, 1998.

[4] R.E. Curto and L.A. Fialkow. *Solution of the truncated complex moment problem for flat data*, volume 119 of *Memoirs of the American Mathematical Society*. 1996.

[5] R.E. Curto and L.A. Fialkow. The truncated complex K-moment problem. *Transactions of the American Mathematical Society*, 352:2825–2855, 2000.

[6] J.W. Demmel, J. Nie and B. Sturmfels. Minimizing polynomials via sums of squares over the gradient ideal. ArXiv:math.OC/0411342 v3, December 2004.

[7] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness*, SAn Francisco, W.H. Freeman & Company, Publishers, 1979.

[8] K. Hägglöf, P.O. Lindberg, and L. Stenvenson. Computing global minima to polynomial optimization problems using gröbner bases. *Journal of Global Optimization*, 7(2):115–125, 1995.

[9] B. Hanzon and D. Jibetean. Global minimization of a multivariate polynomial using matrix methods. *Journal of Global optimization*, 27:1–23, 2003.

[10] D. Henrion and J.-B. Lasserre. GloptiPoly: Global optimization over polynomials with Matlab and SeDuMi. *ACM Trans. Math. Soft.*, 29: 165–194, 2003.

[11] D. Henrion and J.-B. Lasserre. Detecting global optimality and extracting solutions in GloptiPoly. In *Positive Polynomials in Control*, D. Henrion and A. Garulli, eds., Lecture Notes on Control and Information Sciences, Springer Verlag, 2005.

[12] D. Jibetean. *Algebraic optimization with applications to system theory*. PhD thesis. Vrije Universiteit, Amsterdam, 2003. Available at `http:// www.math.vu.nl/res/theses/jibetean.html`.

[13] W. Kaplan. A test for copositive matrices. *Linear Algebra and its Applications*, 313:203–206, 2000.

[14] E. De Klerk, M. Laurent, P. Parrilo. On the equivalence of algebraic approaches to the minimization of forms over the simplex. In *Positive Polynomials in Control*, D. Henrion and A. Garulli, eds., Lecture Notes on Control and Information Sciences, Springer Verlag, 2005.

[15] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[16] J.B. Lasserre. A new hierarchy of SDP-relaxations for polynomial programming. Preprint, 2004.

[17] M. Laurent. Revisiting two theorems of Curto and Fialkow about moment matrices. Preprint, 2004. To appear in *Memoirs of the AMS*.

[18] M. Laurent. Semidefinite representations for finite varieties. Preprint, 2002. To appear in *Mathematical Programming*.

[19] M. Marshall. Optimization of polynomial functions. *Canadian Bulletin of Mathematics*, 46:575–587, 2003.

[20] T.S. Motzkin and E.G. Straus. Maxima for graphs and a a new proof of a theorem of Túran. *Canadian J. Math.*, 17:533-540, 1965.

[21] K.G. Murty and S.N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39:117–129, 1987.

[22] P.A. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, May 2000.

[23] P.A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming B*, 96:293–320, 2003.

[24] P. Parrilo. An explicit construction of distinguished representations of polynomials nonnegative over finite sets. Preprint, ETH, Zürich, 2002.

[25] P.A. Parrilo and B. Sturmfels. Minimizing polynomial functions. In S. Basu, L. Gonzales-Vega (eds.), *Algorithmic and Quantitative Real Algebraic Geometry, DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, Vol. 60, 2003.

[26] M. Putinar. Positive polynomials on compact semialgebraic sets. *Indiana. Univ. Math. J.*, 42:969–984, 1993.

[27] B. Reznick. Uniform denominators in Hilbert's 17th problem. *Mathematische Zeitschrift*, 220(1):75–97, 1995.

[28] N. Shor. Class of global minimum bounds of polynomial functions. *Translated from Kibernetica*, 6:9–11, 1987.

D. Jibetean, TUe, Postbus 513, 5600 MB Eindhoven, The Netherlands.
E-mail address: djibetean@win.tue.nl

M. Laurent, CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands.
E-mail address: monique@cwi.nl