

OPTIMIZATION OF POLYNOMIALS ON COMPACT SEMIALGEBRAIC SETS

MARKUS SCHWEIGHOFER

ABSTRACT. A basic closed semialgebraic subset S of \mathbb{R}^n is defined by simultaneous polynomial inequalities $g_1 \geq 0, \dots, g_m \geq 0$. We give a short introduction to Lasserre's method for minimizing a polynomial f on a compact set S of this kind. It consists of successively solving tighter and tighter convex relaxations of this problem which can be formulated as semidefinite programs. We give a new short proof for the convergence of the optimal values of these relaxations to the infimum f^* of f on S which is constructive and elementary. In the case where f possesses a *unique* minimizer x^* , we prove that every sequence of "nearly" optimal solutions of the successive relaxations gives rise to a sequence of points in \mathbb{R}^n converging to x^* .

1. INTRODUCTION TO LASSERRE'S METHOD

Throughout the paper, we suppose $1 \leq n \in \mathbb{N}$ and abbreviate (X_1, \dots, X_n) by \bar{X} . We let $\mathbb{R}[\bar{X}]$ denote the ring of real polynomials in n indeterminates. Suppose we are given a so called basic closed semialgebraic set, i.e., a set

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

defined by polynomials $g_1, \dots, g_m \in \mathbb{R}[\bar{X}]$. We consider the problem of minimizing a polynomial $f \in \mathbb{R}[\bar{X}]$ on S in the case where S is compact. So we are interested in computing numerically the optimal value

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

and, if possible, an optimal point, i.e., an element of the set

$$S^* := \{x^* \in S \mid \forall x \in S : f(x^*) \leq f(x)\}.$$

Note that $S^* \neq \emptyset$ and $f^* = \min\{f(x) \mid x \in S\}$ unless $S = \emptyset$ since S is assumed to be compact.

We will now outline Lasserre's [Las] approach to solve this problem. It will be convenient to set $g_0 := 1 \in \mathbb{R}[\bar{X}]$. Furthermore, we denote by $\mathbb{R}[\bar{X}]^2$ the set of all squares p^2 of polynomials $p \in \mathbb{R}[\bar{X}]$, by $\mathbb{R}[\bar{X}]^2 g_i$ the set of all $p^2 g_i$, by $\sum \mathbb{R}[\bar{X}]^2 g_i$

1991 *Mathematics Subject Classification*. Primary 90C26, 13J30; Secondary 44A60, 14P10, 11E25.

Key words and phrases. nonconvex optimization, positive polynomial, sum of squares, moment problem, Positivstellensatz, semidefinite programming.

This work was partially done during the author's one year stay at Université de Rennes 1 supported by the European network RAAG (EC contract HPRN-CT-2001-00271). Part of the work was also supported by the project "Darstellung positiver Polynome" (Deutsche Forschungsgemeinschaft, Kennung 214371).

the set of all finite sums of such elements, and so on. In real algebra, the set

$$(1) \quad \begin{aligned} M &:= \sum \mathbb{R}[\bar{X}]^2 + \sum \mathbb{R}[\bar{X}]^2 g_1 + \cdots + \sum \mathbb{R}[\bar{X}]^2 g_m \\ &= \left\{ \sum_{i=0}^m \sigma_i g_i \mid \sigma_i \in \sum \mathbb{R}[\bar{X}]^2 \right\} \subseteq \mathbb{R}[\bar{X}] \end{aligned}$$

is called the quadratic module generated by g_1, \dots, g_m [PD, 5.1.7]. It is a set of polynomials that are nonnegative on S and possess a very nice certificate for this property. Lasserre's method works only if a certain assumption is satisfied. It can be expressed in several different ways that are equivalent by the following theorem.

Theorem 1 (Schmüdgen). *The following are equivalent:*

(i) *There exist finitely many $p_1, \dots, p_s \in M$ such that the set*

$$\{x \in \mathbb{R}^n \mid p_1(x) \geq 0, \dots, p_s(x) \geq 0\}$$

(which contains S) is compact and $\prod_{i \in I} p_i \in M$ for all $I \subseteq \{1, \dots, s\}$.

(ii) *There exists some $p \in M$ such that $\{x \in \mathbb{R}^n \mid p(x) \geq 0\}$ is compact.*

(iii) *There exists an $N \in \mathbb{N}$ such that $N - \sum_{i=1}^n X_i^2 \in M$.*

(iv) *For all $p \in \mathbb{R}[\bar{X}]$, there is some $N \in \mathbb{N}$ such that $N \pm p \in M$.*

It is clear that (iv) \implies (iii) \implies (ii) \implies (i). We don't include the easy proof of (iii) \implies (iv) [PD, 5.1.13][S1, 2.1] since this implication comes out as a byproduct in this article: Soon we will impose (iii) as a general assumption under which we will be able to prove Theorem 3 below which together with the compactness of S implies immediately (iv).

The essential part of Theorem 1 is the implication (i) \implies (iii). This is a weak form of Schmüdgen's Theorem [Sch, Corollary 3] because $N - \sum_{i=1}^n X_i^2 > 0$ holds on the compact set defined by the p_i when $N \in \mathbb{N}$ is big enough. The reader acquainted with real algebra should not be confused by the fact that in Schmüdgen's work there is no quadratic module appearing: When proving (i) \implies (iii) you can assume without loss of generality that M is even a so-called *preorder* (the one generated by the p_i). We will not include a proof of this weak form of Schmüdgen's Theorem and instead state our assumption in the seemingly stronger form (iii). We do this for the following reasons:

- Once assuming (iii), we have a new short proof of Theorem 3 (and thus the convergence of Lasserre's procedure to f^*) that uses only elementary analysis (see Section 2). It completely avoids any arguments from functional analysis used in Putinar's original proof [Put] or from real algebra used in the other proof of Jacobi [Jac][PD].
- In many practical applications, an $N \in \mathbb{N}$ is known such that S is contained in the closed ball around 0 with radius \sqrt{N} . Then the condition (iii) can be assured by adding the redundant inequality $N - \sum_{i=1}^n X_i^2 \geq 0$ in the definition of S . This increases the number m of g_i 's only by one.
- Whether it is decidable that the equivalent conditions of Theorem 1 are satisfied is not known and subject to current research. But (iii) seems to be the condition the most suitable for algorithmic exploitation among the given ones.
- Condition (i) is interesting if it is known that S is compact, but no ball containing S is known explicitly. It can be satisfied by adding the redundant inequalities $\prod_{i \in I} g_i \geq 0$ for all $I \subseteq \{1, \dots, m\}$ of cardinality $\#I \geq 2$ to the

definition of S . However, this gives rise to an exponential growth of the number of inequalities, which is fatal from the complexity point of view unless the number m of g_i 's is very small.

We refer readers interested in the proof of the implication (i) \implies (iii) to [PD, 5.1.17] or [S1, 2.2]. These are expositions of Wörmann's nice algebraic proof [BW, Theorem 4] that don't use a sledgehammer to crack a nut (the sledgehammer is [BW, Corollary 1]). Theorem 1 was first proved by Schmüdgen in his pioneering work [Sch] where he proved even Putinar's Theorems 2 and 3 in the special case where g_1, \dots, g_m are the $m = 2^s$ products $\prod_{i \in I} p_i$ of some polynomials p_1, \dots, p_s . In this context, we should also mention that deep work of Jacobi and Prestel gives criteria for the conditions of Theorem 1 to be satisfied that are even more applicable than condition (i) [JP][PD, 6.3]. We end the discussion around Theorem 1 once and for all by assuming from now on the following condition, which is really stronger than the compactness of S [PD, 6.3.1].

General assumption.

$$\exists N \in \mathbb{N} : N - \sum_{i=1}^n X_i^2 \in M$$

A fundamental obstacle when one tries to find f^* is that the set S and the function $f|_S : S \rightarrow \mathbb{R}$ are in general far from being convex. Of course, S need not even be connected. The first step towards Lasserre's method is to convexify the problem by brute force. There are two ways of doing this.

The first one is to exchange the points of S by probability measures on S . Every point $x \in S$ can be identified with the Dirac measure δ_x at x , and a convex combination of probability measures is again a probability measure. Obviously, we have

$$(2) \quad f^* = \inf \left\{ \int f d\mu \mid \mu \in \mathcal{M}^1(S) \right\}.$$

Here, for any set $A \subseteq \mathbb{R}^n$, $\mathcal{M}^1(A)$ denotes the set of all probability measures on A . To be precise, A is understood as a measurable space. More exactly, it is equipped with the σ -algebra $\mathcal{B}(A)$ of its Borel sets, i.e., the σ -algebra generated by all sets that are open in the topological space A . Consequently, a probability measure μ on A is a σ -additive map $\mu : \mathcal{B}(A) \rightarrow [0, \infty)$ with $\mu(\emptyset) = 0$ and $\mu(A) = 1$.

The second method of convexification is to take a dual standpoint, namely, to write

$$(3) \quad f^* = \sup \{ a \in \mathbb{R} \mid f - a \geq 0 \text{ on } S \} = \sup \{ a \in \mathbb{R} \mid f - a > 0 \text{ on } S \}.$$

In (2) and (3), we got rid of the usually nonconvex set S by introducing a convex object which is, however, very hard to describe and not suitable for algorithmic purposes — namely, the set $\mathcal{M}^1(S)$ of probability measures on S and the set of polynomials nonnegative (or positive) on S , respectively. The following key theorems of Putinar [Put] will help us to overcome this problem.

Theorem 2 (Putinar). *For any map $L : \mathbb{R}[\bar{X}] \rightarrow \mathbb{R}$, the following are equivalent:*

- (i) L is linear, $L(1) = 1$ and $L(M) \subseteq [0, \infty)$.
- (ii) L is integration with respect to a probability measure μ on S , i.e.,

$$\exists \mu \in \mathcal{M}^1(S) : \forall p \in \mathbb{R}[\bar{X}] : L(p) = \int p d\mu.$$

Theorem 3 (Putinar). *If $p \in \mathbb{R}[\bar{X}]$ satisfies $p > 0$ on S , then $p \in M$.*

Theorem 2 does not really describe $\mathcal{M}^1(S)$, but all real families $(a_\alpha)_{\alpha \in \mathbb{N}^n}$ that are the sequence of moments of a probability measure on S , that is to say,

$$a_\alpha = \int \bar{X}^\alpha d\mu \quad \text{for all } \alpha \in \mathbb{N}^n$$

where $\bar{X}^\alpha := X_1^{\alpha_1} \cdots X_n^{\alpha_n}$. This is clear from the fact that every linear map $L : \mathbb{R}[\bar{X}] \rightarrow \mathbb{R}$ is given uniquely by its values $L(\bar{X}^\alpha)$ on the basis $(\bar{X}^\alpha)_{\alpha \in \mathbb{N}^n}$ of $\mathbb{R}[\bar{X}]$. One says therefore that Theorem 2 is a solution to the *moment problem* on S . For our purposes this is sufficient since the integral of a polynomial depends only on the moments, and we can rewrite Equation (2) as

$$(4) \quad f^* = \inf\{L(f) \mid L : \mathbb{R}[\bar{X}] \rightarrow \mathbb{R} \text{ is linear, } L(1) = 1, L(M) \subseteq [0, \infty)\}.$$

Remark 4. Some articles (e.g., [Put]) prefer the viewpoint of a linear map $L : \mathbb{R}[\bar{X}] \rightarrow \mathbb{R}$ as a family $(L(\bar{X}^\alpha))_{\alpha \in \mathbb{N}^n}$. In other words, they use an infinite matrix representation of L (with respect to the standard basis of $\mathbb{R}[\bar{X}]$). Sometimes this is useful as we will see (especially in the proof) of Theorem 12.

A third way of seeing such an L is by looking at the bilinear forms

$$\mathbb{R}[\bar{X}] \times \mathbb{R}[\bar{X}] \rightarrow \mathbb{R} : (p, q) \mapsto L(pq)$$

(confer proof of Lemma 24) for $L(M) \subseteq [0, \infty)$ means nothing else than these bilinear forms are all positive semidefinite. Again, an infinite matrix representation of these bilinear forms can be useful, see for example [Lau].

Theorem 3 is called Putinar's *Positivstellensatz* in analogy to Hilbert's Nullstellensatz. Of course, it does not really characterize the polynomials positive on S since the polynomials lying in M must only be nonnegative on S but not positive. Also, it does not fully describe the polynomials nonnegative on S since they are not always contained in M [PD, Example before 8.2.3]. Nevertheless, we get together with (3) that

$$(5) \quad f^* = \sup\{a \in \mathbb{R} \mid f - a \in M\}.$$

Putinar proved first Theorem 2 using tools from functional analysis and then deduced Theorem 3 from it. Jacobi found an algebraic approach to these theorems. He proved an abstract generalization of Putinar's Positivstellensatz using tools of real algebra [Jac][PD, 5.3.7]. (In fact, the author observed that it is not hard to get that abstract theorem back from Theorem 3 using compactness arguments.) Putinar's solution to the moment problem can then be derived from his Positivstellensatz using the well-known Riesz Representation Theorem (see Section 3).

In Section 2, we shall present a new proof of Putinar's Positivstellensatz that uses only elementary arguments. Like an earlier approach of the author to Schmüdgen's Positivstellensatz [S1], it is based on an old theorem of Pólya, Theorem 6 below. At first glance, this earlier approach seems to be inferior to the one we shall take here, since it works only for Schmüdgen's Positivstellensatz, which is a special case of Theorem 3. On the other hand, it led to complexity results [S2] that we were despite all efforts not able to reproduce in the more general situation that we encounter here.

The idea is now to relax (4) and (5) by introducing approximations $M_k \subseteq \mathbb{R}[\bar{X}]_k$ of $M \subseteq \mathbb{R}[\bar{X}]$. Here, for $k \in \mathbb{N}$, $\mathbb{R}[\bar{X}]_k$ denotes the \mathbb{R} -vector space consisting of the polynomials $p \in \mathbb{R}[\bar{X}]$ of degree $\deg p \leq k$ (we set $\deg 0 := -\infty$). So we define, in analogy to (1),

$$(6) \quad \begin{aligned} M_k &:= \sum \mathbb{R}[\bar{X}]_{d_0}^2 + \sum \mathbb{R}[\bar{X}]_{d_1}^2 g_1 + \cdots + \sum \mathbb{R}[\bar{X}]_{d_m}^2 g_m \\ &= \left\{ \sum_{i=0}^m \sigma_i g_i \mid \sigma_i \in \sum \mathbb{R}[\bar{X}]^2, \deg(\sigma_i g_i) \leq k \right\} \end{aligned}$$

for arbitrary

$$(7) \quad k \in \mathcal{N} := \{s \in \mathbb{N} \mid s \geq \max\{\deg g_1, \dots, \deg g_m, \deg f\}\}.$$

Here the d_i depend on k and $\deg g_i$, more exactly,

$$(8) \quad d_i := \max\{d \in \mathbb{N} \mid 2d + \deg g_i \leq k\}$$

if $g_i \neq 0$, and (for instance) $d_i := 0$ otherwise. The equality in (6) follows from the fact that the leading form (the highest homogeneous part, i.e., the sum of the monomials of highest degree) of a nonzero square polynomial is a square polynomial and a fortiori a globally nonnegative polynomial. Hence, when adding square polynomials, no cancellation of leading forms can occur because the sum of two or more nonzero globally nonnegative polynomials cannot vanish. In contrast to this, cancellation of leading forms of the terms in the sum $\sum_{i=0}^m \sigma_i g_i$ is a frequent phenomenon on which the validity of Theorem 3 relies. In the expression of a given polynomial p as such a sum, the degree of the terms of the sum sometimes necessarily has to exceed $\deg p$ enormously [Ste] (see also [S2][PD, 8.3]). So M_k should never be confused with $M \cap \mathbb{R}[\bar{X}]_k \supseteq M_k$.

Replacing M by M_k in (4) and (5) motivates the consideration of the following pair of optimization problems for each $k \in \mathcal{N}$.

$$(P_k) \quad \begin{array}{ll} \text{minimize} & L(f) \quad \text{subject to} \quad L : \mathbb{R}[\bar{X}]_k \rightarrow \mathbb{R} \text{ is linear,} \\ & L(1) = 1 \text{ and} \\ & L(M_k) \subseteq [0, \infty) \end{array}$$

$$(D_k) \quad \begin{array}{ll} \text{maximize} & a \quad \text{subject to} \quad a \in \mathbb{R} \text{ and} \\ & f - a \in M_k \end{array}$$

We call (P_k) and (D_k) the *primal* and *dual* relaxation of *order* $k \in \mathcal{N}$, respectively. The *optimal value* of (P_k) , i.e., the infimum over all $L(f)$ where L ranges over all feasible solutions of (P_k) , is denoted by $P_k^* \in \mathbb{R} \cup \{\pm\infty\}$. Analogously, we write $D_k^* \in \mathbb{R} \cup \{\pm\infty\}$ for the *optimal value* of (D_k) , i.e., the supremum of the feasible set $\{a \in \mathbb{R} \mid f - a \in M_k\}$ of (D_k) .

For every $x \in S$, the evaluation at x

$$\varepsilon_x : \mathbb{R}[\bar{X}]_k \rightarrow \mathbb{R} : p \mapsto p(x)$$

is a feasible solution of (P_k) . This shows that $P_k^* \leq f^*$. Moreover, if L is feasible for (P_k) and a for (D_k) , then $L(f) \geq a$ since $f - a \in M_k$ implies $L(f) - a = L(f) - aL(1) = L(f - a) \geq 0$. We therefore get also $D_k^* \leq P_k^*$. It is easy to see that every feasible solution $a \in \mathbb{R}$ of (D_k) is also feasible for (D_{k+1}) , and every feasible solution L of (P_{k+1}) is feasible for (P_k) when restricted to the subspace $\mathbb{R}[\bar{X}]_k$ of $\mathbb{R}[\bar{X}]_{k+1}$ ($k \in \mathcal{N}$). This tells us that the sequences $(D_k)_{k \in \mathcal{N}}$ and $(P_k)_{k \in \mathcal{N}}$ are increasing. Furthermore, $(D_k^*)_{k \in \mathcal{N}}$ and consequently $(P_k^*)_{k \in \mathcal{N}}$ converges to f^* .

Indeed, for any $\varepsilon > 0$, we have for sufficiently large $k \in \mathcal{N}$ that $f - f^* + \varepsilon \in M_k$ by Theorem 3, i.e., $f^* - \varepsilon$ is feasible for (D_k) , whence $f^* - \varepsilon \leq D_k^*$. Summarizing, we have seen that the following is true.

Theorem 5 (Lasserre). *$(D_k^*)_{k \in \mathcal{N}}$ and $(P_k^*)_{k \in \mathcal{N}}$ are increasing sequences that converge to f^* and satisfy $D_k^* \leq P_k^* \leq f^*$ for all $k \in \mathcal{N}$.*

In special cases, something has been proved about the rate of convergence. For example, if $S = \{x \in \mathbb{R}^n \mid g(x) \geq 0\}$ is defined by only one polynomial $g \in \mathbb{R}[\bar{X}]$, then there is a constant $c \in \mathbb{N}$ depending on f and g and a constant $0 < d \in \mathbb{N}$ depending on g such that

$$f^* - D_k^* \leq \frac{c}{\sqrt[d]{k}} \quad \text{for big } k.$$

Much more precise information can be found in [S2, Section 2]. It is based on the author's earlier constructive approach to representations of positive polynomials [S1]. The new approach taken here is simpler and more general as long as one is interested only in the mere existence of representations of positive polynomials. But on the other hand, it seems to be less suited to the purpose of keeping track of complexity. At least, we didn't succeed in proving a similar rate of convergence for the general case.

Lasserre observed that (P_k) and (D_k) (when modified insignificantly) can be easily translated into semidefinite programs and are as such dual to each other. We will recapitulate this in Section 5. Semidefinite programs are generalizations of linear programs and can be solved efficiently. Currently, there exist two implementations of the method [HL][PPP]. Practical experience shows that f^* is often actually reached by D_k^* and P_k^* already for small $k \in \mathcal{N}$.

If $S \neq \emptyset$, one is usually not only interested in finding the minimum value f^* of f on S , but also in obtaining a minimizer $x^* \in S^*$. Section 3 deals with this problem. Speaking very roughly, Theorem 12 shows that the method allows for getting finite families of real numbers that are arbitrarily close to being the family of moments up to some fixed order of some probability measure on S^* . In case $S^* = \{x^*\}$ is a singleton, i.e., f possesses a *unique* minimizer, there is only one probability measure on S^* . Its moments of order one are the coordinates of the unique minimizer x^* , and we can consequently approximate them arbitrarily closely (see Corollary 13). So the method is good for computing minimizers when they are unique. As soon as there are two or more minimizers, symmetries in the problem, which often occur in practice, frequently prevent the algorithm from steering towards one of the minimizers. In the proof of Theorem 12, we use Theorem 2. For convenience of the reader, we include the brief well-known derivation of Theorem 2 from Theorem 3 at the beginning of Section 3.

In Section 4, we investigate the duality of the problems (P_k) and (D_k) . If S has nonempty interior, then there is no duality gap, i.e., $P_k^* = D_k^*$ for all $k \in \mathcal{N}$. This result is essentially due to Lasserre who proved it by applying the duality theory of semidefinite programming. However, the duality of the problems (P_k) and (D_k) is much easier to understand than the duality theory of semidefinite programming in general. We try to give the reader a good visual image of this duality, and then take Marshall's [Mar] approach to it, which gives in our opinion a much better understanding than looking at the duality in general semidefinite programming. We liberate Marshall's exposition from arguments using algebraic geometry and

present it in a less technical framework. The only really new material in Section 4 are some negative examples in the case where S has no interior points.

The author would like to thank Chip Delzell for his numerous useful comments concerning the exposition of the present work.

2. CONVERGENCE TO THE INFIMUM: AN NEW PROOF OF PUTINAR'S THEOREM

In this section, we present our new proof of Theorem 3. Thinking in terms of optimization, this means that we prove Theorem 5, that is, the convergence of $(D_k^*)_{k \in \mathbb{N}}$ (and hence of $(P_k^*)_{k \in \mathbb{N}}$) to f^* . The key is the following theorem proved by Pólya in 1928 [Pól]. Like the proof of all the other statements in this section, its proof requires only very elementary analysis. We recommend [PR] as a reference where in addition a bound on the exponent k can be found.

Theorem 6 (Pólya). *Suppose $F \in \mathbb{R}[\bar{X}]$ is homogeneous and satisfies*

$$F > 0 \text{ on } [0, \infty)^n \setminus \{0\}.$$

Then for all big enough k , the polynomial

$$(X_1 + \cdots + X_n)^k F$$

has only nonnegative coefficients.

Lemma 7. *For all $k \in \mathbb{N}$ and $y \in [0, 1]$,*

$$(y - 1)^{2k} y \leq \frac{1}{2k + 1}.$$

Proof. In fact, by elementary calculus, we have even

$$(y - 1)^{2k} y \leq \left(\frac{1}{2k + 1} - 1 \right)^{2k} \frac{1}{2k + 1} \leq \frac{1}{2k + 1}.$$

□

Suppose now that we are given some $p \in \mathbb{R}[\bar{X}]$ with $p > 0$ on S . We want to show that it lies in M . The idea is to subtract first from p a certain element of M in order to extend the positivity condition from S to a much larger set C . To understand the next lemma, we propose to consider the pointwise limit of the left hand side of (9) below separately at points of $C \cap S$ and at points of $C \setminus S$, using the hypothesis that $g_i \leq 1$ on C together with the preceding lemma.

Lemma 8. *Suppose $C \subseteq \mathbb{R}^n$ is compact and $g_i \leq 1$ on C for all $i \in \{1, \dots, m\}$. Suppose $p \in \mathbb{R}[\bar{X}]$ satisfies $p > 0$ on S . Then there exists $s \in \mathbb{N}$ such that for all sufficiently large $k \in \mathbb{N}$,*

$$(9) \quad p - s \sum_{i=1}^m (1 - g_i)^{2k} g_i > 0 \quad \text{on } C.$$

Proof. Since S is compact, we can choose $\varepsilon > 0$ such that $p > \varepsilon$ on S . Set

$$A := \{x \in C \mid p(x) \leq \varepsilon\}.$$

Then A is compact and disjoint to S , whence we can choose $\delta > 0$ such that

$$(10) \quad \min\{g_1(x), \dots, g_m(x)\} \leq -\delta \quad \text{for } x \in A.$$

It suffices to show that for $s, k \in \mathbb{N}$ satisfying

$$(11) \quad p + \frac{s\delta}{2} > 0 \text{ on } C,$$

$$(12) \quad \frac{\delta}{2} \geq \frac{m-1}{2k+1} \quad \text{and}$$

$$(13) \quad \varepsilon \geq \frac{sm}{2k+1},$$

we have (9). If $x \in A$, then in the sum

$$(14) \quad \sum_{i=1}^m (1 - g_i(x))^{2k} g_i(x)$$

at most $m-1$ terms are nonnegative. By Lemma 7, these nonnegative terms add up to at most $\frac{m-1}{2k+1}$. At least one term is negative, even $\leq -\delta$ by (10). All in all, if we evaluate the left hand side of our claim (9) in a point $x \in A$, then it is

$$\geq p(x) - s \frac{m-1}{2k+1} + s\delta \geq \underbrace{p(x) + \frac{s\delta}{2}}_{>0 \text{ by (11)}} + s \underbrace{\left(\frac{\delta}{2} - \frac{m-1}{2k+1} \right)}_{\geq 0 \text{ by (12)}} > 0.$$

When we evaluate it at a point $x \in C \setminus A$, all terms of the sum (14) might happen to be nonnegative. Again by Lemma 7, they add up to at most $\frac{m}{2k+1}$. But at the same time, the definition of A gives us a good lower bound on $p(x)$, so that the result is

$$> \varepsilon - s \frac{m}{2k+1} \geq 0$$

by (13). □

Now we are able to prove Putinar's Positivstellensatz.

Proof of Theorem 3. In this proof, N is chosen like in the general assumption we have imposed on page 3. Consider the compact set

$$\Delta := \left\{ y \in [0, \infty)^{2n} \mid y_1 + \dots + y_{2n} = 2n \left(N + \frac{1}{4} \right) \right\} \subseteq \mathbb{R}^{2n}$$

and let $C := l(\Delta) \subseteq \mathbb{R}^n$ be its image under the linear map

$$l : \mathbb{R}^{2n} \rightarrow \mathbb{R}^n : y \mapsto \left(\frac{y_1 - y_{n+1}}{2}, \dots, \frac{y_n - y_{2n}}{2} \right).$$

Since $l(\Delta)$ is compact, we can scale each g_i with a positive factor such that $g_i \leq 1$ on C . Thereby no generality is lost since this affects neither S nor M (noting that all positive real numbers are squares). So we can apply the preceding lemma and get $s, k \in \mathbb{N}$ such that

$$q := p - s \sum_{i=1}^m (g_i - 1)^{2k} g_i > 0 \quad \text{on } C.$$

It suffices to show that $q \in M$, and we shall even show that

$$q \in T := \sum \mathbb{R}[\bar{X}]^2 + \sum \mathbb{R}[\bar{X}]^2 \left(N - \sum_{i=1}^n X_i^2 \right) \subseteq M.$$

To start with, write

$$q = \sum_{i=0}^d Q_i$$

where $d := \deg q$ and $Q_i \in \mathbb{R}[\bar{X}]$ is homogeneous of degree i , $i = 0, \dots, d$. Now define

$$F := \sum_{i=0}^d Q_i \left(\frac{Y_1 - Y_{n+1}}{2}, \dots, \frac{Y_n - Y_{2n}}{2} \right) \left(\frac{Y_1 + \dots + Y_{2n}}{2n(N + \frac{1}{4})} \right)^{d-i} \in \mathbb{R}[\bar{Y}]$$

where $\mathbb{R}[\bar{Y}]$ denotes the polynomial ring in $2n$ new indeterminates Y_1, \dots, Y_{2n} . For each $y \in \Delta$, we obtain

$$F(y) = \sum_{i=0}^d Q_i(l(y)) = q(l(y)) > 0,$$

since $l(y) \in l(\Delta) = C$. Since F is a homogeneous polynomial, it has constant sign on each ray emanating by the origin, whence

$$F > 0 \text{ on } [0, \infty)^{2n} \setminus \{0\}.$$

By Pólya's Theorem 6, there is some $e \in \mathbb{N}$ such that

$$G := \left(\frac{Y_1 + \dots + Y_{2n}}{2n(N + \frac{1}{4})} \right)^e F \in \mathbb{R}[\bar{Y}]$$

has only nonnegative coefficients. Now we apply on this polynomial the \mathbb{R} -algebra homomorphism $\varphi : \mathbb{R}[\bar{Y}] \rightarrow \mathbb{R}[\bar{X}]$ defined by

$$Y_i \mapsto \left(N + \frac{1}{4} \right) + X_i, \quad Y_{n+i} \mapsto \left(N + \frac{1}{4} \right) - X_i \quad (i \in \{1, \dots, n\}).$$

Note that $\varphi(Y_i) \in T$ for each $i \in \{1, \dots, 2n\}$ since

$$\left(N + \frac{1}{4} \right) \pm X_i = \sum_{j \neq i} X_j^2 + \left(X_i \pm \frac{1}{2} \right)^2 + \left(N - \sum_{j=1}^n X_j^2 \right) \in T.$$

Noting that T is closed under addition and multiplication, we obtain therefore that

$$\varphi(G) = \varphi(F) = \sum_{i=0}^d Q_i = q$$

is contained in T . □

3. CONVERGENCE TO THE UNIQUE MINIMIZER

Recall that \mathcal{N} is the set of k for which we have defined the relaxations (P_k) and (D_k) , see (7) on page 5. If, for some $k \in \mathcal{N}$, (P_k) happens to possess an optimal solution L that comes from a probability measure μ on S , that is,

$$\forall p \in \mathbb{R}[\bar{X}]_k : L(p) = \int p d\mu,$$

then we can conclude that

$$f^* \leq \int f d\mu = L(f) \leq f^*.$$

Consequently, $L(f) = f^*$ and $\mu(S^*) = 1$. If $S^* = \{x^*\}$ is a singleton, then μ must be the Dirac measure at x^* and

$$(L(X_1), \dots, L(X_n)) = \left(\int X_1 d\mu, \dots, \int X_n d\mu \right) = x^*.$$

In this section, we shall prove that, for high $k \in \mathcal{N}$, we are never terribly far from this nice situation (see Theorem 12). The important Corollary 13 tells us that Lasserre's procedure converges not only to the infimum f^* , but also to the minimizer x^* in the case that it is unique. These results are based on Theorem 2, whose well-established proof we include.

Proof of Theorem 2. The implication (ii) \implies (i) is trivial. To show the reverse, suppose that (i) holds. Consider the ring homomorphism

$$\varphi : \mathbb{R}[\bar{X}] \rightarrow \mathcal{C}(S, \mathbb{R}) : p \mapsto p|_S$$

from the polynomial ring into the ring $\mathcal{C}(S, \mathbb{R})$ of continuous real-valued functions on S . Suppose $p \in \mathbb{R}[\bar{X}]$ satisfies $p \geq 0$ on S . Then $p + \varepsilon \in M$ by Theorem 3 and a fortiori $L(p) + \varepsilon = L(p + \varepsilon) \geq 0$ for all $\varepsilon > 0$. This implies $L(p) \geq 0$. In particular, L vanishes on the kernel of φ and induces therefore a linear map $\bar{L} : \varphi(\mathbb{R}[\bar{X}]) \rightarrow \mathbb{R}$ well defined by $\bar{L}(\varphi(p)) := L(p)$ for all $p \in \mathbb{R}[\bar{X}]$. We equip $\mathcal{C}(S, \mathbb{R})$ with the supremum norm and thus turn it into a normed vector space, noting that $S = \emptyset$ would imply $-1 \in M$, whence $-1 = -L(1) = L(-1) \geq 0$. By the Stone-Weierstrass Approximation Theorem, $\varphi(\mathbb{R}[\bar{X}])$ lies dense in $\mathcal{C}(S, \mathbb{R})$. It is easy to see that $\bar{L}(\varphi(p)) = L(p) \leq \|p\|$ for all $p \in \mathbb{R}[\bar{X}]$. Hence the linear map \bar{L} is (uniformly) continuous. But every map uniformly continuous on a subspace of a metric space extends uniquely to a continuous map on the closure of this subspace. Therefore we may consider \bar{L} as a continuous map on the whole of $\mathcal{C}(S, \mathbb{R})$. Using again the Stone-Weierstrass Theorem, it is easy to see that \bar{L} maps $\mathcal{C}(S, [0, \infty))$ into $[0, \infty)$. Since S is compact, the Riesz Representation Theorem [Rud, 2.14] tells us that \bar{L} is integration with respect to a measure on S . \square

Definition 9. We say that L_k solves (P_k) nearly to optimality ($k \in \mathcal{N}$) if L_k is a feasible solution of (P_k) ($k \in \mathcal{N}$) such that $\lim_{k \rightarrow \infty} L_k(f) = \lim_{k \rightarrow \infty} P_k^*$.

If L_k is an optimal solution of (P_k) for every $k \in \mathcal{N}$, then, of course, L_k solves (P_k) nearly to optimality. The same applies for example if $|L_k(f) - P_k^*| \leq \frac{1}{k}$ for all big k . This is a useful notion because (P_k) might not possess an optimal solution (see Example 22 below), and even if it does, we might not be able to compute it exactly. In view of Theorem 5, we note for later reference the following.

Remark 10. Suppose L_k is a feasible solution of (P_k) ($k \in \mathcal{N}$). Then L_k solves (P_k) nearly to optimality ($k \in \mathcal{N}$) if and only if $\lim_{k \rightarrow \infty} L_k(f) = f^*$.

Notation 11. For $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$, we set

$$|\alpha| := \alpha_1 + \dots + \alpha_n$$

and, for $d \in \mathbb{N}$, we introduce the notation

$$\Lambda(d) := \{\alpha \in \mathbb{N}^n \mid |\alpha| \leq d\}.$$

Note that $(\bar{X}^\alpha)_{\alpha \in \Lambda(d)}$ is a basis of $\mathbb{R}[\bar{X}]_d$ for each $d \in \mathbb{N}$. We now give the exact formulation of the main theorem in this section. The norm on $\mathbb{R}^{\Lambda(d)}$ appearing in the theorem can be anyone.

Theorem 12. *Suppose that $S \neq \emptyset$ and L_k solves (P_k) nearly to optimality ($k \in \mathcal{N}$). Then*

$$\forall d \in \mathbb{N} : \forall \varepsilon > 0 : \exists k_0 \in \mathcal{N} \cap [d, \infty) : \forall k \geq k_0 : \exists \mu \in \mathcal{M}^1(S^*) :$$

$$\left\| \left(L_k(\bar{X}^\alpha) - \int \bar{X}^\alpha d\mu \right)_{\alpha \in \Lambda(d)} \right\| < \varepsilon.$$

Proof. Let $d \in \mathbb{N}$ and $\varepsilon > 0$ be given. For each $k \in \mathcal{N}$, there is some $N_k \in \mathbb{N}$ such that

$$N_k \pm \bar{X}^\alpha \in M_{N_k} \quad \text{for all } \alpha \in \mathbb{N}^n \text{ with } |\alpha| = k$$

by (iv) of Theorem 1 (note that $M = \bigcup_{k \in \mathcal{N}} M_k$ and $N_0 \pm \bar{X}^\alpha \in M_k$ implies $N \pm \bar{X}^\alpha \in M_N$ for all $N \geq \max\{N_0, k\}$). The topological space

$$Z := \prod_{\alpha \in \mathbb{N}^n} [-N_{|\alpha|}, N_{|\alpha|}]$$

is compact by Tychonoff's Theorem. Now we have that

$$(15) \quad \{(a_\alpha)_{\alpha \in \mathbb{N}^n} \in Z \mid a_0 = 1\} \cap$$

$$\begin{aligned} & \bigcap_{p \in M} \left\{ (a_\alpha)_{\alpha \in \mathbb{N}^n} \in Z \mid p = \sum_{\alpha} b_\alpha \bar{X}^\alpha, b_\alpha \in \mathbb{R}, \sum_{\alpha} b_\alpha a_\alpha \geq 0 \right\} \cap \\ & \bigcap_{\mu \in \mathcal{M}^1(S^*)} \left\{ (a_\alpha)_{\alpha \in \mathbb{N}^n} \in Z \mid \left\| \left(a_\alpha - \int \bar{X}^\alpha d\mu \right)_{\alpha \in \Lambda(d)} \right\| \geq \varepsilon \right\} \cap \\ & \bigcap_{\delta > 0} \left\{ (a_\alpha)_{\alpha \in \mathbb{N}^n} \in Z \mid \left| \sum_{\alpha} c_\alpha a_\alpha - f^* \right| \leq \delta \right\} = \emptyset \end{aligned}$$

where $(c_\alpha)_{\alpha \in \mathbb{N}^n}$ is the family of coefficients of f , i.e., $f = \sum_{\alpha} c_\alpha \bar{X}^\alpha$.

Indeed, an element $(a_\alpha)_\alpha$ of these huge intersection would give rise to a linear map $L : \mathbb{R}[\bar{X}] \rightarrow \mathbb{R} : \bar{X}^\alpha \mapsto a_\alpha$, which by Theorem 2 comes from a probability measure $\mu \in \mathcal{M}^1(S)$, i.e.,

$$\forall p \in \mathbb{R}[\bar{X}] : L(p) = \int p d\mu,$$

in other words,

$$(16) \quad \forall \alpha \in \mathbb{N}^n : a_\alpha = \int \bar{X}^\alpha d\mu.$$

Moreover, we would have $|L(f) - f^*| \leq \delta$ for all $\delta > 0$, whence

$$\int (f - f^*) d\mu = L(f - f^*) = L(f) - f^* = 0$$

so that $\mu(S^*) = 1$, and μ can be looked at as a probability measure on S^* , that is, $\mu \in \mathcal{M}^1(S^*)$. Regarding (16), this means that the a_α are the moments of a probability measure on S^* , which is impossible since $(a_\alpha)_{\alpha \in \Lambda(d)}$ has distance $\geq \varepsilon > 0$ to the family of moments up to order d of each probability measure on S^* . This establishes (15).

In fact, we can even find finitely many sets contributing to the intersection in (15) that have no point in common since it is an empty intersection of closed sets in a compact space. In particular, we get a tight version of (15) where we take

the intersection only over certain finitely many $p_1, \dots, p_s \in M$ and a sufficiently small $\delta > 0$ (instead of all $p \in M$ and all $\delta > 0$). Choose $t \in \mathcal{N} \cap [d, \infty)$ such that $p_1, \dots, p_s \in M_t \subseteq \mathbb{R}[\bar{X}]_t$. Choose

$$k_0 \geq \max\{t, N_0, \dots, N_t\}$$

so large that $|L_k(f) - f^*| \leq \delta$ for all $k \geq k_0$, which is possible by Remark 10. Now suppose that $k \geq k_0$ and define $(a_\alpha)_{\alpha \in \mathbb{N}^n}$ by

$$a_\alpha := \begin{cases} L_k(\bar{X}^\alpha) & \text{if } |\alpha| \leq t, \\ 0 & \text{otherwise.} \end{cases}$$

Then $(a_\alpha)_\alpha \in Z$ since, for $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq t$,

$$N_{|\alpha|} \pm a_\alpha = N_{|\alpha|} \pm L_k(\bar{X}^\alpha) = L_k(N_{|\alpha|} \pm \bar{X}^\alpha) \geq 0.$$

Here the inequality is implied by $N_{|\alpha|} \pm \bar{X}^\alpha \in M_{N_{|\alpha|}} \subseteq M_k$, which follows in turn from $N_{|\alpha|} \leq k_0 \leq k$. Furthermore, $a_0 = L_k(1) = 1$ and, for $b_\alpha \in \mathbb{R}$ such that $\sum_\alpha b_\alpha \bar{X}^\alpha \in \{p_1, \dots, p_s\}$,

$$\begin{aligned} \sum_\alpha b_\alpha a_\alpha &= \sum_\alpha b_\alpha L_k(\bar{X}^\alpha) = L_k\left(\sum_\alpha b_\alpha \bar{X}^\alpha\right) \in \{L_k(p_1), \dots, L_k(p_s)\} \\ &\subseteq L_k(M_t) \subseteq L_k(M_k) \subseteq [0, \infty). \end{aligned}$$

Because of $\deg f \leq t$ (recall $t \in \mathcal{N}$), we have moreover

$$\left| \sum_\alpha c_\alpha a_\alpha - f^* \right| = |L_k(f) - f^*| \leq \delta,$$

and the tight version of (15) implies the existence of $\mu \in \mathcal{M}^1(S^*)$ such that

$$\left\| \left(a_\alpha - \int \bar{X}^\alpha d\mu \right)_{\alpha \in \Lambda(d)} \right\| < \varepsilon.$$

But here we can replace a_α by $L_k(\bar{X}^\alpha)$ since $d \leq t$. \square

Corollary 13. *Suppose that $S^* = \{x^*\}$ is a singleton and L_k solves (P_k) nearly to optimality ($k \in \mathcal{N}$). Then*

$$\lim_{k \rightarrow \infty} (L_k(X_1), \dots, L_k(X_n)) = x^*.$$

Proof. We set $d = 1$ in the preceding theorem and note that $\mathcal{M}^1(S^*)$ contains only the Dirac measure at the point x^* . \square

As a nice application, we can approximate unique minimizers of polynomials on polytopes not only from below but also from above. This allows one to eventually detect with certitude that f^* has almost been reached when successively solving the relaxations (P_k) . Note that the general assumption from page 3 is automatically satisfied under the hypothesis that S is compact and $\deg g_i \leq 1$ for all $i \in \{1, \dots, m\}$ [PD, 6.5.3].

Corollary 14. *Suppose all the g_i have degree ≤ 1 , f has a unique minimizer on the polytope S and L_k solves (P_k) nearly to optimality ($k \in \mathcal{N}$). Then for each $k \in \mathcal{N}$,*

$$[L_k(f), f(L_k(X_1), \dots, L_k(X_n))]$$

is an interval containing f^* , and the left and right endpoints of these intervals converge to f^* for $k \rightarrow \infty$, respectively.

Proof. Taking into account Theorem 5, it suffices to show that

$$(f(L_k(X_1), \dots, L_k(X_n)))_{k \in \mathcal{N}}$$

is a sequence, which is bounded from below by f^* and converges to f^* . The convergence follows immediately from the preceding corollary. To see that f^* is a lower bound, observe that

$$g_i(L_k(X_1), \dots, L_k(X_n)) = L_k(g_i) \geq 0,$$

whence $(L_k(X_1), \dots, L_k(X_n)) \in S$ for all $k \in \mathcal{N}$. \square

4. DUALITY

Throughout this section, we fix some $k \in \mathcal{N}$ (\mathcal{N} has been defined in (7) on page 5). The kernel of any feasible solution of (P_k) is a hyperplane of the finite-dimensional vector space $\mathbb{R}[\bar{X}]_k$ with the following properties:

- (i) The convex cone M_k is contained in one of the two closed halfspaces of $\mathbb{R}[\bar{X}]_k$ defined by the hyperplane.
- (ii) The vector $1 \in \mathbb{R}[\bar{X}]_k$ lies not on the hyperplane.

Conversely, all hyperplanes fulfilling (i) and (ii) are the kernel of a feasible solution of (P_k) .

Now the following image is helpful. Suppose at the point $f \in \mathbb{R}[\bar{X}]_k$ there is a source that produces subatomic particles and sends them on the way in direction of the vector $-1 \in \mathbb{R}[\bar{X}]_k$. Every particle has constant speed and covers a distance of 1 per time unit (this will correspond to the constraint $L(1) = 1$ in (P_k)). You are interested in the age D_k^* of such a particle when it leaves the cone M_k (suppose that $D_k^* \in [0, \infty)$ in order to make this interpretation work). The only measuring instrument you have is a detector that catches the particles with a big metalplate of which we think as a hyperplane. When the instrument detects a particle it can determine its age. Since the cone M_k is a material body, the metalplate can touch it but not interpenetrate it, i.e., it satisfies constraint (i). If we brought the metalplate in a position where it contains the vector $1 \in \mathbb{R}[\bar{X}]_k$, then it would detect particles of all ages. So we will obey constraint (ii) while measuring. So P_k^* can be seen as the age you can measure by moving around the metalplate, and D_k^* as the actual age of the particles when they leave the cone.

This visualization nicely illustrates “weak duality” $P_k^* \geq D_k^*$, since a particle cannot be detected before it leaves the cone. It also suggests that “strong duality” $P_k^* = D_k^*$ should hold in many cases because it should often be possible to catch a particle just when it leaves the cone. However, you can imagine that there might be problems when the particle moves along the boundary of M_k , and M_k is not closed. In this section, we will discuss these matters in detail.

Notation 15. For $s \in \mathbb{N}$, we denote by $\mathbb{R}^{s \times s}$ the vector space of real $s \times s$ matrices. Defining the scalar product $\langle A, B \rangle$ of two such matrices A and B by

$$(17) \quad \langle A, B \rangle := \sum_{i,j=1}^s A_{ij} B_{ij},$$

makes $\mathbb{R}^{s \times s}$ into an Euclidean vector space. We write $S\mathbb{R}^{s \times s}$ for its subspace of symmetric matrices. By $S\mathbb{R}_+^{s \times s}$, we mean the set of symmetric *positive semidefinite* $s \times s$ matrices. It is a closed convex cone of $S\mathbb{R}^{s \times s}$.

Often, we will index rows and columns of matrices by elements of $\Lambda(d)$ for some $d \in \mathbb{N}$ (recall Notation 11). We denote the corresponding objects by $\mathbb{R}^{\Lambda(d) \times \Lambda(d)}$, $S\mathbb{R}^{\Lambda(d) \times \Lambda(d)}$ and $S\mathbb{R}_+^{\Lambda(d) \times \Lambda(d)}$. We write $A(\alpha, \beta)$ (instead of $A_{\alpha\beta}$) for the entry of a matrix $A \in \mathbb{R}^{\Lambda(d) \times \Lambda(d)}$ at line $\alpha \in \Lambda(d)$ and column $\beta \in \Lambda(d)$. Accordingly, we consider the vector space $\mathbb{R}^{\Lambda(d)}$ consisting of vectors v with entries $v(\alpha)$, $\alpha \in \Lambda(d)$.

Finally, we extend the definition of $\langle A, B \rangle$ by (17) to matrices A and B with polynomial entries.

Lemma 16. *If $A \in S\mathbb{R}_+^{s \times s}$, then $A = \sum_{i=0}^s v_i v_i^T$ for some vectors $v_1, \dots, v_s \in \mathbb{R}^s$.*

Proof. As A is symmetric, there exists an (orthogonal) matrix $B \in \mathbb{R}^{s \times s}$ and a real diagonal matrix D such that $A = BDB^T$. The entries of D are the eigenvalues of A , which are all nonnegative by hypothesis. So there exists a real diagonal matrix \sqrt{D} such that $(\sqrt{D})^2 = D$. Therefore we get $A = (B\sqrt{D})(B\sqrt{D})^T = CC^T$ where $C := B\sqrt{D}$. Choosing v_1, \dots, v_s as the columns of C yields the desired expression of A . \square

The following lemma relates positive semidefinite matrices to sums of squares of polynomials. More about this well-known connection can be found in [CLR].

Lemma 17. *For all $d \in \mathbb{N}$,*

$$\begin{aligned} \sum \mathbb{R}[\bar{X}]_d^2 &= \left\{ \langle (\bar{X}^{\beta+\gamma})_{(\beta, \gamma) \in \Lambda(d) \times \Lambda(d)}, G \rangle \mid G \in S\mathbb{R}_+^{\Lambda(d) \times \Lambda(d)} \right\} \\ &= \{q_1^2 + \dots + q_s^2 \mid q_1, \dots, q_s \in \mathbb{R}[\bar{X}]_d\} \end{aligned}$$

where $s := \dim \mathbb{R}[\bar{X}]_d = \#\Lambda(d)$ denotes the dimension of the \mathbb{R} -vector space $\mathbb{R}[\bar{X}]_d$.

Proof. To see that the first set is contained in the second one, note that the second one is closed under addition. So it suffices to show that it contains q^2 for each $q \in \mathbb{R}[\bar{X}]_d$. The polynomial q defines a vector $v \in \mathbb{R}^{\Lambda(d)}$ by setting $v(\alpha)$ to be the coefficient of \bar{X}^α in q for $\alpha \in \Lambda(d)$. Then $G := vv^T \in S\mathbb{R}^{\Lambda(d) \times \Lambda(d)}$ is positive semidefinite and

$$\begin{aligned} \langle (\bar{X}^{\beta+\gamma})_{(\beta, \gamma) \in \Lambda(d) \times \Lambda(d)}, G \rangle &= \sum_{\beta, \gamma \in \Lambda(d)} \bar{X}^{\beta+\gamma} G(\beta, \gamma) \\ &= \sum_{\beta, \gamma \in \Lambda(d)} \bar{X}^{\beta+\gamma} v(\beta)v(\gamma) = \left(\sum_{\beta \in \Lambda(d)} v(\beta) \bar{X}^\beta \right) \left(\sum_{\gamma \in \Lambda(d)} v(\gamma) \bar{X}^\gamma \right) = q^2. \end{aligned}$$

To show that the second set is contained in the last one, first consider the case that $G = vv^T$ for some vector $v \in \mathbb{R}^{\Lambda(k)}$. If we now set $q := \sum_{\alpha \in \Lambda(k)} v(\alpha) \bar{X}^\alpha$, then the above equation holds again. In the general case, use Lemma 16 and argue similarly. Finally, it is trivial that the last set is contained in the first one. \square

Note that $\mathbb{R}[\bar{X}]_k$ carries a natural topology, which is induced by any norm on it. The proof of the next result is essentially contained in [PS, 2.6], see also [Mar]. We repeat the proof without using results from algebraic geometry. The next remark should prevent misunderstandings concerning the notion of ‘‘interior’’ appearing in the statement.

Remark 18. The interior of S is understood in the usual topological sense, i.e., the set of all interior points of S where an interior point is the center of a nonempty open (full-dimensional) ball contained in S . Note that

$$S' := \{x \in \mathbb{R}^n \mid g_1(x) > 0, \dots, g_m(x) > 0\}$$

is contained in this interior. Examples where this inclusion is proper are trivial to find. Nevertheless, $S' = \emptyset$ implies that the interior of S is empty, too, provided that none of the g_i is the zero polynomial. Indeed, $S' = \emptyset$ would imply that the polynomial $g_1 \cdots g_m$ vanishes on S . If the interior of S were nonempty, then $g_1 \cdots g_m = 0$ would follow.

Theorem 19 (Powers and Scheiderer). *If S has nonempty interior, then M_k is closed in $\mathbb{R}[\bar{X}]_k$.*

Proof. By (6) and Lemma 17, the convex cone M_k is the image under the continuous map

$$\varphi : \begin{cases} \mathbb{R}[\bar{X}]_{d_0}^{\Lambda(d_0)} \times \cdots \times \mathbb{R}[\bar{X}]_{d_m}^{\Lambda(d_m)} \rightarrow \mathbb{R}[\bar{X}]_k \\ ((q_{0\alpha})_{\alpha \in \Lambda(d_0)}, \dots, (q_{m\alpha})_{\alpha \in \Lambda(d_m)}) \mapsto \sum_{i=0}^m \sum_{\alpha \in \Lambda(d_i)} q_{i\alpha}^2 g_i. \end{cases}$$

This map is quadratically homogeneous, i.e., $\varphi(\lambda z) = \lambda^2 \varphi(z)$ for all $\lambda \in \mathbb{R}$ and z in the domain of φ . Let us assume without loss of generality that none of the g_i is the zero polynomial. Then there is some $x \in S$ with $g_i(x) > 0$ for all $i \in \{1, \dots, m\}$, confer the preceding remark. All g_i are positive on a neighborhood U of x . For z in the domain of φ , we have that $\varphi(z) \geq 0$ on U , but $\varphi(z)$ cannot vanish on U unless $z = 0$. In other words, the map φ is injective. Now equip the \mathbb{R} -vector space on which φ is defined with an arbitrary norm. Then the image V of the unit sphere with respect to this norm is a compact set not containing the origin. From the homogeneity of φ and the fact that M_k is a cone, it is easy to see that M_k is the set of nonnegative multiples of elements of V . Finally, it follows from $0 \notin V$ that M_k is closed. Indeed, any sequence in M_k is of the form $(\lambda_i v_i)_{i \in \mathbb{N}}$ where $\lambda_i \in [0, \infty)$ and $0 \neq v_i \in V$ for all $i \in \mathbb{N}$. Suppose that $p = \lim_{i \rightarrow \infty} \lambda_i v_i$. We have to show that $p \in M_k$. Since V is compact, the sequence $(v_i)_{i \in \mathbb{N}}$ has a convergent subsequence. Without loss of generality we may assume that $(v_i)_{i \in \mathbb{N}}$ itself is convergent, say $v := \lim_{i \rightarrow \infty} v_i$. Then $v \in V$ and a fortiori $v \neq 0$. Now the limit

$$\lim_{i \rightarrow \infty} \lambda_i = \lim_{i \rightarrow \infty} \frac{\|\lambda_i v_i\|}{\|v_i\|} = \frac{\|\lim_{i \rightarrow \infty} \lambda_i v_i\|}{\|\lim_{i \rightarrow \infty} v_i\|} = \frac{\|p\|}{\|v\|}$$

exists, so that

$$\lim_{i \rightarrow \infty} \lambda_i v_i = \left(\lim_{i \rightarrow \infty} \lambda_i \right) \left(\lim_{i \rightarrow \infty} v_i \right) = \frac{\|p\|}{\|v\|} v \in [0, \infty)V = M_k.$$

□

Following Marshall [Mar], we draw the following corollary. Note that a feasible solution L of (P_k) has to obey the constraint $L(1) = 1$. So the corollary is not really just an instance of the well-known fact that the bi-dual of a closed convex cone is the cone itself.

Corollary 20 (Marshall). *If S has nonempty interior, then M_k is the intersection of all halfspaces $\{p \in \mathbb{R}[\bar{X}]_k \mid L(p) \geq 0\}$ where L ranges over all feasible solutions of (P_k) .*

Proof. Consider an arbitrary $p \in \mathbb{R}[\bar{X}]_k \setminus M_k$. We have to show that there is a feasible solution of (P_k) such that $L(p) < 0$.

Fix some scalar product on the \mathbb{R} -vector space $\mathbb{R}[\bar{X}]_k$. Since M_k is closed by the preceding theorem, we are sure that there exists $q \in M_k$ with minimal distance to p . Choose a linear map $L_0 : \mathbb{R}[\bar{X}]_k \rightarrow \mathbb{R}$ whose kernel is the hyperplane perpendicular to $p - q$ satisfying $L_0(p - q) < 0$. Note that $L_0(q) = 0$ either because of $q = 0$ or because q is the orthogonal projection of p on the line $\mathbb{R}q$. So we have $L_0(p) = L_0(p - q) + L_0(q) = L_0(p - q) < 0$.

It is easy to see that $L_0(M_k) \subseteq [0, \infty)$. Indeed, if there were a $q' \in M_k$ such that $L_0(q') < 0$, then all points q'' on the line segment from q' to q in a neighborhood of q would be even closer to p than it is q . At the same time, the convexity of M_k would imply $q'' \in M_k$, contradicting the choice of q .

Of course, we have $L_0(1) \geq 0$ since $1 \in M_k$. If $L_0(1)$ happens to be positive, then we can scale L_0 with a positive factor in order to obtain an L with the desired properties. If $L_0(1) = 0$, then we choose any point $x \in S$ and set $L := \varepsilon_x + \lambda L_0$ for sufficiently large $\lambda \in [0, \infty)$ where $\varepsilon_x : \mathbb{R}[\bar{X}]_k \rightarrow \mathbb{R}$ is evaluation at the point x . \square

We can now proof strong duality. In the case that (D_k) has a feasible solution, this has been first proved by Lasserre using the duality theory of semidefinite programming [Las].

Corollary 21 (Lasserre). *Suppose that S has nonempty interior. Then $P_k^* = D_k^*$. In addition, if (D_k) has a feasible solution, then it has also an optimal solution.*

Proof. Suppose that $a \in \mathbb{R}$ is such that $f - a \notin M_k$. Then Corollary 20 ensures the existence of a feasible solution L of (P_k) such that $L(f - a) < 0$, whence $D_k^* \leq P_k^* \leq L(f) < a$. This implies all our claims. \square

The next example shows that (P_k^*) might fail to have an optimal solution, even if S has nonempty interior.

Example 22. We consider an example in $n := 2$ variables that we denote by X and Y instead of X_1 and X_2 . To prepare our example, consider a linear map $L : \mathbb{R}[X, Y]_2 \rightarrow \mathbb{R}$ satisfying $L(1) = 1$. We claim that the condition

$$(18) \quad L((aX + bY + c)^2) \geq 0 \quad \text{for all } a, b, c \in \mathbb{R}$$

is equivalent to the three simultaneous inequalities

$$(19) \quad (L(XY) - L(X)L(Y))^2 \leq \underbrace{(L(X^2) - L(X)^2)}_{\geq 0} \underbrace{(L(Y^2) - L(Y)^2)}_{\geq 0}.$$

Indeed, using a scaling argument, we can split up Condition (18) into the two conditions

$$(20) \quad L((X + bY + c)^2) \geq 0 \quad \text{for all } b, c \in \mathbb{R}, \text{ and}$$

$$(21) \quad L((Y + c)^2) \geq 0 \quad \text{for all } c \in \mathbb{R}.$$

We can rewrite (21) as

$$c^2 + 2L(Y)c + L(Y^2) \geq 0 \quad \text{for all } c \in \mathbb{R},$$

which is equivalent to

$$(22) \quad (L(Y))^2 \leq L(Y^2).$$

In a similar way, we see that (20) is equivalent to

$$(L(X + bY))^2 \leq L((X + bY)^2), \text{ i.e.,}$$

$$(L(X))^2 + 2bL(X)L(Y) + b^2(L(Y))^2 \leq L(X^2) + 2bL(XY) + b^2L(Y^2)$$

for all $b \in \mathbb{R}$. Now take all the terms to the right hand side and look at it as a (perhaps degenerate) parabola in b . This parabola opens up by (22). Writing down the condition that this parabola has at most one zero (taking care of degeneration) and combining this with (22), shows the equivalence of (18) and (19).

Having finished the preparation, consider now the set S defined by $m := 5$ polynomials $g_1 = XY - 1$, $g_2 = 2 + X$, $g_3 = 2 - X$, $g_4 = 2 + Y$ and $g_5 = 2 - Y$. It is the filled region in the following picture and clearly has interior points.

Note that our general assumption we made on page 3 is satisfied since $4(4 - X^2) = (g_2 + g_3)g_2g_3 = g_2^2g_3 + g_3^2g_2 \in M$, analogously $4(4 - Y^2) \in M$, and therefore $8 - (X^2 + Y^2) \in M$. We look at the second primal and dual relaxation, (P_2) and (D_2) . We observe that

$$M_2 = \sum \mathbb{R}[X, Y]_1^2 + [0, \infty)(XY - 1) + [0, \infty)(2 + X) + [0, \infty)(2 - X) \\ + [0, \infty)(2 + Y) + [0, \infty)(2 - Y).$$

Therefore, the feasible solutions of (P_2) are the linear maps $L : \mathbb{R}[X, Y]_2 \rightarrow \mathbb{R}$ satisfying $L(1) = 1$, Condition (19), $L(XY) \geq 1$ and $(L(X), L(Y)) \in [-2, 2] \times [-2, 2]$. Suppose that we want to minimize $f := X^2$. For any feasible solution L of (P_2) , we have that $L(f) > 0$. Otherwise, like in dominoes, all the subexpressions of (19) involving X would have to vanish, contradicting $L(XY) \geq 1$. On the other hand, for all $\varepsilon > 0$, the linear map $L : \mathbb{R}[X, Y]_2 \rightarrow \mathbb{R}$ defined by

$$L(X^2) = \varepsilon^2, \quad L(Y^2) = \frac{1}{\varepsilon^2}, \quad L(XY) = 1, \quad L(X) = L(Y) = 0 \quad \text{and} \quad L(1) = 1$$

is feasible for (P_k) . This proves that $P_2^* = 0$, and (P_2) has no optimal solution.

We conclude the section with an example where a big duality gap occurs. By Theorem 21, S cannot have interior points in such an example.

Example 23. Take again $n := 2$ variables X and Y . The set $S = [-1, 1] \times \{0\}$ defined by the $m := 3$ polynomials $g_1 := -Y^2$, $g_2 := 1 + X$ and $g_3 := 1 - X$ has empty interior. The general assumption from page 3 is trivially satisfied since

$$2(1 - (X^2 + Y^2)) = (g_2 + g_3)g_2g_3 + 2g_1 = g_2^2g_3 + g_3^2g_2 + (\sqrt{2})^2g_1 \in M.$$

This time, we want to minimize $f := XY$ on this set. Once more, we look at the second relaxations. It is easy to see that

$$M_2 = \sum \mathbb{R}[X, Y]_1^2 - [0, \infty)Y^2 + [0, \infty)(1 + X) + [0, \infty)(1 - X).$$

Every feasible solution L of (P_2) satisfies $L(\varepsilon X^2 + f) \geq 0$ for all $\varepsilon > 0$ since

$$\varepsilon X^2 + f = \left(\sqrt{\varepsilon} X + \frac{1}{2\sqrt{\varepsilon}} Y \right)^2 - \frac{1}{4\varepsilon} Y^2 \in M_2.$$

But then $\varepsilon L(X^2) + L(f) \geq 0$ for all $\varepsilon > 0$, i.e., $L(f) \geq 0$. This shows $0 \leq P_2^* \leq f^* = 0$, whence $P_2^* = f^* = 0$.

On the other hand, $D_2^* = -\infty < 0 = P_2^*$ because there is no $a \in \mathbb{R}$ such that $f - a \in M_2$. If there were $s \in \mathbb{N}$, $a, b_i, c_i, d_i, e_i \in \mathbb{R}$ such that

$$f - a = \sum_{i=1}^s (b_i X + c_i Y + d_i)^2 - e_1 Y^2 + e_2(1 + X) + e_3(1 - X),$$

then comparing the coefficients of X^2 on both sides would yield $\sum_{i=1}^s b_i^2 = 0$. Hence the coefficient of XY on the right hand side would be zero.

5. FORMULATION AS A SEMIDEFINITE PROGRAM

We fix again some $k \in \mathcal{N}$ (as defined in (7) on page 5). In this section, we show how to solve the problems (P_k) and (D_k) using semidefinite programming. This translation of (P_k) and (D_k) into semidefinite programs has become a quite common technique now (see, e.g., [Las, Mar, PPP]). Though this short section therefore does not contain new ideas, we include it in order to make this article a complete introduction to Lasserre's method. Also, other articles use quite different notations and a technically different setup. (For example, Lasserre [Las] considers the relaxations (P_k) and (D_k) only for even k . This is in our opinion not natural, at least if one of the g_i has odd degree.) Semidefinite programs are generalizations of linear programs, which can be solved efficiently (see for example [Tod]). Before we translate (P_k) and (D_k) into semidefinite programs, we will carry out three modifications, which are harmless. First, we will pass over from f to $f - f(0)$. This amounts just to adding the same constant to the goal functions of both problems. Second, we will negate both goal functions and at the same time exchange minimization and maximization. Third, we will exchange primal and dual. So we will forget about (P_k) and (D_k) , and instead look at the following pair of modified problems (P_k^{mod}) and (D_k^{mod}) .

$$(P_k^{\text{mod}}) \quad \text{minimize} \quad a \quad \text{subject to} \quad \begin{array}{l} a \in \mathbb{R} \text{ and} \\ (f - f(0)) + a \in M_k \end{array}$$

$$(D_k^{\text{mod}}) \quad \text{maximize} \quad -L(f - f(0)) \quad \text{subject to} \quad \begin{array}{l} L : \mathbb{R}[\bar{X}]_k \rightarrow \mathbb{R} \text{ is linear,} \\ L(1) = 1 \text{ and} \\ L(M_k) \subseteq [0, \infty) \end{array}$$

Define $d_0, \dots, d_m \in \mathbb{N}$ like in (8) and recall Notations 11 and 15. The key lemma is the following.

Lemma 24. *Suppose $L : \mathbb{R}[\bar{X}]_k \rightarrow \mathbb{R}$ is a linear map. Then $L(M_k) \subseteq [0, \infty)$ if and only if the $m + 1$ matrices*

$$(L(\bar{X}^{\beta+\gamma} g_i))_{(\beta,\gamma) \in \Lambda(d_i) \times \Lambda(d_i)} \quad (i \in \{0, \dots, m\})$$

are positive semidefinite. Moreover,

$$M_k = \left\{ \sum_{i=0}^m \langle (\bar{X}^{\beta+\gamma} g_i)_{(\beta,\gamma) \in \Lambda(d_i) \times \Lambda(d_i)}, G_i \rangle \mid G_0, \dots, G_m \in S\mathbb{R}_+^{\Lambda(d_i) \times \Lambda(d_i)} \right\}.$$

Proof. The first part follows from the observation that the $m + 1$ matrices under consideration represent the bilinear forms

$$\mathbb{R}[\bar{X}]_{d_i} \times \mathbb{R}[\bar{X}]_{d_i} \rightarrow \mathbb{R} : (p, q) \mapsto L(pqg_i) \quad (i \in \{0, \dots, m\})$$

with respect to the bases $(\bar{X}^\alpha)_{\alpha \in \Lambda(d_i)}$ of $\mathbb{R}[\bar{X}]_{d_i}$. The positive semidefiniteness of such a matrix expresses the positive semidefiniteness of the corresponding bilinear form, that is, $L(p^2g_i) \geq 0$ for all $p \in \mathbb{R}[\bar{X}]_{d_i}$. The second part follows easily from (6) and Lemma 17. \square

Using this Lemma, we can reformulate (P_k^{mod}) and (D_k^{mod}) using positive semidefinite matrices.

$$(P_k^{\text{I}}) \quad \begin{array}{ll} \min. & a \\ \text{s.t.} & a \in \mathbb{R}, G_0, \dots, G_m \in S\mathbb{R}_+^{\Lambda(d_i) \times \Lambda(d_i)} \text{ and} \\ & \sum_{i=0}^m \langle (\bar{X}^{\beta+\gamma}g_i)_{(\beta,\gamma) \in \Lambda(d_i) \times \Lambda(d_i)}, G_i \rangle = f - f(0) + a \end{array}$$

$$(D_k^{\text{I}}) \quad \begin{array}{ll} \max. & -L(f - f(0)) \\ \text{s.t.} & L : \mathbb{R}[\bar{X}]_k \rightarrow \mathbb{R} \text{ is linear, } L(1) = 1 \text{ and} \\ & (L(\bar{X}^{\beta+\gamma}g_i))_{(\beta,\gamma) \in \Lambda(d_i) \times \Lambda(d_i)} \text{ is positive semidefinite, } i = 0, \dots, m \end{array}$$

The matrix $(L(\bar{X}^{\beta+\gamma}))_{(\beta,\gamma) \in \Lambda(d_0, d_0)}$ appearing in (D_k^{I}) (recall that $g_0 = 1$) is often called a *moment matrix* of L . This is because, if L is integration with respect to a measure, then its entries are moments of this measure. Such matrices seem to be very interesting objects (see [CF, Lau] and the references therein). The same is true for the so called *localizing matrices* $(L(\bar{X}^{\beta+\gamma}g_i))_{(\beta,\gamma) \in \Lambda(d_i) \times \Lambda(d_i)}$, $i = 1, \dots, m$.

Now for $i \in \{0, \dots, m\}$ and $\alpha \in \Lambda(k)$, define a matrix $A_{\alpha i} \in S\mathbb{R}^{\Lambda(d_i) \times \Lambda(d_i)}$ by requiring

$$(23) \quad \bar{X}^{\beta+\gamma}g_i = \sum_{\alpha \in \Lambda(k)} A_{\alpha i}(\beta, \gamma) \bar{X}^\alpha,$$

that is, $A_{\alpha i}(\beta, \gamma)$ is the coefficient of \bar{X}^α in $\bar{X}^{\beta+\gamma}g_i$. Also, we define b_α to be the coefficient of \bar{X}^α in f for each $0 \neq \alpha \in \Lambda(k)$, i.e.,

$$(24) \quad f - f(0) = \sum_{\alpha \in \Lambda(k) \setminus \{0\}} b_\alpha \bar{X}^\alpha.$$

We plug (23) and (24) in (P_k^{I}) and get:

$$(P_k^{\text{II}}) \quad \begin{array}{ll} \text{minimize} & a \\ \text{subject to} & a \in \mathbb{R}, G_0, \dots, G_m \in S\mathbb{R}_+^{\Lambda(d_i) \times \Lambda(d_i)} \text{ and} \\ & \sum_{\alpha \in \Lambda(k)} \bar{X}^\alpha \sum_{i=0}^m \langle A_{\alpha i}, G_i \rangle = \sum_{\alpha \in \Lambda(k) \setminus \{0\}} b_\alpha \bar{X}^\alpha + a \end{array}$$

$$(D_k^{\text{II}}) \quad \begin{array}{ll} \text{maximize} & \sum_{\alpha \in \Lambda(k) \setminus \{0\}} -b_\alpha L(\bar{X}^\alpha) \\ \text{subject to} & L : \mathbb{R}[\bar{X}]_k \rightarrow \mathbb{R} \text{ is linear, } L(1) = 1 \text{ and} \\ & \sum_{\alpha \in \Lambda(k)} L(\bar{X}^\alpha) A_{\alpha i} \text{ is positive semidefinite, } i = 0, \dots, m \end{array}$$

Finally, we rewrite the polynomial equality constraint in (P_k^{II}) by splitting it up, coefficient by coefficient, and incorporating the equation for the constant coefficient into the goal function. Concerning the dual problem (D_k^{II}) , we exploit that a linear

map $L : \mathbb{R}[\bar{X}]_k \rightarrow \mathbb{R}$ with $L(1) = 1$ can be identified with its values $y_\alpha := -L(\bar{X}^\alpha)$, $0 \neq \alpha \in \Lambda(k)$.

$$\begin{aligned} (P_k^{\text{sdp}}) \quad & \min. \quad \sum_{i=0}^m \langle A_{0i}, G_i \rangle \\ & \text{s.t.} \quad G_0, \dots, G_m \in S\mathbb{R}_+^{\Lambda(d_i) \times \Lambda(d_i)} \quad \text{and} \\ & \quad \sum_{i=0}^m \langle A_{\alpha i}, G_i \rangle = b_\alpha, \quad 0 \neq \alpha \in \Lambda(k) \end{aligned}$$

$$\begin{aligned} (D_k^{\text{sdp}}) \quad & \max. \quad \sum_{\alpha \in \Lambda(k) \setminus \{0\}} b_\alpha y_\alpha \\ & \text{s.t.} \quad y_\alpha \in \mathbb{R}, \quad 0 \neq \alpha \in \Lambda(k), \quad \text{and} \\ & \quad A_{0i} - \sum_{\alpha \in \Lambda(k) \setminus \{0\}} y_\alpha A_{\alpha i} \text{ is positive semidefinite, } i = 0, \dots, m \end{aligned}$$

This is in fact a primal-dual pair of semidefinite programs with matrices in block diagonal structure (see [Tod]). For each $\alpha \in \Lambda(k)$, you could define A_α to be the matrix consisting of the $m+1$ diagonal blocks $A_{\alpha i}$, $i = 0, \dots, m$, and G to be a positive semidefinite matrix variable of the same size. Then (P_k^{sdp}) could be written as the minimization of $\langle A_0, G \rangle$ subject to $\langle A_\alpha, G \rangle = b_\alpha$ for $0 \neq \alpha \in \Lambda(k)$. Since semidefinite programming solvers usually take advantage of such a block diagonal structure, we stick to the above formulation.

REFERENCES

- [BW] R. Berr, T. Wörmann: Positive polynomials on compact sets, *Manuscr. Math.* **104**, No. 2, 135–143 (2001).
- [CF] R. E. Curto, L. Fialkow: The truncated complex K -moment problem, *Trans. Am. Math. Soc.* **352**, No. 6, 2825–2855 (2000)
- [CLR] M. D. Choi, T. Y. Lam, B. Reznick: Sums of squares of real polynomials, *Proc. Symp. Pure Math.* **58**, Part 2, 103–126 (1995)
- [HL] D. Henrion, J. B. Lasserre: GloptiPoly: Global Optimization over Polynomials with Matlab and SeDuMi, *ACM Trans. Math. Soft.* **29**, 165–194 (2003)
- [Jac] T. Jacobi: A representation theorem for certain partially ordered commutative rings, *Math. Z.* **237**, No. 2, 259–273 (2001)
- [JP] T. Jacobi, A. Prestel: Distinguished representations of strictly positive polynomials, *J. Reine Angew. Math.* **532**, 223–235 (2001)
- [Las] J. B. Lasserre: Global optimization with polynomials and the problem of moments, *SIAM J. Optim.* **11**, No. 3, 796–817 (2001)
- [Lau] M. Laurent: Revisiting two theorems of Curto and Fialkow on moment matrices, to appear in *Proc. Am. Math. Soc.*
- [Mar] M. Marshall: Optimization of polynomial functions, *Canad. Math. Bull.* **46**, 575–587 (2003)
- [PD] A. Prestel, C. Delzell: Positive polynomials, Springer Monographs in Mathematics, Berlin: Springer (2001)
- [Pól] G. Pólya: Über positive Darstellung von Polynomen, *Vierteljahresschrift der Naturforschenden Gesellschaft in Zürich* **73** (1928), 141–145, reprinted in: *Collected Papers*, Volume 2, 309–313, Cambridge: MIT Press (1974)
- [PPP] S. Prajna, A. Papachristodoulou, P. A. Parrilo, P. Seiler: SOSTOOLS: Sum of Squares Optimization Toolbox for MATLAB, to appear in *Positive Polynomials in Control*, Springer (2004)
- [PR] V. Powers, B. Reznick: A new bound for Pólya’s theorem with applications to polynomials positive on polyhedra, *J. Pure Appl. Algebra* **164**, No.1–2, 221–229 (2001)
- [PS] V. Powers, C. Scheiderer: The moment problem for non-compact semialgebraic sets, *Adv. Geom.* **1**, No. 1, 71–88 (2001)
- [Put] M. Putinar: Positive polynomials on compact semi-algebraic sets, *Indiana Univ. Math. J.* **42**, No. 3, 969–984 (1993)
- [Rud] W. Rudin: Real and complex analysis, McGraw-Hill Series in Higher Mathematics, New York: McGraw-Hill Book Company (1966)
- [S1] M. Schweighofer: An algorithmic approach to Schmüdgen’s Positivstellensatz, *J. Pure Appl. Algebra* **166**, No. 3, 307–319 (2002)

- [S2] M. Schweighofer: On the complexity of Schmüdgen's Positivstellensatz, *J. Complexity* **20**, 529–543 (2004)
- [Sch] K. Schmüdgen: The K -moment problem for compact semi-algebraic sets, *Math. Ann.* **289**, No. 2, 203–206 (1991)
- [Ste] G. Stengle: Complexity estimates for the Schmüdgen Positivstellensatz, *J. Complexity* **12**, No. 2, 167–174 (1996)
- [Tod] M. Todd: Semidefinite Optimization, *Acta Numerica* **10**, 515–560 (2001)

UNIVERSITÄT KONSTANZ, FACHBEREICH MATHEMATIK UND STATISTIK, 78457 KONSTANZ, ALLE-
MAGNE

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