Introduction to concepts and advances in polynomial optimization

Martin Mevissen

September 2007

Institute for Operations Research

ETH Zurich

Contents

1 Introduction

Polynomial optimization and the problem of global nonnegativity of polynomials are active fields of research and remain in the focus of researchers from various areas as real algebra, semidefinite programming and operator theory. Shor [25] was the first who introduced the idea of applying a convex optimization technique to minimize an unconstrained multivariate polynomial. Also, Nesterov [17] was one of the first who discussed to exploit the duality of moment cones and cones of nonnegative polynomials in a convex optimization framework. He showed the characterization of a moment cone by linear matrix inequalities, i.e., semidefinite constraints, in case the elements of the corresponding cone of nonnegative polynomials can be written as sum of squares. The next milestone in minimizing multivariate polynomials was given by Lasserre [11], who realized to apply recent real algebraic results by Putinar [22] to construct a sequence of semidefinite program relaxations whose optima converge to the optimum of a polynomial optimization problem. Another approach to apply real algebraic results to attempt the problem of nonnegativity of polynomials was introduced by Parrilo [18]. Recent developments in polynomial optimization are for

instance the exploitation of sparsity in order to achieve strong numerical improvements by proposing a sequence of sparse SDP relaxations by Waki, Kim, Kojima and Muramatsu [27], and other approaches to characterize the polynomial optimization problem by semidefinite programs via finite varieties by Laurent [14].

We attempt to solve the following **polynomial optimization problem**:

$$\begin{aligned} \min \quad & p(x) \\ \text{s.t.} \quad & g_i(x) \geq 0 \quad \forall i = 1, \ldots, m \end{aligned} \tag{1.1}$$

where $p, g_1, \ldots, g_m \in \mathbb{R}[x]$. Problem (1.1) can also be written as

$$\min_{x \in K} p(x) \tag{1.2}$$

where K the closed semialgebraic set that is defined by the polynomials $g_1, \ldots, g_m$. Let $p^\star$ denote the optimal value of problem (1.2) and $K^\star := \{x^\star \in K \mid \forall x \in K : \ p(x^\star) \leq p(x)\}$. Since K compact, $K^\star \neq \emptyset$, if $K \neq \emptyset$.

The structure of this paper is outlined as follows: After introducing some notations in chapter 2, will discuss the most important concepts for characterizations of globally nonnegative polynomials and polynomials nonnegative or positive on closed semialgebraic sets in chapter 3 at first. In particular we will present the connections of sum of squares characterizations by Parrilo and the Positivstellensätze by Schmüdgen and Putinar. In chapter 4 we will discuss the problem of moments which is closely interlinked to the problem of nonnegativity of polynomials and some applications of the related concepts by Bertsimas and Popescu. In chapter 5 we will present Lasserre's approach to solve the polynomial optimization problem (1.1) by constructing a sequence of convergent semidefinite programming relaxations. Also, we discuss further topics as the problem of extracting the global minimizers of (1.1) and an approach to exploit sparsity in polynomial optimization problems by Waki, Kim, Kojima and Muramatsu. Finally, we introduce an alternative approach by Laurent to attempt problem (1.1) in case its complex variety is finite.

2   Notations

We briefly introduce notions and objects from real algebra that will be used in the subsequent sections. A monograph which provides a detailed insight in these objects, the real algebraic concepts and their relations is given in the book 'Positive Polynomials' [21] by Prestel and Delzell.

Let $\mathbb{R}[x]$ denote the ring of polynomials with coefficients in $\mathbb{R}$ where $x \in \mathbb{R}^n$, and $\sum \mathbb{R}[x]^2$ the convex cone of elements in $\mathbb{R}[x]$ that can be written as **sums of squares of polynomials** in $\mathbb{R}[x]$,

$$\sum \mathbb{R}[x]^2 = \left\{ p \in \mathbb{R}[x]; \ p = \sum_{i=1}^r p_i^2, \ p_i \in \mathbb{R}[x] \ \text{for some } r \in \mathbb{N} \right\}.$$

A **closed semialgebraic set** is denoted by

$$K = \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\},$$

where $g_1, \ldots, g_m \in \mathbb{R}[x]$.

If K the semialgebraic set defined by $g_1, \ldots, g_m \in \mathbb{R}[x]$, let $M(K)$ be the **quadratic module** generated by $g_1, \ldots, g_m$, i.e.

$$M(K) = \sum \mathbb{R}[x]^2 + g_1 \sum \mathbb{R}[x]^2 + \ldots + g_m \sum \mathbb{R}[x]^2.$$

Let $\Sigma^2 \langle g_1, \ldots, g_m \rangle$

$$\begin{aligned} &:= \sum \mathbb{R}[x]^2 + g_1 \sum \mathbb{R}[x]^2 + \ldots + g_m \sum \mathbb{R}[x]^2 \quad + g_1 g_2 \sum \mathbb{R}[x]^2 + \ldots + g_1 g_2 \cdots g_m \sum \mathbb{R}[x]^2 \\ &:= M(K) \hspace{11.5em} + g_1 g_2 \sum \mathbb{R}[x]^2 + \ldots + g_1 g_2 \cdots g_m \sum \mathbb{R}[x]^2 \end{aligned}$$

be the **multiplicative convex cone** generated by $\sum \mathbb{R}[x]^2$ and $g_1, \ldots, g_m$.

Let $O(g_1, \ldots, g_m)$ denote the **multiplicative monoid** generated by $g_1, \ldots, g_m$, i.e. the set of finite products of the elements $g_1, \ldots, g_m$:

$$O(g_1, \ldots, g_m) = \{\prod_{i=1}^{r} t_i \mid t_i \in \{g_1, \ldots, g_m\} \text{ for } i \in \{1, \ldots, m\} \text{ and } r \in \mathbb{N}\}.$$

Let $I(g_1, \ldots, g_m)$ denote the **ideal** generated by $g_1, \ldots, g_m$, i.e. the set

$$\begin{aligned} I(g_1, \ldots, g_m) \quad &:= \langle g_1, \ldots, g_m \rangle_{\mathbb{R}[x]} \quad = \{\textstyle\sum_{i=1}^{m} f_i g_i, \ f_i \in \mathbb{R}[x]\} \\ &= \mathbb{R}[x] + g_1 \mathbb{R}[x] + \ldots + g_m \mathbb{R}[x]. \end{aligned}$$

In the language of linear algebra $I(g_1, \ldots, g_m)$ can be understood as a $\mathbb{R}[x]-$submodule of the $\mathbb{R}[x]$-module $\mathbb{R}[x]$. Recall, a $R-$module is a generalization of the notion of a $R-$vector space, with R being a ring with 1 instead of R being a field.

$\mathbb{R}[x]_\omega \subseteq \mathbb{R}[x]$ denotes the set of all real polynomials of degree less than $\omega \in \mathbb{N}$ and

$$\Lambda(\omega) := \{\alpha \in \mathbb{N}^n : \mid \alpha \mid \leq \omega\},$$

the set of all multivariate indices of degree less than or equal to $\omega \in \mathbb{N}$.

Given a set $\mathcal{A} \subseteq \mathbb{N}^n$ of multivariate indices, the **monomial vector** $u(x, \mathcal{A})$ is defined as

$$u(x, \mathcal{A}) = (x^\alpha \mid \alpha \in \mathcal{A}),$$

in fact its elements form a basis for $\mathbb{R}[x, \mathcal{A}] := \{p \in \mathbb{R}[x] \mid \text{supp}(p) \subseteq \mathcal{A}\}$. The length of the vector $u(x, \Lambda(\omega))$ of all monomials of degree less or equal than ω is denoted as

$$s(\omega) = \begin{pmatrix} n + \omega \\ \omega \end{pmatrix},$$

also, it equals the dimension of $\mathbb{R}[x]_\omega$. $u(x)$ denotes the (infinite dimensional) basis of $\mathbb{R}[x]$. Obviously holds $\mathbb{R}[x]_\omega = \mathbb{R}[x, \Lambda(\omega)]$.

3 Positive Polynomials

3.1 Decomposition of globally nonnegative polynomials

The origin of research in characterizing nonnegative and positive polynomials lies in Hilbert's 17th problem, whether it is possible to express a nonnegative rational function as sum of squares of rational functions. This question was answered positively by Artin in 1927. Furthermore the question arises, whether it is possible to express any nonnegative polynomial as sum of squares of polynomials. In case of univariate polynomials the answer to this question is yes, as stated in the following theorem.

Theorem 3.1 *Let $p \in \mathbb{R}[x]$, $x \in \mathbb{R}$. Then, $p(x) \geq 0$ for all $x \in \mathbb{R}$ if and only if $p \in \sum \mathbb{R}[x]^2$.*

Proof " $\Leftarrow$ " : Trivial.

" $\Rightarrow$ " : Let $p(x) \geq 0$ for all $x \in \mathbb{R}$. It is obvious that $\deg(p) = 2k$ for some $k \in \mathbb{N}$. Then, the real roots of of $p(x)$ should have even multiplicity, otherwise $p(x)$ would alter its sign in a neighborhood of a root. Let λ_i, $i = 1, \ldots, r$ be its real roots with corresponding multiplicity $2m_i$. Its complex roots can be arranged in conjugate pairs, $a_j + Ib_j$, $a_j - Ib_j$, $j = 1, \ldots, h$. Then,

$$p(x) = C \prod_{i=1}^{r} (x - \lambda_i)^{2m_i} \prod_{j=1}^{h} ((x - a_j)^2 + b_j^2).$$

Note that the leading coefficient C needs to be positive. Thus, by expanding the terms in the products, we see that $p(x)$ can be written as a sum of squares of polynomials, of the form

$$p(x) = \sum_{i=0}^{k} \left(\sum_{j=0}^{k} v_{ij} x^j \right)^2. \qquad \square$$

Nevertheless, Hilbert himself noted already that not every nonnegative polynomial can be written as sum of squares. For instance the Motzkin form M,

$$M(x, y, z) = x^4 y^2 + x^2 y^4 + z^6 - 3x^2 y^2 z^2$$

is nonnegative but not sum of squares. In fact Hilbert gave a complete characterization of the cases where nonnegativity and the existence of a sum of squares decomposition are equivalent.

Definition 3.2 *A **form** is a polynomial where all the monomials have the same total degree m. $P_{n,m}$ denotes the set of nonnegative forms of degree m in n variables, $\Sigma_{n,m}$ the set of forms p such that $p = \Sigma_k h_k^2$, where h_k are forms of degree $\frac{m}{2}$.*

There is a correspondance between forms in n with power m and polynomials in $n-1$ variables with degree less or equal to m. In fact, a form in n variables of degree m can be dehomogenized to a polynomial in $n-1$ variables by fixing any of the n variables to the constant value 1. Conversely, given a polynomial in $n-1$ variables in can be homogenized by multiplying each monomial by powers of a new variable such that the degree of all monomials equals m. Obviously, $\Sigma_{n,m} \subseteq P_{n,m}$ holds for all n and m. The following Theorem is due to Hilbert.

Theorem 3.3 $\Sigma_{n,m} \subseteq P_{n,m}$ *holds with equality only in the following cases:*

(i) *Bivariate forms: $n = 2$,*

(ii) *Quadratic forms: $m = 2$,*

(iii) *Ternary quartic forms: $n = 3$, $m = 4$.*

We interpret the three cases in Theorem 3.3 in terms of polynomials. The first one corresponds to the equivalence of nonnegativity and sum of squares condition in the univariate case as in Theorem (3.1). The second one is the case of quadratic polynomials, where the sum of squares decomposition follows from an eigenvalue/eigenvector factorization. The third case corresponds to quartic polynomials in two variables.

Relevance of sum of squares characterizations Recall that the constraints of our original polynomial optimization problem are nonnegativity constraints for polynomials of the type $g_i(x) \geq 0 \quad (i = 1, \ldots, m)$. The question, whether a given polynomial is globally nonnegative is decidable, for instance by the Tarski-Seidenberg decision procedure [2]. Nonetheless, regarding complexity, the general problem of testing global nonnegativity of a polynomial function is NP-hard [16], if the degree of the polynomial is at least four. Therefore it is reasonable to substitute the nonnegativity constraints by expressions that can be decided easier. It was shown by Parrilo that the decision whether a polynomial is sum of squares is equivalent to a semidefinite program as stated in the following theorem.

Theorem 3.4 *The existence of a sum of squares decomposition of a polynomial in n variables of degree 2d can be decided by solving a semidefinite programming feasibility problem [18]. If the polynomial is dense, the dimensions of the matrix inequality are equal to*
$$\begin{pmatrix} n+d \\ d \end{pmatrix} \times \begin{pmatrix} n+d \\ d \end{pmatrix}.$$

Proof Let $p \in \mathbb{R}[x]$ with degree $2d$. Recall $u(x, \Lambda(d))$ denotes the ordered vector of monomials $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$ with $\sum_{i=1}^{n} \alpha_i \leq d$. The length of $u(x, \Lambda(d))$ is $s := s(d) = \begin{pmatrix} n+d \\ d \end{pmatrix}$.

Claim: $p \in \sum \mathbb{R}[x]^2$ if and only if $\exists V \in \mathbb{S}_+^s$ such that $p = u(x, \Lambda(d))^T V u(x, \Lambda(d))$.

Pf: $\Rightarrow$: $p \in \sum \mathbb{R}[x]^2$, i.e.

$$p = \sum_{i=1}^{r} q_i^2 = \sum_{i=1}^{r} (w_i^T u(x, \Lambda(d)))^2 = u(x, \Lambda(d))^T \left(\sum_{i=1}^{r} w_i w_i^T \right) u(x, \Lambda(d)).$$

Thus, $V = \sum_{i=1}^{r} w_i w_i^T$ and $V \in \mathbb{S}_+^s$.

$\Leftarrow$: As $V \in \mathbb{S}_+^s$ there exists a Cholesky factorization $V = WW^T$, where $W \in \mathbb{R}^{s \times s}$ and let w_i denote the ith column of W. We have

$$p = u(x, \Lambda(d))^T V u(x, \Lambda(d)) = \sum_{i=1}^{s} w_i w_i^T u(x, \Lambda(d)) = \sum_{i=1}^{s} (w_i^T u(x, \Lambda(d)))^2,$$

i.e., $p \in \mathbb{R}[x]$. Thus, the claim follows.

Expanding the quadratic form gives $p = \sum_{i,j=1}^{s} V_{i,j} u(x, \Lambda(d))_i u(x, \Lambda(d))_j$. Equating the coefficients in this expression with the coefficients of the corresponding monomials in the original form for p generates a set of linear equalities for the variables $V_{i,j}$ $(i, j = 1, \ldots, s)$. Adding the constraint $V \in \mathbb{S}_+^s$ to those linear equality constraints, we obtain conditions for p which are equivalent to claiming $p \in \sum \mathbb{R}[x]^2$. Therefore, the decidability problem whether $p \in \sum \mathbb{R}[x]^2$ is equivalent to a semidefinite programming feasibility problem. $\square$

3.2 Decomposition of polynomials positive on closed semialgebraic sets

Real algebraic geometry deals with the analysis of the real solution set of a system of polynomial equations. The main difference to algebraic geometry in the complex case lies in the fact that $\mathbb{R}$ is not algebraically closed. One of the main results of real algebra are the **Positivstellensätze** which provide certificates in the case a semialgebraic set is empty. Improved versions of the Positivstellensätze can be obtained in case of compact semialgebraic sets.

3.2.1 General semialgebraic sets

The **Positivstellensatz** below is due to Stengle; a proof can be found in [2].

Theorem 3.5 (Stengle) *Let* $(f_j)_{j=1,\ldots,t}$, $(g_k)_{k=1,\ldots,m}$, $(h_l)_{l=1,\ldots,k}$ *be finite families of polynomials in* $\mathbb{R}[x]$. *The following properties are equivalent:*

(i)
$$
\left\{ x \in \mathbb{R}^n \ \middle| \ \begin{array}{ll} g_j(x) \geq 0, & j = 1, \ldots, m \\ f_s(x) \neq 0, & s = 1, \ldots, t \\ h_i(x) = 0, & i = 1, \ldots, k \end{array} \right\} = \emptyset.
$$

(ii) *There exist* $g \in \Sigma^2 \langle g_1, \ldots, g_m \rangle$, $f \in O(f_1, \ldots, f_t)$, *the multiplicative monoid generated by* $f_1, \ldots, f_t$, $h \in I(h_1, \ldots, h_k)$, *the ideal generated by* $h_1, \ldots, h_k$, *such that* $g + f^2 + h = 0$.

To understand the differences between the real and the complex case, and the use of the Positivstellensatz 3.5 consider the following example.

Example 3.6 *Consider the very simple standard quadratic equation*

$$ x^2 + ax + b = 0. $$

By the fundamental theorem of algebra, the equation has always solutions on $\mathbb{C}$. *For the case when the solution is required to be real, the solution set will be empty if and only if the discriminant* D *satisfies*

$$ D := b - \frac{a^2}{4} > 0. $$

In this case taking

$$ g := \left(\frac{1}{\sqrt{D}} (x + \frac{a}{2}) \right)^2, \quad f := 1, \quad h := -\frac{1}{D}(x^2 + ax + b), $$

the identity $g + f^2 + h = 0$ *is satisfied.*

It is to remark, the Positivstellensatz represents the most general deductive system for which inferences from the given equations can be made. It guarantees the existence of **infeasibility certificates** given by the polynomials f, g and h. For complexity reasons these certificates cannot be polynomial time checkable for every possible instance, unless NP=co-NP. Parrilo showed that it is possible that the problem of finding infeasibility certificates is equivalent to an semidefinite program, if the degree of the possible multipliers is restricted [18].

Theorem 3.7 *Consider a system of polynomial equalities and inequalities as in Theorem 3.5. Then, the search for bounded degree Positivstellensatz infeasibility certificates can be done using semidefinite programming. If the degree bound is chosen to be large enough, then the SDPs will be feasible, and the certificates are obtained from its solution.*

Proof: Consequence of the Positivstellensatz and Theorem 3.4, c.f. [18].

As the feasible set of initial problem (1.2) is a closed semialgebraic set, we are interested in characterizations for these sets and polynomials positive on semialgebraic sets. The Positivstellensatz allows to deduce conditions for the positivity or the nonnegativity of a polynomial over a semialgebraic set. A direct consequence of the Positivstellensatz is the following corollary [2], pp. 92.

Corollary 3.8 *Let* $g_1, \ldots, g_m \in \mathbb{R}[x]$,
$K = \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$ *and* $f \in \Sigma^2 \langle g_1, \ldots, g_m \rangle$. *Then:*

(i) $\forall x \in K \ f(x) \geq 0 \ \Leftrightarrow \ \exists s \in \mathbb{N} \ \exists g, h \in \Sigma^2 \langle g_1, \ldots, g_m \rangle \ s.t. \ fg = f^{2s} + h.$

(ii) $\forall x \in K \ f(x) > 0 \ \Leftrightarrow \ \exists g, h \in \Sigma^2 \langle g_1, \ldots, g_m \rangle \ s.t. \ fg = 1 + h.$

Proof

(i) Apply the Positivstellensatz to the set

$$\{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0, -f(x) \geq 0, f(x) \neq 0\}.$$

(ii) Apply the Positivstellensatz to the set

$$\{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0, -f(x) \geq 0\}. \qquad \square$$

Those conditions for the nonnegativity and positivity of polynomials on semialgebraic sets can be improved under additional assumptions. We present these improved conditions for **compact** semi-algebraic sets in the following section.

3.2.2 Compact semialgebraic sets

It is our aim to characterize polynomials that are positive or nonnegative on compact semialgebraic sets. A first characterization is a theorem due to Schmüdgen [23]:

Theorem 3.9 (Schmüdgen) *Let* $K = \{x \in \mathbb{R}^n; g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$ *be a compact semialgebraic subset of* $\mathbb{R}^n$ *and let* p *be a positive polynomial on* K. *Then* $p \in \Sigma^2 \langle g_1, \ldots, g_m \rangle$.

It was Putinar [22] who simplified this characterization under an additional assumption.

Definition 3.10 *A quadratic module* $M(K)$ *is called* ***archimedean*** *if* $N - \sum_{i=1}^n x_i^2 \in M(K)$ *for some* $N \in \mathbb{N}$.

Theorem 3.11 (Putinar) *Let* p *be a polynomial, positive on the compact semialgebraic set* K *and* $M(K)$ *archimedian, then* $p \in M(K)$.

Thus, under the additional assumption of an archimedian quadratic module $M(K)$, we obtain the stricter characterization $p \in M(K) \subseteq \Sigma^2 \langle g_1, \ldots, g_m \rangle$ instead of $p \in \Sigma^2 \langle g_1, \ldots, g_m \rangle$. A further theorem by Schmüdgen [23] provides equivalent conditions for $M(K)$ being archimedian.

Theorem 3.12 *The following are equivalent:*

(i) *There exist finitely many* $t_1, \ldots, t_s \in M(K)$ *such that the set*

$$\{x \in \mathbb{R}^n \mid t_1(x) \geq 0, \ldots, t_s(x) \geq 0\}$$

(which contains K*) is compact and* $\prod_{i \in I} t_i \in M(K)$ *for all* $I \subset \{1, \ldots, s\}$.

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

(iii) *There exists an* $N \in \mathbb{N}$ *such that* $N - \sum_{i=1}^n x_i^2 \in M(K)$, *i.e.,* $M(K)$ *is archimedian.*

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

Thus, for any polynomial p positive on K, $p \in M(K)$ holds, if one of the conditions in Theorem 3.12 is satisfied. Whether it is decidable that one of the equivalent conditions hold is not known and subject of current research.

Example 3.13 *Consider the compact semialgebraic set*

$$K = \left\{ x \in \mathbb{R}^2 \mid g(x) = 1 - x_1^2 - x_2^2 \geq 0 \right\}.$$

The quadratic module $M(K)$ is archimedian, as $1 - x_1^2 - x_2^2 = 0^2 + 1^2 \cdot g(x) \in M(K)$. The polynomials $f_1(x) := x_1 + 2$ and $x_1^3 + 2$ are positive on K. Thus $f_1, f_2 \in M(K)$ with Theorem 3.11. Their decomposition can be derived as

$$\begin{aligned}
f_1(x) &= x_1 + 2 &= \tfrac{1}{2}(x_1 + 1)^2 + \tfrac{1}{2}x_2^2 + 1 + \tfrac{1}{2}(1 - x_1^2 - x_2^2), \\
f_2(x) &= 2x_1^3 + 3 &= (x_1^3 + 1)^2 + (x_1^2 x_2)^2 + (x_1 x_2)^2 + x_2^2 + 1 + (x_1^4 + x_1^2 + 1)(1 - x_1^2 - x_2^2).
\end{aligned}$$

The next example demonstrates that in general not every polynomial nonnegative on a compact semialgebraic set K is contained in $M(K)$ even if $M(K)$ is archimedian.

Example 3.14 *Consider the compact semialgebraic set*

$$K = \left\{ x \in \mathbb{R} \mid g_1(x) := x^2 \geq 0, \ g_2(x) := -x^2 \geq 0 \right\}.$$

It is obvious that $M(K)$ is archimedian. Also, it is easy to see that there are no $q, r, s \in \sum \mathbb{R}[x]^2$ such that

$$p(x) := x = q(x) + r(x)\,x^2 + s(x)\,(-x^2),$$

although p is nonnegative on K. Nevertheless, the polynomial $p_a \in \mathbb{R}[x]$ defined by $p_a(x) = x + a$ for $a > 0$ can be decomposed as

$$p_a(x) = x + a = \frac{1}{4a}(x + 2a)^2 - \frac{1}{4a}x^2.$$

Thus $p_a \in M(K)$ for all $a > 0$.

The original proof of Theorem 3.11 is due to Putinar [22]. In this proof Putinar applies the separation theorem for convex sets and some arguments from functional analysis. A new proof was found by Schweighofer [24] that avoids the arguments from functional analysis and requires only results from elementary analysis. As Theorem 3.11 is of central importance in Lasserre's approach to solve polynomial optimization problems we will outline Schweighofer's proof of this theorem here. Before we will state Polya's theorem and a lemma that are both applied in his proof.

Theorem 3.15 (Polya) *Suppose $F \in \mathbb{R}[x]$ is homogeneous and satisfies $F > 0$ on $[0, \infty)^n \setminus \{0\}$. Then for all k big enough, the polynomial $(x_1 + \ldots + x_n)^k F$ has only nonnegative coeffcients.*

Proof C.f. [20].

Lemma 3.16 *Suppose $C \subseteq \mathbb{R}^n$ is compact and $g_i \leq 1$ on C for all $i \in \{1, \ldots, m\}$. Suppose $p \in \mathbb{R}[x]$ satisfies $p > 0$ on S. Then there exists $s \in \mathbb{N}$ such that for all sufficiently large $k \in \mathbb{N}$,*

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

Proof C.f. [24].

Proof of Theorem 3.11 Assume $N \in \mathbb{N}$ such that $N - \sum_{i=1}^{n} x_i^2 \in M(K)$. Consider the compact set

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

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

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

Since $l(\Delta)$ is compact, we can scale each g_i with a positive factor such that $g_i \leq 1$ on C. So we an apply Lemma 3.16 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 is sufficient to show that $q \in M(K)$, and we shall even show that

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

In order to show that, write $q = \sum_{i=0}^{d} Q_i$ where $d := \deg q$ and $Q_i \in \mathbb{R}[x]$ is homogeneous of degree i, $i = 0, \ldots, d$. Define

$$F := \sum_{i=0}^{d} Q_i \left(\frac{y_1 - y_{n+1}}{2}, \ldots, \frac{y_n - y_{2n}}{2} \right) \left(\frac{y_1 + \ldots + y_{2n}}{2n(N + \frac{1}{4})} \right)^{d-i} \in \mathbb{R}[y]$$

where $y \in \mathbb{R}^{2n}$. For each $y \in \Delta$, we obtain

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

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 $F > 0$ on $[0, \infty)^{2n} \setminus \{0\}$. By Polyas's Theorem 3.15, there is some $e \in \mathbb{N}$ such that

$$G := \left(\frac{y_1 + \ldots + y_{2n}}{2n(N + \frac{1}{4})} \right)^e F \in \mathbb{R}[y]$$

has only nonnegative coefficients. Now we apply on this polynomial the $\mathbb{R}-$algebra homomorphism $\phi : \mathbb{R}[y] \to \mathbb{R}[x]$ defined by

$$y_i \mapsto (N + \frac{1}{4}) + x_i, \quad y_{n+i} \mapsto (N + \frac{1}{4}) - x_i \quad (i \in \{1, \ldots, n\}).$$

Note that $\phi(y_i) \in T$ for each $i \in \{1, \ldots, 2n\}$ 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

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

is contained in T. $\qquad$ $\square$

Remark 3.17 *Obviously it holds, given a compact semialgebraic set K, any positive polynomial on K belongs to the cone $M(K)$ if and only if $M(K)$ is archimedian.*

Theorem 3.11 is called **Putinar's Positivstellensatz**. Obviously, it does not really characterize the polynomials positive on K since the polynomials in $M(K)$ must only be nonnegative on K. Also, it does not fully describe the polynomials nonnegative on K since they are not always contained in $M(K)$. But it is Theorem 3.11 that is exploited by Lasserre in order to attempt the polynomial optimization problem, as we will see in chapter 5.

4 Moment problems

A field closely interlinked with the theory of nonnegative polynomials and sum of squares decompositions is the problem of moments. As we will see, many questions in the field of moment problems can be answered by applying convex and in particular semidefinite optimization methods. Therefore particular moment problems are tractable and can be solved in appropriate time. In particular, we are able to exploit the real algebraic results from the previous chapter in case of K-moment problems, if K a closed or compact semialgebraic set.

4.1 The general moment problem

The **moment problem** is the following:
Given a sequence $y = (y_\alpha)_{\alpha \in \mathbb{N}^n}$, does there exist a Borel measure $\mu \in \mathcal{M}(\mathbb{R}^n)$ such that y_α is the α-th moment of μ, i.e. $y_\alpha = \int x^\alpha \, d\mu$, for all $\alpha \in \mathbb{N}_0^n$?
The key connection between moment problems and semidefinite optimization lays in the notion of a **feasible moment vector**.

Definition 4.1 $y = (y_\alpha)_{\alpha \leq \omega}$ *is a **feasible (n, ω, Ω)-moment vector**, if there is a Borel measure $\mu \in \mathcal{M}(\Omega)$ whose moments are given by y, that is $y_\alpha = \int x^\alpha \, d\mu$ for all $\alpha \leq \omega$.*

A notion that is useful in order to derive conditions for a moment vector to be feasible is is given in the next definition.

Definition 4.2 *Let $y = (y_\alpha)_{\alpha \leq k} = (y_{\alpha_1}, \ldots, y_{\alpha_t})$ be a vector. The **Hankel matrix** $H(y)$ is defined by*

$$H(y) = \begin{pmatrix} y_{\alpha_1} & y_{\alpha_2} & \cdots & y_{\alpha_{t-1}} & y_{\alpha_t} \\ y_{\alpha_2} & y_{\alpha_3} & \cdots & y_{\alpha_t} & y_{\alpha_{t+1}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_{\alpha_{t-1}} & y_{\alpha_t} & \cdots & y_{\alpha_{2(t-1)}} & y_{\alpha_{2t-1}} \\ y_{\alpha_t} & y_{\alpha_{t+1}} & \cdots & y_{\alpha_{2t-1}} & y_{\alpha_{2t}} \end{pmatrix}.$$

The problem whether a given y is a feasible moment vector has been completely characterized in the **univariate case** $(n = 1)$.

Theorem 4.3 *(a) The vector $(y_1, \ldots, y_{2n+1})$ is a feasible $(1, 2n+1, \mathbb{R}^+)$-moment vector if and only if the following matrices are positive semidefinite:*

$$H\left((1, y_1, \ldots, y_{2n})\right) = \begin{pmatrix} 1 & y_1 & \cdots & y_n \\ y_1 & y_2 & \cdots & y_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & y_{n+1} & \cdots & y_{2n} \end{pmatrix} \succcurlyeq 0,$$

$$H\left((y_1, \ldots, y_{2n+1})\right) = \begin{pmatrix} y_1 & y_2 & \cdots & y_{n+1} \\ y_2 & y_3 & \cdots & y_{n+2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n+1} & y_{n+2} & \cdots & y_{2n+1} \end{pmatrix} \succcurlyeq 0.$$

(b) The vector $(y_1, \ldots, y_{2n})$ is a feasible $(1, 2n, \mathbb{R})$-moment vector if and only if

$$H\left((1, y_1, \ldots, y_{2n})\right) \succcurlyeq 0.$$

Proof C.f. [3], pp. 65.

In the **multivariate case**, there are known necessary conditions for a vector y to be a feasible $(n, \omega, \mathbb{R}^n)$-vector that also involve the semidefiniteness conditions for Hankel matrices derived from y. However, these conditions are not sufficient. The complexity whether a vector y is a feasible $(n, \omega, \mathbb{R}^n)$-moment vector has not been resolved, yet. A typical necessary condition is given by the following theorem.

Theorem 4.4 *Let σ be a feasible $(2, 2k, \mathbb{R}^2)$-moment vector. Then $H(\sigma) \succcurlyeq 0$ hold for the Hankel matrix H of σ.*

Proof: C.f. [3], pp. 204. Let $p : \mathbb{R}^2 \to \mathbb{R}$ be a polynomial of degree k with coefficients $c_{i,j}$,

$$p(x) = \sum_{i=0}^{k} \sum_{j=0}^{k-i} c_{i,j} x_1^i x_2^j,$$

and let μ be a Borel measure on $\mathbb{R}^2$ with moments $\sigma_{i,j}$. Then, $E_\mu\left[p(x)^2\right]$ can be expressed as a quadratic form in c,

$$0 \le E\left[p(x)^2\right] = c^T H(\sigma) c,$$

where H is the Hankel matrix of σ. Thus, $H(\sigma) \succcurlyeq 0$. $\quad\square$
Nevertheless, characterizations are known in some special cases.

Theorem 4.5 *A vector $y = (M, \Gamma)$ is a feasible $(n, 2, \mathbb{R}^n)$-vector if and only if the following matrix is positive semidefinite:*

$$S = \begin{pmatrix} 1 & M^T \\ M & \Gamma \end{pmatrix} \succcurlyeq 0.$$

Proof " $\subseteq$ " : Suppose (M, Γ) is a feasible $(n, 2, \mathbb{R}^n)$-moment vector. Then, there exists a Borel measure $\mu \in \mathcal{M}(\mathbb{R}^n)$ such that $M = E_\mu[x]$ and $\Gamma = E_\mu[xx^T]$. The matrix $(x - M)(x - M)^T$ is positive semidefinite. Taking expections with respect to μ, we obtain that

$$E\left[(x - M)(x - M)^T\right] = \Gamma - MM^T \succcurlyeq 0,$$

which expresses the fact that the covariance matrix needs to be positive semdefinite. Obviously, $\Gamma - MM^T \succcurlyeq 0$ if and only if $S \succcurlyeq 0$.

" $\supseteq$ " : If $S \succcurlyeq 0$, then $\Gamma - MM^T \succcurlyeq 0$. Let μ be an Borel measure on $\mathcal{M}(\mathbb{R}^n)$ with $M = E_\mu[x]$ and covariance matrix $\Gamma - MM^T$. Therefore $y := (M, \Gamma)$ is a feasible $(n, 2, \mathbb{R}^n)$−moment vector.□

For various applications of moment problems and their semidefinite characterizations or approximations see chapter 16 in [28].

4.2 The K-moment problem

The **K-moment problem** is given by:

Let K be a closed semialgebraic set and $y = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ be a sequence, does there exist a Borel measure $\mu \in \mathcal{M}(K)$ supported on K such that y_α is the α-th moment of μ, i.e. $y_\alpha = \int x^\alpha \, d\mu$, for all $\alpha \in \mathbb{N}_0^n$?

A convenient notion in order to characterize this problem is given by the next definition.

Definition 4.6 *Let $y = (y_{\alpha_1}, y_{\alpha_2}, \ldots, y_{\alpha_{s(2r)}})$ with $y_{\alpha_1} = 1$ be a sequence of length $s(2r) = \begin{pmatrix} n + 2r \\ 2r \end{pmatrix}$. The **moment matrix** $M_r(y)$ is constructed as follows. $M_r(y)(i, 1) = y_{\alpha_i}$ and $M_r(y)(1, j) = y_{\alpha_j}$ for $i, j \in \{1, \ldots, s(r)\}$, and $M_r(y)(i, j) = y_{\alpha_i + \alpha_j}$ for $i, j \in \{2, \ldots, s(r)\}$. For instance, if $n = 2$ and $r = 2$, $M_2(y)$ is given by*

$$M_2(y) = \begin{pmatrix} 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{pmatrix}.$$

Moreover $M_r(y)$ defines a bilinear form $\langle ., . \rangle_y$ on $\mathbb{R}[x]_r$ by

$$\langle q(x), v(x) \rangle_y := \langle q, M_r(y)v \rangle, \ q(x), v(x) \in \mathbb{R}[x]_r,$$

and if y is a sequence of moments of some measure μ_y, then

$$\langle q, M_r(y)q \rangle = \int q(x)^2 \, d\mu_y \geq 0,$$

so that $M_r(y) \succcurlyeq 0$.

*Let $p(x) \in \mathbb{R}[x]$ with coeffcient vector $(p_\beta)_{\beta \in \mathbb{N}_0^n}$, we define the **localizing matrix** $M_r(py)$ by*

$$M_r(py)(i, j) = \sum_\beta p_\beta y_{\alpha_i + \alpha_j + \beta}.$$

For instance, with

$$M_1(y) = \begin{pmatrix} 1 & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{pmatrix} \quad \text{and } p(x) = a - x_1^2 - x_2^2,$$

we obtain

$$M_1(py) = \begin{pmatrix} a - y_{20} - y_{02} & ay_{10} - y_{30} - y_{12} & ay_{01} - y_{21} - y_{03} \\ ay_{10} - y_{30} - y_{12} & ay_{20} - y_{40} - y_{22} & ay_{11} - y_{31} - y_{13} \\ ay_{01} - y_{21} - y_{03} & ay_{11} - y_{31} - y_{13} & ay_{02} - y_{22} - y_{04} \end{pmatrix}.$$

Under the condition that μ_y is a Borel measure with moment sequence y, which is supported on $\{p(x) \geq 0\}$, it holds $M_r(py) \succcurlyeq 0$.

Schmüdgen [23] gave a complete characterization of the K-moment problem in case the semialgebraic set K is compact.

Theorem 4.7 *Let the closed semialgebraic set K defined by $g_1, \ldots, g_m \in \mathbb{R}[x]$ be compact. Then a sequence $y = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ is a K-moment sequence if and only if $M_r((y_\alpha)_{\alpha \leq 2r}) \succcurlyeq 0$ for all $r \in \mathbb{N}$ and $M_r(((g_{j_1} \cdots g_{j_k} y)_\alpha)_{\alpha \leq 2r}) \succcurlyeq 0$ for all possible choices $j_1, \ldots, j_k$ of pairwise different numbers from $\{1, \ldots, m\}$ and for all $r \in \mathbb{N}$.*

This theorem by Schmüdgen characterizes infinite moment sequences. In real case problems, we are often facing the case of a truncated moment sequence. For instance a truncated sequence $(y_\alpha)_{\alpha \leq \omega}$ is given, which reflects our knowledge of the moments up to order ω of an unknown distribution. Furthermore dealing with an infinite sequence of semidefiniteness conditions leads to numerically intractable problems. For this reason, representations for truncated moment vectors y are in demand. The truncated complex K-moment problem was analyzed in detail by Curto and Fialkow [5]. It is to be stated that the multivariate case differs from the univariate case as it might happen that for a vector y the moment matrix $M_r(y) \succ 0$ for some $r \in \mathbb{N}$, but there exists no representing measure μ_y. Thus, $M_r(y) \succcurlyeq 0$ for a finite number of $r \in \mathbb{N}$ does in general not imply there exists a representing Borel measure μ_y. Nevertheless, we know by Schmüdgens Theorem 4.7 that the converse holds. Another solution of the K-moment problem for compact semialgebraic sets is Theorem 5.1, which is presented in the next chapter and applied in Lasserre's approach to solve problem (1.2). Also, Curto and Fialkow provided solutions of the moment problem in case the rank of the infinite matrix $M(y)$ is finite, which we apply in section 6.3.

As mentioned before, moment problems arise in settings where partial moment data (moments up to a certain order) of a unknown distribution are known and the class of distributions fitting to this truncated moment sequence is to be examined. Some of these problems can be solved by semidefinite programming characterizations. We will present a problem by Bertsimas and Popescu [1] as an example of those problems.

4.3 Optimal bounds in probability

Suppose $\sigma \in \mathbb{R}^{s(\omega,n)}$ is a vector of moments up to order $\omega \in \mathbb{N}$ and $\mu \in \mathcal{M}_P(\mathbb{R}^n)$ is a probability measure with $\sigma_\alpha = \int_\Omega x^\alpha d\mu$ for all $\alpha \leq \omega$. The (n, ω, Ω)-bound problem is to find the best possible upper (or lower) bounds on $\mu(S)$ for arbitrary events $S \subseteq \Omega$, i.e. maximize (or minimize) $\mu(S)$ over all probability measures μ that fit to the moment data σ.

Definition 4.8 α *is a **tight upper bound** on $\mu(S)$, and denoted by $\sup_{\mu \sim \sigma} \mu(S)$ if,*

(a) *it is an upper bound, i.e., $\mu(S) \le \alpha$ for all $\mu \in \mathcal{M}(\Omega)$, $\mu \sim \sigma$;*

(b) *it cannot be improved, i.e., for any $\epsilon > 0$ there is a $\mu_\epsilon \sim \sigma$ for which $\mu_\epsilon(S) > \alpha - \epsilon$.*

First, we formulate the problem of finding a tight upper bound for $\mu(S)$ for μ corresponding to a fixed truncated moment vector σ as a primal-dual pair of problems. With f the density corresponding to a probability measure μ, we can write the (n, ω, Ω)-problem as

$$
\begin{aligned}
Z_P = \quad &\max && \int_S f(z)\, dz \\
&\text{s.t.} && \int_\Omega z^\alpha f(z)\, dz = \sigma_\alpha, && \forall\, \alpha \le \omega, \\
& && f(z) \ge 0, && \forall\, z \in \Omega \subseteq \mathbb{R}^n.
\end{aligned}
\tag{4.1}
$$

Note that if problem (4.1) is feasible, σ is a feasible moment vector, and any feasible distribution $f(z)$ is a σ−feasible distribution.

As a dual to problem (4.1), the following problem can be derived.

$$
\begin{aligned}
Z_D = \quad &\min && \sum_{\alpha \le \omega} u_\alpha \sigma_\alpha = E\left[g(X)\right] \\
&\text{s.t.} && g(x) = \sum_{\alpha \le \omega} u_\alpha x^\alpha \in \mathbb{R}\left[x\right],\ \deg g = \omega, \\
& && g(x) \ge \chi_S(x),\ \forall\, x \in \Omega \subseteq \mathbb{R}^n,
\end{aligned}
\tag{4.2}
$$

where $\chi_S(x)$ the indicator function of S. Under some mild restrictions it can be shown that weak and strong duality hold for this pair of optimization problems.

4.3.1 The univariate case

We will show that the best tight bounds of the upper bound problem in the univariate case can be derived as solutions of a semidefinite program.

Given the first k moments $(M_1, \ldots, M_k)$, $M_0 = 1$ of a measure $\mu \in \mathcal{M}(\Omega)$ we attempt the dual problem (4.2) in order to determine the tight bound for $\mu(S)$.

$$
\begin{aligned}
&\min && \sum_{r=0}^{k} y_r M_r \\
&\text{s.t.} && \sum_{r=0}^{k} y_r x^r \ge 1, && \forall\, x \in S \\
& && \sum_{r=0}^{k} y_r x^r \ge 0, && \forall\, x \in \Omega.
\end{aligned}
\tag{4.3}
$$

In case S and Ω are intervals in the real line we provide a theorem that is the basis for expressing (4.3) as a semidefinite program.

Theorem 4.9 (a) *The polynomial $g(x) = \sum_{r=0}^{2k} y_r x^r$ satisfies $g(x) \ge 0$ for all $x \in \mathbb{R}$ if and only if there exists a positive semidefinite matrix V such that*

$$
y_r = \sum_{i,j:\, i+j=r} v_{ij}, \qquad r = 0, \ldots, 2k, \qquad V \succcurlyeq 0.
$$

(b) *The polynomial $g(x) = \sum_{r=0}^{k} y_r x^r$ satisfies $g(x) \ge 0$ for all $x \ge 0$ if and only if there exists a positive semidefinite matrix V, such that*

$$
\begin{aligned}
0 &= \sum_{i,j:\, i+j=2l-1} v_{ij}, & l &= 1, \ldots, k, \\
y_l &= \sum_{i,j:\, i+j=2l} v_{ij}, & l &= 0, \ldots, k, \\
V &\succcurlyeq 0.
\end{aligned}
$$

15

(c) The polynomial $g(x) = \sum_{r=0}^{k} y_r x^r$ satisfies $g(x) \geq 0$ for all $x \in [0, a]$ if and only if there exists a positive semidefinite matrix V, such that

$$
\begin{aligned}
0 &= \textstyle\sum_{i,j\,:\,i+j=2l-1} v_{ij}, \quad l = 1, \ldots, k, \\
\textstyle\sum_{r=0}^{l} y_r \binom{k-r}{l-r} a^r &= \textstyle\sum_{i,j\,:\,i+j=2l} v_{ij}, \qquad l = 0, \ldots, k, \\
V &\succcurlyeq 0.
\end{aligned}
$$

(d) The polynomial $g(x) = \sum_{r=0}^{k} y_r x^r$ satisfies $g(x) \geq 0$ for all $x \in [a, \infty)$ if and only if there exists a positive semidefinite matrix V, such that

$$
\begin{aligned}
0 &= \textstyle\sum_{i,j\,:\,i+j=2l-1} v_{ij}, \quad l = 1, \ldots, k, \\
\textstyle\sum_{r=l}^{k} y_r \binom{r}{l} a^r &= \textstyle\sum_{i,j\,:\,i+j=2l} v_{ij}, \qquad l = 0, \ldots, k, \\
V &\succcurlyeq 0.
\end{aligned}
$$

(e) The polynomial $g(x) = \sum_{r=0}^{k} y_r x^r$ satisfies $g(x) \geq 0$ for all $x \in (-\infty, a]$ if and only if there exists a positive semidefinite matrix V, such that

$$
\begin{aligned}
0 &= \textstyle\sum_{i,j\,:\,i+j=2l-1} v_{ij}, \quad l = 1, \ldots, k, \\
\textstyle\sum_{r=0}^{k-l} y_r \binom{k-r}{l} a^r &= \textstyle\sum_{i,j\,:\,i+j=2l} v_{ij}, \qquad l = 0, \ldots, k, \\
V &\succcurlyeq 0.
\end{aligned}
$$

(f) The polynomial $g(x) = \sum_{r=0}^{k} y_r x^r$ satisfies $g(x) \geq 0$ for all $x \in [a, b]$ if and only if there exists a positive semidefinite matrix V, such that

$$
\begin{aligned}
0 &= \textstyle\sum_{i,j\,:\,i+j=2l-1} v_{ij}, \quad l = 1, \ldots, k, \\
\textstyle\sum_{m=0}^{l} \sum_{r=m}^{k+m-l} y_r \binom{r}{m} \binom{k-r}{l-m} a^{r-m} b^m &= \textstyle\sum_{i,j\,:\,i+j=2l} v_{ij}, \qquad l = 0, \ldots, k, \\
V &\succcurlyeq 0.
\end{aligned}
$$

Proof (a) follows immediately with the equivalence of nonnegativity of polynomials and the existence of sum of squares decompositions in the univariate case and Theorem 3.4. (b) - (f) follow from (a), c.f. [28], pp. 488. $\square$

The following theorem is a direct consequence of Theorem 4.9. It provides a description of problem (4.3) as semidefinite programs for specific cases of $S \subseteq \Omega$ and $\Omega \subseteq \mathbb{R}$.

Theorem 4.10 *Given the first k moments $(M_1, \ldots, M_k)$, $M_0 = 1$ of a probability measure $\mu \in \mathcal{M}(\Omega)$ we obtain the following tight upper bounds:*

1. *If $\Omega = \mathbb{R}^+$, the tight upper bound on $\mu([a, \infty))$ is given as the solution of the semidef-*

inite optimization problem

$$
\begin{aligned}
\min \quad & \sum_{r=0}^{k} y_r M_r \\
s.t. \quad & 0 = \sum_{i,j:\,i+j=2l-1} v_{ij}, & l = 1,\dots,k, \\
& (y_0 - 1) + \sum_{r=1}^{k} y_r \binom{r}{l} a^r = v_{00}, \\
& \sum_{r=l}^{k} y_r \binom{r}{l} a^r = \sum_{i,j:\,i+j=2l} v_{ij}, & l = 1,\dots,k, \qquad (4.4) \\
& 0 = \sum_{i,j:\,i+j=2l-1} z_{ij}, & l = 1,\dots,k, \\
& \sum_{r=0}^{l} y_r \binom{k-r}{l-r} a^r = \sum_{i,j:\,i+j=2l} z_{ij}, & l = 0,\dots,k, \\
& V, Z \succcurlyeq 0.
\end{aligned}
$$

If $\Omega = \mathbb{R}$, the tight upper bound on $\mu\left([a,\infty)\right)$ is given as the solution of the semidefinite optimization problem (4.4), where the next to the last constraint is replaced by

$$
\sum_{r=0}^{k-l} y_r \binom{k-r}{l} a^r = \sum_{i,j:\,i+j=2l} z_{ij}, \quad l = 0,\dots,k.
$$

2. *If $\Omega = \mathbb{R}^+$, the tight upper bound on $\mu\left([a,b]\right)$ is given as the solution of the semidefinite optimization problem*

$$
\begin{aligned}
\min \quad & \sum_{r=0}^{k} y_r M_r \\
s.t. \quad & 0 = \sum_{i,j:\,i+j=2l-1} v_{ij}, & l = 1,\dots,k, \\
& \sum_{m=0}^{l} \sum_{r=m}^{k+m-l} y_r \binom{r}{m} \binom{k-r}{l-m} a^{r-m} b^m \\
& \quad = \binom{k}{l} + \sum_{i,j:\,i+j=2l} v_{ij}, & l = 0,\dots,k, \\
& 0 = \sum_{i,j:\,i+j=2l-1} z_{ij}, & l = 1,\dots,k, \qquad (4.5) \\
& \sum_{r=0}^{l} y_r \binom{k-r}{l-r} a^r = \sum_{i,j:\,i+j=2l} z_{ij}, & l = 0,\dots,k, \\
& 0 = \sum_{i,j:\,i+j=2l-1} u_{ij}, & l = 1,\dots,k, \\
& \sum_{r=l}^{k} y_r \binom{r}{l} b^r = \sum_{i,j:\,i+j=2l} u_{ij}, & l = 0,\dots,k, \\
& V, Z, U \succcurlyeq 0.
\end{aligned}
$$

If $\Omega = \mathbb{R}$, the tight upper bound on $\mu\left([a,b]\right)$ is given as the solution of the semidefinite optimization problem (4.5) where the fourth set of constraints is replaced by

$$
\sum_{r=0}^{k-l} y_r \binom{k-r}{l} a^r = \sum_{i,j:\,i+j=2l} z_{ij}, \quad l = 0,\dots,k.
$$

Proof Apply Theorem 4.9, (b)-(f), c.f. [28], pp. 492.

4.3.2 The multivariate case

In the multivariate case we do not have the equivalence of nonnegativity of polynomials and the existence of sum of squares decompositions. In fact, most instances of the (n,ω,Ω)-bound problem are NP-hard. For instance the $(n,2,\mathbb{R}^n_+)-$ and $(n,k,\mathbb{R}^n)-$ bound

problems are NP-hard for $k \geq 0$ [1]. Nevertheless, in case S and Ω are semialgebraic sets and other particular cases, it is possible to construct a sequence of semidefinite programs whose optima converge to the optimum of problem (4.2). The construction is in spirit of the relaxations for polynomial optimization problems by Lasserre [11]. It is based on Putinars Positivstellensatz, Theorem 3.11.

Suppose Ω and S are **compact semialgebraic sets** given by

$$
\begin{aligned}
\Omega &= \left\{ x \in \mathbb{R}^n \mid \omega_i(x) = \sum_{\kappa \in \Lambda(l)} \omega_\kappa^i x^\kappa \geq 0, \ i = 1, \ldots, r \right\}, \\
S &= \left\{ x \in \mathbb{R}^n \mid s_i(x) = \sum_{\kappa \in \Lambda(t)} s_\kappa^i x^\kappa \geq 0, \ i = 1, \ldots, m \right\},
\end{aligned}
$$

that is Ω and S are semialgebraic sets. Let $\delta_{\kappa,0} = 1$ if $\kappa = 0$, and zero, otherwise.

Theorem 4.11 *Assume $M(\Omega)$ and $M(S)$ are archimedian, i.e. the conditions of Putinar's Positivstellensatz are satisfied. Then, for every $\epsilon > 0$ there exists a nonnegative integer $d \in \mathbb{Z}_+$ (representing the degree of the polynomials in Putinar's Positivstellensatz), such that the objective function value Z_D in problem (4.2) satisfies $\mid Z_D - Z_D(d) \mid \leq \epsilon$, where $Z_D(d)$ is the value of the following semidefinite program:*

$$
Z_D(d) = \min \sum_{\kappa \in \Lambda(k)} y_\kappa \sigma_\kappa
$$

$$
\begin{aligned}
s.t. \quad & y_\kappa - \delta_{\kappa,0} = q_\kappa^0 + \sum_{i=1}^m \sum_{\eta \in \Lambda(d), \theta \in \Lambda(t), \eta + \theta = \kappa} q_\eta^i s_\theta^i && \forall \, \kappa \in \Lambda(k), \\
& 0 = q_\kappa^0 + \sum_{i=1}^m \sum_{\eta \in \Lambda(d), \theta \in \Lambda(t), \eta + \theta = \kappa} q_\eta^i s_\theta^i && \forall \, \kappa \in \Lambda(t+d) \setminus \Lambda(k), \\
& y_\kappa = p_\kappa^0 + \sum_{i=1}^r \sum_{\eta \in \Lambda(d), \theta \in \Lambda(l), \eta + \theta = \kappa} p_\eta^i \omega_\theta^i && \forall \, \kappa \in \Lambda(k), \\
& 0 = p_\kappa^0 + \sum_{i=1}^r \sum_{\eta \in \Lambda(d), \theta \in \Lambda(l), \eta + \theta = \kappa} p_\eta^i \omega_\theta^i && \forall \, \kappa \in \Lambda(l+d) \setminus \Lambda(k), \\
& q_\kappa^i = \sum_{\eta,\theta \in \Lambda(d), \eta + \theta = \kappa} q_{\eta,\theta}^i && \forall \, \kappa \in \Lambda(d), \ i = 0, 1, \ldots, m, \\
& p_\kappa^i = \sum_{\eta,\theta \in \Lambda(d), \eta + \theta = \kappa} p_{\eta,\theta}^i && \forall \, \kappa \in \Lambda(d), \ i = 0, 1, \ldots, r, \\
& Q^i = \left[q_{\eta,\theta}^i \right]_{\eta,\theta \in \Lambda(d)} \succcurlyeq 0, && i = 0, 1, \ldots, m, \\
& P^i = \left[p_{\eta,\theta}^i \right]_{\eta,\theta \in \Lambda(d)} \succcurlyeq 0, && i = 0, 1, \ldots, r.
\end{aligned}
$$

$$(4.6)$$

Proof We follow the proof in [1]. First remark that the value of the dual problem (4.2) equals that of the following strict inequality formulation:

$$
Z_D = \inf_{y \in \mathbb{R}^{|\Lambda(k)|}} \sum_{\kappa \in \Lambda(k)} y_\kappa \sigma_\kappa
$$

$$
\textbf{s.t.} \quad
\begin{aligned}
& g(x) = \sum_{\kappa \in \Lambda(k)} y_\kappa x^\kappa > 1 && \forall \, x \in S, \\
& g(x) = \sum_{\kappa \in \Lambda(k)} y_\kappa x^\kappa > 0 && \forall \, x \in \Omega.
\end{aligned}
\qquad (4.7)
$$

This problem may not admit an optimal solution. However, for any $\epsilon > 0$ there exists a feasible polynomial g_ϵ resulting in an objective value that is less than $Z_D + \epsilon$. As $g_\epsilon > 0$ on Ω it follows with Putinar's Theorem 3.11

$$
g_\epsilon = p_0(x) + \sum_{i=1}^m p_i(x) \omega_i(x),
$$

where p_i sum of squares, i.e. there exist some $P^i \succcurlyeq 0$ such that $p_i(x) = f^T P^i f$. f denotes the vector f monomials up to a certain degree d_0. Writing

$$p_i(x) = \sum_{\eta \in \Lambda(d)} p_\eta^i x^\eta, \qquad \omega_i(x) = \sum_{\theta \in \Lambda(l)} \omega_\theta^i x^\theta,$$

we have that $g_\epsilon(x) > 0$ for $x \in \Omega$ if and only if

$$g_\epsilon(x) := \sum_{\kappa \in I} y_\kappa x^\kappa = \sum_{\eta \in \Lambda(d)} p_\eta^0 x^\eta + \sum_{i=1}^r \left(\sum_{\eta \in \Lambda(d)} p_\eta^i x^\eta \right) \left(\sum_{\theta \in \Lambda(l)} \omega_\theta^i x^\theta \right).$$

Comparing coefficients, we obtain the third and fourth sets of linear constraints in (4.6), corresponding to the degree d_0.

Similarly, one can translate the feasibility constraint $g_\epsilon(x) - 1 > 0$ for all $x \in S$ into the first two sets of constraints in problem (4.6), corresponding to a certain degree d_1. It follows that the vector of coefficients y of the polynomial g_ϵ is feasible for problem (4.6) with degree $d = \max(d_0, d_1)$, so $Z_D(\epsilon)(d) \leq Z_D + \epsilon$.

On the other hand, for any d, problem (4.7) is a restriction of problem (4.6) for feasible polynomials of degree d, so $Z_D(d) \geq Z_D$, and the desired result follows. $\square$

This theorem provides an asymptotically exact sequence of semidefinite programs of the (n, ω, Ω)-bound problem. As a drawback we have to notice, the sizes of these formulations are not bounded by a polynomial in n, even for fixed k, as they depend on the degree d of the polynomials appearing in (4.6). This can be understood by the NP-hardness of the general (n, ω, Ω)-bound problem. Nevertheless, we can obtain increasingly better upper bounds on Z_P by solving semdefinite problems of size polynomially bounded in n for fixed ω. As every polynomial that can be expressed as sum of squares of polynomials of degree d, can also be expressed as sum of squares of polynomials of degree $d + 1$, it holds

$$Z_P = Z_D \leq Z_D(d) \leq \ldots \leq Z_D(2) \leq Z_D(1).$$

Example 4.12 *Consider the $(2, 1, \Omega)-$bound problem, where*

$$\begin{aligned} \Omega &= \left\{ x \in \mathbb{R}^2 \mid \omega_1(x) := -x_1^2 - x_2^2 + 2 \geq 0 \right\}, \\ S &= \left\{ x \in \mathbb{R}^2 \mid s_1(x) := -x_1^2 - x_2^2 + 2 \geq 0, \; s_2(x) := x_1^2 + x_2^2 - 1 \geq 0 \right\}. \end{aligned}$$

Thus, $k = 1$, $l = 2$, $t = 2$, $r = 1$, $m = 2$ and

$$\begin{aligned} \omega^1 &= (2, 0, 0, -1, 0, -1)^T, \\ s^1 &= (2, 0, 0, -1, 0, -1)^T, \\ s^2 &= (-1, 0, 0, 1, 0, 1)^T. \end{aligned}$$

It is immediate, that $M(\Omega)$ and $M(S)$ are archimedian. We illustrate the semidefinite program $Z_D(1)$ (i.e. $d = 1$) which approximates the optimal value $Z_P = Z_D$ of the

$(2, 1, \Omega)-$ *bound problem.*

$$
\begin{aligned}
Z_D(1) &= \min y_{(0,0)} + y_{(1,0)}\sigma_{(1,0)} + y_{(0,1)}M_2\sigma_{(0,1)} \\
y_{(0,0)} - 1 &= q^0_{(0,0)} + 2q^1_{(0,0)} - q^2_{(0,0)}, \\
y_{(1,0)} &= q^0_{(1,0)} + 2q^1_{(1,0)} - q^2_{(1,0)}, \\
y_{(0,1)} &= q^0_{(0,1)} + 2q^1_{(0,1)} - q^2_{(0,1)}, \\
0 &= q_{(2,0)} - q^1_{(0,0)} + q^2_{(0,0)}, \\
0 &= q_{(1,1)}, \\
0 &= q_{(0,2)} - q^1_{(0,0)} + q^2_{(0,0)}, \\
0 &= q_{(3,0)} - q^1_{(1,0)} + q^2_{(1,0)}, \\
0 &= q_{(2,1)} - q^1_{(0,1)} + q^2_{(0,1)}, \\
0 &= q_{(1,2)} - q^1_{(1,0)} + q^2_{(1,0)}, \\
0 &= q_{(0,3)} - q^1_{(0,1)} + q^2_{(0,1)}, \\
y_{(0,0)} &= p^0_{(0,0)} + 2p^1_{(0,0)}, \\
y_{(1,0)} &= p^0_{(1,0)} + 2p^1_{(1,0)}, \\
y_{(0,1)} &= p^0_{(0,1)} + 2p^1_{(0,1)}, \\
0 &= p^0_{(2,0)} - p^1_{(0,0)}, \\
0 &= p^0_{(1,1)}, \\
0 &= p^0_{(0,2)} - p^1_{(0,0)}, \\
0 &= p^0_{(3,0)} - p^1_{(1,0)}, \\
0 &= p^0_{(2,1)} - p^1_{(0,1)}, \\
0 &= p^0_{(1,2)} - p^1_{(1,0)}, \\
0 &= p^0_{(0,3)} - p^1_{(0,1)}, \\
q^i_\kappa &= \textstyle\sum_{\eta,\theta \in \Lambda(1), \eta+\theta=\kappa} q^i_{\eta,\theta} && \forall \kappa \in \Lambda(1),\ i = 0,1,2 \\
p^i_\kappa &= \textstyle\sum_{\eta,\theta \in \Lambda(1), \eta+\theta=\kappa} p^i_{\eta,\theta} && \forall \kappa \in \Lambda(1),\ i = 0,1, \\
Q^1,\ Q^2,\ Q^3 &\succcurlyeq 0, \\
P^1,\ P^2 &\succcurlyeq 0.
\end{aligned}
$$

In addition to the semialgebraic case it is also possible to derive explicit tight bounds in particular cases. We study two of these cases.

The first is the $(n, 1, \mathbb{R}^n_+)$-bound problem   where $S \subseteq \mathbb{R}^n_+$ is a convex set. Let $M > 0$ be a moment vector, which represents the vector of means of a multivariate distribution.

Theorem 4.13 *The tight $(n, 1, \mathbb{R}^n_+)$-upper bound for an arbitrary convex set S is given by:*

$$
\sup_{\mu \sim M} \mu(S) = \min\left(1,\ \max_{i=1,\ldots,n} \frac{M_i}{\inf x \in S_i x_i}\right), \tag{4.8}
$$

where $S_i = \left\{ x \in S \mid \frac{x_i}{M_i} \geq \frac{x_j}{M_j}\ \forall j \neq i \right\}$ is the convex subset of S for which the mean-rescaled ith coordinate is largest.

Proof C.f. [1].
When we specialize the bound from Theorem 4.13 for the set
$S = \{x \mid x_i \geq (1+\delta_i)M_i,\ \forall i = 1,\ldots,n\}$, we obtain

$$
\sup_{\mu \sim M} \mu\left(X_1 \geq (1+\delta_1)M_1, \ldots, X_n \geq (1+\delta_n)M_n\right) = \min_{i=1,\ldots,n} \frac{1}{1+\delta_i}, \tag{4.9}
$$

20

which represents a multidimensional generalization of Markov's inequality. In the univariate case it is exactly Markov's inequality

$$\sup_{\mu \sim M} \mu(X \geq (1+\delta)M) = \frac{1}{1+\delta}.$$

The second case is the $(n, 2, \mathbb{R}^n)$-bound problem . Let M denote the vector of known first moments. Instead of assuming the matrix of second moments $\int xx^T \, d\mu$ is known, we assume the matrix of centralized second moments to be known, i.e. the covariance matrix $\Gamma = \int (x - M)(x - M)^T \, d\mu$ of μ is known. That implies $\Gamma \in \mathbb{S}_+^n$. Marshall and Olkin [15] solved the $(n, 2, \mathbb{R}^n)-$bound problem in case S is a convex set:

Theorem 4.14 *The tight $(n, 2, \mathbb{R}^n)$-upper bound for a convex event S is given by*

$$\sup_{\mu \sim (M,\Gamma)} \mu(S) = \frac{1}{1+d^2}, \tag{4.10}$$

where $d^2 = \inf_{x \in S}(x - M)^T \Gamma^{-1}(x - M)$ is the squared distance from M to the set S, under the norm induced by the matrix Γ^{-1}.

Proof: C.f. [15]
This bound can be understood as a multivariate generalization of Chebyshev's inequality. If $\Gamma^{-1}M_\delta$, the tight bound can be expressed as

$$\sup_{\mu \sim (M,\Gamma)} \mu\left(X > (1+\delta)M\right) = \frac{1}{1 + (\delta M)^T \Gamma^{-1}(\delta M)}. \tag{4.11}$$

It is interesting to note that it improves Chebyshev's inequality in the univariate case. If we define the coefficient of variation: $C_M^2 = \frac{M_2 - M_1^2}{M_1^2}$. Chebyshev's inequality is given by

$$\mu\left(X > (1+\delta)M_1\right) \leq \frac{C_M^2}{\delta^2},$$

where as the tight bound (4.11) is stronger:

$$\mu\left(X > (1+\delta)M_1\right) \leq \frac{C_M^2}{C_M^2 + \delta^2}.$$

5 Polynomial optimization

The idea to apply convex optimization techniques to solve polynomial optimization problems was first proposed in the pioneering work of Shor [25]. Shor introduced lower bounds for the global minimum of a polynomial function p. These bounds are derived by minimizing a quadratic function subject to quadratic constraints. Also Nesterov discussed the minimizaion of univariate polynomials and mentioned the problem of minimizing multivariate polynomials in [17]. It was Lasserre [11] who first realized the possibility to apply Putinar's Positivstellensatz, Theorem 3.11, to solve a broader class of polynomial optimization problems, that goes beyond the case where $p - p^\star$ can be described as sum of squares of polynomials.

First, we introduce Lasserre's approach to derive semidefinite relaxations for the minimizing a polynomial over a semialgebraic set as Putinar's theorem is applied directly there. Second, we present the unconstrained case, which was considered by Shor for a particular case first. Since semialgebraic sets enter through the backdoor, in order to be able to apply Putinar's Positivstellensatz, we present it after the constrained case.

5.1 Lasserre's relaxation in the constrained case

After studying positivity and nonnegativity of polynomials and the related problem of moments, we attempt the inital polynomial optimization problem (1.2) over a compact semialgebraic set K,

$$\min_{x \in K} p(x).$$

One of the major obstacles for finding the optimum $p^\star$ is the fact that the set K and the function p are far from being convex. It is the idea of Lasserre' s approach [11] to convexify problem (1.2). We outline this procedure of convexification. It has to be emphasized that Lasserre's approach is based on two assumptions. First we require the semi-algebraic set K to be **compact**, and second we assume $M(K)$ is **archimedian**. These assumptions imply, we are able to apply Putinar's Positivstellensatz to polynomials positive on K.

At first we note,

$$p^\star = \sup\left\{a \in \mathbb{R} \mid p - a \geq 0 \text{ on } K\right\} = \sup\left\{a \in \mathbb{R} \mid p - a > 0 \text{ on } K\right\}. \tag{5.1}$$

Since we assume that $M(K)$ archimedian, we apply Theorem 3.11 to (5.1). Thus

$$p^\star \leq \sup\left\{a \in \mathbb{R} \mid p - a \in M(K)\right\} \leq \sup\left\{a \in \mathbb{R} \mid p - a \geq 0 \text{ on } K\right\} = p^\star.$$

Finally we obtain

$$p^\star = \sup\left\{a \in \mathbb{R} \mid p - a \in M(K)\right\}. \tag{5.2}$$

As a second approach, we note for the minimum $p^\star$ of (1.1) holds

$$p^\star = \inf\left\{\int p \, d\mu \mid \mu \in \mathcal{M}_P(K)\right\}, \tag{5.3}$$

where $\mathcal{M}_P(K) \subseteq \mathcal{M}(K)$ denotes the set of all Borel measures on K which are also probability measures. $' \leq '$ holds since $p(x) \geq p^\star$ on K implies $\int p \, d\mu \geq p^\star$. And $' \geq '$ follows as each x feasible in (1.1) corresponds to a $\mu = \delta_x \in \mathcal{M}(K)$, where δ_x the Dirac measure at x.

In order to get rid of the set $\mathcal{M}(K)$ in 5.3 we exploit the following theorem by Putinar [22].

Theorem 5.1 *For any map $L : \mathbb{R}[x] \to \mathbb{R}$, the following are equivalent:*

(i) L is linear, $L(1) = 1$ and $L(M(K)) \subseteq [0, \infty)$.

(ii) L is integration with respect to a probability measure μ on K, i.e.,

$$\exists \, \mu \in \mathcal{M}_P(K) : \, \forall p \in \mathbb{R}[x] : \, L(p) = \int p \, d\mu.$$

Proof C.f. [24], pp. 10: The implication $(ii) \Rightarrow (i)$ is trivial. To show the converse, suppose that (i) holds. Consider the ring homomorphism

$$\phi : \mathbb{R}[x] \to \mathcal{C}(K, \mathbb{R}) : p \mapsto p \mid_K$$

from the polynomial ring into the ring $\mathcal{C}(K, \mathbb{R})$ of continuous real-valued functions on K. Suppose $p \in \mathbb{R}[x]$ satisfies $p \geq 0$ on K. Then $p + \epsilon \in M(K)$ by Theorem 3.11 and a fortiori $L(p) + \epsilon = L(p + \epsilon) \geq 0$ for all $\epsilon > 0$. This implies $L(p) \geq 0$. In particular, L vanishes on the kernel of ϕ and induces therefore a linear map $\bar{L} : \phi(\mathbb{R}[x]) \to \mathbb{R}$ well defined by $\bar{L}(\phi(p)) := L(p)$ for all $p \in \mathbb{R}[x]$. We equip $\mathcal{C}(K, \mathbb{R})$ with the supremum norm and thus turn it into a normed vector space, noting that $K = \emptyset$ would imply $-1 \in M(K)$, whence $-1 = -L(1) = L(-1) \geq 0$. By the Stone-Weierstrass Approximation Theorem, $\phi(\mathbb{R}[x])$ lies dense in $\mathcal{C}(K, \mathbb{R})$. It is easy to see that $\bar{L}(\phi(p)) = L(p) \leq \parallel p \parallel$ for all $p \in \mathbb{R}[x]$. 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 $\mathcal{C}(K, \mathbb{R})$. Using again the Stone-Weierstrass Theorem, it is easy to see that $\bar{L}$ maps $\mathcal{C}(K, [0, \infty))$ into $[0, \infty)$. Since K is compact, the Risz Representation Theorem tells us that $\bar{L}$ is integration with respect to a measure on K. $\square$

This theorem does not really characterize $\mathcal{M}_P(K)$, but all real families $(y_\alpha)_{\alpha \in \mathbb{N}^n}$ that are sequences of moments of probability measures on K, i.e.,

$$y_\alpha = \int x^\alpha \, d\mu \qquad \forall \, \alpha \in \mathbb{N}^n,$$

where $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. This statement is true, as every linear map $L : \mathbb{R}[x] \to \mathbb{R}$ is given uniquely by its values $L(x^\alpha)$ on the basis $(x^\alpha)_{\alpha \in \mathbb{N}^n}$ of $\mathbb{R}[x]$. With Theorem 5.1 we obtain

$$p^\star = \inf \{ L(f) \mid L : \mathbb{R}[x] \to \mathbb{R} \text{ is linear}, \, L(1) = 1, \, L(M(K)) \subseteq [0, \infty) \}. \qquad (5.4)$$

Recall (5.2)

$$p^\star = \sup \{ a \in \mathbb{R} \mid f - a \in M(K) \}.$$

Thus (5.4) can be understood as a primal approach to the original problem (1.1) and (5.2) as a dual approach. Due to complexity reasons it is necessary to introduce relaxations to these primal-dual pair of optimization problems, in order to solve the problem (1.1). Therefore we approximate $M(K)$ by the sets $M_\omega(K) \subseteq \mathbb{R}[x]$, where $M_\omega(K) := \left\{ \sum_{i=0}^m \sigma_i g_i \mid \sigma_i \in \sum \mathbb{R}[x]^2, \deg(\sigma_i g_i) \leq 2\omega \right\}$ for an

$$\omega \in \mathcal{N} := \{ s \in \mathbb{N} \mid s \geq \omega_{\max} := \max \{ \omega_0, \omega_1, \ldots, \omega_m \} \},$$

$\omega_i := \lceil \frac{\deg g_i}{2} \rceil$ $(i = 1, \ldots, m)$, $\omega_0 := \lceil \frac{\deg p}{2} \rceil$. Replacing $M(K)$ by $M_\omega(K)$ motivates to consider the following pair of optimization problems for a $\omega \in \mathcal{N}$.

$$
\begin{array}{llll}
(P_\omega) & \min \quad L(f) & \text{subject to} & L : \mathbb{R}[x]_{2\omega} \to \mathbb{R} \text{ is linear}, \\
& & & L(1) = 1 \text{ and} \\
& & & L(M_\omega(K)) \subseteq [0, \infty) \\
(D_\omega) & \max \quad a & \text{subject to} & a \in \mathbb{R} \text{ and} \\
& & & p - a \in M_\omega(K)
\end{array} \qquad (5.5)
$$

The optimal values of (P_ω) and (D_ω) are denoted by $P_\omega^\star$ and $D_\omega^\star$, respectively. The parameter $\omega \in \mathcal{N}$ is called the **relaxation order** of (5.5). It determines the size of the relaxations (P_ω) and (D_ω) to (1.2) and therefore also the numerical effort that is necessary to solve them.

Theorem 5.2 (Lasserre) *Assume $M(K)$ is archimedian. $(P_\omega^\star)_{\omega \in \mathcal{N}}$ and $(D_\omega^\star)_{\omega \in \mathcal{N}}$ are increasing sequences that converge to $p^\star$ and satisfy $D_\omega^\star \leq P_\omega^\star \leq p^\star$ for all $\omega \in \mathcal{N}$. Moreover, if $p - p^\star \in M(K)$, then $D_\omega^\star = P_\omega^\star = p^\star$ for a sufficiently large relaxation order ω, i.e. strong duality holds.*

Proof Since the feasible set of (5.4) is a subset of the feasible set of (P_ω), $P_\omega^\star \leq p^\star$. Moreover, if L feasible for (P_ω) and a for (D_ω), $L(p) \geq a$ holds since $p - a \in M_\omega(K)$ implies $L(p) - a = L(p) - aL(1) = L(p - a) \geq 0$. Thus $D_\omega^\star \leq P_\omega^\star$. Obviously, a feasible a for (D_ω) is feasible for $(D_{\omega+1})$, and every feasible L of $(P_{\omega+1})$ is feasible for (P_ω). This implies $(P_\omega^\star)_{\omega \in \mathcal{N}}$ and $(D_\omega^\star)_{\omega \in \mathcal{N}}$ are increasing. Futhermore, as for any $\epsilon > 0$ there exists a sufficiently large $\omega \in \mathcal{N}$ such that $p - p^\star + \epsilon \in M_\omega(K)$ by Theorem 3.11, i.e. $p^\star - \epsilon$ feasible for (D_ω), the convergence follows. If $p - p^\star \in M(K)$, $p - p^\star \in M_\omega(K)$ for ω sufficiently large. Thus $p^\star$ feasible for (D_ω) und therefore $D_\omega^\star = P_\omega^\star = p^\star$. $\square$

If $M(K)$ not archimedian, we are still able to exploit Schmuedgen's Positivstellensatz to characterize $p - a$ in (D_ω).

As a next step we follow the observation of Lasserre and translate (D_ω) and (P_ω) to a pair of primal-dual semidefinite programs. We will exploit the following key lemma [24].

Lemma 5.3 *Suppose $L : \mathbb{R}[x] \to \mathbb{R}$ is a linear map. Then $L(M_\omega) \subseteq [0, \infty)$ if and only if the $m + 1$ matrices*

$$M_{\omega-\omega_i}(L(xg_i)) \succcurlyeq 0, \qquad \forall i \in \{0, \ldots, m\}.$$

Moreover,

$$M_\omega(K) = \left\{ \sum_{i=0}^{m} \langle M_{\omega-\omega_i}(xg_i), G_i \rangle \mid G_0, \ldots, G_m \in \mathbb{S}_+^{s(\omega-\omega_i)} \right\}.$$

Proof C.f. [24], p. 19.

Using this Lemma, we reformulate (5.5) as

$$
\begin{array}{lll}
(P_\omega^{\mathrm{I}}) & \min & L(p) \\
& \text{s.t.} & L : \mathbb{R}[x]_{2\omega} \to \mathbb{R} \text{ is linear }, \ L(1) = 1 \text{ and} \\
& & M_{\omega-\omega_i}(L(xg_i)) \succcurlyeq 0, \ i = 0, \ldots, m \\
(D_\omega^{\mathrm{I}}) & \max & a \\
& \text{s.t.} & a \in \mathbb{R}, \ G_0 \in \mathbb{S}_+^{s(\omega)}, G_i \in \mathbb{S}_+^{s(\omega-\omega_i)} \text{ for } i = 1, \ldots, m \text{ and} \\
& & \sum_{i=0}^{m} \langle M_{\omega-\omega_i}(xg_i), G_i \rangle = p - a
\end{array}
\tag{5.6}
$$

Sort the monomials in the polynomials $x^{\beta+\gamma}$ and define for $i \in \{1, \ldots, m\}$ ($g_0 \equiv 1$) and $\alpha \in \Lambda(2\omega)$ matrices $B_\alpha \in \mathbb{S}^{s(\omega)}$ and $C_{\alpha i} \in \mathbb{S}^{s(\omega-\omega_i)}$ such that

$$M_{\omega-\omega_0}(x) = M_\omega(x) = \sum_{\alpha \in \Lambda(2\omega)} B_\alpha x^\alpha, \qquad M_{\omega-\omega_i}(xg_i) = \sum_{\alpha \in \Lambda(2\omega)} C_{\alpha i}(\beta, \gamma) x^\alpha.$$

Also we define b_α to be the coefficient of x^α in f for each $0 \neq \alpha \in \Lambda(2\omega)$, i.e.,

$$p = \sum_{\alpha \in \Lambda(2\omega) \setminus \{0\}} b_\alpha x^\alpha.$$

Using the B_α, $C_{\alpha i}$ and the b_α we obtain

$$
\begin{aligned}
(P_\omega^{\mathrm{II}}) \quad &\min \quad \sum_{\alpha \in \Lambda(2\omega) \setminus \{0\}} b_\alpha L(x^\alpha) \\
&\text{s.t.} \quad L : \mathbb{R}\,[x]_{2\omega} \to \mathbb{R} \text{ is linear }, \ L(1) = 1 \text{ and} \\
&\qquad \sum_{\alpha \in \Lambda(2\omega)} L(x^\alpha) B_\alpha \succcurlyeq 0, \\
&\qquad \sum_{\alpha \in \Lambda(2\omega)} L(x^\alpha) C_{\alpha i} \succcurlyeq 0, \ i = 1, \dots, m \\
(D_\omega^{\mathrm{II}}) \quad &\max \quad a \\
&\text{s.t.} \quad a \in \mathbb{R}, \ G_0 \in \mathbb{S}_+^{s(\omega)}, \ G_i \in \mathbb{S}_+^{s(\omega - \omega_i)} \text{ for } i = 1, \dots, m \text{ and} \\
&\qquad \sum_{\alpha \in \Lambda(2\omega)} x^\alpha \left(\langle B_\alpha, G_0 \rangle + \sum_{i=1}^m \langle C_{\alpha i}, G_i \rangle \right) = \sum_{\alpha \in \Lambda(2\omega) \setminus \{0\}} b_\alpha x^\alpha - a
\end{aligned}
\tag{5.7}
$$

Exploiting the fact that a linear function $L : \mathbb{R}\,[x]_{2\omega} \to \mathbb{R}$ with $L(1) = 1$ can be identified with its values $y_\alpha := L(x^\alpha)$, $0 \neq \alpha \in \Lambda(2\omega)$ and $y_0 = 1$, (5.7) can be written as the following pair of primal-dual semidefinite programs

$$
\begin{aligned}
(P_\omega^{\mathrm{SDP}}) \quad &\min \quad \sum_{\alpha \in \Lambda(2\omega) \setminus \{0\}} b_\alpha y_\alpha \\
&\text{s.t.} \quad y_\alpha \in \mathbb{R}, \ 0 \neq \alpha \in \Lambda(2\omega), \text{ and} \\
&\qquad B_0 + \sum_{\alpha \in \Lambda(2\omega) \setminus \{0\}} y_\alpha B_\alpha \succcurlyeq 0, \\
&\qquad C_{0i} + \sum_{\alpha \in \Lambda(2\omega) \setminus \{0\}} y_\alpha C_{\alpha i} \succcurlyeq 0, \ i = 1, \dots, m \\
(D_\omega^{\mathrm{SDP}}) \quad &\max \quad -G_0(1,1) - \sum_{i=1}^m \langle C_{0i}, G_i \rangle \\
&\text{s.t.} \quad a \in \mathbb{R}, \ G_0 \in \mathbb{S}_+^{s(\omega)}, \ G_i \in \mathbb{S}_+^{s(\omega - \omega_i)} \text{ for } i \in \{1, \dots, m\} \text{ and} \\
&\qquad \langle B_\alpha, G_0 \rangle + \sum_{i=1}^m \langle C_{\alpha i}, G_i \rangle = b_\alpha, \ 0 \neq \alpha \in \Lambda(2\omega)
\end{aligned}
\tag{5.8}
$$

Given $(y_\alpha)_{\alpha \in \Lambda(2\omega)}$ feasible for $(D_\omega^{\mathrm{SDP}})$, we define a linear map $L : \mathbb{R}\,[x]_{2\omega} \to \mathbb{R}$ with $L(1) = 1$ and $L(x^\alpha) := -y_\alpha$ for $0 \neq \alpha \in \Lambda(2\omega)$. Obviously, L is feasible for (D_ω^{II}). Thus the optima of $(P_\omega^{\mathrm{SDP}})$, $(D_\omega^{\mathrm{SDP}})$ and $(P_\omega^{\mathrm{mod}})$, $(D_\omega^{\mathrm{mod}})$ coincide and with Theorem 5.2 follows that $(P_\omega^{\mathrm{SDP}})^\star$, $(D_\omega^{\mathrm{SDP}})^\star$ converge to $p^\star$ for $\omega \to \infty$. It is known that semidefinite programs can be solved in polynomial time efficiently.

5.2 Lasserre's relaxation in the unconstrained case

The procedure to derive a sequence of convergent SDP relaxations in the case of an unconstrained polynomial optimization problem

$$
\min_{x \in \mathbb{R}^n} p(x),
\tag{5.9}
$$

where $p \in \mathbb{R}\,[x]$ and $p^\star := \min_x p(x)$, is similar to the constrained case that we discussed before. Let p be of even degree $2l$, otherwise $\inf p = -\infty$. Furthermore, we will exploit the characterization of sum of squares decompositions by semidefinite matrices and Putinar's Positivstellensatz. In order to apply this theorem, it is necessary to construct an appropriate semialgebraic set as will be shown later.

First, we derive the following relaxation,

$$
\begin{aligned}
p^\star &= \inf \left\{ \int p \, d\mu \mid \mu \in \mathcal{M}_P(\mathbb{R}^n) \right\} \\
&\geq \inf \left\{ L(p) \mid L : \mathbb{R}\,[x] \to \mathbb{R}, \ L(1) = 1, \ M_l\,(L(x)) \in \mathbb{S}_+^{s(l)} \right\}.
\end{aligned}
\tag{5.10}
$$

We order the expression $M_l\,(L(x))$ and introduce symmetric matrices $B_\alpha \in \mathbb{S}^{s(l)}$ such that $M_l\,(L(x)) = \sum_{\alpha \in \Lambda(2l)} B_\alpha L(x^\alpha)$. Finally we identify $y_\alpha = L(x^\alpha)$ for $\alpha \in \Lambda(2l) \setminus \{0\}$ and

$y_0 = 1$ to obtain a relaxation for (5.9)

$$(P_l) \quad \min \quad \sum_\alpha p_\alpha y_\alpha$$
$$\text{s.t.} \quad \sum_{\alpha \neq 0} y_\alpha B_\alpha \succcurlyeq -B_0. \tag{5.11}$$

Analogous to the constrained case we can also apply another approach to (5.9),

$$p^\star = \sup \left\{ a \in \mathbb{R} \mid p(x) - a \geq 0 \,\forall\, x \in \mathbb{R}^n \right\} \geq \sup \left\{ a \in \mathbb{R} \mid p(x) - a \in \sum \mathbb{R}[x]^2 \right\}$$
$$= \sup \left\{ a \mid p(x) - a = \langle M_l(x), G \rangle, \, G \in \mathbb{S}_+^{s(l)} \right\}. \tag{5.12}$$

Thus, we derive another relaxation to problem (5.9),

$$(D_l) \quad \max \quad -G(1,1)$$
$$\text{s.t.} \quad \langle B_\alpha, G \rangle = p_\alpha, \quad \alpha \neq 0 \tag{5.13}$$
$$G \succcurlyeq 0.$$

With the duality theory of convex optimization it can be shown easily, that the two convex programs (5.11) and (5.13) are dual to each other. In case (5.13) has a feasible solution, strong duality holds, that is

$$P_l^\star = D_l^\star.$$

The idea of the following theorem was proposed by Shor [25] first. The presented version is due to Lasserre [11].

Theorem 5.4 (Shor) *If the nonnegative polynomial $p - p^\star$ is a sum of squares of polynomials, then (5.9) is equivalent to (5.11). More precisely, $p^\star = Z_P$ and, if $x^\star$ is a global minimizer of (5.9), then the vector*

$$y^\star := \left(x_1^\star, \ldots, x_n^\star, (x_1^\star)^2, x_1^\star x_2^\star, \ldots, (x_1^\star)^{2m}, \ldots, (x_n^\star)^{2m} \right)$$

is a minimizer of (5.11).

Next, we treat the **general case**, that is, when $p - p^\star$ is not sum of squares. As mentioned at the beginning we have to construct a semialgebraic set in order to be able to apply Putinar's Positivstellensatz. Suppose we know that a global minimizer $x^\star$ of $p(x)$ has norm less than a for some $a > 0$, that is, $p(x^\star) = p^\star$ and $\| x^\star \|_2 \leq a$. Then, with $x \to q_a(x) = a^2 - \| x \|_2^2$, we have $p(x) - p^\star \geq 0$ on $K_a := \{ q_a(x) \geq 0 \}$. $M(K_a)$ is obviously archimedian, as the condition (iii) in Theorem 3.12 is satisfied for $N = a^2$. Now, we can use that every polynomial f, strictly positive on the semialgebraic set K_a is contained in the quadratic module $M(K_a)$.
For every $\omega \geq l$, consider the following semidefinite program

$$(P_\omega^a) \quad \min \sum_\alpha p_\alpha y_\alpha,$$
$$M_\omega(y) \succcurlyeq 0, \tag{5.14}$$
$$M_{\omega-1}(q_a y) \geq 0.$$

Writing $M_{\omega-1}(q_a y) = \sum_\alpha y_\alpha D_\alpha$, for appropriate matrices D_α ($| \alpha | \leq 2\omega$), the dual of (P_ω^a) is the semidefinite program

$$(D_\omega^a) \quad \max -G(1,1) - a^2 H(1,1),$$
$$\langle G, B_\alpha \rangle + \langle H, D_\alpha \rangle = p_\alpha, \, \alpha \neq 0. \tag{5.15}$$

Then, the following theorem is due to Lasserre [11].

Theorem 5.5 (Lasserre) *Given (P^a_ω) and (D^a_ω) for some $a > 0$ such that $\| x^\star \|_2 \leq a$ at some global minimizer $x^\star$. Then*

(a) *as $\omega \to \infty$, one has*

$$\inf(P^a_\omega) \uparrow p^\star.$$

Moreover, for ω sufficiently large, there is no duality gap between (P^a_ω) and its dual (D^a_ω), and (D^a_ω) is solvable.

(b) $\min(P^a_\omega) = p^\star$ *if and only if $p - p^\star \in M_\omega(K_a)$. In this case, the vector*

$$y^\star := \left(x^\star_1, \ldots, x^\star_n, (x^\star_1)^2, x^\star_1 x^\star_2, \ldots, (x^\star_1)^{2\omega}, \ldots, (x^\star_n)^{2\omega}\right)$$

is a minimizer of (P^a_ω). In addition, $\max(P^a_\omega) = \min(D^a_\omega)$.

Proof

(a) From $x^\star \in K_a$ and with

$$y^\star := \left(x^\star_1, \ldots, x^\star_n, (x^\star_1)^2, x^\star_1 x^\star_2, \ldots, (x^\star_1)^{2\omega}, \ldots, (x^\star_n)^{2\omega}\right)$$

it follows that $M_\omega(y^\star), M_{\omega-1}(q_a y^\star) \succcurlyeq 0$ so that $y^\star$ is feasible for (P^a_ω) and thus $\inf(P^a_\omega) \leq p^\star$.
Now, fix $\epsilon > 0$ arbitrary. Then, $p - p^\star + \epsilon > 0$ and therefore, with Theorem 3.11 there is some N_0 such that

$$p - p^\star + \epsilon = \sum_{i=1}^{r_1} q_i(x)^2 + q(x) \sum_{j=1}^{r_2} t_j(x)^2$$

for some polynomials $q_i(x)$, $i = 1, \ldots, r_1$, of degree at most N_0, and some polynomials $t_j(x)$, $j = 1, \ldots, r_2$, of degree at most $N_0 - 1$. Let $q_i \in \mathbb{R}^{s(N_0)}, t_j \in \mathbb{R}^{s(N_0-1)}$ be the corresponding vectors of coefficients, and let

$$G := \sum_{i=1}^{r_1} q_i q_i^T, \quad Z := \sum_{j=1}^{r_2} t_j t_j^T$$

so that $G, H \succcurlyeq 0$. It is immediate to check that (G, H) feasible for (D^a_ω) with value $-G(1,1) - a^2 H(1,1) = (p^\star - \epsilon)$. From weak duality follows convergence as

$$p^\star - \epsilon \leq \inf(P^a_\omega) \leq p^\star.$$

For strong duality and for (b), c.f. [11]. $\square$

5.3 Global minimizer

Usually one is not only interested in finding the minimum value $p^\star$ of p on K, but also in obtaining a global minimizer $x^\star \in K^\star$ with $p(x^\star) = p^\star$. It will be shown that in Lasserre's procedure not only $(P^\star_\omega)$ converges to the infimum $p^\star$, but also a convergence to the minimizer $x^\star$ of (1.2) in case it is unique.

Definition 5.6 L_ω *solves (P_ω) nearly to optimality $(\omega \in \mathcal{N})$ if L_ω is a feasible solution of (P_ω) $(\omega \in \mathcal{N})$ such that $\lim_{\omega \to \infty} L_\omega(p) = \lim_{\omega \to \infty} P^\star_\omega$.*

This notation is useful because (P_ω) might not possess an optimal solution, and even if it does, we might not be able to compute it exactly. For an example, c.f. [24], Example 22. Obviously L_ω solves (P_ω) nearly to optimality ($\omega \in \mathcal{N}$) if and only if $\lim_{\omega \to \infty} L_\omega(f) = p^\star$. The following theorem is the basis for the convergence to a minimizer in case $K^\star$ is a singleton.

Theorem 5.7 *Suppose $K \neq \emptyset$ and L_ω solves (P_ω) nearly to optimality ($\omega \in \mathcal{N}$). Then*

$$\forall d \in \mathbb{N}: \ \forall \epsilon > 0: \exists k_0 \in \mathcal{N} \cap [d, \infty): \forall k \geq k_0: \exists \mu \in \mathcal{M}(K^\star): \left\| \left(L_\omega(x^\alpha) - \int x^\alpha \, d\mu \right)_{\alpha \in \Lambda(2d)} \right\| < \epsilon.$$

Proof Schweighofer, p. 11.
In the convenient case where $K^\star$ is a singleton it is possible to guarantee convergence of the minimizer:

Corollary 5.8 *$K^\star = \{x^\star\}$ is a singleton and L_ω solves (P_ω) nearly to optimality ($\omega \in \mathcal{N}$). Then*

$$\lim_{\omega \to \infty} (L_\omega(x_1), \dots, L_\omega(x_n)) = x^\star.$$

Proof We set $d = 1$ in Theorem 5.7 and note that $\mathcal{M}(K^\star)$ contains only the Dirac measure $\delta_{x^\star}$ at the point $x^\star$. It is possible to apply Corollary 5.8 to certify that $p^\star$ has almost been reached after succesively solving the relaxations (P_ω).

Corollary 5.9 *Suppose $M(K)$ is archimedian, p has a unique minimizer on the compact semialgebraic set K and L_ω solves (P_ω) nearly to optimality for all $\omega \in \mathcal{N}$. Then holds for all $\omega \in \mathcal{N}$,*

$$L_\omega(p) \leq p^\star \leq p(L_\omega(x_1), \dots, L_\omega(x_n)),$$

and the lower and upper bounds for $p^\star$ converge to $p^\star$ for $\omega \to \infty$.

Proof $L_\omega(p) \leq p^\star$ follows from Theorem 5.2. The convergence of $p(L_\omega(x_1), \dots, L_\omega(x_n))$ is a consequence of Corollary 5.8. To see that $p^\star$ is a lower bound, observe that

$$g_i (L_\omega(x_1), \dots, L_\omega(x_n)) = L_\omega(g_i) \geq 0,$$

whence $(L_\omega(x_1), \dots, L_\omega(x_n)) \in K$ for all $k \in \mathcal{N}$. $\quad \square$
The case where several optimal solutions exist is more difficult to handle. In fact, as soon as there are two or more global minimizers, it often occurs that symmetry in the problem prevents the nearly optimal solutions of the SDP relaxations to converge to a particular minimizer. Henrion and Lasserre developed an algorithm [8] to extract all optimal solutions in case $K^\star$ is finite and the condition

$$\mathrm{rank} M_\omega(y^\star) = \mathrm{rank} M_{\omega - \omega_{\max}}(y^\star) \tag{5.16}$$

is satisfied. In the case where the feasible set can be written

$$K = \{x \in \mathbb{R}^n \mid h_i(x) = 0, \ i = 1, \dots, k: \ g_j(x) \geq 0, \ j = 1, \dots, m\}$$

and the ideal $I(h_1, \dots, h_k)$ is zero-dimensional and radical (c.f. chapter 6), condition (5.16) is guaranteed to be satisfied. Beside that particular case it remains unclear when this sufficient condition (5.16) holds. Nevertheless, it is possible to derive **Karush-Kuhn-Tucker conditions for global optimality** and the global minimizers.

Theorem 5.10 *Let K be a compact semialgebraic set defined by the inequalities $g_i(x) \geq 0$, $i = 1, \ldots, m$. Assume that $x^\star \in K$ is a global minimizer of (1.1). If $p - p^\star$ can be written*

$$p(x) - p^\star = \sum_{i=1}^{r_0} q_i(x)^2 + \sum_{k=1}^{m} g_k(x) \sum_{j=1}^{r_k} t_{kj}(x)^2, \quad x \in \mathbb{R}^n,$$

for some polynomials $q_i(x)$, $t_{kj}(x)$, $i = 1, \ldots, r_0$, $k = 1, \ldots, m$, $j = 1, \ldots, r_k$, i.e. $p - p^\star \in M(K)$, then

$$
\begin{aligned}
0 &= g_k(x^\star) \left[\sum_{j=1}^{r_k} t_{kj}(x^\star)^2 \right], \quad k = 1, \ldots, m. \\
\nabla p(x^\star) &= \sum_{k=1}^{m} \nabla g_k(x^\star) \left[\sum_{j=1}^{r_k} t_{kj}(x^\star)^2 \right].
\end{aligned}
$$

Moreover, if there exist associated Lagrange Karush-Kuhn-Tucker multipliers $\lambda^\star \in \mathbb{R}_+^m$ and if the gradients $\nabla g_k(x^\star)$ are linearly independent, then

$$\sum_{j=1}^{r_k} t_{kj}(x^\star)^2 = \lambda_k^\star, \quad k = 1, \ldots, r.$$

Proof As $x^\star$ is a global minimizer of (1.1), it follows from $p(x^\star) - p^\star$ that

$$0 = \sum_{i=1}^{r_0} q_i(x^\star)^2 + \sum_{k=1}^{m} g_k(x^\star) \sum_{j=1}^{r_k} t_{kj}(x^\star)^2$$

so that

$$0 = q_i(x^\star), \; i = 1, \ldots, r_0, \;\; \text{and} \; 0 = g_k(x^\star) \sum_{j=1}^{r_k} t_{kj}(x^\star)^2, \; k = 1, \ldots, m.$$

Moreover,

$$\nabla p(x^\star) = \sum_{k=1}^{m} \nabla g_k(x^\star) \sum_{j=1}^{r_k} t_{kj}(x^\star)^2 = \sum_{k=1}^{m} \lambda_k^\star \nabla g_k(x^\star)$$

so that the claim follows from the linear independence of the $\nabla g_k(x^\star)$. $\quad\square$

Therefore, the condition $p - p^\star \in M(K)$ can be viewed as a global optimality condition of Karush-Kuhn-Tucker type, where the multipliers are now nonnegative polynomials instead of nonnegative constants. In contrast to the usual local Karush-Kuhn-Tucker optimality conditions, the polynomial multiplier associated to a constraint $g_k(x) \geq 0$, nonactive at $x^\star$, is not identically null, but vanishes at $x^\star$. If $p - p^\star \notin M(K)$, we still have that $p - p^\star + \epsilon \in M(K)$ for every $\epsilon > 0$. Of course, the degrees of q_i and t_{kj} in the representation of $p - p^\star + \epsilon$ depend on ϵ, but we have

$$\lim_{\epsilon \to 0} \sum_{i=1}^{r_0(\epsilon)} q_i(x^\star)^2 = 0 \;\; \text{and} \;\; \lim_{\epsilon \to 0} \sum_{j=1}^{r_k(\epsilon)} t_{kj}(x^\star)^2 = 0$$

for every k such that $g_k(x^\star) > 0$.

5.4 Sparse polynomial optimization

The SDP relaxation method by Lasserre is very appealing as a theoretical result as it allows to approximate the solutions of polynomial optimization problems (1.1) as closely as desired

by solving a finite sequene of SDP relaxations. However, since the size of the SDP relaxation grows as $\binom{n+\omega}{\omega}$, it is too difficult to solve even medium scaled problems. In many problems of type (1.1), the involved polynomials p, $g_1, \ldots, g_m$ are sparse. Waki, Kojima, Kim and Muramatsu constructed a sequence of SDP and SOS relaxations that exploits the sparsity of polynomial optimization problems [27]. This method showed strong numerical efforts in comparision to Lasserre's SDP relaxations. The convergence of the sparse SDP relaxations to the optimum of the original problem (1.1) was shown by Lasserre [12] and Kojima and Muramatsu [10] recently. In the following, we give an outline of the sparse SDP relaxation method.

Let the polynomial optimization problem be given as in (1.2),

$$\min_{x \in K} p(x),$$

where K is a compact semialgebraic set defined by the m inequality constraints $g_1 \geq 0, \ldots, g_m \geq 0$. We will construct a sequence of SDP relaxations to this polynomial optimization problem, which exploits the sparsity of it, in case only few of the n variables occur in some inequality constraint g_j or some monomial of the objective f together. Under a certain condition on the sparsity pattern of the problem, the optima of these SDP relaxations converge to the optimum of the polynomial optimization problem (1.2).

First, let $\{1, \ldots, n\}$ be the union $\cup_{k=1}^{q} I_k$ of subsets $I_k \subset \{1, \ldots, n\}$, such that every g_j, $j \in \{1, \ldots, m\}$ is only concerned with variables $\{x_i \mid i \in I_k\}$ for some k. And it is required the objective p can be written as $p = p_1 + \ldots + p_q$ where each p_k uses only variables $\{x_i \mid i \in I_k\}$. In order to obtain the convergence results in the later section, we impose the following condition:

Assumption 1: For all $k = 1, \ldots, q-1, \quad I_{k+1} \cap \bigcup_{j=1}^{k} I_j \subseteq I_s$ for some $s \leq k$. $\quad$ (5.17)

One way to construct these subsets $\{I_k\}$ is the procedure via the chordal extension of the correlative sparsity pattern graph by Waki et. al [27], which we will mention later. Consider a graph $G = (V, E)$, given by $V = \{1, \ldots, n\}$ and $E = \{\{i, j\} \mid \{i, j\} \subseteq I_k$ for some $k\}$. In context of graph theory, the property of Assumption 1 is called the **running intersection property**.

Note that 5.17 is always satsfied for $q = 2$. Since property (5.17) depends on the ordering, it can be satisfied possibly after some relabelling of the $\{I_k\}$. In order to tackle the sparse SDP relaxations we introduce some further definitions.

Definition 5.11 *Given a subset I of $\{1, \ldots, n\}$ we define the sets*

$$\begin{aligned} \mathcal{A}^I \quad &= \left\{ \alpha \in \mathbb{N}^n : \ \alpha_i = 0 \ \text{if } i \notin I \right\}, \\ \mathcal{A}_\omega^I \quad &= \left\{ \alpha \in \mathbb{N}^n : \ \alpha_i = 0 \ \text{if } i \notin I \ \text{and } \textstyle\sum_{i \in I} \alpha_i \leq \omega \right\}. \end{aligned}$$

*Then, we define $\mathbb{R}[x, \mathcal{G}] := \{f \in \mathbb{R}[x] : \ supp(f) \subseteq \mathcal{G}\}$. Also the **restricted moment matrix** $M_r(y, I)$ **and localizing matrices** $M_r(gy, I)$ are defined for $I \subseteq \{1, \ldots, n\}$, $r \in \mathbb{N}$ and $g \in \mathbb{R}[x]$. They are obtained from $M_r(y)$ and $M_r(gy)$ by retaining only those rows (and columns) $\alpha \in \mathbb{N}^n$ of $M_r(y)$ and $M_r(gy)$ with $supp(\alpha) \subseteq \mathcal{A}_r^I$. In doing so, $M_r(y, I)$ and $M_r(gy, I)$ can be interpreted as moment and localizing matrices with rows and columns indexed in the canonical basis $u(x, \mathcal{A}_r^I)$ of $\mathbb{R}[x, \mathcal{A}_r^I]$. Finally, we denote the set of sum of square polynomials in $\mathbb{R}[x, \mathcal{G}]$ as $\sum \mathbb{R}[x, \mathcal{G}]^2$. In analogy to Theorem 3.4, $\sum \mathbb{R}[x, \mathcal{G}]^2$ can be written as*

$$\sum \mathbb{R}[x, \mathcal{G}]^2 = \left\{ u(x, \mathcal{G})^T V u(x, \mathcal{G}) : V \succeq 0 \right\}.$$

Before proposing the sparse SDP relaxations we add two assumptions:

Assumption 2: Let $K \subseteq \mathbb{R}^n$ be a closed semialgebraic set. Then, there is $M > 0$ such that $\| x \|_\infty < M$ for all $x \in K$.

This assumption implies $\| x(I_k) \|_\infty^2 < n_k M^2$, $k = 1, \ldots, q$, where $x(I_k) := \{x_i \mid i \in I_k\}$, and therefore we add to K the q redundant quadratic constraints

$$g_{m+k}(x) := n_k M^2 - \| x(I_k) \| \geq 0, \ k = 1, \ldots, q,$$

and set $m' = m + q$, so that K is now defined by:

$$K := \left\{ x \in \mathbb{R}^n \mid g_j(x) \geq 0, \ j = 1, \ldots, m' \right\}. \tag{5.18}$$

Notice that $g_{m+k} \in \mathbb{R}\left[x, \mathcal{A}_2^{I_k}\right]$ for every $k = 1, \ldots, q$. With Assumption 2, K is a **compact semialgebraic set**. Assumption 2 is also needed to guarantee the quadratic module $M(K)$ is archimedian, the condition of Putinar's Positivstellensatz.

Assumption 3: Let $K \subseteq \mathbb{R}^n$ as in (5.18). The index set $J = \{1, \ldots, m'\}$ is partitioned into q disjoint sets J_k, $k = 1, \ldots, q$, and the collections $\{I_k\}$ and $\{J_k\}$ satisfy:

1. For every $j \in J_k$, $g_j \in \mathbb{R}\left[x, \mathcal{A}^{I_k}\right]$, that is, for every $j \in J_k$, the constraint $g_j(x) \geq 0$ is only concerned with the variables $x(I_k)$. Equivalently, viewing g_j as a polynomial in $\mathbb{R}[x]$, $g_{j\alpha} \neq 0 \Rightarrow \operatorname{supp}(\alpha) \in \mathcal{A}^{I_k}$.

2. The objective function $p \in \mathbb{R}[x]$ can be written

$$p = \sum_{k=1}^{q} p_k, \ \text{ with } p_k \in \mathbb{R}\left[x, \mathcal{A}^{I_k}\right], \ \ k = 1, \ldots, q.$$

   Equivalently, $f_\alpha \neq 0 \Rightarrow \operatorname{supp}(\alpha) \in \cup_{k=1}^{q} \mathcal{A}^{I_k}$.

Example 5.12 *with $n = 6$ and $m = 6$, let*

$$g_1(x) = x_1 x_2 - 1, \quad g_2(x) = x_1^2 + x_2 x_3 - 1, \quad g_3(x) = x_2 + x_3^2 x_4,$$

and

$$g_4(x) = x_3 + x_5, \quad g_5(x) = x_3 x_6, \quad g_6(x) = x_2 x_3.$$

Then we can construct $\{I_k\}$ and $\{J_k\}$ for $q = 4$ with

$$\begin{aligned} I_1 &= \{1, 2, 3\}, & I_2 &= \{2, 3, 4\}, & I_3 &= \{3, 5\}, & I_4 &= \{3, 6\}, \\ J_1 &= \{1, 2, 6\}, & J_2 &= \{3\}, & J_3 &= \{4\}, & J_4 &= \{5\}. \end{aligned}$$

It is easy to check, that Assumption 1 and 3 are satisfied. In case the $\{I_k\}$ satisfying Assumption 1 are not that easy to detect, we apply the procedure by Waki et.al. [27] via the chordal extension of the **correlative sparsity pattern (csp) graph** of problem (1.2). In case we apply this procedure, the I_k are derived as the maximal cliques C_k of the chordal extension of the csp graph G. The vertex set V of G is $V = \{1, \ldots, n\}$, and (i, j) is a edge of G if and only if x_i and x_j occur in some inequality constraint together or they occur in the same monomial of the objective. The csp graph G represents the sparsity structure of the polynomial optimization problem (1.2). We determine the maximal cliques C_l of the chordal extension of G since it is an NP-hard problem to determine the maximal cliques of an arbitrary graph.

Next, we will construct sparse SDP relaxations in analogy to the dense SDP relaxations

which we studied before. For each $j = 1, \ldots, m'$ write $\omega_j = \lceil \frac{\deg g_j}{2} \rceil$. Then, with $\omega \in \mathcal{N}$ consider the following semidefinite program

$$
\begin{aligned}
(P_\omega^{sp}) \quad \inf_y \quad & \sum_\alpha p_\alpha y_\alpha \\
\text{s.t.} \quad & M_\omega(y, I_k) \succcurlyeq 0, \quad k = 1, \ldots, q, \\
& M_{\omega - \omega_j}(g_j y, I_k) \succcurlyeq 0, \quad j \in J_k;\ k = 1, \ldots, q, \\
& y_0 = 1.
\end{aligned}
\tag{5.19}
$$

Program (5.19) is well defined under Assumption 3, and it is easy to see that it is an SDP relaxation of problem (1.2). Setting $I_k = C_k$ the maximal cliques of the chordal extended csp graph, the SDP relaxations (5.19) are stronger than the original relaxations proposed in [27] as we added the q redundant constraints $g_{m+k} \geq 0$ for $k = 1, \ldots, q$. There are symmetric matrices $\{B_\alpha^k\}$ and $\{C_\alpha^{jk}\}$ such that

$$
\begin{aligned}
M_\omega(y, I_k) &= \sum_{\alpha \in \mathbb{N}^n} y_\alpha B_\alpha^k, \quad k = 1, \ldots, q, \\
M_{\omega - \omega_j}(g_j y, I_k) &= \sum_{\alpha \in \mathbb{N}^n} y_\alpha C_\alpha^{jk}, \quad k = 1, \ldots, q,\ j \in J_k,
\end{aligned}
\tag{5.20}
$$

with $B_\alpha^k = 0$ and $C_\alpha^{jk} = 0$ whenever $\operatorname{supp}(\alpha) \notin \mathcal{A}^{I_k}$. Then we can rewrite (5.19) as

$$
\begin{aligned}
(P_\omega^{sp}) \quad \inf_y \quad & \sum_\alpha p_\alpha y_\alpha \\
\text{s.t.} \quad & \sum_{0 \neq \alpha \in \mathbb{N}^n} y_\alpha B_\alpha^k \succcurlyeq -B_0^k, \quad k = 1, \ldots, q, \\
& \sum_{0 \neq \alpha \in \mathbb{N}^n} y_\alpha C_\alpha^{jk} \succcurlyeq -C_0^{jk}, \quad j \in J_k;\ k = 1, \ldots, q,
\end{aligned}
\tag{5.21}
$$

and we derive the dual of this semidefinite program as

$$
\begin{aligned}
(D_\omega^{sp}) \quad \sup_{Y_k, Z_{jk}, \lambda} \quad & \lambda \\
& \sum_{k:\, \alpha \in \mathcal{A}^{I_k}} \left[\langle Y_k, B_\alpha^k \rangle + \sum_{j \in J_k} \langle Z_{jk}, C_\alpha^{jk} \rangle \right] + \lambda \delta_{\alpha 0} = p_\alpha \quad \forall\, \alpha \in \Gamma_\omega, \\
& Y_k, Z_{jk} \succcurlyeq 0, \qquad\qquad\qquad\qquad\qquad\qquad j \in J_k,\ k = 1, \ldots, q,
\end{aligned}
\tag{5.22}
$$

where $\Gamma_\omega := \left\{ \alpha \in \mathbb{N}^n : \alpha \in \bigcup_{k=1}^q \mathcal{A}^{I_k};\ \mid \alpha \mid \leq 2\omega \right\}$. Following Lasserre [12] and applying the characterization of sum of squares polynomials as semidefinite forms we can transform (D_ω^{sp}) to

$$
\begin{aligned}
(D_\omega^{sos}) \quad \sup_{t_k, t_{jk}, \lambda} \quad & \lambda \\
\text{s.t.} \quad & p - \lambda = \sum_{k=1}^q \left(t_k + \sum_{j \in J_k} t_{jk} g_j \right), \\
& t_k,\ t_{jk} \in \sum \mathbb{R}\left[x, \mathcal{A}_\omega^{I_k} \right]^2 \qquad j \in J_k,\ k = 1, \ldots, q,
\end{aligned}
\tag{5.23}
$$

We define the **generalized Lagrangian function** L of (1.2) as

$$
L(x, \phi) = p(x) - \sum_{k=1}^n \phi_k(x) g_k(x),
$$

where $\phi_k \in \sum \mathbb{R}\left[x, \mathcal{A}_{\omega - \omega_k}^{I_k} \right]^2$, as introduced by Kim et. al. in [9]. Moreover, if we neglect the p additional redundant contraints $g_{m+k} \geq 0$, we are able to transform (5.23) into

$$
\begin{aligned}
(D_\omega^{kkw}) \quad \sup \quad & \lambda \\
\text{s.t.} \quad & L(x, \phi) - \lambda \in \sum_{k=1}^q \sum \mathbb{R}\left[x, \mathcal{A}_\omega^{I_k} \right]^2, \\
& \phi \in \left\{ (\phi_1, \ldots, \phi_m) : \phi_k \in \sum \mathbb{R}\left[x, \mathcal{A}_{\omega - \omega_k}^{I_k} \right]^2,\ k = 1, \ldots, m \right\},
\end{aligned}
\tag{5.24}
$$

which is basically the sum of squares relaxation that was also derived by Waki, Kim, Kojima and Muramatsu in [27].

In order to understand the improved efficiency of the sparse SDP relaxations, let us compare the computational complexity of the dense relaxation $(P_\omega^{\mathrm{SDP}})$ and the sparse relaxation (P_ω^{sp}). The number of variables in (P_ω^{sp}) is bounded by $\sum_{k=1}^{q} \binom{n_k + 2\omega}{\omega}$. Supposed $n_k \approx \frac{n}{q}$ for all k, the number of variables is bounded by $O(q(\frac{n}{q})^{2\omega})$, a strong improvement compared with $O(n^{2\omega})$, the number of variables in $(P_\omega^{\mathrm{SDP}})$.

Also in (P_ω^{sp}) there are p LMI constraints of size $O((\frac{n}{q})^\omega)$ and $m + q$ LMI constraints of size $O((\frac{n}{q})^{\omega - \omega_{\max}})$, to be compared with a single LMI constraint of size $O(n^\omega)$ and m LMI constraints of size $O(n^{\omega - \omega_{\max}})$ in $(P_\omega^{\mathrm{SDP}})$.

5.4.1 Convergence

The convergence of the optima of the sparse SDP relaxations (P_ω^{sp}) was shown by Lasserre [12].

Theorem 5.13 *Let $p^\star$ denote the global minimum of (1.2) and let Assumption 1-3 hold. Then:*

(a) $\inf(P_\omega^{sp}) \uparrow p^\star$ *as $\omega \to \infty$.*

(b) *If K has nonempty interior, then strong duality holds and (D_ω^{sp}) solvable for sufficiently large ω, i.e., $\inf(P_\omega^{sp}) = \max(D_\omega^{sp})$.*

(c) *Let y^ω be a nearly optimal solution of (P_ω^{sp}), with e.g.*

$$\sum_\alpha p_\alpha y_\alpha \leq \inf(P_\omega^{sp}) + \frac{1}{\omega}, \quad \forall \omega \geq \omega_0,$$

and let $\hat{y}^\omega := \{y_\alpha^\omega : \mid \alpha \mid = 1\}$. If (1.2) has a unique global minimizer $x^\star \in K$, then $\hat{y}^\omega \to x^\star$ as $\omega \to \infty$.

Proof C.f. [12].

5.4.2 Sparse version of Putinar's Positivstellensatz

As a by product of Theorem 5.13, a sparse version of Putinar's Positivstellensatz was obtained by Lasserre in case K has nonempty interior [12]. A more general result, avoiding the assumption K has nonempty interior, was given by Kojima and Muramtsu [10]:

Theorem 5.14 *Let Assumption 1 hold. Furthermore, assume*

$$K_j := \left\{ x(I_j) \in \mathbb{R}^{|I_j|} \mid g_j(x(I_j)) \geq 0 \right\}$$

are nonempty and compact for $j \in \{1, \ldots, m\}$ and

$$K = \{x \in \mathbb{R}^n \mid x(I_j) \in K_j \ (j = 1, \ldots, m)\} = \{x \in \mathbb{R}^n \mid g_j(x(I_j)) \geq 0 \ (j = 1, \ldots, m)\}$$

is nonempty. Finally, assume

$$\forall j \in \{1, \ldots, m\} \ \exists p_j \in M(K_j), s.t. \ \left\{ x_{I_j} \mid p_j(x(I_j)) \right\} \ is \ compact$$

hold, i.e., $M(K_j)$ archimedian for all $j \in \{1, \ldots, m\}$. Then any $f \in \sum_{j=1}^m \mathbb{R}\,[x(I_j)]$ positive on K belongs to $\sum_{j=1}^m M(K_j)$.

Proof C.f. [10]. Another proof of the sparse version of Putinar's Positivstellensatz and the convergence result of the sparse SDP relaxation was also given by Grimm, Netzer and Schweighofer [7].

6 Equality constrained polynomial optimization problems

The original polynomial optimization problem (1.1) contains the inequality constrained case only. Certainly, an equality constraint $h(x) = 0$ can be included by adding the inequality constraints $h(x) \geq 0$ and $-h(x) \geq 0$. Nevertheless, more advanced methods to deal with semialgebraic sets containing equality constraints have been proposed.

6.1 Representation of nonnegative polynomials

As discussed in the first chapter, there is a representation of polynomials strictly positive on a compact semialgebraic set K as elements of $M(K)$. But an explicit representation of polynomials nonnegative on a closed or compact semialgebraic set has not been found yet. As mentioned before, the decision whether a general polynomial is nonnegative is an NP-hard problem. Nonetheless, Parrilo [19] found a simple construction for sum of square representations of polynomials nonnegative on finite sets described by polynomial equalities and inequalities.

Let a basic semialgebraic set be of the form

$$S = \{x \in \mathbb{R}^n, \ g_i(x) \geq 0, \ i = 0, \ldots, m, \ h_j(x) = 0, \ j = 1, \ldots, k\},$$

where $g_i, h_j \in \mathbb{R}[x_1, \ldots, x_n]$.

In order to construct Parrilo's representations the following notations are needed. The **complex variety** $V_{\mathbb{C}}$ of a sequence of equality constraints is defined as the set $V_{\mathbb{C}} = \{x \in \mathbb{C}^n \mid h_i(x) = 0 \ \forall i \in \{1, \ldots, k\}\}$. Given the ideal $I = I(h_1, \ldots, h_k)$, we define

$$\sqrt{I} := \{f \in \mathbb{C}[x] \mid f^m \in I \text{ for some } m \in \mathbb{N} \setminus \{0\}\}.$$

Obviously $I \subseteq \sqrt{I}$ holds for any ideal I. The ideal I is said to be **radical** if $I = \sqrt{I}$. The result by Parrilo is stated in the following theorem.

Theorem 6.1 *Let the ideal $I(h_1, \ldots, h_k)$ be radical and its complex variety $V_{\mathbb{C}}(I)$ be finite. If $p \in \mathbb{R}[x]$ is nonnegative over S, then there exists a representation*

$$p(x) = s_0(x) + \sum_i s_i(x) g_i(x) + \sum_j \lambda_j(x) h_j(x), \tag{6.1}$$

where $s_i \in \sum \mathbb{R}[x]^2$, $\lambda_i \in \mathbb{R}[x]$.

Proof Theorem 6.1 is a consequence of the constructions resulting from two algorithms. We will present the first algorithm, which considers the purely equality constraint case only. The generalization to the equality and inequality constrained case, the second algorithm can be found in [18].

Algorithm 1 Given $p(x)$, with $v_i \in (V_{\mathbb{C}} \cap \mathbb{R}^n) \Rightarrow p(v_i) \geq 0$, compute an affine representation certifying nonnegativity.

1. Find $\mid V_{\mathbb{C}} \mid$ polynomials $p_i(x)$, that satisfy $p_i(v_j) = \delta_{ij}$, where δ_{ij} is the Kronecker delta function. The polynomials p_i are essentially the "indicator function" of the point v_i, taking there the value one and vanishing at the remaining points. They can be easily found using Lagrange interpolation, or given a basis for the quotient ring, by solving a $\mid V_{\mathbb{C}} \mid \times \mid V_{\mathbb{C}} \mid$ system of linear equations.

2. For every point $v_i \in (V_{\mathbb{C}} \cap \mathbb{R}^n)$, or a pair of complex conjugate solutions $v_i, v_j \in \mathbb{C}^n$ ($v_j = v_i^\star$), define polynomials

$$q_i := \gamma p_i \qquad \text{or} \qquad q_i := \gamma p_i + \gamma^\star p_j,$$

respectively, where $\gamma = \sqrt{p(v_i)}$. Notice that $q_i \in \mathbb{R}[x]$: in the first case, because v_i is real and $\gamma \geq 0$, and in the second, as a consequence of the complex-conjugate symmetry. With these definitions, it holds that $p(x) = \sum_i q_i^2(x) \ \forall x \in V_{\mathbb{C}}$, where only one term per pair of complex conjugate roots appears in the sum. Since the ideal I is radical, it follows that

$$p \equiv \sum_i q_i^2 \mod I.$$

Notice that p is then a sum of squares in the quotient ring.

3. To put the expression in the standard form 6.1, choose a basis for the quotient, and reduce the q_i modulo the ideal, to obtain:

$$p(x) = \sum_i \hat{q}_i^2 + \lambda_i(x) h_i(x).$$

Example 6.2 *Consider the polynomial and constraints:*

$$p := x + y^2 - z^2 + 1, \qquad \begin{cases} h_1 := xy - z = 0 \\ h_2 := yz - x = 0 \\ h_3 := zx - y = 0 \end{cases}$$

The ideal $I(h_1, h_2, h_3)$ is radical, with the corresponding variety having five isolated real elements, namely

$$v_1 = (0,0,0), \ v_2 = (1,1,1), \ v_3 = (1,-1,-1), \ v_4 = (-1,1,-1), \ v_5 = (-1,-1,1).$$

We construct the interpolating polynomials p_i, already reduced to the basis of the quotient $\mathbb{R}[x] \mod I(h_1, h_2, h_3)$ and the corresponding functional values:

$$\begin{aligned}
p_1 &= 1 - z^2, & p(v_1) &= 1, \\
p_2 &= (z^2 + z + x + y)/4, & p(v_2) &= 2, \\
p_3 &= (z^2 - z + x - y)/4, & p(v_3) &= 2, \\
p_4 &= (z^2 - z - x + y)/4, & p(v_4) &= 0, \\
p_5 &= (z^2 + z - x - y)/4, & p(v_5) &= 0.
\end{aligned}$$

And then the representation

$$p = p_1^2 + 2p_2^2 + 2p_3^2 + h_1(5z^3 - 3z)/4 + h_2(x - 5xz^2 - 4)/4 + h_3(-3y - 2z - 5zx)/4$$

is obtained directly.

In opposite to the approaches in [11], [18], [24] or [27], this approach does not apply any of the convexity-based techniques, but is purely algebraic. It might occur that much more concise represenations than the one provided by Algorithm 1 exist.

6.2 Polynomial optimization via finite varieties

As mentioned, the equality inquality constrained case of a polynomial optimization problem is a special case of the inequality constrained one, as the constraint $h(x) = 0$ is equivalent to the constraints $h(x) \geq 0$ and $-h(x) \geq 0$. But also more advanced approaches which exploit the algebraic geometry of inequality constraints have been proposed. For instance, Laurent [14] proposed an approach to solve equality and inequality constrained polynomial optimization problems by new semidefinite representations of finite complex varieties. We will present her approach in the following.

Consider the problem

$$
\begin{aligned}
\inf \quad & p(x) \\
\text{s.t.} \quad & h_i(x) = 0, \quad i \in \{1, \ldots, k\}, \\
& g_j(x) \geq 0, \quad j \in \{1, \ldots, m\},
\end{aligned}
\tag{6.2}
$$

where $p, h_1, \ldots, h_k, g_1, \ldots, g_m \in \mathbb{R}[x_1, \ldots, x_n]$. Recall, the **complex variety** $V_{\mathbb{C}}(h_1, \ldots, h_k)$ is defined as

$$
\begin{aligned}
V_{\mathbb{C}}(h_1, \ldots, h_k) := V_{\mathbb{C}}(I(h_1, \ldots, h_k)) \quad &= \{x \in \mathbb{C}^n \mid h_1(x) = 0, \ldots, h_k(x) = 0\} \\
&= \{x \in \mathbb{C}^n \mid f(x) = 0 \, \forall f \in I(h_1, \ldots, h_k)\},
\end{aligned}
$$

the **real variety** $V_{\mathbb{R}}(I)$ as

$$
V_{\mathbb{R}}(I) := V_{\mathbb{C}}(I) \cap \mathbb{R}^n
$$

and the closed semialgebraic set K as

$$
K := V_{\mathbb{R}}(I(h_1, \ldots, h_k)) \cap \{x \mid g_1 \geq 0, \ldots, g_m \geq 0\}.
$$

Finally let I denote the ideal generated by $\{h_1, \ldots, h_k\}$ and $p^\star = \inf_{x \in K} p(x)$.

Laurent derives a characterization of polynomials nonnegative on (6.2)'s feasible set K under the two assumptions, that $V_{\mathbb{C}}(I(h_1, \ldots, h_k))$ is finite and that $I(h_1, \ldots, h_k)$ is a **radical ideal**, i.e.

$$
I = I(V_{\mathbb{C}}(h_1, \ldots, h_k)) := \{f \in \mathbb{R}[x] \mid f(x) = 0 \, \forall x \in V\}.
$$

Definition 6.3 *A set $\mathcal{B} = \{f_1, \ldots, f_N\}$ of polynomials forms a **basis for** $\mathbb{R}[x]/I$, if for every polynomial f, there exists a unique set of real numbers $\lambda_1^{(f)}, \ldots, \lambda_N^{(f)}$ such that $f - \sum_{i=1}^N \lambda_i^{(f)} f_i \in I$. When $\mathcal{B}$ contains only monomials, we call $\mathcal{B}$ a **monomial basis**. The polynomial $res_{\mathcal{B}}(f) = \sum_{i=1}^N \lambda_i^{(f)} f_i$ is called the **residue of f modulo** I w.r.t. the basis $\mathcal{B}$ and we set $\lambda^{(f)} := (\lambda_i^{(f)})_{i=1}^N \in \mathbb{R}^{|\mathcal{B}|}$. Moreover, for $v \in V_{\mathbb{C}}(I)$, define the vector $\xi_v^{\mathcal{B}} := (f_i(v))_{i=1}^N \in \mathbb{R}^{|\mathcal{B}|}$; thus $f(v) = (\lambda^{(f)})^T \xi_v^{\mathcal{B}}$.*

Following Lasserre's procedure of convexifying problem (1.1) as in chapter 5, and (6.2), respectively, it holds

$$
p^\star = \inf \left\{ p^T y \mid y \text{ has a representing measure } \mu \text{ supported by } K \right\}.
\tag{6.3}
$$

Under the assumption K is a finite set, every probability measure μ supported by K is atomic, i.e., μ can be written as $\mu = \sum_{v \in F} \lambda_v \delta_v$, where $\lambda_v \geq 0$, $\sum_{v \in K} \lambda_v = 1$, and δ_v is the Dirac measure at v. Then the moment of order α of μ is equal to $\sum_{v \in K} \lambda_v v^\alpha$. The next lemma is a consequence of the moment theory discussed in chapter 4.

Lemma 6.4 *(i) If $y \in \mathbb{R}^{\mathbb{Z}^n_+}$ has a representing measure μ, then the infinite moment matrix $M(y) \succcurlyeq 0$. Moreover, if $p \in \mathbb{R}[x]$ such that $M(y)p = 0$, then the support of μ is contained in the set of zeros of $p(x)$.*

(ii) Let $g \in \mathbb{R}[x]$ and $F := \{x \in \mathbb{R}^n \mid g(x) \geq 0\}$. If $y \in \mathbb{R}^{\mathbb{Z}^n_+}$ has a representing measure supported by F, then the infinite localizing matrix $M(gy) \succcurlyeq 0$.

Proof

(i) For $p \in \mathbb{R}[x]$,

$$p^T M(y)p = \sum_{\alpha,\beta} p_\alpha, p_\beta y_{\alpha+\beta} = \sum_{\alpha+\beta} p_\alpha p_\beta \int x^{\alpha+\beta} \mu(dx) = \int p(x)^2 \mu(dx) \geq 0.$$

(ii) For $p \in \mathbb{R}[x]$, $p^T M(gy)p = \int g(x)p(x)^2 \mu(dx) \geq 0$. $\square$

Define the following bound for $p^\star$:

$$\begin{aligned}
\mu^\star := \quad &\inf \quad p^T y \\
&\text{s.t.} \quad M(y) \succcurlyeq 0, \\
&\qquad\quad M(h_i y) = 0 \quad \forall 1 \leq i \leq k, \\
&\qquad\quad M(g_j y) \succcurlyeq 0 \quad \forall 1 \leq j \leq m, \\
&\qquad\quad y_0 = 1.
\end{aligned} \tag{6.4}$$

By Lemma 6.4, $\mu^\star \leq p^\star$. In fact equality $\mu^\star = p^\star$ holds with the following proposition.

Proposition 6.5 *Let K be the feasible set of (6.2) and assume $V_{\mathbb{C}}(I(h_1, \ldots, h_m))$ is finite. The following two assertions are equivalent for sequences $y \in \mathbb{R}^{\mathbb{Z}^n_+}$:*

(i) y has a representing measure supported by K.

(ii) $M(y) \succcurlyeq 0, M(h_i y) = 0 \ (i = 1, \ldots, m), \ M(g_j y) \succcurlyeq 0 \ (j = 1, \ldots, k)$.

Proof The proposition is a consequence of the general Theorem 4.7 by Schmüdgen. $\square$

For the following assume $V_{\mathbb{C}}(I)$ is finite. Thus $\mathbb{R}[x]/I$ is finite dimensional and $\mathcal{B} = (f_1, \ldots, f_N)$ denotes its basis. Since $h = 0 \ \forall h \in I(h_1, \ldots, h_k)$ on the feasible set K of (6.2), problem (6.2) remains unchanged if we replace the objective p by its residue $\sum_{i=1}^N \lambda_i^{(p)} f_i(x)$ modulo I w.r.t. $\mathcal{B}$. Thus,

$$\begin{aligned}
p^\star = \quad &\min \quad (\lambda^{(p)})^T \xi_v^{\mathcal{B}} \\
&\text{s.t.} \quad v \in K \\
= \quad &\min \quad y^T \lambda^{(f)} \\
&\text{s.t.} \quad y \in P_{\mathcal{B}}(K) := \mathrm{conv}(\xi_v^{\mathcal{B}} \mid v \in K)
\end{aligned} \tag{6.5}$$

We will derive a semidefinite representation for the polytope $P_{\mathcal{B}}(K)$, which can be understood as a finite analogue of the formulation (6.4). It does not require the explicit knowledge of the complex variety $V_{\mathbb{C}}(I)$ but only the knowledge of a basis $\mathcal{B}$ of $\mathbb{R}[x]/I$. Let $y = (y_1, \ldots, y_N) \in \mathbb{R}^{|\mathcal{B}|}$ be a vector, its **combinatorial moment matrix** $M_{\mathcal{B}}(y)$ is the $|\mathcal{B}| \times |\mathcal{B}|$-matrix indexed by $\mathcal{B}$, whose (f_i, f_j)th entry is $y^T \lambda^{(f_i f_j)} = \sum_{k=1}^N \lambda_k^{(f_i f_j)} y_k$ for $f_i f_j \in \mathcal{B}$. Given a polynomial $h \in \mathbb{R}[x]$, define $hy := M_{\mathcal{B}}(y)\lambda^{(h)} \in \mathbb{R}^{|\mathcal{B}|}$. Let U denote the $N \times |\mathbb{Z}^n_+|$-matrix whose rows are indexed by $\mathcal{B}$ (resp. by $\mathbb{Z}^n_+$) and whose $(i, \alpha)-$entry is

equal to $\lambda_i^{(x^\alpha)}$, i.e., the αth column of U contains the coordinates of the residue of x^α in the basis $\mathcal{B}$. For $h \in \mathbb{R}[x]$ with residue $\sum_{i=1}^N \lambda_i^{(h)} f_i$ obviously holds

$$\lambda^{(h)} = Uh. \tag{6.6}$$

Given a vector $y \in \mathbb{R}^{|\mathcal{B}|}$, define its **extension** as

$$\tilde{y} := U^T y \in \mathbb{R}^{\mathbb{Z}_+^n} \tag{6.7}$$

Hence, $\tilde{y}_\alpha$ can be interpreted as a linearization of the residue of x^α modulo I w.r.t. $\mathcal{B}$.

Lemma 6.6 *Let $y \in \mathbb{R}^{|\mathcal{B}|}$ and let $\tilde{y} \in \mathbb{R}^{\mathbb{Z}_+^n}$ be its extension.*

(i) *$M(\tilde{y}) = U^T M_\mathcal{B}(y) U$. Hence, when $\mathcal{B}$ is a monomial basis, $M_\mathcal{B}(y)$ is the principal submatrix of $M(\tilde{y})$ indexed by $\mathcal{B}$.*

(ii) *The extension of hy is equal to $h\tilde{y}$.*

(iii) *$I \subseteq \mathrm{Ker} M(\tilde{y})$.*

(iv) *For a polynomial h holds $h^T \tilde{y} = (\lambda^{(h)})^T y$.*

Proof C.f. [14]

Theorem 6.7 *The following assertions are equivalent for $y \in \mathbb{R}^{|\mathcal{B}|}$ and its extension $\tilde{y} \in \mathbb{R}^{\mathbb{Z}_+^n}$.*

(i) *The vector y belongs to the cone generated by the vectors $\xi_v^\mathcal{B} = (f_i(v))_{i=1}^N$ ($v \in K$).*

(ii) *$M_\mathcal{B}(y) \succcurlyeq 0$, $M_\mathcal{B}(g_j y) \succcurlyeq 0$ ($j = 1, \ldots, m$).*

(iii) *$M(\tilde{y}) \succcurlyeq 0$, $M(h_i \tilde{y}) = 0$ ($i = 1, \ldots, k$), $M(g_j \tilde{y}) \succcurlyeq 0$ ($j = 1, \ldots, m$).*

(iv) *The vector $\tilde{y}$ belongs to the cone generated by the vectors $\xi_v = (v^\alpha)_{\alpha \in \mathbb{Z}_+^n}$ ($v \in K$).*

Proof

(i)$\Rightarrow$(ii): Let $y := \xi_v^\mathcal{B}$ for a $v \in K$. Then, $M_\mathcal{B}(y) = yy^T \succcurlyeq 0$. Indeed, the (f_i, f_j)th entry of yy^T is $f_i(v) f_j(v)$, while the (f_i, f_j)th entry of $M_\mathcal{B}(y)$ is equal to $\sum_{k=1}^N \lambda_k^{(f_i, f_j)} f_k(v)$ and thus to $f_i(v) f_j(v)$, since $f_i f_j \equiv \sum_{k=1}^N \lambda_k^{(f_i f_j)} f_k$ modulo I. Moreover, $M_\mathcal{B}(h_j y) = g_j(v) yy^T \succcurlyeq 0$, for $1 \le j \le k$.

(ii)$\Rightarrow$(iii): As $M(\tilde{y}) = U^T M_\mathcal{B}(y) U$, it follows that $M(\tilde{y}) \succcurlyeq 0$. For $j = 1, \ldots, k$, $h_j \tilde{y} = M(\tilde{y}) h_j = 0$, since $I \subseteq \mathrm{Ker} M(\tilde{y})$ (by Lemma 6.6 (iii)). For $j = 1, \ldots, m$, $M(g_j \tilde{y}) = M(\tilde{g_j y}) = U^T M_\mathcal{B}(h_j y) U \succcurlyeq 0$ (by Lemma 6.6 (i),(ii)).

(iii)$\Rightarrow$(iv): Follows from Proposition 6.5.

(iv)$\Rightarrow$(i): Say, $\tilde{y} = \sum_{v \in K} a_v \xi_v$ with $a_v \ge 0$. Let $i = 1, \ldots, N$. As $\tilde{y} = U^T y$, $f_i^T \tilde{y} = f_i^T (U^T y) = (U f_i)^T y = y_i$. On the other hand, as $f_i^T \xi_v = f_i(v) = (\xi_v^\mathcal{B})_i$, $f_i^T \tilde{y} = \sum_{v \in K} a_v f_i(v)$ is the ith coordinate of $\sum_{v \in K} a_v \xi_v^\mathcal{B}$. Hence, $y = \sum_{v \in K} a_v \xi_v^\mathcal{B}$. $\square$

The proposed semidefinite representation of problem (6.2) is a direct consequence from Theorem 6.7.

Corollary 6.8 *Assume that $V_{\mathbb{C}}(I)$ is finite and let $\mathcal{B}$ be a monomial basis of $\mathbb{R}[x]/I$ containing the constant monomial 1. For convenience, the same symbol $\mathcal{B}$ denotes the set of exponents β for which $x^\beta \in \mathcal{B}$. Then, problem (6.2) is equivalent to*

$$
\begin{aligned}
\min \quad & r^T y \\
s.t. \quad & M_{\mathcal{B}}(y) \succcurlyeq 0, \\
& M_{\mathcal{B}}(g_j y) \succcurlyeq 0, \quad j \in \{1, \ldots, m\} \\
& y_0 = 1,
\end{aligned} \tag{6.8}
$$

where $r(x) = \sum_{\beta \in \mathcal{B}} r_\beta x^\beta$ is the residue of the polynomial p w.r.t. $\mathcal{B}$, and y_0 is the coordinate of y indexed by 1.

Proof Follows directly from Theorem 6.7 since, for $y = \sum_v a_v \xi_v^{\mathcal{B}}$,:

$$
\sum_v a_v = 1 \Leftrightarrow y_0 = 1. \qquad \square
$$

Although Corollary 6.8 was formulated for a monomial basis due to simplicity reasons, the formulation can be generalized to arbitrary bases for $\mathbb{R}[x]/I$. Next let us consider the dual of the semidefinite program (6.8). Setting $h_0(x) = 1$, $M_{\mathcal{B}}(h_i y) = \sum_{\beta \in \mathcal{B}} C_\beta^i y_\beta$ for $i = 0, \ldots, k$ and $M_{\mathcal{B}}(g_j y) = \sum_{\beta \in \mathcal{B}} C_\beta^{m+j} y_\beta$ for $j = 1, \ldots, m$, the dual semidefinite program to (6.8) reads:

$$
\begin{aligned}
\rho^\star := \quad \sup \quad & r_0 - \langle C_0^0, Z_0 \rangle - \sum_{j=k+1}^{m+k} \langle C_0^j, Z_j \rangle \\
s.t. \quad & \langle C_\beta^0, Z_0 \rangle + \sum_{j=k+1}^{m+k} \langle C_\beta^j, Z_j \rangle = r_\beta \; (\beta \in \mathcal{B} \setminus \{0\}) \\
& Z_0, Z_{k+1}, \ldots, Z_{m+k} \succcurlyeq 0.
\end{aligned} \tag{6.9}
$$

It is easy to verify that (6.9) is equivalent to

$$
\begin{aligned}
\rho^\star = \quad \sup \quad & \rho \\
s.t. \quad & r(x) - \rho = (\sum_{i_0} q_{0,i_0}^2) + \sum_{j=1}^m g_j(\sum_{i_j} q_{j,i_j}^2) + q \\
& q_{j,i_j} \in \mathbb{R}^{\mathcal{B}} \text{ and } q \in I
\end{aligned} \tag{6.10}
$$

and thus to the program

$$
\rho^\star = \sup \rho \text{ s.t. } f(x) - \rho \in M(K). \tag{6.11}
$$

With Theorem 3.11 holds $p^\star = \rho^\star$, i.e. there is no duality gap between (6.8) and its dual (6.9).

It is to emphasize that (6.8) is a semidefinite characterization of the polynomial optimization problem (6.2) whose minimum coincides with the minimum $p^\star$ of (6.2), whereas (5.8) is only a semidefinite approximation of problem (1.1) whose solutions converge to the minimum of (1.1) for relaxation order $\omega \to \infty$. On the other hand, from a complexity point of view, the finite semidefinite formulation (6.8) for problem (6.2) is not very useful, since it involves matrices of size $|\mathcal{B}| \times |\mathcal{B}|$, with $|\mathcal{B}| \geq |V_{\mathbb{R}}(I)|$ being at least as large as the size of the feasible set. Therefore, it might be appropriate to consider semidefinite approximations by restricting ourselves to some principal submatrix $M_{\mathcal{A}}(y)$ of $M_{\mathcal{B}}(y)$ instead of the full combinatorial moment matrix.

6.3 Semidefinite characterization and computation of real radical ideals

An approach to apply moment relaxation methods and semidefinite programming to compute the zero-dimensional real radical ideal $I(V_{\mathbb{R}}(I))$ and the real variety $V_{\mathbb{R}}(I)$ of an ideal $I \subseteq \mathbb{R}[x]$ was proposed by Lasserre, Laurent and Rostalski [13]. It can be understood as an extension to some of the concepts of the previous section 6.2. This approach is of particular interest, as it computes the points in the real variety $V_{\mathbb{R}}(I)$ without computing the complex variety $V_{\mathbb{C}}(I)$ beforehand. It is based on a semidefinite characterization of the ideal $I(V_{\mathbb{R}}(I))$, where $I \subset \mathbb{R}[x]$ is an ideal defined by a set of generators $h_1, \ldots, h_k \in \mathbb{R}[x]$ satisfying $\mid V_{\mathbb{R}}(I) \mid < \infty$. As we will see, all information needed to compute $I(V_{\mathbb{R}}(I))$ and $V_{\mathbb{R}}(I)$ is contained in the (infinite) moment matrix $M(y)$ (where $y \in \mathbb{R}^{\mathbb{N}^n}$), whose entries depend on the generating polynomials $h_1, \ldots, h_k$, and who is required to be positive semidefinite. In the following we will outline this approach und the proposed algorithm to compute $I(V_{\mathbb{R}}(I))$ and $V_{\mathbb{R}}(I)$.

In addition to the notations in sections 6.1 and 6.2 we introduce the real radical

$$\sqrt[\mathbb{R}]{I} := \left\{ p \in \mathbb{R}[x] \mid p^{2m} + \sum_j g_j^2 \in I \text{ for some } q_j \in \mathbb{R}[x], \, m \in \mathbb{N} \setminus \{0\} \right\}.$$

$I := I(h_1, \ldots, h_k) = \langle h_1, \ldots, h_k \rangle$ is said to be **real radical** if $I = \sqrt[\mathbb{R}]{I}$. Furthermore $\mathcal{B} \subset \{x^\alpha \mid \alpha \in \mathbb{N}^n\}$ is called an **order ideal** if $\mathcal{B}$ is stable under division, i.e., for all $a, b \in \{x^\alpha \mid \alpha \in \mathbb{N}^n\}$, $b \in \mathcal{B}$, $a \mid b$ implies $a \in \mathcal{B}$. With a **Real Nullstellensatz** [2] it holds $\sqrt[\mathbb{R}]{I} = I(V_{\mathbb{R}}(I))$ for any ideal $I \subseteq \mathbb{R}[x]$. Given the ideal $I \subseteq \mathbb{R}[x]$, consider the quotient space $\mathbb{R}[x]/I$. With $h \in \mathbb{R}[x]$ the **multiplication operator**

$$m_h : \; \mathbb{R}[x]/I \to \mathbb{R}[x]/I, \, [f] \mapsto [hf]$$

is well defined. Furthermore, the cardinality of $V_{\mathbb{C}}(I)$ and the dimension of the vector space $\mathbb{R}[x]/I$ are interlinked by the following theorem [26].

Theorem 6.9 *For an ideal I in $\mathbb{R}[x]$, $\mid V_{\mathbb{C}}(I) \mid < \infty \; \Leftrightarrow \; \dim \mathbb{R}[x]/I < \infty$. Moreover, $\mid V_{\mathbb{C}}(I) \mid \leq \dim \mathbb{R}[x]/I$, with equality if and only if I is radical.*

Now, assume $\mid V_{\mathbb{C}}(I) \mid < \infty$. Then let $\mathcal{B}$ be a basis of $\mathbb{R}[x]/I$, $N := \dim \mathbb{R}[x]/I$ and $h \in \mathbb{R}[x]$, and let χ_h denote the matrix of the multiplication operator m_h with respect to $\mathcal{B}$, i.e., writing $\text{res}_{\mathcal{B}}(hb_j) = \sum_{i=1}^{N} \lambda_i^{(hb_j)} b_i$, the jth column of χ_h is the vector $(\lambda_i^{(hb_j)})_{i=1}^{N}$. The following well known result [6] relates the points of the variety $V_{\mathbb{C}}(I)$ to the eigenvalues and eigenvectors of χ_h.

Theorem 6.10 *Let $h \in \mathbb{R}[x]$ and $v \in V_{\mathbb{C}}(I)$, set $\zeta_v^{\mathcal{B}} := (b_i(v))_{i=1}^{N}$. The set $\{h(v) \mid v \in V_{\mathbb{C}}(I)\}$ is the set of eigenvalues of χ_h and $\chi_h^T \zeta_v^{\mathcal{B}} = h(v) \zeta_v^{\mathcal{B}}$ for all $v \in V_{\mathbb{C}}(I)$.*

When the matrix χ_h is **non-derogatory** (i.e. all its eigenspaces are 1-dimensional) one can recover the points $v \in V_{\mathbb{C}}(I)$ from the eigenvectors of χ_h^T. Moreover, if I is radical, then $N = \mid V_{\mathbb{C}}(I) \mid$.

In order to outline the algorithm for computing $V_{\mathbb{R}}(I)$ and $I(V_{\mathbb{R}}(I))$, further notations and results are necessary. Given an order ideal $\mathcal{B} = \{b_1, \ldots, b_N\} \subseteq \{x^\alpha \mid \alpha \in \mathbb{N}^n\}$, the **border of $\mathcal{B}$** is the set

$$\partial\mathcal{B} := \left\{ x_i x^\beta \mid x^\beta \in \mathcal{B}, \, i = 1, \ldots, n \right\} \setminus \mathcal{B} = \{c_1, \ldots, c_H\}.$$

A set of polynomials $G = \{g_1, \ldots, g_H\}$ is called a $\mathcal{B}-$**border prebasis** if each g_j is of the form

$$g_j = c_j - \sum_{i=1}^{N} a_{ij} b_i \quad \text{for some } a_{ij} \in \mathbb{R}.$$

The set $G \subseteq I$ is said to be $\mathcal{B}-$**border basis** of I if $\mathcal{B}$ is linearly independent in $\mathbb{R}[x]/I$, i.e. if $\mathcal{B}$ is a linear basis of $\mathbb{R}[x]/I$; in that case G generates the ideal I. When G is a $\mathcal{B}-$border prebasis, one can mimic the construction of the multiplication matrices that were defined previously. Fix $k \in \{1, \ldots, n\}$. The **formal multiplication matrix** χ_k is the $N \times N$ matrix whose ith column is defined as follows. If $x_k b_i \in \mathcal{B}$, say, $x_k b_i = b_r$, then the ith column of χ_k is the standard unit vector e_r. Otherwise, $x_k b_i \in \partial\mathcal{B}$, say, $x_k b_i = c_j$, then the ith column of χ_k is the vector $(a_{ij})_{i=1}^{N}$. Moreover, the exploitation of the following theorem [6] is of central importance for the proposed algorithm.

Theorem 6.11 *Let $\mathcal{B} \subseteq \{x^\alpha \mid \alpha \in \mathbb{N}^n\}$ be an order ideal, let G be a $\mathcal{B}-$border prebasis with associated formal multiplication matrices $\chi_1, \ldots, \chi_n$, and let $\langle G \rangle$ be the ideal generated by G. Then, G is a border basis of $\langle G \rangle$ if and only if the matrices $\chi_1, \ldots, \chi_n$ commute pairwise. In that case, $\mathcal{B}$ is a linear basis of $\mathbb{R}[x]/\langle G \rangle$ and the matrix χ_k represents the multiplication operator m_{x_k} of $\mathbb{R}[x]/\langle G \rangle$ with respect to the basis $\mathcal{B}$.*

As already mentioned, the moment matrix $M(y^\mu)$ of a feasible moment vector $y^\mu \in \mathbb{R}^{\mathbb{N}^n}$ contains all information necessary to compute $I(V_\mathbb{R}(I))$. Given $h_1, \ldots, h_k \in \mathbb{R}[x]$, define

$$d_j := \lceil \frac{\deg h_j}{2} \rceil, \qquad d := \max_{j=1,\ldots,k} d_j$$

and

$$K_t^\mathbb{R} := \left\{ y \in \mathbb{R}^{\mathbb{N}_{2t}^n} \mid y_0 = 1, \ M_t(y) \succcurlyeq 0, M_{t-d_j}(h_j y) = 0 \ (j = 1, \ldots, k) \right\}.$$

Then, $K_t^\mathbb{R}$ is a convex set which contains the vectors $\zeta_{2t,v} := (v^\alpha)_{\alpha \in \mathbb{N}_{2t}^{2n}}$ for all $v \in V_\mathbb{R}(I)$. Two fundamental results of Curto and Fialkow [4] are the basis for the approach proposed by Lasserre, Laurent and Rostalski. Their theorems state, that it is sufficient to show rank conditions and positive semidefiniteness of the truncated moment matrices $M_t(y)$, in order to obtain a positive semidefinite infinite moment matrix $M(\tilde{y})$, in case the rank of $M(\tilde{y})$ is finite. A further notation is required before stating these theorems. Given a Hermitian matrix A and a principal submatrix B of A, A is said to be a **flat extension** of B if $\text{rank} A = \text{rank} B$; then $A \succcurlyeq 0 \Leftrightarrow B \succcurlyeq 0$.

Theorem 6.12 *If $M(y) \succcurlyeq 0$ and $\text{rank} M(y) < \infty$, then $y = \sum_{v \in W} \lambda_v \zeta_v$ for some finite set $W \subseteq \mathbb{R}^n$ and $\lambda_v > 0$, $\mid W \mid = \text{rank} M(y)$, and $\text{Ker} M(y) = I(W)$*

Theorem 6.13 *If $M_t(y) \succcurlyeq 0$ and $\text{rank} M_t(y) = \text{rank} M_{t-1}(y)$, then y can be extended in a unique way to $\tilde{y} \in \mathbb{R}^n$ such that $M(\tilde{y})$ is a flat extension of $M_t(y)$ (and thus $M(\tilde{y}) \succcurlyeq 0$).*

Proposition 6.14 *$\text{Ker} M(y)$ is an ideal in $\mathbb{R}[x]$, which is real radical if $M(y) \succcurlyeq 0$. Assume $M(y) \succcurlyeq 0$ and $\text{rank} M(y) = \text{rank} M_{t-1}(y)$ for some integer $t \geq 1$. Then, $\text{Ker} M(y) = \langle \text{Ker} M_t(y) \rangle$ and, for $\mathcal{B} \subseteq \{x^\alpha \mid \alpha \in \mathbb{N}^n\}$, $\mathcal{B}$ indexes a (maximum) nonsingular principal submatrix of $M(y)$ if and only if $\mathcal{B}$ is a (maximum) linearly independent subset of $\mathbb{R}[x]/\text{Ker} M^\mathbb{R}(y)$.*

Proof Sketch:
Use $\text{vec}(h)^T M(y) \text{vec}(pq) = \text{vec}(hq)^T M(y) \text{vec}(p)$ for $p, q, h \in \mathbb{R}[x]$ to show that $\text{Ker} M(y)$ is

an ideal. To show that $\mathrm{Ker}M(y)$ is real radical, it is sufficient to show that if $\sum_{i=1}^{k} p_i^2 \in \mathrm{Ker}M(y)$ for some $p_i \in \mathbb{R}[x]$, then $p_i \in \mathrm{Ker}M(y)$.

Assume $\mathrm{rank}M(y) = \mathrm{rank}M_{t-1}(y) =: r$ and set $J := \langle \mathrm{Ker}M_t(y) \rangle$. Obviously, $J \subseteq \mathrm{Ker}M(y)$; we show equality. For this, let $\mathcal{B} \subseteq \Lambda(t-1)$ index an $r \times r$ nonsingular principal submatrix of $M(y)$. We show that, for all $\alpha \in \mathbb{N}^n$, $x^\alpha \in \mathrm{Span}_{\mathbb{R}}(\mathcal{B}) + J$, using induction on $\mid \alpha \mid$. This holds for $\mid \alpha \mid \le t$ by the definition of $\mathcal{B}$. Assume $\mid \alpha \mid \ge t+1$ and write $x^\alpha = x_i x^\delta$. By the induction assumption, $x^\delta = \sum_{x^\beta \in \mathcal{B}} c_\beta x^\beta + q$ where $q \in J$, $c_\beta \in \mathbb{R}$. Thus, $x^\alpha = \sum_{x^\beta \in \mathcal{B}} c_\beta x_i x^\beta + x_i q$. Here, $x_i q \in J$ and $x_i x^\beta \in \mathrm{Span}_{\mathbb{R}}(\mathcal{B}) + J$ since $\deg(x_i x^\beta) \le t$, which implies $x^\alpha \in \mathrm{Span}_{\mathbb{R}}(\mathcal{B}) + J$. Thus we have shown that $\mathbb{R}[x] = \mathrm{Span}_{\mathbb{R}}(\mathcal{B}) + J$. As $\mathrm{Ker}M(y) \cap \mathrm{Span}_{\mathbb{R}}(\mathcal{B}) = \{0\}$, this implies easily that $\mathrm{Ker}M(y) = \langle \mathrm{Ker}M_t(y) \rangle$.

For $\mathcal{B} \subseteq \{x^\alpha \mid \alpha \in \mathbb{N}^n\}$, it is obvious that $\mathcal{B}$ indexes a nonsingular submatrix of $M(y)$ if and only if $\mathcal{B}$ is linearly independent in $\mathbb{R}[x]/\mathrm{Ker}M(y)$. The last statement of the lemma now follows since $\dim \mathbb{R}[x]/\mathrm{Ker}M(y) = r$ (as $\mathrm{Ker}M(y)$ is radical and using the identity $\mid V_{\mathbb{C}}(\mathrm{Ker}M(y)) \mid = \mathrm{rank}M(y)$ from Theorem 6.12). $\square$

Proposition 6.14 is the basis for the following key proposition.

Proposition 6.15 *Assume that $V_{\mathbb{R}}(I)$ is finite. If*

$$M(y) \succcurlyeq 0, \ M(h_j y) = 0 \ (j = 1, \dots, k)$$

then the kernel of $M(y)$ is a real radical ideal, $\mathrm{rank}M(y) \le \mid V_{\mathbb{R}}(I) \mid$ and $I(V_{\mathbb{R}}(I)) \subseteq \mathrm{Ker}M(y)$, with equality if and only if $M(y)$ has maximum rank, equal to $\mid V_{\mathbb{R}}(I) \mid$.

Proof By Proposition 6.14, $J := \mathrm{Ker}M(y)$ is a real radical ideal, since $M(y) \succcurlyeq 0$. As $0 = M(h_j y) = M(y)\mathrm{vec}(h_j)$ for all j, we have $I \subseteq J$, which implies that $V_{\mathbb{R}}(J) \subseteq V_{\mathbb{R}}(I)$ is finite. As J is real radical, we deduce that $V_{\mathbb{C}}(J) = V_{\mathbb{R}}(J) \subseteq \mathbb{R}^n$. Hence J is zero-dimensional and $I(V_{\mathbb{R}}(I)) \subseteq J$ since $V_{\mathbb{C}}(J) \subseteq V_{\mathbb{R}}(I)$. Set $r := \dim\mathbb{R}[x]/J = \mid V_{\mathbb{C}}(J) \mid \le \mid V_{\mathbb{R}}(I) \mid$. Let $\mathcal{B} \subset \{x^\alpha \mid \alpha \in \mathbb{N}^n\}$ be a linear basis of $\mathbb{R}[x]/J$, $\mid \mathcal{B} \mid = r$. Then the columns of $M(y)$ indexed by $\mathcal{B}$ form a basis of the column space of $M(y)$ and thus $\mathrm{rank}M(y) = r$. Moreover, $r = \mid V_{\mathbb{R}}(I) \mid$ if and only if $V_{\mathbb{C}}(J) = V_{\mathbb{R}}(I)$ which in turn is equivalent to $J = I(V_{\mathbb{R}}(I))$. Now, this maximum rank $\mid V_{\mathbb{R}}(I) \mid$ is reached by the sequence $y := y^\mu = \sum_{v \in V_{\mathbb{R}}(I)} \lambda_v \zeta_v > 0$ which indeed satisfies $M(y) \succcurlyeq 0$ and $M(h_j y) = 0 \ (j = 1, \dots, m)$. $\square$

Note that the cardinality of $V_{\mathbb{R}}(I)$ and $V_{\mathbb{C}}(I)$ may differ significantly as can be seen in the following two examples.

Example 6.16 *Let $I \subseteq \mathbb{R}[x]$ be generated by $h_i = x_i(x_i^2+1)$ $(i = 1, \dots, n)$. Then, $V_{\mathbb{R}}(I) = \{0\}$, $\mid V_{\mathbb{C}}(I) \mid = 3^n, d_i = 2$ for all i. Assume y satisfies $M_3(y) \succcurlyeq 0$ and $M_1(h_i y) = 0$ $(i = 1, \dots, n)$. Then $M_1(h_i y) = 0$ implies $y_{4e_i} = -y_{2e_i}$ and $M_3(y) \succcurlyeq 0$ implies $y_{2e_i}, y_{4e_i} \ge 0$ which in turn implies $y_\alpha = 0$ for all $\alpha \ne 0$ with $\mid \alpha \mid \le 5$ $(e_1, \dots, e_n$ denote the standard vectors in $\mathbb{R}^n$). Hence $\mathrm{rank}M_0(y) = \mathrm{rank}M_2(y) = 1$. In fact, $\mathrm{Ker}M_1(y)$ is spanned by $x_1, \dots, x_n$, the generators of $V_{\mathbb{R}}(I)$.*

Example 6.17 *Let $I \subseteq \mathbb{R}[x_1, x_2]$ be generated by $h = x_1^2 + x_2^2$. Then $V_{\mathbb{R}}(I) = \{0\}$ and $V_{\mathbb{C}}(I) = \{(x_1, x_2) \mid x_1 = \pm i x_2\}$ is infinite. Then $M_0(hy) = 0$ gives $y_{2e_1} + y_{2e_2} = 0$ which, together with $M_1(y) \succcurlyeq 0$, implies $y_\alpha = 0$ for $\alpha \ne 0$. Hence the maximum rank of $M_1(y)$ is equal to 1 and $\mathrm{Ker}M_1(y)$ is spanned by x_1, x_2, the generators of $V_{\mathbb{R}}(I)$.*

Proposition 6.15 guarantees that $I(V_{\mathbb{R}}(I)) = \mathrm{Ker}M(y)$ in case positive semidefiniteness conditions for $M(y)$ hold and $M(y)$ has maximum rank among all feasible moment matrices $M(z)$. Also, this proposition provides a basis for the real radical ideal $I(V_{\mathbb{R}}(I))$. Nevertheless, for computational issues it is not possible to handle an infinite matrix as $M(y)$ is.

Therefore, truncated moment matrices $M_t(y)$ are considered. The following proposition states, that it is sufficient to show rank conditions for those truncated moment matrices, in order to characterize the real radical $I(V_{\mathbb{R}}(I))$.

Proposition 6.18 *Let $t \geq d$ and $y \in K_t^{\mathbb{R}}$ for which $\operatorname{rank} M_t(y)$ is maximum. Assume that either*

$$\operatorname{rank} M_s(y) = \operatorname{rank} M_{s-1}(y) \qquad \text{for some } 2d \leq s \leq t$$

or

$$\operatorname{rank} M_s(y) = \operatorname{rank} M_{s-d}(y) \qquad \text{for some } d \leq s \leq t$$

holds. Then $I(V_{\mathbb{R}}(I)) = \operatorname{Ker} M(y) = \langle \operatorname{Ker} M_s(y) \rangle$ (and one can find $V_{\mathbb{R}}(I)$). Moreover, $\operatorname{rank} M_s(y) = \operatorname{rank} M(y) = |\, V_{\mathbb{R}}(I) \,|$.

Proof C.f. [13].

In feasible sets of polynomial optimization problems, as for instance 6.2, occur constraints of both types $h(x) = 0$ and $g(x) \geq 0$. Therefore, consider the closed semialgebraic set $L := \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m \geq 0\}$ and define the set

$$K_{t,L}^{\mathbb{R}} := K_t^{\mathbb{R}} \cap \left\{ y \mid M_{t-d_{k+j}}(g_j y) \succcurlyeq 0 \ (j = 1, \ldots, m) \right\}$$

for $t \geq d$. In analogy to Proposition 6.18, a characterization for $I(V_{\mathbb{R}}(I) \cap L)$ can be derived as well.

Proposition 6.19 *Let $t \geq d$ and $y \in K_{t,L}^{\mathbb{R}}$ for which $\operatorname{rank} M_t(y)$ is maximum. Assume for some $d \leq s \leq t$ holds*

$$\operatorname{rank} M_s(y) = \operatorname{rank} M_{s-d}(y).$$

Then $I(V_{\mathbb{R}}(I) \cap L) = \langle \operatorname{Ker} M_s(y) \rangle$.

To conclude the results that are necessary to derive the proposed algorithm, we show that one of the rank conditions in Proposition 6.18 and Proposition 6.19 is satisfied for t large enough, in case $V_{\mathbb{R}}(I)$ is finite.

Proposition 6.20 *Assume $|\, V_{\mathbb{R}}(I) \,| < \infty$.*

(i) *If $V_{\mathbb{R}}(I) = \emptyset$ then $K_{t,L}^{\mathbb{R}} = \emptyset$ for t large enough.*

(ii) *If $V_{\mathbb{R}}(I) \neq \emptyset$ then, for t large enough, there exists $d \leq s \leq t$ such that $\operatorname{rank} M_s(y) = \operatorname{rank} M_{s-d}(y)$ for all $y \in K_{t,L}^{\mathbb{R}}$.*

Proof C.f. [13].

Hence, it is possible to detect the existence of real solutions via the following criterion:

$$V_{\mathbb{R}}(I) = \emptyset \quad \Leftrightarrow \quad K_t^{\mathbb{R}} = \emptyset \text{ for some } t.$$

Proposition 6.20 remains valid under the weaker assumption $|\, V_{\mathbb{R}}(I) \cap L \,| < \infty$, if in the definition of the set $K_{t,L}^{\mathbb{R}}$, we add the constraints $M_{t-d_e}(p_e y) \succcurlyeq 0$ for $e \in \{0,1\}^k$, after setting $p_e := \prod_{i=1}^m g_i^{e_i}$.

Given the theorems and propositions 6.9 - 6.20 we are now able to outline the algorithm that computes $I(V_{\mathbb{R}}(I))$ and $V_{\mathbb{R}}(I)$. The algorithm consists of five main parts: For a given order $t \geq d$,

(i) Find an element $y \in K_t^{\mathbb{R}}$ maximizing the rank of $M_t(y)$.

(ii) Check the ranks of the principal submatrices of $M_t(y)$ and search for a submatrix $M_s(y)$ of $M_t(y)$ satisfying the rank conditions of Proposition 6.18 and 6.19, respectively.

(iii) Compute a basis for the column space of $M_s(y)$ and the quotient space $\mathbb{R}[x]/\langle \mathrm{Ker}\, M_s(y) \rangle$, for a suitable $1 \le s \le t$.

(iv) Compute the formal multiplication matrices χ_{x_i}.

(v) Construct a basis for the ideal $\langle \mathrm{Ker}\, M_s(y) \rangle$.

The first task (i) can be cast as a problem of finding a feasible solution of a semidefinite program, that has maximum rank. We consider the semidefinite program

$$p^\star = \min \; 1 \quad \text{s.t.} \; M_t(y) \succcurlyeq 0, \; M_{t-d_j}(h_j y) = 0 \; (j = 1, \ldots, m), \; y_0 = 1. \tag{6.12}$$

If we assume that 6.12 is strictly feasible, interior point algorithms construct sequences of points on the central path, which have the property of converging to the optimum solution of maximum rank.

For checking the rank condition (ii) and computing a basis for the column space of $M_s(y)$ and the quotient space $\mathbb{R}[x]/\langle \mathrm{Ker}\, M_s(y) \rangle$ (iii) a singular value decomposition of the submatrices $M_s(y)$ of $M_t(y)$ may be used. $M_s(y)$ can be decomposed as $M_s(y) = U \Sigma V^T$, where U and V are orthogonal matrices. The rank of $M_s(y)$ equals the number of positive entries of the diagonal matrix Σ. The columns of U form a basis $\mathcal{B} = \{b_1, \ldots, b_r\}$ of the column space of $M_s(y)$.

In order to compute the formal multiplication matrices χ_{x_i}, we exploit that

$$x_i b_j - \sum_{k=1}^{r} \lambda_k^{(x_i b_j)} b_k \in \mathrm{Ker}\, M_s(y),$$

for all $i = 1, \ldots, n$, $j = 1, \ldots, r$, as the $\mathrm{rank}\, M_s(y) = \mathrm{rank}\, M_{s-1}(y)$. Then the vector $(\lambda_k^{(x_i b_j)})_{k=1}^r$ is the jth column of the (formal) multiplication matrix χ_{x_i}. As shown in [13], χ_{x_i} can be computed as

$$\chi_{x_i} = M_\mathcal{B}^{-1} P_{x_i},$$

where $M_\mathcal{B}$ the principal submatrix of $M_s(y)$ indexed by $\mathcal{B}$ and P_{x_i} the submatrix of $M_s(y)$ whose rows are indexed by $\mathcal{B}$ and whose columns are indexed by the set $x_i \mathcal{B} := \{x_i b_j \mid j = 1, \ldots, r\}$. Theorem 6.10 implies the sets $\{v_i \mid v \in V_\mathbb{R}(I)\}$ are the sets of eigenvalues of the matrices χ_{x_i}. Thus, we are able to extract the real variety $V_\mathbb{R}(I)$.

To construct a linear basis for the ideal $\langle \mathrm{Ker}\, M_s(y) \rangle$ (iv) it is again possible to exploit the singular value decomposition of the truncated moment matrix $M_s(y) = U \Sigma V^T$. Indeed, the columns of the matrix V form an orthonormal basis of $\langle \mathrm{Ker}\, M_s(y) \rangle$. However, a drawback of this basis is that it is usually highly overdetermined and has a large cardinality, equal to $|\Lambda(s)| - \mathrm{rank}\, M_s(y)$. Other methods to construct a border or a Gröbner basis for $\langle \mathrm{Ker}\, M_s(y) \rangle$ are described in [13].

From the computational point of view solving the semidefinite program 6.12 is the most expensive part of this algorithm. As potential sparsity in the underlying generating polynomials $h_1, \ldots, h_m$ is not considered, the size of the resulting SDP (6.12) restricts the applicability of the algorithm to middle-scaled cases at the moment.

7   Open questions

Characterization of nonnegative and positive polynomials The formulation of the relaxation (5.8) for problem (1.2) by Lasserre and also Theorem 4.11 by Bertsimas and Popescu rely heavily on Putinar's Positivstellensatz, Theorem 3.11. Nonetheless, Theorem 3.11 does neither fully characterize the polynomials positive on the compact semialgebraic set K, since the polynomials in $M(K)$ are not necessarily positive on K, nor does it characterize the polynomials nonnegative on K, since they are not always contained in $M(K)$. In fact, Theorem 3.11 is the most recent in a series of partial characterizations of polynomials positive on a compact semialgebraic set; another partial characterization was presented in Theorem 3.9 by Schmüdgen. It is an open question, whether it will be possible to derive further (partial) characterizations of the polynomials positive and nonnegative on compact semialgebraic or closed semialgebraic sets. The complexity of those potential characterizations will determine whether it is useful to apply them to attempt the polynomial optimization problem (1.2). However, we are facing the fact that the decision whether a given polynomial is positive or nonnegative on $\mathbb{R}^n$ is NP-hard.

Convergence analysis of dense and sparse SDP relaxations As introduced Lasserre constructed a sequence of dense semidefinite relaxations (5.8) whose optima converge to the optimum of the original polynomial optimization problem (1.2) as the relaxation order ω tends to infinity. Although this result is appealing in theory, it was observed that this approach yields numerical difficulties for middle scaled problems already. This problem was partially relieved by constructing sequences of sparse SDP relaxations that exploit potential correlative sparsity in the polynomial optimization problem (1.2). Compared to the dense relaxations, the sparse SDP relaxations improved the efficiency drastically, in case of correlative sparsity in (1.2). Nevertheless, also the size of the sparse SDP relaxations increases rapidly in the relaxation order ω. Due to the limited capacity of present SDP solvers we are only able to process the SDP relaxations for small values of ω. Although in many examples the minimum of (1.2) is attained by the SDP relaxation for small choices of ω already, say $\omega \in \{\omega_{\max}, \ldots, \omega_{\max} + 3\}$, it is not possible to predict which optimization problem can be approximated suffciently close with a small relaxation order ω. The question remains if it is possible to characterize a class of polynomial optimization problems that can be solved approximated sufficiently close within a fixed number ω of SDP relaxations.

Extracting global minimizers Usually we are not only interested in the minimum of (1.2) but also its global minimizer. As mentioned in chapter 5.3, the convergence of the optimal solutions of Lasserre's relaxations (1.2) to a global optimizer of (1.2) can be guaranteed in case (1.2) has a unique solution. The case, where (1.2) has more than one optimal solution, is more difficult. Henrion and Lasserre established a procedure to detect all optimal solutions of (1.2), in case its optimal set $K^\star$ is finite. For the reason the computational effort of this procedure is very high, its applicability is restricted to small-scale problems only. In fact, the sparse SDP relaxation by Waki, Kim, Kojima and Muramatsu forces the polynomial optimization problem to have a unique solution, in order to preserve the improved computational effort. Techniques to determine all global minimizers of (1.2) efficiently remain in high demand.

References

[1] D. Bertsimas, I. Popescu, *Optimal inequalities in probability theory: A convex optimization approach*, SIAM J. Opt., 15 (2005), No. 3, pp. 780-804

[2] J. Bochnak, M. Coste, and M.-F. Roy, *Real Algebraic Geometry*, Springer-Verlag (1998)

[3] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge University Press (2004)

[4] R.E. Curto, and L.A. Fialkow, *Solution of the truncated complex moment problem for flat data*, Mem. Amer. Math. Soc. Vol. 119 (1996), No. 568

[5] R.E. Curto, and L.A. Fialkow, *The truncated complex K-moment problem*, Transactions of the American mathematical society, Vol. 352 (2000), 6, pp. 2825-2855

[6] A. Dickenstein, and I.Z. Emiris (eds.), *Solving Polynomial Equations: Foundations, Algorithms, and Applications*, Algorithms and Computation in Mathematics 14 (2005), Springer-Verlag

[7] D. Grimm, T. Netzer, and M. Schweighofer, *A note on the representation of positive polynomials with structured sparsity*, to appear in Archiv der Mathematik

[8] D. Henrion, J.B. Lasserre, *Detecting global optimality and extracting solutions in GloptiPoly*, Chapter in D. Henrion, A. Garulli, editors, *Positive polynomials in control. Lecture Notes in Control and Information science*, Springer Verlag (2005), Berlin

[9] S. Kim, M. Kojima, and H. Waki, *Generalized Lagrangian duals and sums of squares relaxations of sparse polynomial optimization problems*, SIAM Journal on Optimization, 15 (2005), pp. 697-719

[10] M. Kojima, and M. Muramatsu, *A Note on Sparse SOS and SDP Relaxations for Polynomial Optimization Problems over Symmetric Cones*, Research Report B-421, Dep. of Math. and Comp. Sciences, Tokyo Institute of Technology (2006)

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

[12] J.B. Lasserre, *Convergent SDP-Relaxations in Polynomial Optimization with Sparsity*, SIAM Journal on Optimization, 17 (2006), No. 3, pp. 822-843.

[13] J.B. Lasserre, M. Laurent, and P. Rostalski, *Semidefinite characterization and computation of zero-dimensional real radical ideals*, to appear in Foundations of Computational Mathematics

[14] M. Laurent, *Semidefinite representations for finite varieties*, Mathematical Programming, 109 (2007), pp. 1-26

[15] A. Marshall, and I. Olkin, *Multivariate Chebyshev inequalities*, Ann. Math. Statist., 31 (1960), pp. 1001-1014

[16] K.G. Murty and S.N. Kabadi, *Some NP-complete problems in quadratic and nonlinear programming*, Mathematical Programming, 39 (1987), pp. 117-129

[17] Y. Nesterov, *Squared functional systems and optimization problems*, in J.B.G. Frenk, C. Roos, T. Terlaky, and S. Zhang, editors, *High Performance Optimization*, pp. 405-440. Kluwer Academic Publishers (2000)

[18] P.A. Parrilo, *Semidefinite programming relaxations for semialgebraic problems*, Math. Programming, 96 (2003), pp. 293 - 320.

[19] P.A. Parrilo, *An explicit construction of distinguished representations of polynomials nonnegative over finite sets*, IfA Technical Report AUT02-02 (2002)

[20] V. Powers, B. Reznick, *A new bound for Polya's theorem with applications to polynomials positive on polyhedra*, J. Pure Appl. Algebra 164 (2001), No. 1-2, 221-229

[21] A. Prestel, and C.N. Delzell, *Positive Polynomials*, Springer-Verlag (2001)

[22] M. Putinar, *Positive Polynomials on Compact Semi-algebraic Sets*, Indiana Univ. Math. Journal 42 (1993), No. 3, pp. 969-984

[23] K. Schmüdgen, *The K-moment problem for compact semi-algebraic sets*, Math. Ann. 289 (1991), pp. 203-206

[24] M. Schweighofer, *Optimization of polynomials on compact semialgebraic sets*, SIAM J. Optimization 15 (2005), pp. 805-825

[25] N.Z. Shor, *Class of global minimum bounds of polynomial functions*, Cybernetics, 23 (1987), 6, 731-734

[26] H.J. Stetter, *Numerical Polynomial Algebra*, SIAM (2004), Philadelphia

[27] H. Waki, S. Kim, M. Kojima, and M. Muramatsu, *Sums of Squares and Semidefinite Program Relaxations for Polynomial Optimization Problems with structured sparsity*, SIAM Journal on Optimization, 17 (2006), pp. 218-242.

[28] H. Wolkowicz, R. Saigal, and L. Vandenberghe, *Handbook of semidefinite programming*, Kluwer's International Series (2000)