SUMS OF SQUARES, MOMENT MATRICES AND OPTIMIZATION OVER POLYNOMIALS

MONIQUE LAURENT*

Updated version: February 6, 2010

Abstract. We consider the problem of minimizing a polynomial over a semialgebraic set defined by polynomial equations and inequalities, which is NP-hard in general. Hierarchies of semidefinite relaxations have been proposed in the literature, involving positive semidefinite moment matrices and the dual theory of sums of squares of polynomials. We present these hierarchies of approximations and their main properties: asymptotic/finite convergence, optimality certificate, and extraction of global optimum solutions. We review the mathematical tools underlying these properties, in particular, some sums of squares representation results for positive polynomials, some results about moment matrices (in particular, of Curto and Fialkow), and the algebraic eigenvalue method for solving zero-dimensional systems of polynomial equations. We try whenever possible to provide detailed proofs and background.

Key words. positive polynomial, sum of squares of polynomials, moment problem, polynomial optimization, semidefinite programming

AMS(MOS) subject classifications. 13P10, 13J25, 13J30, 14P10, 15A99, 44A60, 90C22, 90C30

Contents

[*] Centrum Wiskunde & Informatica (CWI), Science Park 123, 1098 XG Amsterdam, Netherlands. Email: M.Laurent@cwi.nl.

Note. This is an updated version of the article *Sums of Squares, Moment Matrices and Polynomial Optimization*, published in *Emerging Applications of Algebraic Geometry*, Vol. 149 of IMA Volumes in Mathematics and its Applications, M. Putinar and S. Sullivant (eds.), Springer, pages 157-270, 2009.

1. Introduction.

This survey focuses on the following *polynomial optimization problem*: Given polynomials $p, g_1, \ldots, g_m \in \mathbb{R}[\mathbf{x}]$, find

$$p^{\min} := \inf_{x\in\mathbb{R}^n} p(x) \quad \text{subject to } g_1(x) \geq 0, \ldots, g_m(x) \geq 0, \tag{1.1}$$

the infimum of p over the basic closed semialgebraic set

$$K := \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}. \tag{1.2}$$

Here $\mathbb{R}[\mathbf{x}] = \mathbb{R}[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ denotes the ring of multivariate polynomials in the n-tuple of variables $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. This is a hard, in general nonconvex, optimization problem. The objective of this paper is to survey relaxation methods for this problem, that are based on relaxing positivity over K by sums of squares decompositions, and the dual theory of moments. The polynomial optimization problem arises in numerous applications. In the rest of the Introduction, we present several instances of this problem, discuss the scope of the paper, and give some preliminaries about polynomials and semidefinite programming.

1.1. The polynomial optimization problem. We introduce several instances of problem (1.1).

The unconstrained polynomial minimization problem. This is the problem

$$p^{\min} = \inf_{x \in \mathbb{R}^n} p(x), \tag{1.3}$$

of minimizing a polynomial p over the full space $K = \mathbb{R}^n$. We now mention several problems which can be cast as instances of the unconstrained polynomial minimization problem.

Testing matrix copositivity. An $n \times n$ symmetric matrix M is said to be *copositive* if $x^T M x \geq 0$ for all $x \in \mathbb{R}^n_+$; equivalently, M is copositive if and only if $p^{\min} = 0$ in (1.3) for the polynomial $p := \sum_{i,j=1}^n \mathbf{x}_i^2 \mathbf{x}_j^2 M_{ij}$. Testing whether a matrix is not copositive is an NP-complete problem [111].

The partition problem. The partition problem asks whether a given sequence $a_1, \ldots, a_n$ of positive integer numbers can be partitioned, i.e., whether $x^T a = 0$ for some $x \in \{\pm 1\}^n$. Equivalently, the sequence can be partitioned if $p^{\min} = 0$ in (1.3) for the polynomial $p := (\sum_{i=1}^n a_i \mathbf{x}_i)^2 + \sum_{i=1}^n (\mathbf{x}_i^2 - 1)^2$. The partition problem is an NP-complete problem [45].

The distance realization problem. Let $d = (d_{ij})_{ij \in E} \in \mathbb{R}^E$ be a given set of scalars (distances) where E is a given set of pairs ij with $1 \leq i < j \leq n$. Given an integer $k \geq 1$ one says that d is realizable in $\mathbb{R}^k$ if there exist vectors $v_1, \ldots, v_n \in \mathbb{R}^k$ such that $d_{ij} = \|v_i - v_j\|$ for all $ij \in E$. Equivalently, d is realizable in $\mathbb{R}^k$ if $p^{\min} = 0$ for the polynomial $p := \sum_{ij \in E} (d_{ij}^2 - \sum_{h=1}^k (\mathbf{x}_{ih} - \mathbf{x}_{jh})^2)^2$ in the variables $\mathbf{x}_{ih}$ ($i = 1, \ldots, n, h = 1, \ldots, k$). Checking whether d is realizable in $\mathbb{R}^k$ is an NP-complete problem, already for dimension $k = 1$ (Saxe [142]).

Note that the polynomials involved in the above three instances have degree 4. Hence the unconstrained polynomial minimization problem is a hard problem, already for degree 4 polynomials, while it is polynomial time solvable for degree 2 polynomials (cf. Section 3.2). The problem (1.1) also contains (0/1) linear programming.

(0/1) Linear programming. Given a matrix $A \in \mathbb{R}^{m \times n}$ and vectors $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, the linear programming problem can be formulated as

$$\min \ c^T x \ \text{ s.t. } Ax \leq b,$$

thus it is of the form (1.1) where the objective function and the constraints are all linear (degree at most 1) polynomials. As is well known it can be solved in polynomial time (cf. e.g. [146]). If we add the quadratic constraints $x_i^2 = x_i$ ($i = 1, \ldots, n$) we obtain the 0/1 linear programming problem:

$$\min \ c^T x \ \text{ s.t. } Ax \leq b, \ x_i^2 = x_i \ \forall i = 1, \ldots, n,$$

well known to be NP-hard.

The stable set problem. Given a graph $G = (V, E)$, a set $S \subseteq V$ is said to be stable if $ij \notin E$ for all $i, j \in S$. The stable set problem asks for the maximum cardinality $\alpha(G)$ of a stable set in G. Thus it can be formulated as

$$\alpha(G) = \max_{x\in\mathbb{R}^V} \sum_{i\in V} x_i \ \text{ s.t. } x_i + x_j \leq 1 \ (ij \in E),\ x_i^2 = x_i \ (i \in V) \tag{1.4}$$

$$= \max_{x\in\mathbb{R}^V} \sum_{i\in V} x_i \ \text{ s.t. } x_i x_j = 0 \ (ij \in E),\ x_i^2 - x_i = 0 \ (i \in V). \tag{1.5}$$

Alternatively, using the theorem of Motzkin-Straus [109], the stability number $\alpha(G)$ can be formulated via the program

$$\frac{1}{\alpha(G)} = \min\ x^T(I + A_G)x \ \text{ s.t. } \sum_{i\in V} x_i = 1,\ x_i \geq 0 \ (i \in V). \tag{1.6}$$

Using the characterization mentioned above for copositive matrices, one can derive the following further formulation for $\alpha(G)$

$$\alpha(G) = \inf\ t \ \text{ s.t. } t(I + A_G) - J \ \text{ is copositive,} \tag{1.7}$$

which was introduced in [37] and further studied e.g. in [51] and references therein. Here, J is the all ones matrix, and A_G is the adjacency matrix of G, defined as the $V \times V$ 0/1 symmetric matrix whose (i, j)th entry is 1 precisely when $i \neq j \in V$ and $ij \in E$. As computing $\alpha(G)$ is an NP-hard problem (see, e.g., [45]), we see that problem (1.1) is NP-hard already in the following two instances: the objective function is linear and the constraints are quadratic polynomials (cf. (1.5)), or the objective function is quadratic and the constraints are linear polynomials (cf. (1.6)). We will use the stable set problem and the following max-cut problem in Section 8.2 to illustrate the relaxations methods for polynomial problems in the 0/1 (or ±1) case.

The max-cut problem. Let $G = (V, E)$ be a graph and $w_{ij} \in \mathbb{R}$ $(ij \in E)$ be weights assigned to its edges. A *cut* in G is the set of edges $\{ij \in E \mid i \in S, j \in V \setminus S\}$ for some $S \subseteq V$ and its weight is the sum of the weights of its edges. The max-cut problem, which asks for a cut of maximum weight, is NP-hard [45]. Note that a cut can be encoded by $x \in \{\pm 1\}^V$ by assigning $x_i = 1$ to nodes $i \in S$ and $x_i = -1$ to nodes $i \in V \setminus S$ and the weight of the cut is encoded by the function $\sum_{ij\in E}(w_{ij}/2)(1 - x_i x_j)$. Therefore the max-cut problem can be formulated as the polynomial optimization problem

$$\mathrm{mc}(G, w) := \max \sum_{ij\in E} (w_{ij}/2)(1 - x_i x_j) \ \text{ s.t. } x_1^2 = 1, \ldots, x_n^2 = 1. \tag{1.8}$$

1.2. The scope of this paper. As the polynomial optimization problem (1.1) is NP-hard, several authors, in particular Lasserre [78, 79, 80], Nesterov [112], Parrilo [121, 122], Parrilo and Sturmfels [125], Shor [155, 156, 157, 158], have proposed to approximate the problem (1.1) by a hierarchy of convex (in fact, semidefinite) relaxations. Such relaxations can be constructed using representations of nonnegative polynomials as sums of squares of polynomials and the dual theory of moments. The paradigm underlying this approach is that, while testing whether a polynomial is nonnegative is a hard problem, testing whether a polynomial is a sum of squares of polynomials can be formulated as a semidefinite programming problem. Now, efficient algorithms exist for solving semidefinite programs (to any arbitrary precision). Thus approximations for the infimum of p over a semialgebraic set K can be computed efficiently. Moreover, under some assumptions on the set K, asymptotic (sometimes even finite) convergence to $p^{\min}$ can be proved and one may be able to compute global minimizers of p over K. For these tasks the interplay between positive polynomials and sums of squares of polynomials on the one hand, and the dual objects, moment sequences and matrices on the other hand, plays a significant role. The above is a rough sketch of the theme of this survey paper. Our objective is to introduce the main theoretical tools and results needed for proving the various properties of the approximation scheme, in particular about convergence and extraction of global minimizers. Whenever possible we try to provide detailed proofs and background.

The link between positive (nonnegative) polynomials and sums of squares of polynomials is a classic question which goes back to work of Hilbert at the end of the nineteenth century. As Hilbert himself already realized not every nonnegative polynomial can be written as a sum of squares of polynomials; he in fact characterized the cases when this happens (cf. Theorem 3.4). This was the motivation for Hilbert's 17th problem, posed in 1900 at the International Congress of Mathematicians in Paris, asking whether every nonnegative polynomial can be written as a sum of squares of *rational* functions. This was later in 1927 answered in the affirmative by E. Artin whose work lay the foundations for the field of *real algebraic geometry*. Some of the milestone results include the Real Nullstellensatz which is the real analogue of Hilbert's Nullstellensatz for the complex field, the Positivstellensatz and its refinements by Schmüdgen and by Putinar, which are most relevant to our optimization problem. We will present a brief exposition on this topic in Section 3.

The study of positive polynomials is intimately linked to the theory of moments, via the following duality relation: A sequence $y \in \mathbb{R}^{\mathbb{N}^n}$ is the sequence of moments of a nonnegative measure μ on $\mathbb{R}^n$ (i.e. $y_\alpha = \int x^\alpha \mu(dx)$ $\forall \alpha \in \mathbb{N}^n$) if and only if $y^T p := \sum_\alpha y_\alpha p_\alpha \geq 0$ for any polynomial $p = \sum_\alpha p_\alpha \mathbf{x}^\alpha \in \mathbb{R}[\mathbf{x}]$ which is nonnegative over $\mathbb{R}^n$. Characterizing moment sequences is a classical problem, relevant to operator theory and several other areas in mathematics (see e.g. [1, 77] and references therein).

Indeed, sequences of moments of nonnegative measures correspond to positive linear functionals on $\mathbb{R}[\mathbf{x}]$; moreover, the linear functionals that are positive on the cone of sums of squares correspond to the sequences y whose moment matrix $M(y) := (y_{\alpha+\beta})_{\alpha,\beta\in\mathbb{N}^n}$ is positive semidefinite. Curto and Fialkow have accomplished a systematic study of the *truncated* moment problem, dealing with sequences of moments up to a given order. We will discuss some of their results that are most relevant to polynomial optimization in Section 5 and refer to [28, 29, 30, 31, 41] and further references therein for detailed information.

Our goal in this survey is to provide a tutorial on the real algebraic tools and the results from moment theory needed to understand their application to polynomial optimization, mostly on an elementary level to make the topic accessible to non-specialists. We obviously do not pretend to offer a comprehensive treatment of these areas for which excellent accounts can be found in the literature and we apologize for all omissions and imprecisions. For a more advanced exposition on positivity and sums of squares and links to the moment problem, we refer in particular to the article by Scheiderer [144], to the survey article by Helton and Putinar [58], and to the monographs by Prestel and Delzell [133] and by Marshall [103, 106].

1.3. Preliminaries on polynomials and semidefinite programs.

We introduce here some notation and preliminaries about polynomials, matrices and semidefinite programs. We will introduce further notation and preliminaries later on in the text when needed.

1.3.1. Polynomials. Throughout, $\mathbb{N}$ denotes the set of nonnegative integers and we set $\mathbb{N}^n_t := \{\alpha \in \mathbb{N}^n \mid |\alpha| := \sum_{i=1}^n \alpha_i \leq t\}$ for $t \in \mathbb{N}$. $\mathbb{R}[\mathbf{x}_1, \dots, \mathbf{x}_n]$ denotes the ring of multivariate polynomials in n variables, often abbreviated as $\mathbb{R}[\mathbf{x}]$ where $\mathbf{x}$ stands for the n-tuple $(\mathbf{x}_1, \dots, \mathbf{x}_n)$. Throughout we use the boldfaced letters $\mathbf{x}_i, \mathbf{x}, \mathbf{y}, \mathbf{z}$, etc., to denote *variables*, while the letters $x_i, x, y, z, \dots$ stand for real valued scalars or vectors. For $\alpha \in \mathbb{N}^n$, $\mathbf{x}^\alpha$ denotes the monomial $\mathbf{x}_1^{\alpha_1} \cdots \mathbf{x}_n^{\alpha_n}$ whose degree is $|\alpha| := \sum_{i=1}^n \alpha_i$. $\mathbb{T}^n := \{\mathbf{x}^\alpha \mid \alpha \in \mathbb{N}^n\}$ is the set of all monomials and, for $t \in \mathbb{N}$, $\mathbb{T}^n_t := \{\mathbf{x}^\alpha \mid \alpha \in \mathbb{N}^n_t\}$ is the set of monomials of degree $\leq t$. Consider a polynomial $p \in \mathbb{R}[\mathbf{x}]$, $p = \sum_{\alpha\in\mathbb{N}^n} p_\alpha \mathbf{x}^\alpha$, where there are only finitely many nonzero p_α's. When $p_\alpha \neq 0$, $p_\alpha \mathbf{x}^\alpha$ is called a term of p. The degree of p is $\deg(p) := \max(t \mid p_\alpha \neq 0 \text{ for some } \alpha \in \mathbb{N}^n_t)$ and throughout we set

$$d_p := \lceil \deg(p)/2 \rceil \quad \text{for } p \in \mathbb{R}[\mathbf{x}]. \tag{1.9}$$

For the set $K = \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \dots, g_m(x) \geq 0\}$ from (1.2), we set

$$d_K := \max(d_{g_1}, \dots, d_{g_m}). \tag{1.10}$$

We let $\mathbb{R}[\mathbf{x}]_t$ denote the set of polynomials of degree $\leq t$.

A polynomial $p \in \mathbb{R}[\mathbf{x}]$ is said to be homogeneous (or a form) if all its terms have the same degree. For a polynomial $p \in \mathbb{R}[\mathbf{x}]$ of degree d,

$p = \sum_{|\alpha|\le d} p_\alpha \mathbf{x}^\alpha$, its homogenization is the polynomial $\tilde{p} \in \mathbb{R}[\mathbf{x}, \mathbf{x}_{n+1}]$ defined by $\tilde{p} := \sum_{|\alpha|\le d} p_\alpha \mathbf{x}^\alpha \mathbf{x}_{n+1}^{d-|\alpha|}$.

For a polynomial $p \in \mathbb{R}[\mathbf{x}]$, $p = \sum_\alpha p_\alpha \mathbf{x}^\alpha$, $\mathrm{vec}(p) := (p_\alpha)_{\alpha\in\mathbb{N}^n}$ denotes its sequence of coefficients in the monomial basis of $\mathbb{R}[\mathbf{x}]$; thus $\mathrm{vec}(p) \in \mathbb{R}^\infty$, the subspace of $\mathbb{R}^{\mathbb{N}^n}$ consisting of the sequences with finitely many nonzero coordinates. Throughout the paper we often identify a polynomial p with its coordinate sequence $\mathrm{vec}(p)$ and, for the sake of compactness in the notation, we often use the letter p instead of $\mathrm{vec}(p)$; that is, we use the same letter p to denote the polynomial $p \in \mathbb{R}[\mathbf{x}]$ and its sequence of coefficients $(p_\alpha)_\alpha$. We will often deal with matrices indexed by $\mathbb{N}^n$ or $\mathbb{N}^n_t$. If M is such a matrix, indexed say by $\mathbb{N}^n$, and $f, g \in \mathbb{R}[\mathbf{x}]$, the notation $f^T M g$ stands for $\mathrm{vec}(f)^T M \mathrm{vec}(g) = \sum_{\alpha,\beta} f_\alpha g_\beta M_{\alpha,\beta}$. In particular, we say that a polynomial f lies in the kernel of M if $Mf := M\mathrm{vec}(f) = 0$, and $\mathrm{Ker}\, M$ can thus be seen as a subset of $\mathbb{R}[\mathbf{x}]$. When $\deg(p) \le t$, $\mathrm{vec}(p)$ can also be seen as a vector of $\mathbb{R}^{\mathbb{N}^n_t}$, as $p_\alpha = 0$ whenever $|\alpha| \ge t+1$.

For a subset $A \subseteq \mathbb{R}^n$, $\mathrm{Span}_{\mathbb{R}}(A) := \{\sum_{j=1}^m \lambda_j a_j \mid a_j \in A, \lambda_j \in \mathbb{R}\}$ denotes the linear span of A, and $\mathrm{conv}(A) := \{\sum_{j=1}^m \lambda_j a_j \mid a_j \in A, \lambda_j \in \mathbb{R}_+, \sum_j \lambda_j = 1\}$ denotes the convex hull of A. Throughout $e_1, \ldots, e_n$ denote the standard unit vectors in $\mathbb{R}^n$, i.e. $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 at the ith position. Moreover $\overline{z}$ denotes the complex conjugate of $z \in \mathbb{C}$.

1.3.2. Positive semidefinite matrices. For an $n \times n$ real symmetric matrix M, the notation $M \succeq 0$ means that M is positive semidefinite, i.e. $x^T M x \ge 0$ for all $x \in \mathbb{R}^n$. Here are several further equivalent characterizations: $M \succeq 0$ if and only if any of the equivalent properties (1)-(3) holds.

(1) $M = VV^T$ for some $V \in \mathbb{R}^{n\times n}$; such a decomposition is sometimes known as a Gram decomposition of M. Here V can be chosen in $\mathbb{R}^{n\times r}$ where $r = \mathrm{rank}\, M$.

(2) $M = (v_i^T v_j)_{i,j=1}^n$ for some vectors $v_1, \ldots, v_n \in \mathbb{R}^n$. Here the v_i's may be chosen in $\mathbb{R}^r$ where $r = \mathrm{rank}\, M$.

(3) All eigenvalues of M are nonnegative.

The notation $M \succ 0$ means that M is positive definite, i.e. $M \succeq 0$ and $\mathrm{rank}\, M = n$ (equivalently, all eigenvalues are positive). When M is an infinite matrix, the notation $M \succeq 0$ means that every finite principal submatrix of M is positive semidefinite. Sym_n denotes the set of symmetric $n\times n$ matrices and PSD_n the subset of positive semidefinite matrices; PSD_n is a convex cone in Sym_n. $\mathbb{R}^{n\times n}$ is endowed with the usual inner product

$$\langle A, B\rangle = \mathrm{Tr}(A^T B) = \sum_{i,j=1}^n a_{ij} b_{ij}$$

for two matrices $A = (a_{ij}), B = (b_{ij}) \in \mathbb{R}^{n\times n}$. As is well known, the cone PSD_n is self-dual, since PSD_n coincides with its dual cone $(\mathrm{PSD}_n)^* := \{A \in \mathrm{Sym}_n \mid \langle A, B\rangle \ge 0 \ \forall B \in \mathrm{PSD}_n\}$.

1.3.3. Flat extensions of matrices. The following notion of *flat extension* of a matrix will play a central role in the study of moment matrices with finite atomic measures, in particular, in Section 5.

Definition 1.1. *Let X be a symmetric matrix with block form*

$$X = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}. \tag{1.11}$$

One says that X is a flat extension *of A if* $\operatorname{rank} X = \operatorname{rank} A$ *or, equivalently, if $B = AW$ and $C = B^T W = W^T AW$ for some matrix W. Obviously, if X is a flat extension of A, then $X \succeq 0 \Longleftrightarrow A \succeq 0$.*

We recall for further reference the following basic properties for positive semidefinite matrices. Recall first that, for $M \in \mathrm{PSD}_n$ and $x \in \mathbb{R}^n$, $x \in \operatorname{Ker} M$ (i.e. $Mx = 0$) $\Longleftrightarrow x^T Mx = 0$.

Lemma 1.2. *Let X be a symmetric matrix with block form (1.11), where A is $p \times p$ and B is $p \times q$.*

(i) *If $X \succeq 0$ or if* $\operatorname{rank} X = \operatorname{rank} A$*, then* $x \in \operatorname{Ker} A \Longrightarrow \begin{pmatrix} x \\ 0 \end{pmatrix} \in \operatorname{Ker} X$.

(ii) *If* $\operatorname{rank} X = \operatorname{rank} A$*, then* $\operatorname{Ker} X = \operatorname{Ker}(A\ B)$.

(iii) *If $X \succeq 0$, A is nonsingular if and only if $\begin{pmatrix} A \\ B^T \end{pmatrix}$ is nonsingular.*

(iv) *If $X \succeq 0$, then each column b of B belongs to the range $\mathcal{R}(A)$ of A, where $\mathcal{R}(A) := \{Au \mid u \in \mathbb{R}^p\}$.*

Proof. (i) $Ax = 0 \Longrightarrow 0 = x^T Ax = \begin{pmatrix} x^T & 0 \end{pmatrix} X \begin{pmatrix} x \\ 0 \end{pmatrix}$, which implies $X \begin{pmatrix} x \\ 0 \end{pmatrix} = 0$ if $X \succeq 0$. If $\operatorname{rank} X = \operatorname{rank} A$, then $B = AW$ for some matrix W and thus $B^T x = 0$, giving $X \begin{pmatrix} x \\ 0 \end{pmatrix} = 0$.
(ii) Obviously, $\operatorname{rank} X \geq \operatorname{rank}(A\ B) \geq \operatorname{rank} A$. If $\operatorname{rank} X = \operatorname{rank} A$, equality holds throughout, which implies $\operatorname{Ker} X = \operatorname{Ker}(A\ B)$.
(iii) follows directly from (i).
(iv) As $A \succeq 0$, $\mathcal{R}(A) = (\operatorname{Ker} A)^\perp$; hence it suffices to show $b \in (\operatorname{Ker} A)^\perp$, which follows easily using (i). □

1.3.4. Semidefinite programs. Consider the program

$$p^* := \sup_{X \in \mathrm{Sym}_n} \langle C, X \rangle \text{ s.t. } X \succeq 0,\ \langle A_j, X \rangle = b_j \ (j = 1, \ldots, m) \tag{1.12}$$

in the matrix variable X, where we are given $C, A_1, \ldots, A_m \in \mathrm{Sym}_n$ and $b \in \mathbb{R}^m$. This is the standard (primal) form of a semidefinite program; its dual semidefinite program reads:

$$d^* := \inf_{y \in \mathbb{R}^m} b^T y \text{ s.t. } \sum_{j=1}^m y_j A_j - C \succeq 0 \tag{1.13}$$

in the variable $y \in \mathbb{R}^m$. Obviously,

$$p^* \leq d^*, \tag{1.14}$$

known as *weak duality.* Indeed, if X is feasible for (1.12) and y is feasible for (1.13), then $0 \leq \langle X, \sum_{j=1}^m y_j A_j - C \rangle = b^T y - \langle C, X \rangle$. One crucial issue in duality theory is to identify sufficient conditions that ensure equality in (1.14), i.e. a *zero duality gap*, in which case one speaks of *strong duality.* We say that (1.12) is *strictly feasible* when there exists $X \succ 0$ which is feasible for (1.12); analogously (1.13) is strictly feasible when there exists y feasible for (1.13) with $\sum_{j=1}^m y_j A_j - C \succ 0$.

THEOREM 1.3. *If the primal program (1.12) is strictly feasible and its dual (1.13) is feasible, then $p^* = d^*$ and (1.13) attains its supremum. Analogously, if (1.13) is strictly feasible and (1.12) is feasible, then $p^* = d^*$ and (1.12) attains its infimum.*

Semidefinite programs are convex programs. As one can test in polynomial time whether a given rational matrix is positive semidefinite (using e.g. Gaussian elimination), semidefinite programs can be solved in polynomial time to any fixed precision using the ellipsoid method (cf. [50]). Algorithms based on the ellipsoid method are however not practical since their running time is prohibitively high. Interior-point methods turn out to be the method of choice for solving semidefinite programs in practice; they can find an approximate solution (to any given precision) in polynomially many iterations and their running time is efficient in practice for medium sized problems. There is a vast literature devoted to semidefinite programming and interior-point algorithms; cf. e.g. [113], [136], [164], [167], [176].

We will use (later in Section 6.6) the following geometric property of semidefinite programs. We formulate the property for the program (1.12), but the analogous property holds for (1.13) as well.

LEMMA 1.4. *Let $\mathcal{R} := \{X \in \text{PSD}_n \mid \langle A_j, X \rangle = b_j \ (j = 1, \ldots, m)\}$ denote the feasible region of the semidefinite program (1.12). If $X^* \in \mathcal{R}$ has maximum rank, i.e.* $\operatorname{rank} X^* = \max_{X \in \mathcal{R}} \operatorname{rank} X$, *then* $\operatorname{Ker} X^* \subseteq \operatorname{Ker} X$ *for all $X \in \mathcal{R}$. In particular, if X^* is an optimum solution to (1.12) for which* $\operatorname{rank} X^*$ *is maximum, then* $\operatorname{Ker} X^* \subseteq \operatorname{Ker} X$ *for any other optimum solution X.*

Proof. Let $X^* \in \mathcal{R}$ for which $\operatorname{rank} X^*$ is maximum and let $X \in \mathcal{R}$. Then $X' := \frac{1}{2}(X^* + X) \in \mathcal{R}$, with $\operatorname{Ker} X' = \operatorname{Ker} X^* \cap \operatorname{Ker} X \subseteq \operatorname{Ker} X^*$. Thus equality $\operatorname{Ker} X' = \operatorname{Ker} X^*$ holds by the maximality assumption on $\operatorname{rank} X^*$, which implies $\operatorname{Ker} X^* \subseteq \operatorname{Ker} X$. The last statement follows simply by adding the constraint $\langle C, X \rangle = p^*$ to the description of the set $\mathcal{R}$. □

Geometrically, what the above lemma says is that the maximum rank matrices in $\mathcal{R}$ correspond to the matrices lying in the relative interior of the convex set $\mathcal{R}$. And the maximum rank optimum solutions to the program

(1.12) are those lying in the relative interior of the optimum face defined by the equation $\langle C, X\rangle = p^*$, of the feasible region $\mathcal{R}$. As a matter of fact primal-dual interior-point algorithms that follow the so-called central path to solve a semidefinite program return a solution lying in the relative interior of the optimum face (cf. [176] for details). Thus (under certain conditions) it is easy to return an optimum solution of maximum rank; this feature will be useful for the extraction of global minimizers to polynomial optimization problems (cf. Section 6.6). In contrast it is *hard* to find optimum solutions of *minimum* rank. Indeed it is easy to formulate hard problems as semidefinite programs with a rank condition. For instance, given a sequence $a \in \mathbb{N}^n$, the program

$$p^* := \min\ \langle aa^T, X\rangle \ \text{ s.t. } X \succeq 0,\ X_{ii} = 1\ (i = 1, \ldots, n), \operatorname{rank} X = 1$$

solves the partition problem introduced in Section 1.1. Indeed any $X \succeq 0$ with diagonal entries all equal to 1 and with rank 1 is of the form $X = xx^T$ for some $x \in \{\pm 1\}^n$. Therefore, the sequence $a = (a_1, \ldots, a_n)$ can be partitioned precisely when $p^* = 0$, in which case any optimum solution $X = xx^T$ gives a partition of a, as $a^T x = \sum_{i=1}^n a_i x_i = 0$.

1.4. Contents of the paper. We provide in Section 2 more detailed algebraic preliminaries about polynomial ideals and varieties and the resolution of systems of polynomial equations. This is relevant to the problem of extracting global minimizers for the polynomial optimization problem (1.1) and can be read separately. Then the rest of the paper is divided into two parts. Part 1 contains some background results about positive polynomials and sums of squares (Section 3) and about the theory of moments (Section 4), and more detailed results about (truncated) moment matrices, in particular, from Curto and Fialkow (Section 5). Part 2 presents the application to polynomial optimization; namely, the main properties of the moment/SOS relaxations (Section 6), some further selected topics dealing in particular with approximations of positive polynomials by sums of squares and various approaches to unconstrained polynomial minimization (Section 7), and exploiting algebraic structure to reduce the problem size (Section 8).

2. Algebraic preliminaries.

We group here some preliminaries on polynomial ideals and varieties, and on the eigenvalue method for solving systems of polynomial equations. For more information, see, e.g., [8, 23, 25, 26, 161].

2.1. Polynomial ideals and varieties.

Let $\mathcal{I}$ be an ideal in $\mathbb{R}[\mathbf{x}]$; that is, $\mathcal{I}$ is an additive subgroup of $\mathbb{R}[\mathbf{x}]$ satisfying $fg \in \mathcal{I}$ whenever $f \in \mathcal{I}$ and $g \in \mathbb{R}[\mathbf{x}]$. Given $h_1, \ldots, h_m \in \mathbb{R}[\mathbf{x}]$,

$$(h_1, \ldots, h_m) := \Big\{ \sum_{j=1}^m u_j h_j \mid u_1, \ldots, u_m \in \mathbb{R}[\mathbf{x}] \Big\}$$

denotes the *ideal generated by* $h_1, \ldots, h_m$. By the finite basis theorem, any ideal in $\mathbb{R}[\mathbf{x}]$ admits a finite set of generators. Given an ideal $\mathcal{I} \subseteq \mathbb{R}[\mathbf{x}]$, define

$$V_{\mathbb{C}}(\mathcal{I}) := \{x \in \mathbb{C}^n \mid f(x) = 0 \ \forall f \in \mathcal{I}\}, \ V_{\mathbb{R}}(\mathcal{I}) := V_{\mathbb{C}}(\mathcal{I}) \cap \mathbb{R}^n;$$

$V_{\mathbb{C}}(\mathcal{I})$ is the *(complex) variety* associated to $\mathcal{I}$ and $V_{\mathbb{R}}(\mathcal{I})$ is its *real variety.* Thus, if $\mathcal{I}$ is generated by $h_1, \ldots, h_m$, then $V_{\mathbb{C}}(\mathcal{I})$ (resp., $V_{\mathbb{R}}(\mathcal{I})$) is the set of common complex (resp., real) zeros of $h_1, \ldots, h_m$. Observe that $V_{\mathbb{C}}(\mathcal{I})$ is closed under complex conjugation, i.e., $\overline{v} \in V_{\mathbb{C}}(\mathcal{I})$ for all $v \in V_{\mathbb{C}}(\mathcal{I})$, since $\mathcal{I}$ consists of polynomials with real coefficients. When $V_{\mathbb{C}}(\mathcal{I})$ is finite, the ideal $\mathcal{I}$ is said to be *zero-dimensional.* Given $V \subseteq \mathbb{C}^n$,

$$\mathcal{I}(V) := \{f \in \mathbb{R}[\mathbf{x}] \mid f(v) = 0 \ \forall v \in V\}$$

is the *vanishing ideal* of V. Moreover,

$$\sqrt{\mathcal{I}} := \{f \in \mathbb{R}[\mathbf{x}] \mid f^k \in \mathcal{I} \text{ for some integer } k \geq 1\}$$

is the *radical* of $\mathcal{I}$ and

$$\sqrt[\mathbb{R}]{\mathcal{I}} := \{f \in \mathbb{R}[\mathbf{x}] \mid f^{2k} + \sum_{j=1}^{m} p_j^2 \in \mathcal{I} \ \text{ for some } k \geq 1, \ p_1, \ldots, p_m \in \mathbb{R}[\mathbf{x}]\}$$

is the *real radical* of $\mathcal{I}$. The sets $\mathcal{I}(V)$, $\sqrt{\mathcal{I}}$ and $\sqrt[\mathbb{R}]{\mathcal{I}}$ are again ideals in $\mathbb{R}[\mathbf{x}]$. Obviously, for an ideal $\mathcal{I} \subseteq \mathbb{R}[\mathbf{x}]$,

$$\mathcal{I} \subseteq \sqrt{\mathcal{I}} \subseteq \mathcal{I}(V_{\mathbb{C}}(\mathcal{I})), \ \mathcal{I} \subseteq \sqrt[\mathbb{R}]{\mathcal{I}} \subseteq \mathcal{I}(V_{\mathbb{R}}(\mathcal{I})).$$

The following celebrated results relate (real) radical and vanishing ideals.

THEOREM 2.1. *Let $\mathcal{I}$ be an ideal in $\mathbb{R}[\mathbf{x}]$.*

(i) **(Hilbert's Nullstellensatz)** *(see, e.g., [25, §4.1])* $\sqrt{\mathcal{I}} = \mathcal{I}(V_{\mathbb{C}}(\mathcal{I}))$.

(ii) **(The Real Nullstellensatz)** *(see, e.g., [13, §4.1])* $\sqrt[\mathbb{R}]{\mathcal{I}} = \mathcal{I}(V_{\mathbb{R}}(\mathcal{I}))$.

The ideal $\mathcal{I}$ is said to be *radical* when $\mathcal{I} = \sqrt{\mathcal{I}}$, and *real radical* when $\mathcal{I} = \sqrt[\mathbb{R}]{\mathcal{I}}$. Roughly speaking, the ideal $\mathcal{I}$ is radical if all points of $V_{\mathbb{C}}(\mathcal{I})$ have single multiplicity. For instance, the ideal $\mathcal{I} := (\mathbf{x}^2)$ is not radical since $V_{\mathbb{C}}(\mathcal{I}) = \{0\}$ and $\mathbf{x} \in \mathcal{I}(V_{\mathbb{C}}(\mathcal{I})) \setminus \mathcal{I}$. Obviously, $\mathcal{I} \subseteq \mathcal{I}(V_{\mathbb{C}}(\mathcal{I})) \subseteq \mathcal{I}(V_{\mathbb{R}}(\mathcal{I}))$. Hence, $\mathcal{I}$ real radical $\Longrightarrow$ $\mathcal{I}$ radical. Moreover,

$$\mathcal{I} \text{ real radical with } |V_{\mathbb{R}}(\mathcal{I})| < \infty \Longrightarrow V_{\mathbb{C}}(\mathcal{I}) = V_{\mathbb{R}}(\mathcal{I}) \subseteq \mathbb{R}^n. \tag{2.1}$$

Indeed, $\mathcal{I}(V_{\mathbb{C}}(\mathcal{I})) = \mathcal{I}(V_{\mathbb{R}}(\mathcal{I}))$ implies $V_{\mathbb{C}}(\mathcal{I}(V_{\mathbb{C}}(\mathcal{I}))) = V_{\mathbb{C}}(\mathcal{I}(V_{\mathbb{R}}(\mathcal{I})))$. Now, $V_{\mathbb{C}}(\mathcal{I}(V_{\mathbb{C}}(\mathcal{I}))) = V_{\mathbb{C}}(\mathcal{I})$, and $V_{\mathbb{C}}(\mathcal{I}(V_{\mathbb{R}}(\mathcal{I}))) = V_{\mathbb{R}}(\mathcal{I})$ since $V_{\mathbb{R}}(\mathcal{I})$ is an algebraic subset of $\mathbb{C}^n$ as it is finite. We will often use the following characterization of (real) radical ideals which follows directly from the (Real) Nullstellensatz:

$$\begin{array}{c} \mathcal{I} \text{ is radical (resp., real radical)} \\ \Longleftrightarrow \\ \text{The only polynomials vanishing at all points of } V_{\mathbb{C}}(\mathcal{I}) \\ \text{(resp., all points of } V_{\mathbb{R}}(\mathcal{I})) \text{ are the polynomials in } \mathcal{I}. \end{array} \tag{2.2}$$

The following lemma gives a useful criterion for checking whether an ideal is (real) radical.

Lemma 2.2. *Let $\mathcal{I}$ be an ideal in $\mathbb{R}[\mathbf{x}]$.*

(i) *$\mathcal{I}$ is radical if and only if*

$$\forall f \in \mathbb{R}[\mathbf{x}] \quad f^2 \in \mathcal{I} \Longrightarrow f \in \mathcal{I}. \tag{2.3}$$

(ii) *$\mathcal{I}$ is real radical if and only if*

$$\forall p_1, \ldots, p_m \in \mathbb{R}[\mathbf{x}] \quad \sum_{j=1}^{m} p_j^2 \in \mathcal{I} \Longrightarrow p_1, \ldots, p_m \in \mathcal{I}. \tag{2.4}$$

Proof. The 'only if' part is obvious in (i), (ii); we prove the 'if part'.

(i) Assume that (2.3) holds. Let $f \in \mathbb{R}[\mathbf{x}]$. We show $f^k \in \mathcal{I} \Longrightarrow f \in \mathcal{I}$ using induction on $k \geq 1$. Let $k \geq 2$. Using (2.3), we deduce $f^{\lceil k/2 \rceil} \in \mathcal{I}$. As $\lceil k/2 \rceil \leq k-1$, we deduce $f \in \mathcal{I}$ using the induction assumption.

(ii) Assume that (2.4) holds. Let $f, p_1, \ldots, p_m \in \mathbb{R}[\mathbf{x}]$ such that $f^{2k} + \sum_{j=1}^m p_j^2 \in \mathcal{I}$; we show that $f \in \mathcal{I}$. First we deduce from (2.4) that $f^k, p_1, \ldots, p_m \in \mathcal{I}$. As (2.4) implies (2.3), we next deduce from the case (i) that $f \in \mathcal{I}$. □

We now recall the following simple fact about interpolation polynomials, which we will need at several occasions in the paper.

Lemma 2.3. *Let $V \subseteq \mathbb{C}^n$ with $|V| < \infty$. There exist polynomials $p_v \in \mathbb{C}[\mathbf{x}]$ (for $v \in V$) satisfying $p_v(v) = 1$ and $p_v(u) = 0$ for all $u \in V \setminus \{v\}$; they are known as* Lagrange interpolation polynomials at the points of *V. Assume moreover that V is closed under complex conjugation, i.e., $V = \overline{V} := \{\overline{v} \mid v \in V\}$. Then we may choose the interpolation polynomials in such a way that they satisfy $p_{\overline{v}} = \overline{p_v}$ for all $v \in V$ and, given scalars a_v ($v \in V$) satisfying $a_{\overline{v}} = \overline{a_v}$ for all $v \in V$, there exists $p \in \mathbb{R}[\mathbf{x}]$ taking the prescribed values $p(v) = a_v$ at the points $v \in V$.*

Proof. Fix $v \in V$. For $u \in V$, $u \neq v$, pick an index $i_u \in \{1, \ldots, n\}$ for which $u_{i_u} \neq v_{i_u}$ and define the polynomial $p_v := \prod_{u \in V \setminus \{v\}} \frac{\mathbf{x}_{i_u} - u_{i_u}}{v_{i_u} - u_{i_u}}$. Then the polynomials p_v ($v \in V$) satisfy the lemma. If $\overline{V} = V$, then we can choose the interpolation polynomials in such a way that $p_{\overline{v}} = \overline{p_v}$ for all $v \in V$. Indeed, for $v \in V \cap \mathbb{R}^n$, simply replace p_v by its real part and, for $v \in V \setminus \mathbb{R}^n$, pick p_v as before and choose $p_{\overline{v}} := \overline{p_v}$. Finally, if $a_{\overline{v}} = \overline{a_v}$ for all $v \in V$, then the polynomial $p := \sum_{v \in V} a_v p_v$ has real coefficients and satisfies $p(v) = a_v$ for $v \in V$. □

The algebraic tools just introduced here permit to show the following result of Parrilo [123], giving a sum of squares decomposition for every polynomial nonnegative on a finite variety assuming radicality of the associated ideal.

Theorem 2.4. [123] *Consider the semialgebraic set*

$$K := \{x \in \mathbb{R}^n \mid h_1(x) = 0, \ldots, h_{m_0}(x) = 0, g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}, \tag{2.5}$$

where $h_1, \ldots, h_{m_0}, g_1, \ldots, g_m \in \mathbb{R}[\mathbf{x}]$ *and* $m_0 \geq 1$, $m \geq 0$. *Assume that the ideal* $\mathcal{I} := (h_1, \ldots, h_{m_0})$ *is zero-dimensional and radical. Then every nonnegative polynomial on* K *is of the form* $u_0 + \sum_{j=1}^m u_j g_j + q$, *where* $u_0, u_1, \ldots, u_m$ *are sums of squares of polynomials and* $q \in \mathcal{I}$.

Proof. Partition $V := V_{\mathbb{C}}(\mathcal{I})$ into $S \cup T \cup \overline{T}$, where $S = V \cap \mathbb{R}^n$, $T \cup \overline{T} = V \setminus \mathbb{R}^n$. Let p_v $(v \in V_{\mathbb{C}}(\mathcal{I}))$ be interpolation polynomials at the points of V, satisfying $p_{\overline{v}} = \overline{p_v}$ for $v \in T$ (as in Lemma 2.3). We first show the following fact: If $f \in \mathbb{R}[\mathbf{x}]$ is nonnegative on the set S, then $f = \sigma + q$ where σ is a sum of squares of polynomials and $q \in \mathcal{I}$. For this, for $v \in S \cup T$, let $\gamma_v = \sqrt{f(v)}$ be a square root of $f(v)$ (thus, $\gamma_v \in \mathbb{R}$ if $v \in S$) and define the polynomials $q_v \in \mathbb{R}[\mathbf{x}]$ by $q_v := \gamma_v p_v$ for $v \in S$ and $q_v := \gamma_v p_v + \overline{\gamma_v p_v}$ for $v \in T$. The polynomial $f - \sum_{v \in S \cup T} (q_v)^2$ vanishes at all points of V; hence it belongs to $\mathcal{I}$, since $\mathcal{I}$ is radical. This shows that $f = \sigma + q$, where σ is a sum of squares and $q \in \mathcal{I}$.

Suppose now that $f \in \mathbb{R}[\mathbf{x}]$ is nonnegative on the set K. In view of Lemma 2.3, we can construct polynomials $s_0, s_1, \ldots, s_m \in \mathbb{R}[\mathbf{x}]$ taking the following prescribed values at the points in V: If $v \in V \setminus S$, or if $v \in S$ and $f(v) \geq 0$, $s_0(v) := f(v)$ and $s_j(v) := 0$ $(j = 1, \ldots, m)$. Otherwise, $v \notin K$ and thus $g_{j_v}(v) < 0$ for some $j_v \in \{1, \ldots, m\}$; then $s_{j_v}(v) := \frac{f(v)}{g_{j_v}(v)}$ and $s_0(v) = s_j(v) := 0$ for $j \in \{1, \ldots, m\} \setminus \{j_v\}$. By construction, each of the polynomials $s_0, s_1, \ldots, s_m$ is nonnegative on S. Using the above result, we can conclude that $s_j = \sigma_j + q_j$, where σ_j is a sum of squares and $q_j \in \mathcal{I}$, for $j = 0, 1, \ldots, m$. Now the polynomial $q := f - s_0 - \sum_{j=1}^m s_j g_j$ vanishes at all points of V and thus belongs to $\mathcal{I}$. Therefore, $f = s_0 + \sum_{j=1}^m s_j g_j + q = \sigma_0 + \sum_{j=1}^m \sigma_j g_j + q'$, where $q' := q + q_0 + \sum_{j=1}^m q_j g_j \in \mathcal{I}$ and σ_0, σ_j are sums of squares of polynomials. □

2.2. The quotient algebra $\mathbb{R}[\mathbf{x}]/\mathcal{I}$.

Given an ideal $\mathcal{I}$ in $\mathbb{R}[\mathbf{x}]$, the elements of the quotient space $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ are the cosets $[f] := f + \mathcal{I} = \{f + q \mid q \in \mathcal{I}\}$. $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ is a $\mathbb{R}$-vector space with addition $[f]+[g] = [f+g]$ and scalar multiplication $\lambda[f] = [\lambda f]$, and an algebra with multiplication $[f][g] = [fg]$, for $\lambda \in \mathbb{R}$, $f, g \in \mathbb{R}[\mathbf{x}]$. Given $h \in \mathbb{R}[\mathbf{x}]$, the '*multiplication by h operator*'

$$\begin{array}{rccc} m_h: & \mathbb{R}[\mathbf{x}]/\mathcal{I} & \longrightarrow & \mathbb{R}[\mathbf{x}]/\mathcal{I} \\ & f + \mathcal{I} & \longmapsto & fh + \mathcal{I} \end{array} \tag{2.6}$$

is well defined. As we see later in Section 2.4, multiplication operators play a central role in the computation of the variety $V_{\mathbb{C}}(\mathcal{I})$. In what follows we often identify a subset of $\mathbb{R}[\mathbf{x}]$ with the corresponding subset of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ consisting of the cosets of its elements. For instance, given $\mathcal{B} = \{b_1, \ldots, b_N\} \subseteq \mathbb{R}[\mathbf{x}]$, if the cosets $[b_1], \ldots, [b_N]$ generate $\mathbb{R}[\mathbf{x}]/\mathcal{I}$, i.e., if any $f \in \mathbb{R}[\mathbf{x}]$ can be written as $\sum_{j=1}^{N} \lambda_j b_j + q$ for some $\lambda \in \mathbb{R}^N$ and $q \in \mathcal{I}$, then we also say by abuse of language that the set $\mathcal{B}$ itself is generating in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. Analogously, if the cosets $[b_1], \ldots, [b_N]$ are pairwise distinct and form a linearly independent subset of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$, i.e., if $\sum_{j=1}^{N} \lambda_j b_j \in \mathcal{I}$ implies $\lambda = 0$, then we say that $\mathcal{B}$ is linearly independent in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$.

Theorem 2.6 below relates the cardinality of $V_{\mathbb{C}}(\mathcal{I})$ and the dimension of the quotient vector space $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. This is a classical result (see, e.g., [25]), which we will use repeatedly in our treatment. The following simple fact will be used in the proof.

Lemma 2.5. *Let $\mathcal{I} \subseteq \mathbb{R}[\mathbf{x}]$ with $|V_{\mathbb{C}}(\mathcal{I})| < \infty$. Partition $V_{\mathbb{C}}(\mathcal{I})$ into $V_{\mathbb{C}}(\mathcal{I}) = S \cup T \cup \overline{T}$ where $S = V_{\mathbb{C}}(\mathcal{I}) \cap \mathbb{R}^n$, and let p_v be interpolation polynomials at the points of $V_{\mathbb{C}}(\mathcal{I})$ satisfying $p_{\overline{v}} = \overline{p_v}$ for all $v \in V_{\mathbb{C}}(\mathcal{I})$. The set $\mathcal{L} := \{p_v \ (v \in S), \mathrm{Re}(p_v), \mathrm{Im}(p_v) \ (v \in T)\}$ is linearly independent in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ and generates $\mathbb{R}[\mathbf{x}]/\mathcal{I}(V_{\mathbb{C}}(\mathcal{I}))$.*

Proof. Assume $\sum_{v \in S} \lambda_v p_v + \sum_{v \in T} \lambda_v \mathrm{Re}(p_v) + \lambda'_v \mathrm{Im}(p_v) \in \mathcal{I}$. Evaluating this polynomial at $v \in V_{\mathbb{C}}(\mathcal{I})$ yields that all scalars λ_v, λ'_v are 0. Thus $\mathcal{L}$ is linearly independent in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. Given $f \in \mathbb{R}[\mathbf{x}]$, the polynomial $f - \sum_{v \in V_{\mathbb{C}}(\mathcal{I})} f(v) p_v$ lies in $\mathcal{I}(V_{\mathbb{C}}(\mathcal{I}))$. Now, $\sum_{v \in V_{\mathbb{C}}(\mathcal{I})} f(v) p_v = \sum_{v \in S} f(v) p_v + \sum_{v \in T} 2\,\mathrm{Re}(f(v) p_v)$ can be written as a linear combination of $\mathrm{Re}(p_v)$ and $\mathrm{Im}(p_v)$. This implies that $\mathcal{L}$ generates $\mathbb{R}[\mathbf{x}]/\mathcal{I}(V_{\mathbb{C}}(\mathcal{I}))$. □

Theorem 2.6. *An ideal $\mathcal{I} \subseteq \mathbb{R}[\mathbf{x}]$ is zero-dimensional (i.e., $|V_{\mathbb{C}}(\mathcal{I})| < \infty$) if and only if the vector space $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ is finite dimensional. Moreover, $|V_{\mathbb{C}}(\mathcal{I})| \leq \dim\ \mathbb{R}[\mathbf{x}]/\mathcal{I}$, with equality if and only if the ideal $\mathcal{I}$ is radical.*

Proof. Assume $k := \dim \mathbb{R}[\mathbf{x}]/\mathcal{I} < \infty$. Then, the set $\{1, \mathbf{x}_1, \ldots, \mathbf{x}_1^k\}$ is linearly dependent in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. Thus there exist scalars $\lambda_0, \ldots, \lambda_k$ (not all zero) for which the polynomial $f := \sum_{h=0}^{k} \lambda_h \mathbf{x}_1^h$ belongs to $\mathcal{I}$. Thus, for $v \in V_{\mathbb{C}}(\mathcal{I})$, $f(v) = 0$, which implies that v_1 takes only finitely many values. Applying the same reasoning to the other coordinates, we deduce that $V_{\mathbb{C}}(\mathcal{I})$ is finite.

Assume now $|V_{\mathbb{C}}(\mathcal{I})| < \infty$. Say, $\{v_1 \mid v \in V_{\mathbb{C}}(\mathcal{I})\} = \{a_1, \ldots, a_k\}$. Then the polynomial $f := \prod_{h=1}^{k} (\mathbf{x}_1 - a_h)$ belongs to $\mathcal{I}(V_{\mathbb{C}}(\mathcal{I}))$. By Theorem 2.1, $f \in \sqrt{\mathcal{I}}$, i.e., $f^{m_1} \in \mathcal{I}$ for some integer $m_1 \geq 1$. Hence the set $\{[1], [\mathbf{x}_1], \ldots, [\mathbf{x}_1^{km_1}]\}$ is linearly dependent in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ and thus, for some integer $n_1 \geq 1$, $[\mathbf{x}_1^{n_1}]$ lies in $\mathrm{Span}_{\mathbb{R}}([1], \ldots, [\mathbf{x}_1^{n_1 - 1}])$. Similarly, for any other coordinate $\mathbf{x}_i$, $[\mathbf{x}_i^{n_i}] \in \mathrm{Span}_{\mathbb{R}}([1], \ldots, [\mathbf{x}_i^{n_i - 1}])$ for some integer $n_i \geq 1$. From this one can easily derive that the set $\{[\mathbf{x}^\alpha] \mid 0 \leq \alpha_i \leq n_i - 1 \ (1 \leq i \leq n)\}$ generates $\mathbb{R}[\mathbf{x}]/\mathcal{I}$, which shows that $\dim \mathbb{R}[\mathbf{x}]/\mathcal{I} < \infty$.

Assume $V_{\mathbb{C}}(\mathcal{I})$ is finite and let $\mathcal{L}$ be as in Lemma 2.5. As $\mathcal{L}$ is linearly independent in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ with $|\mathcal{L}| = |V_{\mathbb{C}}(\mathcal{I})|$ we deduce that $\dim \mathbb{R}[\mathbf{x}]/\mathcal{I} \geq |V_{\mathbb{C}}(\mathcal{I})|$. Moreover, if $\mathcal{I}$ is radical then $\mathcal{I} = \mathcal{I}(V_{\mathbb{C}}(\mathcal{I}))$ and thus $\mathcal{L}$ is also generating in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$, which implies $\dim \mathbb{R}[\mathbf{x}]/\mathcal{I} = |V_{\mathbb{C}}(\mathcal{I})|$. Finally, if $\mathcal{I}$ is not radical, there exists a polynomial $f \in \mathcal{I}(V_{\mathbb{C}}(\mathcal{I})) \setminus \mathcal{I}$ and it is easy to verify that the set $\mathcal{L} \cup \{f\}$ is linearly independent in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. □

For instance, the ideal $\mathcal{I} := (\mathbf{x}_i^2 - \mathbf{x}_i \mid i = 1, \ldots, n)$ is radical and zero-dimensional, since $V_{\mathbb{C}}(\mathcal{I}) = \{0,1\}^n$, and the set $\{\prod_{l \in L} \mathbf{x}_l \mid L \subseteq \{1, \ldots, n\}\}$ is a linear basis of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$.

Assume $N := \dim \mathbb{R}[\mathbf{x}]/\mathcal{I} < \infty$ and let $\mathcal{B} = \{b_1, \ldots, b_N\} \subseteq \mathbb{R}[\mathbf{x}]$ be a basis of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$; that is, any polynomial $f \in \mathbb{R}[\mathbf{x}]$ can be written in a unique way as

$$f = \underbrace{\sum_{j=1}^{N} \lambda_j b_j}_{\mathrm{res}_{\mathcal{B}}(f)} + q, \text{ where } q \in \mathcal{I} \text{ and } \lambda \in \mathbb{R}^N;$$

in short, $f \equiv \sum_{j=1}^N \lambda_j b_j \mod \mathcal{I}$. The polynomial $\mathrm{res}_{\mathcal{B}}(f) := \sum_{j=1}^N \lambda_j b_j$ is called the *residue of f modulo $\mathcal{I}$ with respect to the basis $\mathcal{B}$*. In other words, the vector space $\mathrm{Span}_{\mathbb{R}}(\mathcal{B}) := \{\sum_{j=1}^N \lambda_j b_j \mid \lambda \in \mathbb{R}^N\}$ is isomorphic to $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. As recalled in the next section, the set $\mathcal{B}_\succ$ of standard monomials with respect to any monomial ordering is a basis of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$; then the residue of a polynomial f w.r.t. $\mathcal{B}_\succ$ is also known as the *normal form* of f w.r.t. the given monomial ordering. Let us mention for further reference the following variation of Lemma 2.3.

LEMMA 2.7. *Let $\mathcal{I}$ be a zero-dimensional ideal in $\mathbb{R}[\mathbf{x}]$ and let $\mathcal{B}$ be a basis of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. There exist interpolation polynomials p_v at the points of $V_{\mathbb{C}}(\mathcal{I})$, where each p_v is a linear combination of members of $\mathcal{B}$.*

Proof. Given a set of interpolation polynomials p_v, replace p_v by its residue modulo $\mathcal{I}$ with respect to $\mathcal{B}$. □

2.3. Gröbner bases and standard monomial bases. A classical method for constructing a linear basis of the quotient vector space $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ is to determine a Gröbner basis of the ideal $\mathcal{I}$ with respect to some given monomial ordering; then the corresponding set of standard monomials provides a basis of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. We recall here a few basic definitions about monomial orderings, Gröbner bases, and standard monomials. A *monomial ordering* '$\succ$' is a total ordering of the set $\mathbb{T}_n = \{\mathbf{x}^\alpha \mid \alpha \in \mathbb{N}^n\}$ of monomials, which is a well-ordering and satisfies the condition: $\mathbf{x}^\alpha \succ \mathbf{x}^\beta \Longrightarrow \mathbf{x}^{\alpha+\gamma} \succ \mathbf{x}^{\beta+\gamma}$. We also write $a\mathbf{x}^\alpha \succ b\mathbf{x}^\beta$ if $\mathbf{x}^\alpha \succ \mathbf{x}^\beta$ and $a, b \in \mathbb{R} \setminus \{0\}$. Examples of monomial orderings are the *lexicographic order* '$\succ_{lex}$', where $\mathbf{x}^\alpha \succ_{lex} \mathbf{x}^\beta$ if $\alpha > \beta$ for a lexicographic order on $\mathbb{N}^n$, or the *graded lexicographic order* '$\succ_{grlex}$', where $\mathbf{x}^\alpha \succ_{grlex} \mathbf{x}^\beta$ if $|\alpha| > |\beta|$, or $|\alpha| = |\beta|$ and $\mathbf{x}^\alpha \succ_{lex} \mathbf{x}^\beta$. The latter is an example of a *total degree monomial ordering*, i.e., a monomial ordering $\succ$ such that $\mathbf{x}^\alpha \succ \mathbf{x}^\beta$ whenever $|\alpha| > |\beta|$.

Fix a monomial ordering $\succ$ on $\mathbb{R}[\mathbf{x}]$. For a nonzero polynomial $f = \sum_\alpha f_\alpha \mathbf{x}^\alpha$, its *terms* are the quantities $f_\alpha \mathbf{x}^\alpha$ with $f_\alpha \neq 0$ and its *leading term* $\mathrm{LT}(f)$ is defined as the maximum $f_\alpha \mathbf{x}^\alpha$ with respect to the given ordering for which $f_\alpha \neq 0$. Let $\mathcal{I}$ be an ideal in $\mathbb{R}[\mathbf{x}]$. Its *leading term ideal* is $\mathrm{LT}(\mathcal{I}) := (\mathrm{LT}(f) \mid f \in \mathcal{I})$ and the set

$$\mathcal{B}_{\succ} := \mathbb{T}_n \setminus \mathrm{LT}(\mathcal{I}) = \{\mathbf{x}^\alpha \mid \mathrm{LT}(f) \text{ does not divide } \mathbf{x}^\alpha \;\; \forall f \in \mathcal{I}\}$$

is the set of *standard monomials.* A finite subset $G \subseteq \mathcal{I}$ is called a *Gröbner basis* of $\mathcal{I}$ if $\mathrm{LT}(\mathcal{I}) = \mathrm{LT}(G)$; that is, if the leading term of every nonzero polynomial in $\mathcal{I}$ is divisible by the leading term of some polynomial in G. Hence $\mathbf{x}^\alpha \in \mathcal{B}_{\succ}$ if and only if $\mathbf{x}^\alpha$ is not divisible by the leading term of any polynomial in G. A Gröbner basis always exists and it can be constructed, e.g., using the algorithm of Buchberger.

Once a monomial ordering $\succ$ is fixed, one can apply the division algorithm. Given nonzero polynomials $f, g_1, \ldots, g_m$, the division algorithm applied to dividing f by $g_1, \ldots, g_m$ produces polynomials $u_1, \ldots, u_m$ and r satisfying $f = \sum_{j=1}^m u_j g_j + r$, no term of r is divisible by $\mathrm{LT}(g_j)$ $(j = 1, \ldots, m)$ if $r \neq 0$, and $\mathrm{LT}(f) \succ \mathrm{LT}(u_j g_j)$ if $u_j \neq 0$. Hence $\deg(f) \geq \deg(u_j g_j)$ if $u_j \neq 0$, when the monomial ordering is a graded lexicographic order. When the polynomials $g_1, \ldots, g_m$ form a Gröbner basis of the ideal $\mathcal{I} := (g_1, \ldots, g_m)$, the remainder r is uniquely determined and r is a linear combination of the set of standard monomials, i.e., $r \in \mathrm{Span}_{\mathbb{R}}(\mathcal{B}_{\succ})$; in particular, $f \in \mathcal{I}$ if and only if $r = 0$. In other words, the set $\mathcal{B}_{\succ}$ of standard monomials is a basis of the quotient vector space $\mathbb{R}[\mathbf{x}]/\mathcal{I}$.

Example 2.8. *Consider the polynomial* $f = \mathbf{x}^2\mathbf{y} + \mathbf{x}\mathbf{y}^2 + \mathbf{y}^2$ *to be divided by the polynomials* $h_1 = \mathbf{x}\mathbf{y} - 1$, $h_2 = \mathbf{y}^2 - 1$. *Fix the lex order with* $\mathbf{x} > \mathbf{y}$. *Then* $\mathrm{LT}(f) = \mathbf{x}^2\mathbf{y}$, $\mathrm{LT}(h_1) = \mathbf{x}\mathbf{y}$, $\mathrm{LT}(h_2) = \mathbf{y}^2$. *As* $\mathrm{LT}(h_1) | \mathrm{LT}(f)$, *we write*

$$f = \mathbf{x}^2\mathbf{y} + \mathbf{x}\mathbf{y}^2 + \mathbf{y}^2 = \underbrace{(\mathbf{x}\mathbf{y} - 1)}_{h_1}(\mathbf{x} + \mathbf{y}) + \underbrace{\mathbf{x} + \mathbf{y}^2 + \mathbf{y}}_{q}.$$

Now $\mathrm{LT}(q) = \mathbf{x}$ *is not divisible by* $\mathrm{LT}(h_1), \mathrm{LT}(h_2)$, *but* $\mathrm{LT}(h_2)$ *divides the term* $\mathbf{y}^2$ *of* q. *Thus write*

$$q = \underbrace{(\mathbf{y}^2 - 1)}_{h_2} + \mathbf{x} + \mathbf{y} + 1.$$

This gives

$$f = h_1(\mathbf{x} + \mathbf{y}) + h_2 + \mathbf{x} + \mathbf{y} + 1. \tag{2.7}$$

No term of the polynomial $r := \mathbf{x} + \mathbf{y} + 1$ *is divisible by* $\mathrm{LT}(h_1), \mathrm{LT}(h_2)$, *thus* r *is the remainder of the division of* f *by* h_1, h_2 *(in that order). If we do the division by* h_2, h_1 *then we get the following decomposition:*

$$f = (\mathbf{x} + 1)h_2 + \mathbf{x}h_1 + 2\mathbf{x} + 1. \tag{2.8}$$

Thus (2.7), (2.8) are two disctinct decompositions of f *of the form*

$$f = \sum_{i=1}^{2} u_i h_i + r$$

where no term of r is divisible by $\mathrm{LT}(h_1), \mathrm{LT}(h_2)$*. Hence the remainder is not uniquely defined. This is because the set* $\{h_1, h_2\}$ *is not a Gröbner basis of the ideal* $\mathcal{I} := (h_1, h_2)$*. Indeed the polynomial*

$$h_3 := \mathbf{y}h_1 - \mathbf{x}h_2 = \mathbf{y}(\mathbf{x}\mathbf{y} - 1) - \mathbf{x}(\mathbf{y}^2 - 1) = \mathbf{x} - \mathbf{y} \in \mathcal{I}$$

and $\mathrm{LT}(h_3) = \mathbf{x}$ *is not divisible by* $\mathrm{LT}(h_1), \mathrm{LT}(h_2)$*. For the given monomial ordering, the set of standard monomials is* $\mathcal{B} = \{1, \mathbf{y}\}$*, the set* $\{h_2, h_3\}$ *is a Gröbner basis of* $\mathcal{I}$*, and* $\dim \mathbb{R}[\mathbf{x}]/\mathcal{I} = 2 = |V_{\mathbb{C}}(\mathcal{I})|$ *with* $V_{\mathbb{C}}(\mathcal{I}) = \{(1, 1), (-1, -1)\}$*.*

2.4. Solving systems of polynomial equations. One of the attractive features of Lasserre's method for minimizing a polynomial over a semialgebraic set is that, when some technical rank condition holds for the optimum solution of the given relaxation, then this relaxation is in fact exact and moreover one can extract global minimizers for the original problem. This extraction procedure requires to solve a system of polynomial equations

$$h_1(x) = 0, \ldots, h_{m_0}(x) = 0,$$

where the ideal $\mathcal{I} := (h_1, \ldots, h_{m_0})$ is zero-dimensional (and in fact radical). This problem has received considerable attention in the literature. We present the so-called eigenvalue method (also known as the Stetter-Möller method [108]) which relates the points of $V_{\mathbb{C}}(\mathcal{I})$ to the eigenvalues of the multiplication operators in the quotient space $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. See, e.g., [23, 40, 161] for a detailed account on this method and various other methods for solving systems of polynomial equations.

Fix a basis $\mathcal{B} = \{b_1, \ldots, b_N\}$ of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ and let M_h denote the matrix of the multiplication operator operator m_h from (2.6) with respect to the basis $\mathcal{B}$. Namely, for $j = 1, \ldots, N$, let $\mathrm{res}_{\mathcal{B}}(hb_j) = \sum_{i=1}^N a_{ij} b_i$ denote the residue of hb_j modulo $\mathcal{I}$ w.r.t. $\mathcal{B}$, i.e.,

$$hb_j - \sum_{i=1}^{N} a_{ij} b_i \in \mathcal{I}; \tag{2.9}$$

then the jth column of M_h is equal to the vector $(a_{ij})_{i=1}^N$. When $h = \mathbf{x}_i$, the multiplication matrices $M_{\mathbf{x}_i}$ $(i = 1, \ldots, n)$ are also known as the *companion matrices* of the ideal $\mathcal{I}$. Theorem 2.9 below shows that the coordinates of the points $v \in V$ can be obtained from the eigenvalues of the companion matrices. As a motivation we first treat the univariate case.

2.4.1. Motivation: The univariate case. Given a univariate polynomial

$$p = \mathbf{x}^d - p_{d-1}\mathbf{x}^{d-1} - \ldots - p_0$$

consider the ideal $\mathcal{I} = (p)$ (obviously zero-dimensional). The set $\mathcal{B} = \{1, \mathbf{x}, \ldots, \mathbf{x}^{d-1}\}$ is a basis of $\mathbb{R}[\mathbf{x}]/(p)$. With respect to $\mathcal{B}$, the multiplication matrix $M_{\mathbf{x}}$ has the form

$$M_{\mathbf{x}} = \begin{pmatrix} 0 & \ldots & 0 & p_0 \\ & & & p_1 \\ & I & & \vdots \\ & & & p_{d-1} \end{pmatrix}$$

where I is the identity matrix of size $(d-1) \times (d-1)$. One can verify that $\det(M_{\mathbf{x}} - tI) = (-1)^d p(t)$. Therefore, the eigenvalues of the companion matrix $M_{\mathbf{x}}$ are precisely the roots of the polynomial p. We now see how this fact extends to the multivariate case.

2.4.2. The multivariate case. The multiplication operators $m_{\mathbf{x}_1}, \ldots, m_{\mathbf{x}_n}$ commute pairwise. Therefore the set $\{M_f \mid f \in \mathbb{R}[\mathbf{x}]\}$ is a commutative algebra of $N \times N$ matrices. For a polynomial $h \in \mathbb{R}[\mathbf{x}]$, $h = \sum_\alpha h_\alpha \mathbf{x}^\alpha$, note that

$$M_h = h(M_{\mathbf{x}_1}, \ldots, M_{\mathbf{x}_n}) = \sum_\alpha h_\alpha (M_{\mathbf{x}_1})^{\alpha_1} \cdots (M_{\mathbf{x}_n})^{\alpha_n} =: h(M),$$
$$M_h = 0 \Longleftrightarrow h \in \mathcal{I}.$$

Based on this, one can easily find the *minimal polynomial* of M_h (i.e. the monic polynomial $p \in \mathbb{R}[\mathbf{t}]$ of smallest degree for which $p(M_h) = 0$). Indeed, for $p = \sum_{i=0}^d p_i \mathbf{t}^i \in \mathbb{R}[\mathbf{t}]$, $p(M_h) = \sum_i p_i (M_h)^i = M_{p(h)} = 0$ if and only if $p(h) \in \mathcal{I}$. Thus one can find the minimal polynomial of M_h by computing the smallest integer d for which the set $\{[1], [h], \ldots, [h^d]\}$ is linearly dependent in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. In particular, the minimal polynomial of $M_{\mathbf{x}_i}$ is the monic generator of the elimination ideal $\mathcal{I} \cap \mathbb{R}[\mathbf{x}_i]$.

Let $p_v \in \mathbb{R}[\mathbf{x}]$ be Lagrange interpolation polynomials at the points of $V_{\mathbb{C}}(\mathcal{I})$. As observed in Lemma 2.7, we may assume that $p_v \in \mathrm{Span}_{\mathbb{R}}(\mathcal{B})$ for all $v \in V_{\mathbb{C}}(\mathcal{I})$. For a polynomial $p \in \mathrm{Span}_{\mathbb{R}}(\mathcal{B})$, $p = \sum_{i=1}^N a_i b_i$ with $a_i \in \mathbb{R}$, let $\mathrm{vec}_{\mathcal{B}}(p) := (a_i)_{i=1}^N$ denote the vector of its coefficients in $\mathcal{B}$. Set $\zeta_{\mathcal{B},v} := (b_i(v))_{i=1}^N \in \mathbb{C}^N$, the vector of evaluations at v of the polynomials in the basis $\mathcal{B}$. Observe that

$$\{\zeta_{\mathcal{B},v} \mid v \in V_{\mathbb{C}}(\mathcal{I})\} \text{ is linearly independent in } \mathbb{C}^N. \tag{2.10}$$

Indeed assume $\sum_{v \in V_{\mathbb{C}}(\mathcal{I})} \lambda_v \zeta_{\mathcal{B},v} = 0$, i.e., $\sum_{v \in V_{\mathbb{C}}(\mathcal{I})} \lambda_v b_i(v) = 0$ for $i = 1, \ldots, N$. As $\mathcal{B}$ is a basis of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$, this implies that $\sum_{v \in V_{\mathbb{C}}(\mathcal{I})} \lambda_v f(v) = 0$ for any $f \in \mathbb{R}[\mathbf{x}]$. Applying this to $f := \mathrm{Re}(p_v), \mathrm{Im}(p_v)$ we find $\lambda_v = 0 \ \forall v$.

Theorem 2.9. **(Stickelberger eigenvalue theorem)** *Let $h \in \mathbb{R}[\mathbf{x}]$. The set $\{h(v) \mid v \in V_{\mathbb{C}}(\mathcal{I})\}$ is the set of eigenvalues of M_h. More precisely,*

$$M_h^T \zeta_{\mathcal{B},v} = h(v) \zeta_{\mathcal{B},v} \quad \forall v \in V_{\mathbb{C}}(\mathcal{I}) \tag{2.11}$$

and, if $\mathcal{I}$ is radical, then

$$M_h \mathrm{vec}_{\mathcal{B}}(p_v) = h(v)\mathrm{vec}_{\mathcal{B}}(p_v) \quad \forall v \in V_{\mathbb{C}}(\mathcal{I}). \tag{2.12}$$

Proof. We first show (2.11). Indeed, $(M_h^T \zeta_{\mathcal{B},v})_j = \sum_{i=1}^N b_j(v) a_{ij}$ is equal to $h(v)b_j(v)$ (using (2.9)). Thus $h(v)$ is an eigenvalue of M_h^T with eigenvector $\zeta_{\mathcal{B},v}$. Note that $\zeta_{\mathcal{B},v} \neq 0$ by (2.10).

We now show (2.12) assuming that $\mathcal{I}$ is radical. Say, $p_v = \sum_{j=1}^N c_j b_j$, i.e., $\mathrm{vec}_{\mathcal{B}}(p_v) = (c_j)_{j=1}^N$. The i-th component of $q := M_h \mathrm{vec}_{\mathcal{B}}(p_v)$ is $q_i = \sum_{j=1}^N a_{ij} c_j$. In order to show $q_i = h(v)c_i$ for all i, it suffices to show that the polynomial $f := \sum_{i=1}^N (q_i - h(v)c_i) b_i$ belongs to $\mathcal{I}$; as $\mathcal{I}$ is radical, this holds if we can show that f vanishes on $V_{\mathbb{C}}(\mathcal{I})$. Now,

$$\begin{aligned} f &= \sum_{i=1}^N (\sum_{j=1}^N a_{ij} c_j - h(v) c_i) b_i = \sum_{j=1}^N c_j (\sum_{i=1}^N a_{ij} b_i) - h(v) \sum_{i=1}^N c_i b_i \\ &= \sum_{j=1}^N c_j (\sum_{i=1}^N a_{ij} b_i - h b_j + h b_j) - h(v) p_v \\ &\equiv \sum_{j=1}^N c_j h b_j - h(v) p_v = (h - h(v)) p_v \mod \mathcal{I} \end{aligned}$$

(using (2.9)). Thus, f vanishes $V_{\mathbb{C}}(\mathcal{I})$ and thus $f \in \mathcal{I}$.

Remains to show that any eigenvalue λ of M_h belongs to the set $h(V_{\mathbb{C}}(\mathcal{I})) := \{h(v) \mid v \in V_{\mathbb{C}}(\mathcal{I})\}$. If $\mathcal{I}$ is radical, this is clear since we have already found $|V_{\mathbb{C}}(\mathcal{I})| = N$ linearly independent eigenvectors $\zeta_{\mathcal{B},v}$ ($v \in V_{\mathbb{C}}(\mathcal{I})$) (by (2.10)). Otherwise, assume $\lambda \notin h(V_{\mathbb{C}}(\mathcal{I}))$. Then the system $h_1(x) = 0, \ldots, h_{m_0}(x) = 0, h(x) - \lambda = 0$ has no solution. By Hilbert's Nullstellensatz (Theorem 2.1), $1 \in (h_1, \ldots, h_{m_0}, h - \lambda)$. That is, $1 = \sum_{j=1}^{m_0} f_j h_j + f(h - \lambda)$ for some polynomials f_j, f. Hence,

$$I = M_1 = M_{\sum_{j=1}^{m_0} f_j h_j + f(h-\lambda)} = \sum_{j=1}^{m_0} M_{f_j h_j} + M_f(M_h - \lambda I) = M_f(M_h - \lambda I)$$

since $M_{f_j h_j} = 0$ as $f_j h_j \in \mathcal{I}$. Thus $M_h - \lambda I$ is nonsingular which means that λ is not an eigenvalue of M_h. ∎

EXAMPLE 2.10. *Consider the ideal* $\mathcal{I} = (h_1, h_2, h_3) \subseteq \mathbb{R}[\mathbf{x}, \mathbf{y}]$ *where*

$$\begin{aligned} h_1 &= \mathbf{x}^2 + 2\mathbf{y}^2 - 2\mathbf{y} \\ h_2 &= \mathbf{x}\mathbf{y}^2 - \mathbf{x}\mathbf{y} \\ h_3 &= \mathbf{y}^3 - 2\mathbf{y}^2 + \mathbf{y}. \end{aligned}$$

Obviously, $V_{\mathbb{C}}(\mathcal{I}) = \{(0,0), (0,1)\}$. *One can show that, with respect to the lexicographic order with* $\mathbf{x} > \mathbf{y}$, *the set* $\{h_1, h_2, h_3\}$ *is a Gröbner basis of*

$\mathcal{I}$. As the leading terms of h_1, h_2, h_3 are $\mathbf{x}^2, \mathbf{x}\mathbf{y}^2, \mathbf{y}^3$, the corresponding set of standard monomials is $\mathcal{B} = \{1, \mathbf{y}, \mathbf{y}^2, \mathbf{x}, \mathbf{x}\mathbf{y}\}$ and $\dim \mathbb{R}[\mathbf{x}, \mathbf{y}]/\mathcal{I} = 5$*. As* $\mathbf{x}^2\mathbf{y} \equiv -2\mathbf{y}^2 + 2\mathbf{y} \mod \mathcal{I}$*, the multiplication matrices read:*

$$M_{\mathbf{x}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & -2 & -2 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}, \quad M_{\mathbf{y}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

and their characteristic polynomials are $\det(M_{\mathbf{x}} - tI) = t^5$, $\det(M_{\mathbf{y}} - tI) = t^2(t-1)^3$.

Example 2.11. *Consider now the ideal $\mathcal{I} = (\mathbf{x}^2, \mathbf{y}^2)$ in $\mathbb{R}[\mathbf{x}, \mathbf{y}]$. Obviously, $V_{\mathbb{C}}(\mathcal{I}) = \{(0,0)\}$, $\{\mathbf{x}^2, \mathbf{y}^2\}$ is a Gröbner basis w.r.t. any monomial ordering, with corresponding set $\mathcal{B} = \{1, \mathbf{x}, \mathbf{y}, \mathbf{x}\mathbf{y}\}$ of standard monomials. Thus* $\dim \mathbb{R}[\mathbf{x}, \mathbf{y}]/\mathcal{I} = 4$*,*

$$M_{\mathbf{x}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad M_{\mathbf{y}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

both with characteristic polynomial t^4.

By Theorem 2.9, the eigenvalues of the companion matrices $M_{\mathbf{x}_i}$ are the coordinates v_i of the points $v \in V_{\mathbb{C}}(\mathcal{I})$. It is however not clear how to put these coordinates together for recovering the full vectors v. For this it is better to use the eigenvectors $\zeta_{\mathcal{B},v}$ of the transpose multiplication matrices. Recall that a square matrix M is *non-derogatory* if all its eigenspaces have dimension 1; that is, if $\dim \mathrm{Ker}(M - \lambda I) = 1$ for each eigenvalue λ of M. The next result follows directly from Theorem 2.9.

Lemma 2.12. *The following holds for a multiplication matrix M_h.*

(i) *If M_h^T is non-derogatory then $h(v)$ ($v \in V_{\mathbb{C}}(\mathcal{I})$) are pairwise distinct.*

(ii) *If $\mathcal{I}$ is radical and $h(v)$ ($v \in V_{\mathbb{C}}(\mathcal{I})$) are pairwise distinct, then M_h^T is non-derogatory.*

2.4.3. Computing $V_{\mathbb{C}}(\mathcal{I})$ with a non-derogatory multiplication matrix. Assume we can find $h \in \mathbb{R}[\mathbf{x}]$ for which the matrix M_h^T is non-derogatory. We can assume without loss of generality that the chosen basis $\mathcal{B}$ of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ contains the constant polynomial $b_1 = 1$. Let λ be an eigenvalue of M_h^T with eigenvector u. By Theorem 2.9, $\lambda = h(v)$ and u is a scalar multiple of $\zeta_{\mathcal{B},v}$ for some $v \in V_{\mathbb{C}}(\mathcal{I})$; by rescaling (i.e. replace u by u/u_1 where u_1 is the component of u indexed by $b_1 = 1$), we may assume $u = \zeta_{\mathcal{B},v}$. If $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{B}$, one can read the coordinates of v directly from the eigenvector u. Otherwise, express $\mathbf{x}_i$ as a linear combination modulo $\mathcal{I}$ of the members of $\mathcal{B}$, say, $\mathbf{x}_i = \sum_{j=1}^N c_j b_j \mod \mathcal{I}$. Then, $v_i = \sum_{j=1}^N c_j b_j(v)$ can be computed from the coordinates of the eigenvector u.

One can show that if there exists some $h \in \mathbb{R}[\mathbf{x}]$ for which M_h^T is non-derogatory, then there exists a linear such polynomial $h = \sum_{i=1}^n c_i \mathbf{x}_i$. Following the strategy of Corless, Gianni and Trager [24], one can find such h by choosing the c_i's at random. Then, with high probability, $h(v)$ ($v \in V_{\mathbb{C}}(\mathcal{I})$) are pairwise distinct. If $\mathcal{I}$ is radical then M_h^T is non-derogatory (by Lemma 2.12). If we succeed to find a non-derogatory matrix after a few trials, we can proceed to compute $V_{\mathbb{C}}(\mathcal{I})$; otherwise we are either unlucky or there exists no non-derogatory matrix. Then one possibility is to compute the radical $\sqrt{\mathcal{I}}$ of $\mathcal{I}$ using, for instance, the following characterization:

$$\sqrt{\mathcal{I}} = (h_1, \ldots, h_m, (p_1)_{red}, \ldots, (p_n)_{red})$$

where p_i is the monic generator of $\mathcal{I} \cap \mathbb{R}[\mathbf{x}_i]$ and $(p_i)_{red}$ is its square-free part. The polynomial p_i can be found in the following way: Let k_i be the smallest integer for which the set $\{[1], [\mathbf{x}_i], \ldots, [\mathbf{x}_i^{k_i}]\}$ is linearly dependent in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. Then the polynomial $\mathbf{x}_i^{k_i} + \sum_{j=0}^{k_i-1} c_j \mathbf{x}_i^j$ lies in $\mathcal{I}$ for some scalars c_j and, by the minimality of k_i, it generates $\mathcal{I} \cap \mathbb{R}[\mathbf{x}_i]$.

EXAMPLE 2.10 (continued). None of $M_{\mathbf{x}}^T$, $M_{\mathbf{y}}^T$ is non-derogatory. Indeed, 0 is the only eigenvalue of $M_{\mathbf{x}}^T$ whose corresponding eigenspace is $\operatorname{Ker} M_{\mathbf{x}}^T = \{u \in \mathbb{R}^5 \mid u_2 = u_3, u_4 = u_5 = 0\}$ with dimension 2 and spanned by $\zeta_{\mathcal{B},(0,0)}$ and $\zeta_{\mathcal{B},(0,1)}$. The eigenspace of $M_{\mathbf{y}}^T$ for eigenvalue 0 is $\operatorname{Ker} M_{\mathbf{y}}^T = \{u \in \mathbb{R}^5 \mid u_2 = u_3 = u_5 = 0\}$ with dimension 2 and spanned by $\zeta_{\mathcal{B},(0,0)}$ and $(0,0,0,1,0)^T$. The eigenspace with eigenvalue 1 is $\operatorname{Ker}(M_{\mathbf{y}}^T - I) = \{u \in \mathbb{R}^5 \mid u_1 = u_2 = u_3, u_4 = u_5\}$ also with dimension 2 and spanned by $\zeta_{\mathcal{B},(0,1)}$ and $(0,0,0,1,1)^T$. Thus, for $h = \mathbf{y}$, this gives an example where $h(v)$ ($v \in V_{\mathbb{C}}(\mathcal{I})$) are pairwise distinct, yet the matrix M_h^T is not non-derogatory. On the other hand, for $h = 2\mathbf{x} + 3\mathbf{y}$,

$$M_h = 2M_{\mathbf{x}} + 3M_{\mathbf{y}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & -3 & 4 & 4 \\ 0 & 3 & 6 & -4 & -4 \\ 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & 3 & 3 \end{pmatrix}$$

and M_h^T is non-derogatory. Indeed, M_h^T has two eigenvalues 0, 3. The eigenspace for the eigenvalue 0 is spanned by $\zeta_{\mathcal{B},(0,0)}$, permitting to extract the root $v = (0,0)$, and the eigenspace for the eigenvalue 3 is spanned by $\zeta_{\mathcal{B},(0,1)}$, permitting to extract the root $v = (0,1)$. □

EXAMPLE 2.11 (continued). In this example *every* matrix M_h^T is derogatory. Indeed, say $h = a + b\mathbf{x} + c\mathbf{y} + d\mathbf{x}^2 + e\mathbf{xy} + \ldots$. Then,

$$M_h = \begin{pmatrix} a & b & c & e \\ 0 & a & 0 & c \\ 0 & 0 & a & b \\ 0 & 0 & 0 & a \end{pmatrix}.$$

Thus a is the only eigenvalue of M_h^T with eigenvector space of dimension at least 2. □

2.4.4. Root counting with Hermite's quadratic form. We recall here how to use the multiplication matrices in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ for counting the complex/real roots of $\mathcal{I}$. (See e.g. [8] for details.)

Multiplicity of roots. We first recall some basic facts about multiplicities of roots. Let $\mathcal{I}$ be a zero-dimensional ideal in $\mathbb{R}[\mathbf{x}]$. When $\mathcal{I}$ is radical, $|V_{\mathbb{C}}(\mathcal{I})| = \dim \mathbb{R}[\mathbf{x}]/\mathcal{I}$ and each root $v \in V_{\mathbb{C}}(\mathcal{I})$ has single multiplicity. When $\mathcal{I}$ is not radical, we have $|V_{\mathbb{C}}(\mathcal{I})| < \dim \mathbb{R}[\mathbf{x}]/\mathcal{I} =: N$. There is a well defined notion of multiplicity $\text{mult}(v)$ for each root $v \in V_{\mathbb{C}}(\mathcal{I})$ such that $N = \sum_{v \in V_{\mathbb{C}}(\mathcal{I})} \text{mult}(v)$. We just sketch the idea and refer e.g. to [40, Chap. 2] or [26] for details.

One can verify that $\mathcal{I}$ can be written as $\mathcal{I} = \bigcap\limits_{v \in V(\mathcal{I})} \mathcal{I}_v$, where

$$\mathcal{I}_v := \{f \in \mathbb{R}[\mathbf{x}] \mid fg \in \mathcal{I} \text{ for some } g \in \mathbb{R}[\mathbf{x}] \text{ with } g(v) \neq 0\}$$

are (primary) ideals with $V_{\mathbb{C}}(\mathcal{I}_v) = \{v\}$ for all $v \in V_{\mathbb{C}}(\mathcal{I})$. Moreover,

$$\mathbb{R}[\mathbf{x}]/\mathcal{I} \simeq \prod_{v \in V_{\mathbb{C}}(\mathcal{I})} \mathbb{R}[\mathbf{x}]/\mathcal{I}_v. \tag{2.13}$$

Then, $\text{mult}(v) := \dim \mathbb{R}[\mathbf{x}]/\mathcal{I}_v$ is called the *multiplicity* of $v \in V_{\mathbb{C}}(\mathcal{I})$, and we have $N = \sum_{v \in V_{\mathbb{C}}(\mathcal{I})} \text{mult}(v)$.

Let us briefly compare this notion of multiplicity to the classical algebraic and geometric multiplicities of eigenvalues. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of an $N \times N$ matrix A and say its characteristic polynomial reads $\det(A - \mathbf{t}I) = \prod_{i=1}^k (\mathbf{t} - \lambda_i)^{m_i}$. Then m_i is the *algebraic multiplicity* of λ_i while $m_i' := \dim \text{Ker}(A - \lambda_i I)$ is its *geometric multiplicity*. Then, $\sum_{i=1}^k m_i = N$. Moreover, $m_i' \leq m_i$ for all i. Thus the sum of the geometric multiplicities is equal to N precisely when A has a full set of N linearly independent eigenvectors. In fact, the algebraic multiplicity of λ_i is equal to the dimension of its associated generalized eigenspace, i.e., $m_i = \dim\{u \in \mathbb{R}^N \mid (A - \lambda_i I)^r u = 0 \text{ for some } r \geq 1\}$.

Given a polynomial $h \in \mathbb{R}[\mathbf{x}]$, consider the multiplication matrix M_h. The characteristic polynomial of M_h is equal to

$$\det(M_h - \mathbf{t}I) = \prod_{v \in V_{\mathbb{C}}(\mathcal{I})} (t - h(v))^{\text{mult}(v)},$$

which follows using (2.13), the fact that $h(v)$ is the only eigenvalue of the multiplication operator by h in $\mathbb{R}[\mathbf{x}]/I_v$, and $\text{mult}(v) = \dim \mathbb{R}[\mathbf{x}]/I_v$. Therefore,

$$\text{Tr}(M_h) = \sum_{v \in V_{\mathbb{C}}(\mathcal{I})} \text{mult}(v) h(v). \tag{2.14}$$

Summarizing, for $v \in V_{\mathbb{C}}(\mathcal{I})$, we have three notions of multiplicity:
- the multiplicity $\mathrm{mult}(v)$ of v as a root of $\mathcal{I}$,
- the algebraic multiplicity of $h(v)$ as eigenvalue of M_h^T, i.e., the exponent of $t - h(v)$ in the characteristic polynomial of M_h^T,
- and the geometric multiplicity of $h(v)$, i.e., the dimension of the (usual) eigenspace corresponding to the eigenvalue $h(v)$ of M_h^T.

LEMMA 2.13. *Assume the values $h(v)$ $(v \in V_{\mathbb{C}}(\mathcal{I}))$ are pairwise distinct. Then, mult(v) is equal to the algebraic multiplicity of $h(v)$. Moreover, if $\mathcal{I}$ is radical, then mult$(v) = 1$ and is equal to both the algebraic and the geometric multiplicities of $h(v)$ as eigenvalue of M_h^T.*

EXAMPLE 2.10 (continued). One may show that $\mathcal{I}_{(0,0)} = (x^2, y)$, $\mathcal{I}_{(0,1)} = (x^2 + 2(y-1), x(y-1), (y-1)^2)$, so that $\dim \mathbb{R}[\mathbf{x}]/\mathcal{I}_{(0,0)} = 2 = \mathrm{mult}(0,0)$, $\dim \mathbb{R}[\mathbf{x}]/\mathcal{I}_{(0,1)} = 3 = \mathrm{mult}(0,1)$. The polynomial $h := y$ takes distinct values at $v = (0,0)$ and $v = (0,1)$ and its characteristic polynomial is $t^2(t-1)^3$. □

Root counting. Given a polynomial $h \in \mathbb{R}[\mathbf{x}]$, consider the following symmetric bilinear form:

$$\begin{aligned} S_h : \quad & \mathbb{R}[\mathbf{x}]/\mathcal{I} \times \mathbb{R}[\mathbf{x}]/\mathcal{I} \longrightarrow \mathbb{R} \\ & (f \mod \mathcal{I}, g \mod \mathcal{I}) \mapsto \mathrm{Tr}(M_{fgh}) \end{aligned}$$

sometimes called the *Hermite form*, since Hermite investigated it in the univariate case. In view of (2.14),

$$\mathrm{Tr}(M_{fgh}) = \sum_{v \in V_{\mathbb{C}}(\mathcal{I})} \mathrm{mult}(v) f(v) g(v) h(v).$$

The matrix of S_h with respect to a given basis $\mathcal{B} = \{b_1, \ldots, b_N\}$ of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ is the $\mathcal{B} \times \mathcal{B}$ matrix with entries $\mathrm{Tr}(M_{b_i b_j h}) = \sum_{v \in V_{\mathbb{C}}(\mathcal{I})} \mathrm{mult}(v) b_i(v) b_j(v) h(v)$ (for $i, j = 1, \ldots, n$); that is,

$$S_h = \sum_{v \in V_{\mathbb{C}}(\mathcal{I})} \mathrm{mult}(v) h(v) \zeta_{\mathcal{B},v} \zeta_{\mathcal{B},v}^T. \tag{2.15}$$

Recall that $\zeta_{\mathcal{B},v} = (b(v))_{b \in \mathcal{B}}$. As S_h is an $N \times N$ real symmetric matrix, it has N real eigenvalues. Denote by $\sigma_+(S_h)$ (resp., $\sigma_-(S_h)$) the number of positive (resp., negative) eigenvalues of S_h.

THEOREM 2.14. *If $\mathcal{I} \subseteq \mathbb{R}[\mathbf{x}]$ is a 0-dimensional ideal and $h \in \mathbb{R}[\mathbf{x}]$, then*

$$\mathrm{rank}(S_h) = |\{v \in V_{\mathbb{C}}(\mathcal{I}) \mid h(v) \neq 0\}|,$$

$$\sigma_+(S_h) - \sigma_-(S_h) = |\{v \in V_{\mathbb{R}}(\mathcal{I}) \mid h(v) > 0\}| - |\{v \in V_{\mathbb{R}}(\mathcal{I}) \mid h(v) < 0\}|.$$

Proof. As $\mathcal{I} \subseteq \mathbb{R}[\mathbf{x}]$, its set of roots can be partitioned into $V_{\mathbb{C}}(\mathcal{I}) = V_{\mathbb{R}}(\mathcal{I}) \cup T \cup \bar{T}$, where $V_{\mathbb{R}}(\mathcal{I}) = V_{\mathbb{C}}(\mathcal{I}) \cap \mathbb{R}^n$, $\bar{T} := \{\bar{v} \mid v \in T\}$, and $T \cup \bar{T} = V_{\mathbb{C}}(\mathcal{I}) \setminus V_{\mathbb{R}}(\mathcal{I})$. Set

$$\rho_+ := |\{v \in V_{\mathbb{R}}(\mathcal{I}) \mid h(v) > 0\}|, \ \rho_- := |\{v \in V_{\mathbb{R}}(\mathcal{I}) \mid h(v) < 0\}|,$$

$$\rho_T := |\{v \in T \mid h(v) \neq 0\}|.$$

We now prove that $\operatorname{rank}(S_h) = \rho_+ + \rho_- + 2\rho_T$, $\sigma_+ := \sigma_+(S_h) = \rho_+ + \rho_T$, and $\sigma_- := \sigma_-(S_h) = \rho_- + \rho_T$.

Let U denote the $N \times |V_{\mathbb{C}}(\mathcal{I})|$ matrix whose columns are the vectors $\zeta_{\mathcal{B},v}$ $(v \in V_{\mathbb{C}}(\mathcal{I}))$ and let D be the diagonal matrix with diagonal entries $\operatorname{mult}(v)h(v)$ $(v \in V_{\mathbb{C}}(\mathcal{I}))$. As U has full column rank, one can complete it to a nonsingular $N \times N$ complex matrix V. Similarly complete D to a $N \times N$ diagonal matrix D_0 by adding zeros. Then, $S_h = U^T D U = V D_0 V^T$ by (2.15). As V is nonsingular, we deduce that the rank of S_h is equal to the rank of D_0 and thus to the number of $v \in V_{\mathbb{C}}(\mathcal{I})$ with $h(v) \neq 0$. This shows $\operatorname{rank}(S_h) = \rho_+ + \rho_- + 2\rho_T$.

Using the above partition of $V_{\mathbb{C}}(\mathcal{I})$, we can write

$$U = \begin{pmatrix} A & B & \bar{B} \end{pmatrix}, \ D = \operatorname{diag}(a \ b \ \bar{b})$$

where A is real valued, A is $N \times |V_{\mathbb{R}}(\mathcal{I})|$, B is $N \times |T|$, $a \in \mathbb{R}^{|V_{\mathbb{R}}(\mathcal{I})|}$, $b \in \mathbb{C}^{|T|}$. Hence,

$$S_h = \underbrace{A \operatorname{diag}(a) A^T}_{A_+A_+^T - A_-A_-^T} + \underbrace{B \operatorname{diag}(b) B^T + \overline{B \operatorname{diag}(b) B^T}}_{EE^T - FF^T}.$$

Here, A_+, A_-, E, F are real valued, A_+ (res. A_-) has ρ_+ (resp. ρ_-) columns and E, F have ρ_T columns. Therefore, we can write

$$S_h = \underbrace{(A_+A_+^T + EE^T)}_{P} - \underbrace{(A_-A_-^T + FF^T)}_{Q} = P - Q$$

where $P, Q \succeq 0$. On the other hand, let $\{u_1, \ldots, u_{\sigma_+}, v_1, \ldots, v_{\sigma_-}, w_1, \ldots, w_r\}$ be an orthonormal basis of $\mathbb{R}^N$, where $\{u_1, \ldots, u_{\sigma_+}\}$ are the eigenvectors for the positive eigenvalues λ_i of S_h, $\{v_1, \ldots, v_{\sigma_-}\}$ are the eigenvectors for the negative eigenvalues μ_i of S_h and $\{w_1, \ldots, w_r\}$ form a basis of $\operatorname{Ker}(S_h)$. Then, $S = S_+ - S_-$, where $S_+ := \sum_{i=1}^{\sigma_+} \lambda_i u_i u_i^T \succeq 0$, $S_- := \sum_{i=1}^{\sigma_-} (-\mu_i) v_i v_i^T \succeq 0$. We have: $S_+ - S_- = P - Q$. Therefore, $S_+ = (S_- + P) - Q$, which implies $\operatorname{Ker}(S_- + P) \subseteq \operatorname{Ker} S_+$ and thus $\operatorname{Span}(u_1, \ldots, u_{\sigma_+}) \cap \operatorname{Ker} P = \{0\}$, giving

$$\sigma_+ \leq \operatorname{rank}(P) \leq \rho_+ + \rho_T.$$

Similarly, $S_- = (S_+ + Q) - P$, implying $\operatorname{Span}(v_1, \ldots, v_{\sigma_-}) \cap \operatorname{Ker} Q = \{0\}$ and thus

$$\sigma_- \leq \operatorname{rank}(Q) \leq \rho_- + \rho_T.$$

This gives: $\operatorname{rank}(S_h) = \sigma_+ + \sigma_- \leq \rho_+ + \rho_- + 2\rho_T$. Hence equality holds, implying $\sigma_+ = \rho_+ + \rho_T$ and $\sigma_- = \rho_- + \rho_T$. □

Using the Hermite form S_1 for the constant polynomial $h = 1$, we can count $V_{\mathbb{C}}(\mathcal{I})$ and $V_{\mathbb{R}}(\mathcal{I})$.

COROLLARY 2.15. *For the polynomial $h = 1$,*

$$\operatorname{rank}(S_1) = |V_{\mathbb{C}}(\mathcal{I})|, \;\; \sigma_+(S_1) - \sigma_-(S_1) = |V_{\mathbb{R}}(\mathcal{I})|.$$

2.5. Border bases and commuting multiplication matrices. As we saw in the previous sections, the multiplication operators $m_{\mathbf{x}_i}$ in the quotient space $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ play a central role to compute the variety $V_{\mathbb{C}}(\mathcal{I})$. We now present a result of Mourrain [110] which relates commuting (abstract) multiplication matrices to border bases of polynomial ideals. We need some definitions.

Let $\mathcal{B} \subseteq \mathbb{T}^n$ be a finite set of monomials. Then $\mathcal{B}$ is said to be *connected to 1* if $1 \in \mathcal{B}$ and any non-constant monomial $m \in \mathcal{B}$ can be written $m = \mathbf{x}_{i_1} \cdots \mathbf{x}_{i_k}$, where $\mathbf{x}_{i_1}, \mathbf{x}_{i_1}\mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_1} \cdots \mathbf{x}_{i_k} \in \mathcal{B}$. Any monomial of the form $\mathbf{x}_i b$, where $b \in \mathcal{B}$ and $\mathbf{x}_i b \notin \mathcal{B}$, is called a *border monomial* of $\mathcal{B}$; their set forms the *border* of $\mathcal{B}$, denoted as $\partial\mathcal{B}$ and defined by

$$\partial\mathcal{B} = \{\mathbf{x}_i b \mid b \in \mathcal{B}, \; i = 1, \ldots, n\} \setminus \mathcal{B}.$$

Say $\mathcal{B} := \{b_1, \ldots, b_N\}$. Assume that, for each border monomial $\mathbf{x}_i b_j \in \partial\mathcal{B}$, we are given a polynomial of the form

$$g^{(ij)} := \mathbf{x}_i b_j - \sum_{h=1}^{N} a_h^{(ij)} b_h \quad \text{where} \;\; a_h^{(ij)} \in \mathbb{R}.$$

Thus, $g^{(ij)}$ permits to express the border monomial $\mathbf{x}_i b_j$ linearly in terms of the monomials in the set $\mathcal{B}$. The set

$$F := \{g^{(ij)} \mid i = 1, \ldots, n, \; j = 1, \ldots, N \;\; \text{with } \mathbf{x}_i b_j \in \partial\mathcal{B}\} \tag{2.16}$$

is called a *rewriting family* for $\mathcal{B}$ [110] (or as a *border prebasis* [70]). Indeed, it permits to 'rewrite' in $\operatorname{Span}(\mathcal{B})$ and modulo the ideal (F), all monomials in the border set $\partial\mathcal{B}$, then inductively all monomials in $\partial(\partial\mathcal{B})$, $\partial(\partial(\partial\mathcal{B}))$, etc. When $1 \in \mathcal{B}$, one can easily verify that $\mathcal{B}$ is a generating set for the quotient space $\mathbb{R}[\mathbf{x}]/(F)$, i.e. all monomials of $\mathbb{T}^n$ can be rewritten in $\operatorname{Span}(\mathcal{B})$ modulo (F). In general such rewriting might not be unique.

When $\mathcal{B}$ is connected to 1, Theorem 2.16 below characterizes unicity of rewriting, i.e. the case when $\mathcal{B}$ is a basis of $\mathbb{R}[\mathbf{x}]/(F)$; in that case, the set F is said to be a *border basis* of the ideal (F).

For this, for each $i = 1, \ldots, n$, consider the linear operator:

$$\begin{array}{rlll} \chi_i : & \mathrm{Span}(\mathcal{B}) & \to & \mathrm{Span}(\mathcal{B}) \\ & b_j & \mapsto & \chi_i(b_j) = \begin{cases} \mathbf{x}_i b_j & \text{if } \mathbf{x}_i b_j \in \mathcal{B}, \\ \sum_{h=1}^N a_h^{(ij)} b_h & \text{if } \mathbf{x}_i b_j \in \partial\mathcal{B} \end{cases} \end{array} \tag{2.17}$$

extended to $\mathrm{Span}(\mathcal{B})$ by linearity. When $\mathcal{B}$ is a basis of $\mathbb{R}[\mathbf{x}]/(F)$, χ_i corresponds to the "multiplication operator by $\mathbf{x}_i$" from $\mathbb{R}[\mathbf{x}]/(F)$ to $\mathbb{R}[\mathbf{x}]/(F)$ and thus the operators $\chi_1, \ldots, \chi_n$ commute pairwise. The next result of [110] shows that the converse implication holds when $\mathcal{B}$ is connected to 1; this was also proved in [70] when $\mathcal{B}$ is closed under taking divisors.

Theorem 2.16. *[110] Let $\mathcal{B} \subseteq \mathbb{T}^n$ be a finite set of monomials which is connected to 1, let F be a rewriting family for $\mathcal{B}$ as in* (2.16)*, and let $\chi_1, \ldots, \chi_n$ be defined as in* (2.17)*. The set $\mathcal{B}$ is a basis of the quotient space $\mathbb{R}[\mathbf{x}]/(F)$ if and only if the operators $\chi_1, \ldots, \chi_n$ commute pairwise.*

As we will see in Section 5.3.2, this result will be useful to prove results about flat extensions of moment matrices. We now give a proof of Theorem 2.16, following the treatment in [70].

Proof. Let $\mathcal{B} = \{b_1, \ldots, b_N\} \subseteq \mathbb{T}^n$ be connected to 1 with, say, $b_1 := 1$; let F be a rewriting family for $\mathcal{B}$ as in (2.16), and let $\chi_1, \ldots, \chi_n$ be the linear operators from $\mathrm{Span}(\mathcal{B})$ to $\mathrm{Span}(\mathcal{B})$ defined in (2.17). Assume that $\chi_1, \ldots, \chi_n$ commute pairwise. We show that $\mathcal{B}$ is a linear basis of $\mathbb{R}[\mathbf{x}]/(F)$. As $\mathcal{B}$ is generating for $\mathbb{R}[\mathbf{x}]/(F)$, it suffices to show that $\dim \mathbb{R}[\mathbf{x}]/(F) \geq |B|$.

As the χ_i's commute, the operator $f(\chi) := f(\chi_1, \ldots, \chi_n)$ is well defined for any polynomial $f \in \mathbb{R}[\mathbf{x}]$. Then $\mathbb{R}[\mathbf{x}]$ acts on $\mathrm{Span}(\mathcal{B})$ by

$$(f, p) \in \mathbb{R}[\mathbf{x}] \times \mathrm{Span}(\mathcal{B}) \mapsto f(\chi)(p) \in \mathrm{Span}(\mathcal{B}).$$

We can define the mapping

$$\begin{array}{rlll} \varphi : & \mathbb{R}[\mathbf{x}] & \to & \mathrm{Span}(\mathcal{B}) \\ & f & \mapsto & f(\chi)(b_1) \end{array}$$

(Recall that $b_1 = 1$). Note that $\varphi(fg) = f(\chi)(g(\chi)(b_1)) = f(\chi)(\varphi(g))$ for all $f, g \in \mathbb{R}[\mathbf{x}]$. Hence $\mathrm{Ker}\,\varphi$ is an ideal in $\mathbb{R}[\mathbf{x}]$. We collect some further properties of φ.

Lemma 2.17. *$\varphi(b_k) = b_k$ for all $b_k \in \mathcal{B}$.*

Proof. The proof is by induction on the degree of $b_k \in \mathcal{B}$. For $b_1 = 1$, $b_1(\chi)$ is the identity and thus $\varphi(b_1) = b_1$. Consider $b_k \in \mathcal{B}$. As $\mathcal{B}$ is connected to 1, $b_k = \mathbf{x}_i b_j$ for some $b_j \in \mathcal{B}$. By the induction assumption, $\varphi(b_j) = b_j$. Then, $\varphi(b_k) = \chi_i(\varphi(b_j)) = \chi_i(b_j)$ is equal to $\mathbf{x}_i b_j = b_k$ by the definition of χ_i. □

LEMMA 2.18. $(F) \subseteq \text{Ker } \varphi$.

Proof. It suffices to show that $\varphi(g^{(ij)}) = 0$ whenever $\mathbf{x}_i b_j \in \partial\mathcal{B}$. We have $\varphi(g^{(ij)}) = \varphi(\mathbf{x}_i b_j) - \sum_{h=1}^N a_h^{(ij)} \varphi(b_h)$, where $\varphi(\mathbf{x}_i b_j) = \chi_i(b_j) = \sum_{h=1}^N a_h^{(ij)} b_h$ (by the definition of χ_i) and $\varphi(b_h) = b_h$ for all h (by Lemma 2.17). This implies $\varphi(g^{(ij)}) = 0$. □

As $(F) \subseteq \text{Ker}\,\varphi$, we obtain $\dim \mathbb{R}[\mathbf{x}]/(F) \geq \dim \mathbb{R}[\mathbf{x}]/\text{Ker } \varphi = |\mathcal{B}|$, where the last equality follows from the fact that φ is onto (by Lemma 2.17). This gives the desired inequality $\dim \mathbb{R}[\mathbf{x}]/(F) \geq |\mathcal{B}|$, thus showing that $\mathcal{B}$ is a linear basis of $\mathbb{R}[\mathbf{x}]/(F)$, which concludes the proof. □

Part 1: Sums of Squares and Moments

3. Positive polynomials and sums of squares.

3.1. Some basic facts. A concept which will play a central role in the paper is the following notion of *sum of squares*. A polynomial p is said to be a *sum of squares of polynomials*, sometimes abbreviated as *'p is SOS'*, if p can be written as $p = \sum_{j=1}^{m} u_j^2$ for some $u_1, \ldots, u_m \in \mathbb{R}[\mathbf{x}]$. Given $p \in \mathbb{R}[\mathbf{x}]$ and $S \subseteq \mathbb{R}^n$, the notation '$p \geq 0$ on S' means '$p(x) \geq 0$ for all $x \in S$', in which case we say that p is nonnegative on S; analogously, $p > 0$ on S means that p is positive on S. We begin with some simple properties of sums of squares.

Lemma 3.1. *If $p \in \mathbb{R}[\mathbf{x}]$ is a sum of squares, then $\deg(p)$ is even and any decomposition $p = \sum_{j=1}^{m} u_j^2$ where $u_j \in \mathbb{R}[\mathbf{x}]$ satisfies $\deg(u_j) \leq \deg(p)/2$ for all j.*

Proof. Assume p is SOS. Then $p(x) \geq 0$ for all $x \in \mathbb{R}^n$ and thus $\deg(p)$ must be even, say $\deg(p) = 2d$. Write $p = \sum_{j=1}^{m} u_j^2$ and let $k := \max_j \deg(u_j)$. Assume $k \geq d+1$. Write each $u_j = \sum_\alpha u_{j,\alpha}\mathbf{x}^\alpha$ as $u_j = a_j + b_j$, where $b_j := \sum_{\alpha||\alpha|=k} u_{j,\alpha}\mathbf{x}^\alpha$ and $a_j := u_j - b_j$. Then $p - \sum_j a_j^2 - 2a_jb_j = \sum_j b_j^2$. Here $\sum_j b_j^2$ is a homogeneous polynomial of degree $2k \geq 2d+2$, while $p - \sum_j a_j^2 - 2a_jb_j$ is a polynomial of degree $\leq 2k-1$, which yields a contradiction. This shows $\deg(u_j) \leq d$ for all j. □

Lemma 3.2. *Let p be a homogeneous polynomial of degree $2d$. If p is SOS, then p is a sum of squares of homogeneous polynomials (each of degree d).*

Proof. Assume $p = \sum_{j=1}^{m} u_j^2$ where $u_j \in \mathbb{R}[\mathbf{x}]$. Write $u_j = a_j + b_j$ where a_j is the sum of the terms of degree d of u_j and thus $\deg(b_j) \leq d-1$. Then, $p - \sum_{j=1}^{m} a_j^2 = \sum_{j=1}^{m} b_j^2 + 2a_jb_j$ is equal to 0, since otherwise the right hand side has degree $\leq 2d-1$ and the left hand side is homogeneous of degree $2d$. □

Lemma 3.3. *Consider a polynomial $p \in \mathbb{R}[\mathbf{x}]$ and its homogenization $\tilde{p} \in \mathbb{R}[\mathbf{x}, \mathbf{x}_{n+1}]$. Then, $p \geq 0$ on $\mathbb{R}^n$ (resp., p SOS) $\Longleftrightarrow$ $\tilde{p} \geq 0$ on $\mathbb{R}^{n+1}$ (resp., $\tilde{p}$ SOS).*

Proof. The 'if part' follows from the fact that $p(x) = \tilde{p}(x,1)$ for all $x \in \mathbb{R}^n$. Conversely, if $p \geq 0$ on $\mathbb{R}^n$ then $d := \deg(p)$ is even and $\tilde{p}(x, x_{n+1}) = x_{n+1}^d \tilde{p}(x/x_{n+1}, 1) = x_{n+1}^d p(x/x_{n+1}) \geq 0$ whenever $x_{n+1} \neq 0$. Thus $\tilde{p} \geq 0$ by continuity. An analogous argument shows that, if $p = \sum_j u_j^2$ with $u_j \in \mathbb{R}[\mathbf{x}]$, then $\tilde{p} = \sum_j \tilde{u}_j^2$, where $\tilde{u}_j$ is the homogenization of u_j. □

3.2. Sums of squares and positive polynomials: Hilbert's result. Throughout the paper,

$$\mathcal{P}_n := \{p \in \mathbb{R}[\mathbf{x}] \mid p(x) \geq 0 \ \forall x \in \mathbb{R}^n\} \tag{3.1}$$

denotes the set of nonnegative polynomials on $\mathbb{R}^n$ (also called *positive semidefinite* polynomials in the literature) and

$$\Sigma_n := \{p \in \mathbb{R}[\mathbf{x}] \mid p \text{ SOS}\} \tag{3.2}$$

is the set of polynomials that are sums of squares; we sometimes omit the index n and simply write $\mathcal{P} = \mathcal{P}_n$ and $\Sigma = \Sigma_n$ when there is no danger of confusion on the number of variables. We also set

$$\mathcal{P}_{n,d} := \mathcal{P}_n \cap \mathbb{R}[\mathbf{x}]_d, \ \Sigma_{n,d} := \Sigma_n \cap \mathbb{R}[\mathbf{x}]_d.$$

Obviously any polynomial which is SOS is nonnegative on $\mathbb{R}^n$; that is,

$$\Sigma_n \subseteq \mathcal{P}_n, \ \Sigma_{n,d} \subseteq \mathcal{P}_{n,d}. \tag{3.3}$$

As is well known (cf. Lemma 3.5), equality holds in (3.3) for $n = 1$ (i.e. for univariate polynomials), but the inclusion $\Sigma_n \subseteq \mathcal{P}_n$ is strict for $n \geq 2$. The following celebrated result of Hilbert [64] classifies the pairs (n, d) for which equality $\Sigma_{n,d} = \mathcal{P}_{n,d}$ holds.

THEOREM 3.4. **Hilbert's theorem.** $\Sigma_{n,d} = \mathcal{P}_{n,d} \iff n = 1$*, or* $d = 2$*, or* $(n, d) = (2, 4)$*.*

We give below the arguments for the equality $\Sigma_{n,d} = \mathcal{P}_{n,d}$ in the two cases $n = 1$, or $d = 2$, which are simple and which were already well known in the late 19th century. In his paper [64] David Hilbert proved that $\mathcal{P}_{2,4} = \Sigma_{2,4}$; moreover he proved that any nonnegative polynomial in $n = 2$ variables with degree 4 is a sum of *three* squares; equivalently, any nonnegative ternary quartic form is a sum of *three* squares. Choi and Lam [20] gave a relatively simple proof for the equality $\mathcal{P}_{2,4} = \Sigma_{2,4}$, based on geometric arguments about the cone $\Sigma_{2,4}$; their proof shows a decomposition into *five* squares. Powers et al. [129] found a new approach to Hilbert's theorem and gave a proof of the *three* squares result in the nonsingular case.

LEMMA 3.5. *Any nonnegative univariate polynomial is a sum of two squares.*

Proof. Assume p is a univariate polynomial and $p \geq 0$ on $\mathbb{R}$. Then the roots of p are either real with even multiplicity, or appear in complex conjugate pairs. Thus $p = c\prod_{i=1}^{r}(\mathbf{x} - a_i)^{2r_i} \cdot \prod_{j=1}^{s}((\mathbf{x} - b_j)^2 + c_j^2)^{s_j}$ for some scalars $a_i, b_j, c_j, c \in \mathbb{R}$, $c > 0$, and $r, s, r_i, s_j \in \mathbb{N}$. This shows that p is SOS. To see that p can be written as a sum of two squares, use the identity $(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2$ (for $a, b, c, d \in \mathbb{R}$). □

LEMMA 3.6. *Any nonnegative quadratic polynomial is a sum of squares.*

Proof. Let $p \in \mathbb{R}[\mathbf{x}]_2$ of the form $p = \mathbf{x}^T Q\mathbf{x} + 2c^T\mathbf{x} + b$, where Q is a symmetric $n \times n$ matrix, $c \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Its homogenization is $\tilde{p} = \mathbf{x}^T Q\mathbf{x} + 2\mathbf{x}_{n+1}c^T\mathbf{x} + b\mathbf{x}_{n+1}^2$, thus of the form $\tilde{p} = \tilde{\mathbf{x}}^T\tilde{Q}\tilde{\mathbf{x}}$, after setting

$\tilde{\mathbf{x}} := \begin{pmatrix} \mathbf{x} \\ \mathbf{x}_{n+1} \end{pmatrix}$ and $\tilde{Q} := \begin{pmatrix} Q & c \\ c^T & b \end{pmatrix}$. By Lemma 3.3, $\tilde{p} \geq 0$ on $\mathbb{R}^{n+1}$ and thus the matrix $\tilde{Q}$ is positive semidefinite. Therefore, $\tilde{Q} = \sum_j u^{(j)}(u^{(j)})^T$ for some $u^{(j)} \in \mathbb{R}^{n+1}$, which gives $\tilde{p} = \sum_j (\sum_{i=1}^{n+1} u_i^{(j)} \mathbf{x}_i)^2$ is SOS and thus p too is SOS (by Lemma 3.3 again). □

According to Hilbert's result (Theorem 3.4), for any pair $(n, d) \neq (2, 4)$ with $n \geq 2$, $d \geq 4$ even, there exists a polynomial in $\mathcal{P}_{n,d} \setminus \Sigma_{n,d}$. Some well known such examples include the Motzkin and Robinson polynomials described below.

Example 3.7. *The polynomial $p := \mathbf{x}_1^2\mathbf{x}_2^2(\mathbf{x}_1^2 + \mathbf{x}_2^2 - 3) + 1$, known as the* ***Motzkin polynomial****, belongs to $\mathcal{P}_{2,6} \setminus \Sigma_{2,6}$. Indeed, $p(x_1, x_2) \geq 0$ if $x_1^2 + x_2^2 \geq 3$. Otherwise, set $x_3^2 := 3 - x_1^2 - x_2^2$. By the arithmetic geometric mean inequality, we have $\frac{x_1^2+x_2^2+x_3^2}{3} \geq \sqrt[3]{x_1^2x_2^2x_3^2}$, giving again $p(x_1, x_2) \geq 0$. One can verify directly that p cannot be written as a sum of squares of polynomials. Indeed, assume $p = \sum_k u_k^2$, where $u_k = a_k\mathbf{x}_1^3 + b_k\mathbf{x}_1^2\mathbf{x}_2 + c_k\mathbf{x}_1\mathbf{x}_2^2 + d_k\mathbf{x}_2^3 + e_k\mathbf{x}_1^2 + f_k\mathbf{x}_1\mathbf{x}_2 + g_k\mathbf{x}_2^2 + h_k\mathbf{x}_1 + i_k\mathbf{x}_2 + j_k$ for some scalars $a_k, \ldots, j_k \in \mathbb{R}$. Looking at the coefficient of $\mathbf{x}_1^6$ in p, we find $0 = \sum_k a_k^2$, giving $a_k = 0$ for all k; analogously $d_k = 0$ for all k. Next, looking at the coefficient of $\mathbf{x}_1^4$ and $\mathbf{x}_2^4$ yields $e_k = g_k = 0$ for all k; then looking at the coefficient of $\mathbf{x}_1^2, \mathbf{x}_2^2$ yields $h_k = i_k = 0$ for all k; finally the coefficient of $\mathbf{x}_1^2\mathbf{x}_2^2$ in p is equal to $-3 = \sum_k f_k^2$, yielding a contradiction. Note that this argument shows in fact that $p - \rho$ is not a sum of squares for any scalar $\rho \in \mathbb{R}$.*

Therefore the ***homogeneous Motzkin form*** *$M := \mathbf{x}_1^2\mathbf{x}_2^2(\mathbf{x}_1^2 + \mathbf{x}_2^2 - 3\mathbf{x}_3^2) + \mathbf{x}_3^6$ is nonnegative but not a sum of squares.*

The polynomial $p := \mathbf{x}_1^6+\mathbf{x}_2^6+\mathbf{x}_3^6-\sum_{1\leq i<j\leq 3}(\mathbf{x}_i^2\mathbf{x}_j^2(\mathbf{x}_i^2+\mathbf{x}_j^2))+3\mathbf{x}_1^2\mathbf{x}_2^2\mathbf{x}_3^2$, known as the ***Robinson form****, is nonnegative but not a sum of squares. See e.g. [139] for details.*

We refer to Reznick [139] for a nice overview and historic discussion of Hilbert's results. More examples of positive polynomials that are not sums of squares can be found e.g. in the recent papers [19], [140] and references therein.

3.3. Recognizing sums of squares of polynomials. We now indicate how to recognize whether a polynomial can be written as a sum of squares via semidefinite programming. The next result was discovered independently by several authors; cf. e.g. [22], [131].

Lemma 3.8. **Recognizing sums of squares.**
Let $p \in \mathbb{R}[\mathbf{x}]$, $p = \sum_{\alpha \in \mathbb{N}_{2d}^n} p_\alpha \mathbf{x}^\alpha$, be a polynomial of degree $\leq 2d$. The following assertions are equivalent.

(i) *p is a sum of squares.*

(ii) *The following system in the matrix variable $X = (X_{\alpha,\beta})_{\alpha,\beta \in \mathbb{N}_d^n}$ is* *feasible:*

$$\left\{ \begin{array}{c} X \succeq 0 \\ \displaystyle\sum_{\beta,\gamma \in \mathbb{N}^n_d \mid \beta+\gamma=\alpha} X_{\beta,\gamma} = p_\alpha \quad (|\alpha| \leq 2d). \end{array} \right. \tag{3.4}$$

Proof. Let $\mathbf{z}_d := (\mathbf{x}^\alpha \mid |\alpha| \leq d)$ denote the vector containing all monomials of degree at most d. Then for polynomials $u_j \in \mathbb{R}[\mathbf{x}]_d$, we have $u_j = \text{vec}(u_j)^T \mathbf{z}_d$ and thus $\sum_j u_j^2 = \mathbf{z}_d^T(\sum_j \text{vec}(u_j)\text{vec}(u_j)^T)\mathbf{z}_d$. Therefore, p is a sum of squares of polynomials if and only if $p = \mathbf{z}_d^T X \mathbf{z}_d$ for some positive semidefinite matrix X. Equating the coefficients of the two polynomials p and $\mathbf{z}_d^T X \mathbf{z}_d$, we find the system (3.4). □

Thus to decide whether the polynomial p can be written as a sum of squares one has to verify existence of a positive semidefinite matrix X satisfying the linear equations in (3.4) and any Gram decomposition of X gives a sum of square decomposition for p. For this reason this method is often called the *Gram-matrix method* in the literature (e.g. [22]). The system (3.4) is a system in the matrix variable X, which is indexed by $\mathbb{N}^n_d$ and thus has size $\binom{n+d}{d}$, and with $\binom{n+2d}{2d}$ equations. Therefore, this system has polynomial size if either n is fixed, or d is fixed. The system (3.4) is a semidefinite program. Thus finding a sum of square decomposition of a polynomial can be done using semidefinite programming. Note also that if p has a sum of squares decomposition then it has one involving at most $|\mathbb{N}^n_d| = \binom{n+d}{d}$ squares.

We now illustrate the method on a small example.

EXAMPLE 3.9. *Suppose we want to find a sum of squares decomposition for the polynomial* $p = \mathbf{x}^4 + 2\mathbf{x}^3\mathbf{y} + 3\mathbf{x}^2\mathbf{y}^2 + 2\mathbf{x}\mathbf{y}^3 + 2\mathbf{y}^4 \in \mathbb{R}[\mathbf{x},\mathbf{y}]_4$. *As p is a form of degree 4, we want to find $X \succeq 0$ indexed by* $\mathbf{x}^2, \mathbf{xy}, \mathbf{y}^2$ *satisfying*

$$p = (\mathbf{x}^2 \ \mathbf{xy} \ \mathbf{y}^2) \underbrace{\begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}}_{X} \begin{pmatrix} \mathbf{x}^2 \\ \mathbf{xy} \\ \mathbf{y}^2 \end{pmatrix}.$$

Equating coefficients:

$$\begin{array}{ll} \mathbf{x}^4 = \mathbf{x}^2 \cdot \mathbf{x}^2 & 1 = a \\ \mathbf{x}^3\mathbf{y} = \mathbf{x}^2 \cdot \mathbf{xy} & 2 = 2b \\ \mathbf{x}^2\mathbf{y}^2 = \mathbf{xy} \cdot \mathbf{xy} = \mathbf{x}^2 \cdot \mathbf{y}^2 & 3 = d + 2c \\ \mathbf{xy}^3 = \mathbf{xy} \cdot \mathbf{y}^2 & 2 = 2e \\ \mathbf{y}^4 = \mathbf{y}^2 \cdot \mathbf{y}^2 & 2 = f \end{array}$$

we find $X = \begin{pmatrix} 1 & 1 & c \\ 1 & 3-2c & 1 \\ c & 1 & 2 \end{pmatrix}$. *Therefore* $X \succeq 0 \iff -1 \leq c \leq 1$. *E.g.*

for $c = -1$, $c = 0$, we find, respectively, the matrix

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \\ 0 & 2 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 1 & \sqrt{\frac{3}{2}} & \sqrt{\frac{1}{2}} \\ 0 & \sqrt{\frac{3}{2}} & -\sqrt{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & \sqrt{\frac{3}{2}} & \sqrt{\frac{3}{2}} \\ 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{pmatrix}$$

giving, respectively, the decompositions $p = (\mathbf{x}^2 + \mathbf{xy} - \mathbf{y}^2)^2 + (\mathbf{y}^2 + 2\mathbf{xy})^2$ *and* $p = (\mathbf{x}^2 + \mathbf{xy})^2 + \frac{3}{2}(\mathbf{xy} + \mathbf{y}^2)^2 + \frac{1}{2}(\mathbf{xy} - \mathbf{y}^2)^2$.

3.4. SOS relaxations for polynomial optimization. Although we will come back to it in detail in Section 6, we already introduce here the SOS relaxations for the polynomial optimization problem (1.1) as this will motivate our exposition later in this section of several representation results for positive polynomials. Note first that problem (1.1) can obviously be reformulated as

$$\begin{aligned} p^{\min} &= \sup \rho \;\text{ s.t. } p - \rho \geq 0 \text{ on } K \\ &= \sup \rho \;\text{ s.t. } p - \rho > 0 \text{ on } K. \end{aligned} \tag{3.5}$$

That is, computing $p^{\min}$ amounts to finding the supremum of the scalars ρ for which $p-\rho$ is nonnegative (or positive) on the set K. To tackle this hard problem it is a natural idea (going back to work of Shor [155, 156, 157], Nesterov [112], Lasserre [78], Parrilo [121, 122]) to replace the nonnegativity condition by some simpler condition, involving sums of squares, which can then be tackled using semidefinite programming.

For instance, in the unconstrained case when $K = \mathbb{R}^n$, consider the parameter

$$p^{\text{sos}} := \sup \rho \;\text{ s.t. } p - \rho \text{ is SOS.} \tag{3.6}$$

As explained in the previous section, the parameter p^{sos} can be computed via a semidefinite program involving a matrix of size $|\mathbb{N}^n_d|$ if p has degree $2d$. Obviously, $p^{\text{sos}} \leq p^{\min}$, but as follows from Hilbert's result (Theorem 3.4), the inequality may be strict. For instance, when p is the Motzkin polynomial considered in Example 3.7, then $p^{\text{sos}} = -\infty < p^{\min} = 0$ as p vanishes at $(\pm 1, \pm 1)$.

In the constrained case, one way to relax the condition '$p - \rho \geq 0$ on K' is by considering a sum of square decomposition of the form $p - \rho = s_0 + \sum_{j=1}^m s_j g_j$ where s_0, s_j are SOS. This yields the parameter:

$$p^{\text{sos}} := \sup \rho \;\text{ s.t. } p - \rho = s_0 + \sum_{j=1}^{m} s_j g_j \text{ with } s_0, s_j \text{ SOS.} \tag{3.7}$$

Again $p^{\text{sos}} \leq p^{\min}$ and, under certain assumption on the polynomials g_j describing the set K (cf. Theorem 3.20 below), equality holds. The above formulation does not lead yet to a semidefinite program, since it is not obvious how to bound the degrees of the polynomials s_0, s_j as cancellation of terms may occur in $s_0 + \sum_j s_j g_j$. To get a semidefinite program one may consider for any integer t with $2t \geq \max(\deg p, \deg(g_1), \ldots, \deg(g_m))$ the parameter

$$p_t^{\text{sos}} := \sup \rho \ \text{ s.t. } \quad \begin{array}{l} p - \rho = s_0 + \sum_{j=1}^m s_j g_j \text{ with } s_0, s_j \in \Sigma, \\ \deg(s_0), \deg(s_j g_j) \leq 2t. \end{array} \tag{3.8}$$

Hence each p_t^{sos} can be computed via a semidefinite program involving matrices of size $|\mathbb{N}_t^n|$, $p_t^{\text{sos}} \leq p_{t+1}^{\text{sos}} \leq p^{\text{sos}} \leq p^{\min}$, and $\lim_{t \to \infty} p_t^{\text{sos}} = p^{\text{sos}}$.

3.5. Convex quadratic optimization. Here we consider problem (1.1) in the convex quadratic case, i.e. when $p, g_1, \ldots, g_m$ are quadratic polynomials and $p, -g_1, \ldots, -g_m$ are convex. Then the semialgebraic set K defined by the g_j's is convex; we also assume that K is compact so that p attains its minimum over K at some $x^* \in K$. Let $J(x^*) := \{j \in \{1, \ldots, m\} \mid g_j(x^*) = 0\}$ for $x^* \in K$, and consider the following (MFCQ) constraint qualification:

$$\exists w \in \mathbb{R}^n \quad \text{for which} \quad w^T \nabla g_j(x^*) > 0 \ \forall j \in J(x^*); \tag{3.9}$$

equivalently, $\sum_{j \in J(x^*)} \lambda_j \nabla g_j(x^*) = 0$ with $\lambda_j \geq 0 \ \forall j$ implies $\lambda_j = 0 \ \forall j$.

LEMMA 3.10. *[78] Consider problem (1.1) where $p, -g_1, \ldots, -g_m$ are quadratic (or linear) convex polynomials and assume that the set K from (1.2) is compact. If there exists a local (thus global) minimizer x^* satisfying (3.9), then $p_1^{sos} = p^{min}$.*

Proof. By assumption, $p = \mathbf{x}^T Q \mathbf{x} + 2c^T \mathbf{x}$, $g_j = \mathbf{x}^T Q_j \mathbf{x} + 2c_j^T \mathbf{x} + b_j$, where Q, Q_j are symmmetric $n \times n$ matrices, $Q, -Q_1, \ldots, -Q_m \succeq 0$, $c, c_j \in \mathbb{R}^n$, $b_j \in \mathbb{R}$. The bound p_1^{sos} is defined by

$$\begin{array}{ll} p_1^{\text{sos}} & = \sup_{\rho, \lambda_j \in \mathbb{R}} \rho \ \text{ s.t. } p - \rho - \sum_{j=1}^m \lambda_j g_j \in \Sigma, \ \lambda_1, \ldots, \lambda_m \geq 0 \\ & = \sup_{\rho, \lambda_j \in \mathbb{R}} \rho \ \text{ s.t. } p - \rho - \sum_{j=1}^m \lambda_j g_j \in \mathcal{P}, \ \lambda_1, \ldots, \lambda_m \geq 0, \end{array}$$

where the last equality follows using Lemma 3.6, It suffices now to show that $p^{\min}$ is feasible for the program defining p_1^{sos}. For this let $x^* \in K$ be a local minimizer of p over the set K satisfying (3.9). Then there exist scalars $\lambda_1, \ldots, \lambda_m \geq 0$ for which the first order Karush-Kuhn-Tucker conditions hold (cf. e.g. [118, §12.5]). That is, $\lambda_j g_j(x^*) = 0 \quad \forall j$ and $\nabla p(x^*) = \sum_j \lambda_j \nabla g_j(x^*)$, implying

$$Qx^* + c = \sum_{j=1}^m \lambda_j (Q_j x^* + c_j). \tag{3.10}$$

We claim that

$$p - p^{\min} - \sum_{j=1}^m \lambda_j g_j = (\mathbf{x} - x^*)^T (Q - \sum_{j=1}^m \lambda_j Q_j)(\mathbf{x} - x^*). \tag{3.11}$$

Indeed, $p - p^{\min} - \sum_{j=1}^m \lambda_j g_j = p - p^{\min} + \sum_j \lambda_j (g_j(x^*) - g_j)$ is equal to $\mathbf{x}^T(Q - \sum_j \lambda_j Q_j)\mathbf{x} + 2(c - \sum_j \lambda_j c_j)^T \mathbf{x} - (x^*)^T(Q - \sum_j \lambda_j Q_j)x^* + 2(\sum_j \lambda_j c_j - c)^T x^*$ which, using (3.10), gives the desired identity (3.11). As $Q - \sum_j \lambda_j Q_j \succeq 0$, (3.11) implies that $p - p^{\min} - \sum_{j=1}^m \lambda_j g_j$ is nonnegative over $\mathbb{R}^n$, which concludes the proof. □

We will see in Section 4.3 (cf. Corollary 4.6) a related result, showing that the moment relaxation of order 1 is exact when assuming that $p, -g_j$ are quadratic convex polynomials (thus with no regularity condition).

3.6. Some representation results for positive polynomials.

3.6.1. Positivity certificates via the Positivstellensatz. A classical result about polynomials is Hilbert's Nullstellensatz which characterizes when a system of polynomials in $\mathbb{C}[\mathbf{x}]$ has a common root in $\mathbb{C}^n$. The next result is sometimes called the *weak* Nullstellensatz, while the result of Theorem 2.1 (i) is Hilbert's *strong* Nullstellensatz.

Theorem 3.11. *(cf. e.g. [25])* **Hilbert's (weak) Nullstellensatz.** *Given polynomials $h_1, \ldots, h_m \in \mathbb{C}[\mathbf{x}]$, the system $h_1(x) = 0, \ldots, h_m(x) = 0$ does not have a common root in $\mathbb{C}^n$ if and only if $1 \in (h_1, \ldots, h_m)$, i.e. $1 = \sum_{j=1}^m u_j h_j$ for some polynomials $u_j \in \mathbb{C}[\mathbf{x}]$.*

As a trivial example, the system $h_1 := \mathbf{x} + 1 = 0$, $h_2 := \mathbf{x}^2 + 1 = 0$ has no common root, which is certified by the identity $1 = h_1\ (1 - \mathbf{x})/2 + h_2/2$. The above result works only in the case of an algebraically closed field (like $\mathbb{C}$). For instance, $\mathbf{x}^2 + 1 = 0$ has no solution in $\mathbb{R}$, but 1 does not belong to the ideal generated by $\mathbf{x}^2 + 1$. A basic property of the real field $\mathbb{R}$ is that $\sum_{i=1}^n a_i^2 = 0 \Longrightarrow a_1 = \ldots = a_n = 0$, i.e. -1 is not a sum of squares in $\mathbb{R}$. These properties are formalized in the theory of *formally real* fields (cf. [17, 133]) and one of the objectives of real algebraic geometry is to understand when systems of real polynomial equations and inequalities have solutions in $\mathbb{R}^n$. An answer is given in Theorem 3.12 below, known as the *Positivstellensatz*, often attributed to Stengle [159], although the main ideas were already known to Krivine [74]. A detailed exposition can be found e.g. in [103], [133]. We need some definitions. Given polynomials $g_1, \ldots, g_m \in \mathbb{R}[\mathbf{x}]$, set $g_J := \prod_{j \in J} g_j$ for $J \subseteq \{1, \ldots, m\}$, $g_\emptyset := 1$. The set

$$T(g_1, \ldots, g_m) := \Big\{ \sum_{J \subseteq \{1,\ldots,m\}} u_J g_J \mid u_J \in \Sigma \Big\}, \tag{3.12}$$

is called the *preordering* on $\mathbb{R}[\mathbf{x}]$ generated by $g_1, \ldots, g_m$. As in (1.2), let $K = \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$.

Theorem 3.12. **Positivstellensatz.** *Given a polynomial $p \in \mathbb{R}[\mathbf{x}]$,*

(i) $p > 0$ on $K \Longleftrightarrow pf = 1 + g$ for some $f, g \in T(g_1, \ldots, g_m)$.

(ii) $p \geq 0$ on $K \Longleftrightarrow pf = p^{2k} + g$ for some $f, g \in T(g_1, \ldots, g_m)$ and $k \in \mathbb{N}$.

(iii) $p = 0$ *on* $K \iff -p^{2k} \in T(g_1, \ldots, g_m)$ *for some* $k \in \mathbb{N}$.

(iv) $K = \emptyset \iff -1 \in T(g_1, \ldots, g_m)$.

COROLLARY 3.13. **Real Nullstellensatz.** *Given* $p, h_1, \ldots, h_m \in \mathbb{R}[\mathbf{x}]$, *p vanishes on* $\{x \in \mathbb{R}^n \mid h_j(x) = 0 \ (j = 1, \ldots, m)\}$ *if and only if* $p^{2k} + s = \sum_{j=1}^m u_j h_j$ *for some* $u_j \in \mathbb{R}[\mathbf{x}]$, $s \in \Sigma$, $k \in \mathbb{N}$.

COROLLARY 3.14. **Solution to Hilbert's 17th problem.** *Given* $p \in \mathbb{R}[\mathbf{x}]$, *if* $p \geq 0$ *on* $\mathbb{R}^n$, *then* $p = \sum_j \left(\frac{a_j}{b_j}\right)^2$ *for some* $a_j, b_j \in \mathbb{R}[\mathbf{x}]$.

Following Parrilo [121, 122], one may interpret the above results in terms of certificates of infeasiblity of certain systems of polynomial systems of equations and inequalities. First, observe that Hilbert's Nullstellensatz can be interpreted as follows: Either a system of polynomial equations is feasible, which can be certified by giving a common solution x; or it is infeasible, which can be certified by giving a Nullstellensatz certificate of the form $1 = \sum_{j=1}^m u_j h_j$. Parrilo makes the analogy with Farkas' lemma for linear programming; indeed, given $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, Farkas' lemma asserts that, either the linear system $Ax \leq b, x \geq 0$ has a solution, or it is infeasible, which can be certified by giving a solution y to the alternative system $A^T y \geq 0$, $y \geq 0$, $y^T b < 0$. (Cf. e.g. [146, §7.3]). This paradigm extends to the real solutions of systems of polynomial inequalities and equations, as the following reformulation of the Positivstellensatz (cf e.g. [17]) shows.

THEOREM 3.15. *Let* f_r $(r = 1, \ldots, s)$, g_l $(l = 1, \ldots, t)$, h_j $(j = 1, \ldots, m)$ *be polynomials in* $\mathbb{R}[\mathbf{x}]$. *Then one of the following holds.*

(i) Either the system $f_r(x) \neq 0$ $(r = 1, \ldots, s)$, $g_l(x) \geq 0$ $(l = 1, \ldots, t)$, $h_j(x) = 0$ $(j = 1, \ldots, m)$ *has a solution in* $\mathbb{R}^n$.

(ii) Or $\prod_{r=1}^s f_r^{2d_r} + \sum_{J \subseteq \{1,\ldots,t\}} s_J g_J + \sum_{j=1}^m u_j h_j = 0$ *for some* $d_r \in \mathbb{N}$, $s_J \in \Sigma$, $u_j \in \mathbb{R}[\mathbf{x}]$.

Thus the Positivstellensatz can be seen as a generalization of Hilbert's Nullstellensatz and of Farkas' lemma (for linear programming) and one can search for bounded degree certificates that the system in Theorem 3.15 (i) has no real solution, using semidefinite programming. See [121, 122] for further discussion and references.

One may try to use the Positivstellensatz to approximate the optimization problem (1.1). Namely, in view of Theorem 3.12 (i), one can replace the condition '$p - \rho > 0$ on K' in (3.5) by the condition '$(p - \rho)f = 1 + g$ for some $f, g \in T(g_1, \ldots, g_m)$' and this remains a formulation for $p^{\min}$. However, although membership in $T(g_1, \ldots, g_m)$ with bounded degrees can be formulated via a semidefinite program, this does not lead to a semidefinite programming formulation for $p^{\min}$ because of the presence of the product ρf where both ρ and f are variables. In the case when the semialgebraic set K is compact one may instead use the following refinement of the Positivstellensatz of Schmüdgen. (See [148] for a more elementary exposition of Schmüdgen's result and [149] for degree bounds.)

Theorem 3.16. *[145]* **(Schmüdgen's theorem)** *Assume the semialgebraic set K in (1.2) is compact. Given $p \in \mathbb{R}[\mathbf{x}]$, if $p > 0$ on K, then $p \in T(g_1, \ldots, g_m)$.*

If we replace the condition '$p > 0$ on K' by '$p \geq 0$ on K', the above result of Schmüdgen does not remain true in general, although it does in the univariate case as we see in Section 3.6.3. Indeed, for most instances of K, there exists a polynomial $p \geq 0$ on $\mathbb{R}^n$ for which $p \notin T(g_1, \ldots, g_m)$. This is the case e.g. when $n \geq 3$ and K has a nonempty interior or, more generally, when the dimension of K (defined as the Krull dimension of $\mathbb{R}[\mathbf{x}]/\mathcal{I}(K)$) is at least 3. This is also the case when $n = 2$ and K contains a 2-dimensional affine cone. (See e.g. [106, Sec.2.6,2.7] for details.)

Schmüdgen's theorem naturally leads to a hierarchy of semidefinite relaxations for $p^{\min}$, as the programs

$$\sup \rho \text{ s.t. } p - \rho = \sum_{J \subseteq \{1,\ldots,m\}} u_J g_J \text{ with } u_J \in \Sigma,\ \deg(u_J g_J) \leq t$$

are semidefinite programs whose optimum values converge to $p^{\min}$ as t goes to ∞. However, a drawback is that Schmüdgen's representation involves 2^m sums of squares, thus leading to possibly quite large semidefinite programs. As proposed by Lasserre [78], one may use the further refinement of Schmüdgen's Positivstellensatz proposed by Putinar [134], which holds under some condition on the compact set K.

3.6.2. Putinar's Positivstellensatz.

Given polynomials $g_1, \ldots, g_m \in \mathbb{R}[\mathbf{x}]$, the set

$$\mathbf{M}(g_1, \ldots, g_m) := \Big\{ u_0 + \sum_{j=1}^m u_j g_j \mid u_0, u_j \in \Sigma \Big\}, \tag{3.13}$$

is called the *quadratic module generated by* $g_1, \ldots, g_m$. (We use the boldface letter $\mathbf{M}$ for a quadratic module $\mathbf{M}(g_1, \ldots, g_m)$ to avoid confusion with a moment matrix $M(y)$.) Consider the condition

$$\exists f \in \mathbf{M}(g_1, \ldots, g_m) \text{ s.t. } \{x \in \mathbb{R}^n \mid f(x) \geq 0\} \text{ is a compact set.} \tag{3.14}$$

Obviously, (3.14) implies that K is compact, since $K \subseteq \{x \mid f(x) \geq 0\}$ for any $f \in \mathbf{M}(g_1, \ldots, g_m)$. Note also that (3.14) is an assumption on the description of K, rather than on the set K itself. Condition (3.14) holds, e.g., if the set $\{x \in \mathbb{R}^n \mid g_j(x) \geq 0\}$ is compact for one of the constraints defining K. It also holds when the description of K contains a set of polynomial equations $h_1 = 0, \ldots, h_{m_0} = 0$ with a compact set of common real roots. If an explicit ball of radius R is known containing K, then it suffices to add the (redundant) constraint $R^2 - \sum_{i=1}^n \mathbf{x}_i^2 \geq 0$ in order to obtain a description of K satisfying (3.14). More detailed information can be found in [65, 133]; e.g. it is shown there that condition (3.14) holds when $m \leq 2$.

As we now see, the condition (3.14) admits in fact several equivalent reformulations. Consider the following conditions

$$\exists N \in \mathbb{N} \text{ for which } N - \sum_{i=1}^{n} \mathbf{x}_i^2 \in \mathbf{M}(g_1, \ldots, g_m), \qquad (3.15)$$

$$\forall p \in \mathbb{R}[\mathbf{x}] \; \exists N \in \mathbb{N} \;\text{ for which } N \pm p \in \mathbf{M}(g_1, \ldots, g_m), \qquad (3.16)$$

$$\begin{array}{c}\exists p_1, \ldots, p_s \in \mathbb{R}[\mathbf{x}] \text{ s.t. } p_I \in \mathbf{M}(g_1, \ldots, g_m) \; \forall I \subseteq \{1, \ldots, s\} \\ \text{and } \{x \in \mathbb{R}^n \mid p_1(x) \geq 0, \ldots, p_s(x) \geq 0\} \text{ is compact.}\end{array} \qquad (3.17)$$

Here we set $p_I := \prod_{i \in I} p_i$ for $I \subseteq \{1, \ldots, s\}$. Schmüdgen [145] proved equivalence of the above conditions (3.14)-(3.17) (see [150] for a discussion and further references).

THEOREM 3.17. *The conditions (3.14), (3.15), (3.16), (3.17) are all equivalent.*

Proof. Obviously, (3.16) $\Longrightarrow$ (3.15) $\Longrightarrow$ (3.14) $\Longrightarrow$ (3.17). One can derive the implication (3.17) $\Longrightarrow$ (3.16) from Schmüdgen's theorem (Theorem 3.16). Indeed, if (3.17) holds, then $K_0 := \{x \in \mathbb{R}^n \mid p_1(x) \geq 0, \ldots, p_s(x) \geq 0\}$ is compact and thus there exists $N > 0$ for which $N \pm p > 0$ on K_0. By Theorem 3.16, $N \pm p = \sum_{I \subseteq \{1, \ldots, s\}} s_I p_I$ for some $s_I \in \Sigma$ and thus $N \pm p \in \mathbf{M}(g_1, \ldots, g_m)$ since each $p_I \in \mathbf{M}(g_1, \ldots, g_m)$.
Finally we give an elementary proof for (3.15) $\Longrightarrow$ (3.16), following [106, §5.2]. Assume $N_0 - \sum_{i=1}^{n} \mathbf{x}_i^2 \in \mathbf{M}(g_1, \ldots, g_m)$. We show that the set

$$P := \{p \in \mathbb{R}[\mathbf{x}] \mid \exists N > 0 \;\text{ s.t. }\; N \pm p \in \mathbf{M}(g_1, \ldots, g_m)\}$$

coincides with $\mathbb{R}[\mathbf{x}]$. Obviously, $\mathbb{R} \subseteq P$ and P is closed by addition. Moreover, P is closed by multiplication which follows from the identity

$$N_1 N_2 + \epsilon pq = \frac{1}{2}\Big((N_1 + \epsilon p)(N_2 + q) + (N_1 - \epsilon p)(N_2 - q)\Big)$$

for $\epsilon = \pm 1$. Finally, P contains each variable $\mathbf{x}_i$, which follows from the identity

$$\frac{N_0 + 1}{2} + \epsilon \mathbf{x}_i = \frac{1}{2}\Big((\mathbf{x}_i + \epsilon)^2 + N_0 - \sum_{j=1}^{n} \mathbf{x}_j^2 + \sum_{j \neq i} \mathbf{x}_j^2\Big)$$

for $\epsilon = \pm 1$. The above facts imply $P = \mathbb{R}[\mathbf{x}]$ and thus (3.16) holds. □

DEFINITION 3.18. *Given $g_1, \ldots, g_m \in \mathbb{R}[\mathbf{x}]$, the quadratic module $\mathbf{M}(g_1, \ldots, g_m)$ is said to be* Archimedean *when the condition (3.16) holds.*

Example 3.19. *For the polynomials $g_i := \mathbf{x}_i - 1/2$ $(i = 1, \ldots, n)$ and $g_{n+1} := 1 - \prod_{i=1}^n \mathbf{x}_i$, the module $\mathbf{M}(g_1, \ldots, g_{n+1})$ is not Archimedean [133, Ex. 6.3.1]. To see it, consider a lexicographic monomial ordering on $\mathbb{R}[\mathbf{x}]$ and define the set M of polynomials $p \in \mathbb{R}[\mathbf{x}]$ satisfying $p = 0$, or $p \neq 0$ whose leading term $p_\alpha \mathbf{x}^\alpha$ satisfies either $p_\alpha > 0$ and $\alpha \neq (1, \ldots, 1) \mod 2$, or $p_\alpha < 0$ and $\alpha = (1, \ldots, 1) \mod 2$. Then M is a quadratic module (cf. Definition 3.43) and $g_1, \ldots, g_{n+1} \in M$, implying $\mathbf{M}(g_1, \ldots, g_{n+1}) \subseteq M$. For any $N \in \mathbb{R}$, the polynomial $N - \sum_{i=1}^n \mathbf{x}_i^2$ does not lie in M, which implies that it also does not lie in $\mathbf{M}(g_1, \ldots, g_{n+1})$. This shows that $\mathbf{M}(g_1, \ldots, g_{n+1})$ is not Archimedean.*

Theorem 3.20. *(Putinar [134]; see also Jacobi and Prestel [65]). Let K be as in (1.2) and assume that the quadratic module $\mathbf{M}(g_1, \ldots, g_m)$ is Archimedean. For $p \in \mathbb{R}[\mathbf{x}]$, if $p > 0$ on K then $p \in \mathbf{M}(g_1, \ldots, g_m)$.*

As noted by Lasserre [78], this implies directly the asymptotic convergence to $p^{\min}$ of the hierarchy of bounds from (3.8). We will come back to this hierarchy in Section 6. We refer to Nie and Schweighofer [117] for degree bounds on representations in $\mathbf{M}(g_1, \ldots, g_m)$. We present a proof for Theorem 3.20 in Section 3.7, which uses the condition (3.16) as definition for $\mathbf{M}(g_1, \ldots, g_m)$ Archimedean.

3.6.3. Representation results in the univariate case. We review here some of the main representation results for positive polynomials in the univariate case. We refer to Powers and Reznick [127] for a detailed treatment as well as historic remarks. As we saw earlier any polynomial nonnegative on $\mathbb{R}$ is a sum of (two) squares. We now consider nonnegative polynomials on a closed interval K. Up to a change of variables, there are two cases to consider: $K = [-1, 1]$ and $K = [0, \infty)$. In the case $K = [0, \infty) = \{x \in \mathbb{R} \mid x \geq 0\}$, the next classical result shows that $\mathcal{P}_K = T(\mathbf{x})$. Throughout the paper

$$\mathcal{P}_K := \{p \in \mathbb{R}[\mathbf{x}] \mid p(x) \geq 0 \ \forall x \in K\}$$

denotes the set of polynomials nonnegative on a set $K \subseteq \mathbb{R}^n$; thus, for $K = \mathbb{R}^n$, $\mathcal{P}_K$ coincides with $\mathcal{P}_n$ introduced earlier in (3.1).

Theorem 3.21. **(Pólya-Szegö)** *Let $n = 1$ and $p \in \mathbb{R}[\mathbf{x}]$. If $p \geq 0$ on $K = [0, \infty)$, then $p = f + \mathbf{x}g$, where $f, g \in \Sigma$ and $\deg(f), \deg(\mathbf{x}g) \leq \deg(p)$. Therefore, $\mathcal{P}_K = T(\mathbf{x})$.*

Proof. As $p \geq 0$ on $[0, \infty)$, every positive root of p has an even degree and complex roots come in conjugate pairs with the same multiplicities. Thus p factors as $p = s_0 x^d (x + a_1) \cdots (x + a_t)$, where $a_1, \ldots, a_t > 0$, $s_0 \in \Sigma$ and $d \in \{0, 1\}$. The result now follows as $\prod_{i=1}^t (x + a_i) = f_0 + x g_0$ for some $f_0, g_0 \in \Sigma$. □

One can then derive results for the case $K = [-1, 1]$ using the following Goursat transform. Given a polynomial $f \in \mathbb{R}[\mathbf{x}]$ with $\deg(f) = m$, its

Goursat transform is the polynomial $\tilde{f}$ defined by

$$\tilde{f}(\mathbf{x}) := (1+\mathbf{x})^m f\Big(\frac{1-\mathbf{x}}{1+\mathbf{x}}\Big).$$

One can easily verify that the Goursat transform of $\tilde{f}$ satisfies: $\tilde{\tilde{f}}(\mathbf{x}) = (1+\mathbf{x})^m \tilde{f}\Big(\frac{1-\mathbf{x}}{1+\mathbf{x}}\Big) = (1+\mathbf{x})^m \Big(1+\frac{1-\mathbf{x}}{1+\mathbf{x}}\Big)^m f(\mathbf{x}) = 2^m f(\mathbf{x})$.

LEMMA 3.22. **(Goursat's lemma)** *Let $f \in \mathbb{R}[\mathbf{x}]$ with* $\deg(f) = m$ *and $\tilde{f} \in \mathbb{R}[\mathbf{x}]$ its Goursat transform.*

(i) $f \geq 0$ on $[-1,1] \iff \tilde{f} \geq 0$ on $[0,\infty)$.

(ii) $f > 0$ on $[-1,1] \iff \tilde{f} > 0$ on $[0,\infty)$ and $f(-1) > 0$ (i.e., $\deg(\tilde{f}) = m$*).*

Proof. Easy verification, noting that $\deg(\tilde{f}) \leq m$ and the coefficient of $\mathbf{x}^m$ in $\tilde{f}$ is equal to $f(-1)$. □

THEOREM 3.23. **(Fekete, Markov-Lukácz)** *Let $n = 1$, $p \in \mathbb{R}[\mathbf{x}]$ and $m :=$* $\deg(p)$*. Assume $p \geq 0$ on $K = [-1,1]$. Then,*

(i) $p = s_0 + s_1(1-\mathbf{x}^2)$, where $s_0, s_1 \in \Sigma$, and $\deg(s_0), \deg(s_1(1-\mathbf{x}^2)) \leq m$ *(resp., $m+1$) if m is even (resp., odd).*

(ii) If m is odd, then $p = s_1(1+\mathbf{x}) + s_2(1-\mathbf{x})$, where $s_1, s_2 \in \Sigma$ and $\deg(s_1(1+\mathbf{x})), \deg(s_2(1-\mathbf{x})) \leq m$*.*

In particular, $\mathcal{P}_K = T(1-\mathbf{x}^2) = T(1-\mathbf{x}, 1+\mathbf{x})$.

Proof. As $p \geq 0$ on $[-1,1]$, its Goursat transform $\tilde{p}$ satisfies $\tilde{p} \geq 0$ on $[0,\infty)$ (by Lemma 3.22). Thus, by Theorem 3.21, $\tilde{p} = \sum_{i=1,2} f_i^2 + \mathbf{x}\sum_{i=1,2} g_i^2$, where $f_i, g_i \in \mathbb{R}[\mathbf{x}]$ with $r_i := \deg(f_i) \leq \frac{m}{2}$ and $s_i := \deg(g_i) \leq \frac{m-1}{2}$. Replacing $\mathbf{x}$ by $\frac{1-\mathbf{x}}{1+\mathbf{x}}$, we get:

$$2^m p = \underbrace{\sum_{i=1,2} (1+\mathbf{x})^{m-2r_i} (\tilde{f}_i)^2}_{=:f} + \underbrace{(1-\mathbf{x}) \sum_{i=1,2} (1+\mathbf{x})^{m-1-2s_i} (\tilde{g}_i)^2}_{=:g}.$$

If m is even, the first term f is a sum of squares with degree at most m, and the second term g can be written as $g = (1+\mathbf{x})s_1$ where s_1 is a sum of squares and $\deg(s_1) \leq m-2$. This shows (i) in the case m even.
If m is odd, then g is a sum of squares and $f = (1+\mathbf{x})s_1$ where s_1 is a sum of squares, and $\deg(g), \deg(s_1) \leq m-1$. Thus (ii) holds. Moreover, using the identities:

$$1+\mathbf{x} = \frac{(1+\mathbf{x})^2}{2} + \frac{1}{2}(1-\mathbf{x}^2),\ 1-\mathbf{x} = \frac{(1-\mathbf{x})^2}{2} + \frac{1}{2}(1-\mathbf{x}^2),$$

we get a decomposition as in (i) with summands of degree at most $m+1$. Then, $\mathcal{P}_K = T(1-\mathbf{x}^2) = T(1-\mathbf{x}, 1+\mathbf{x})$ follow directly from (i), (ii). □

Consider the general case when K is an arbitrary closed semialgebraic subset of $\mathbb{R}$. That is, K is a finite union of intervals, of the form $K =$

$[a_1, b_1] \cup [a_2, b_2] \cup \ldots \cup [a_k, b_k]$, where $-\infty \leq a_1 < b_1 < a_2 < b_2 < \ldots < a_k < b_k \leq \infty$. Then $K = \{x \in \mathbb{R} \mid g(x) \geq 0 \ (g \in G)\}$, where G consists of the polynomials $(\mathbf{x} - b_i)(\mathbf{x} - a_{i+1})$ $(i = 1, \ldots, k-1)$ together with $\mathbf{x} - a_1$ if $a_1 > -\infty$ and with $b_k - \mathbf{x}$ if $b_k < \infty$; G is called the *natural description* of K. E.g. the natural description of $[-1, 1]$ is $G = \{\mathbf{x}+1, 1-\mathbf{x}\}$, and the natural description of $[0, \infty)$ is $G = \{\mathbf{x}\}$.

Theorem 3.24. *(cf. [106, Prop. 2.7.3]) Let K be a finite closed semialgebraic set in $\mathbb{R}$ and let G be its natural description. Then $\mathcal{P}_K = T(G)$.*

3.6.4. Other representation results. Several other representation results for positive polynomials exist in the literature. Let us just briefly mention a few.

Theorem 3.25. *(Pólya [126]; see [128] for a proof). Let $p \in \mathbb{R}[\mathbf{x}]$ be a homogeneous polynomial. If $p > 0$ on the simplex $\{x \in \mathbb{R}^n_+ \mid \sum_{i=1}^n x_i = 1\}$, then there exists $r \in \mathbb{N}$ for which the polynomial $(\sum_{i=1}^n \mathbf{x}_i)^r p$ has all its coefficients nonnegative.*

Theorem 3.26. *(Reznick [138]) Let $p \in \mathbb{R}[\mathbf{x}]$ be a homogeneous polynomial. If $p > 0$ on $\mathbb{R}^n \setminus \{0\}$, then there exists $r \in \mathbb{N}$ for which the polynomial $(\sum_{i=1}^n \mathbf{x}_i^2)^r p$ is a sum of squares.*

Example 3.27. *Consider the 5×5 symmetric matrix whose entries are all equal to 1 except $M_{1,2} = M_{2,3} = M_{3,4} = M_{4,5} = M_{5,1} = -1$ and let $p_M := \sum_{i,j=1}^5 M_{i,j}\mathbf{x}_i^2\mathbf{x}_j^2$. Recall from Section 1.1 that M is copositive precisely when p_M is nonnegative. Parrilo [121] proved that, while p_M is not a SOS, $(\sum_{i=1}^5 \mathbf{x}_i^2)p_M$ is a SOS, which shows that p_M is nonnegative and thus M is copositive.*

As an illustration let us briefly sketch how Pólya's theorem can be used to derive a hierarchy of SOS approximations for the stable set problem. See e.g. [36] for further applications, and [35] for a comparison of the hierarchies based on Putinar's and Pólya's theorems.

Example 3.28. *Consider the stable set problem introduced in Section 1.1 and the formulation (1.7) for $\alpha(G)$. For $t \in \mathbb{R}$ define the polynomial $p_{G,t} := \sum_{i,j \in V} \mathbf{x}_i^2\mathbf{x}_j^2(t(I + A_G) - J)_{i,j}$. For $r \in \mathbb{N}$, the parameters*

$$\inf \ t \ \textit{s.t.} \ \Big(\sum_{i \in V} \mathbf{x}_i^2\Big)^r p_{G,t} \ \ \textit{is SOS} \tag{3.18}$$

provide a hierarchy of upper bounds for $\alpha(G)$. Based on an analysis of Pólya's theorem, de Klerk and Pasechnik [37] proved that the bound from (3.18), after being rounded down to the nearest integer, coincides with $\alpha(G)$ when $r \geq (\alpha(G))^2$. Moreover they conjecture that finite convergence takes place at $r = \alpha(G) - 1$. (See [51] for partial results, also for a comparison of the above parameter with the approximation of $\alpha(G)$ derived via Putinar's theorem, mentioned in Example 8.16).

One can also search for a different type of certificate for positivity of a polynomial p over K defined by polynomial inequalities $g_1 \geq 0, \ldots, g_m \geq 0$; namely of the form

$$p = \sum_{\beta \in \mathbb{N}^m} c_\beta \prod_{j=1}^{m} g_j^{\beta_j} \text{ with finitely many nonzero } c_\beta \in \mathbb{R}_+. \tag{3.19}$$

On the moment side this corresponds to Hausdorff-type moment conditions, and this yields hierarchies of *linear programming* relaxations for polynomial optimization. Sherali and Adams [154] use this type of representation for 0/1 polynomial optimization problems. As an example let us mention Handelman's characterization for positivity over a polytope.

THEOREM 3.29. *(Handelman [54]) let $p \in \mathbb{R}[\mathbf{x}]$ and let $K = \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$ be a polytope, i.e. the g_j's are linear polynomials and K is bounded. If $p > 0$ on K then p has a decomposition (3.19).*

The following result holds for a general compact semialgebraic set K, leading to a hierarchy of LP relaxations for problem (1.1). We refer to Lasserre [81, 83] for a detailed discussion and comparison with the SDP based approach.

THEOREM 3.30. *[74, 75] Assume K is compact and the polynomials $g_1, \ldots, g_m$ satisfy $0 \leq g_j \leq 1$ on K $\forall j$ and, together with the constant polynomial 1, they generate the algebra $\mathbb{R}[\mathbf{x}]$, i.e. $\mathbb{R}[\mathbf{x}] = \mathbb{R}[1, g_1, \ldots, g_m]$. Then any $p \in \mathbb{R}[\mathbf{x}]$ positive on K has a representation of the form*

$$p = \sum_{\alpha,\beta \in \mathbb{N}^n} c_{\alpha\beta} \prod_{j=1}^{m} g_j^{\alpha_j} \prod_{j=1}^{m} (1 - g_j)^{\beta_j}$$

for finitely many nonnegative scalars $c_{\alpha\beta}$.

3.6.5. Sums of squares and convexity. A natural question is what can be said about SOS representations of convex positive polynomials. We present here some results in this direction. We first introduce SOS polynomial matrices and SOS-convexity.

Consider a polynomial matrix $P(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]^{r \times r}$, i.e. a $r \times r$ matrix whose entries are polynomials of $\mathbb{R}[\mathbf{x}]$. We say that $P(\mathbf{x})$ *is positive semidefinite* if $P(x) \succeq 0$ for all $x \in \mathbb{R}^n$. Thus, for $r = 1$, $P(\mathbf{x})$ is a usual (scalar) polynomial and being positive semidefinite means (as usual) being nonnegative on $\mathbb{R}^n$. Hence, $P(\mathbf{x})$ is positive semidefinite if and only if the (scalar) polynomial $\mathbf{y}^T P(\mathbf{x})\mathbf{y} \in \mathbb{R}[\mathbf{x}, \mathbf{y}]$ is nonnegative on $\mathbb{R}^{n+r}$, where $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_r)$ are r new variables. The polynomial matrix $P(\mathbf{x})$ is said to be a *SOS-matrix* if $P(\mathbf{x}) = A(\mathbf{x})^T A(\mathbf{x})$ for some polynomial matrix $A(\mathbf{x})$; equivalently, if the (scalar) polynomial $\mathbf{y}^T P(\mathbf{x})\mathbf{y}$ is a sum of squares in $\mathbb{R}[\mathbf{x}, \mathbf{y}]$. Obviously, if $P(\mathbf{x})$ is a SOS-matrix then $P(\mathbf{x})$ is positive semidefinite. The converse implication does not hold for $n \geq 2$, but it does hold in the univariate case.

This result appears in particular in [66]; we will present a proof at the end of this section following the treatment in [21].

Theorem 3.31. *(cf. [66, 21]) Let $n = 1$ and let $P(\mathbf{x})$ be a symmetric matrix polynomial. Then $P(\mathbf{x})$ is positive semidefinite if and only if $P(\mathbf{x})$ is a SOS-matrix.*

The next lemma mentions some nice properties of SOS-matrices.

Lemma 3.32. *[2] If $P(\mathbf{x})$ is a SOS-matrix, then its determinant is a SOS polynomial and any principal submatrix of $P(\mathbf{x})$ is a SOS matrix.*

Proof. The second property is obvious. We now check that $\det(P(\mathbf{x}))$ is a SOS polynomial. The proof relies on the Cauchy-Binet formula, which states that, for matrices A $(r \times s)$ and B $(s \times r)$, $\det(AB) = \sum_S \det(A_S B_S)$, where S runs over all subsets of size r of $[1, s]$, A_S is the column submatrix of A with columns indexed by S, and B is the row submatrix of B with rows indexed by S. We can write $P(\mathbf{x}) = A(\mathbf{x})^T A(\mathbf{x})$; then, by Cauchy-Binet, $\det P(\mathbf{x}) = \sum_S \det(A(\mathbf{x})_S)^2$ is a sum of squares of polynomials. Similary the determinant of a principal submatrix of $P(\mathbf{x})$ is a sum of squares. □

We now turn to examining some properties of convex polynomials in terms of SOS-matrices. Let C be a convex subset of $\mathbb{R}^n$. A function $f : C \to \mathbb{R}$ is convex on C if $f(\lambda x + (1 - \lambda y)) \leq \lambda f(x) + (1 - \lambda) f(y)$ for all $x, y \in C$ and $\lambda \in [0, 1]$. Recall that, if C is open and f is twice differentiable on C, then f is convex on C if and only if its Hessian matrix

$$\nabla^2 f := \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{i,j=1}^n$$

is positive semidefinite on C, i.e. $\nabla^2 f(x) \succeq 0$ for all $x \in C$. If $\nabla^2 f(x) \succ 0$ for all $x \in C$, then f is strictly convex on C. The complexity of testing whether a polynomial (in n variables of degree 4) defines a convex function is not known (cf. [120]). Helton and Nie [57] propose to replace the positivity condition on the Hessian by a SOS-matrix condition, leading to SOS-convexity, a sufficient condition for convexity, which can be tested efficiently.

Definition 3.33. *[57] A polynomial $f \in \mathbb{R}[\mathbf{x}]$ is said to be* SOS-convex *if its Hessian $\nabla^2 f(\mathbf{x})$ is a SOS-matrix.*

Obviously, for a polynomial $f \in \mathbb{R}[\mathbf{x}]$, SOS-convexity implies convexity. The converse implication holds in the univariate case ($n = 1$, by Theorem 3.31) but not in general for $n \geq 2$. Recently Ahmadi and Parrilo [2] give an example of a polynomial in $n = 2$ variables with degree 8, which is convex but not SOS-convex.

Helton and Nie [57] show the following sufficient condition for a SOS-convex polynomial to be a sum of squares.

Theorem 3.34. *Let $p \in \mathbb{R}[\mathbf{x}]$. If p is SOS-convex and there exists $u \in \mathbb{R}^n$ such that $p(u) = 0$ and $\nabla p(u) = 0$, then p is a sum of squares of polynomials.*

The proof relies on the following lemmas.

LEMMA 3.35. *Let $q \in \mathbb{R}[\mathbf{t}]$ be a univariate polynomial. Then,*

$$q(1) - q(0) - q'(0) = \int_0^1 \int_0^t q''(s)\ ds\ dt.$$

Proof. Direct verification. □

LEMMA 3.36. *Let $P(\mathbf{x})$ be a $r \times r$ polynomial matrix, let $u \in \mathbb{R}^n$, and define the polynomial matrix*

$$\tilde{P}(\mathbf{x}) := \int_0^1 \int_0^t P(u + s(\mathbf{x} - u)) ds dt.$$

If $P(\mathbf{x})$ is a SOS-matrix then $\tilde{P}(\mathbf{x})$ too is a SOS-matrix.

Proof. Let $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_r)$ be new variables. Say $P(\mathbf{x})$ has degree $2d$ and define $\mathbf{z}[\mathbf{x},\mathbf{y}] := (\mathbf{y}_i \mathbf{x}^\alpha)_{\substack{i=1,\dots,r\\ \alpha \in \mathbb{N}^n_d}}$. By assumption, the polynomial $\mathbf{y}^T P(\mathbf{x})\mathbf{y}$ is a sum of squares in $\mathbb{R}[\mathbf{x},\mathbf{y}]$. That is, $\mathbf{y}^T P(\mathbf{x})\mathbf{y} = \mathbf{z}[\mathbf{x},\mathbf{y}]^T A^T A \mathbf{z}[\mathbf{x},\mathbf{y}]$ for some matrix A. If we replace $\mathbf{x}$ by $u + s(\mathbf{x} - u)$ in $\mathbf{z}[\mathbf{x},\mathbf{y}]$, we can write $\mathbf{z}[u + s(\mathbf{x} - u), \mathbf{y}] = C(s,u)\mathbf{z}[\mathbf{x},\mathbf{y}]$ for some matrix $C(s,u)$. Thus,

$$\begin{aligned}
\mathbf{y}^T \tilde{P}(\mathbf{x})\mathbf{y} &= \textstyle\int_0^1 \int_0^t \big(\mathbf{y}^T P(u + s(\mathbf{x} - u))\mathbf{y}\big)\ dsdt \\
&= \textstyle\int_0^1 \int_0^t \big(\mathbf{z}[\mathbf{x},\mathbf{y}]^T C(s,u)^T A^T A C(s,u)\mathbf{z}[\mathbf{x},\mathbf{y}]\big)\ dsdt \\
&= \mathbf{z}[\mathbf{x},\mathbf{y}]^T \underbrace{\Big(\int_0^1 \int_0^t C(s,u)^T A^T A C(s,u) dsdt\Big)}_{:=M} \mathbf{z}[\mathbf{x},\mathbf{y}] \\
&= \mathbf{z}[\mathbf{x},\mathbf{y}]^T B^T B \mathbf{z}[\mathbf{x},\mathbf{y}],
\end{aligned}$$

where we use the fact that $M = B^T B$ for some matrix B since $M \succeq 0$. This shows that $\tilde{P}(\mathbf{x})$ is a SOS-matrix. □

LEMMA 3.37. *Let $p \in \mathbb{R}[\mathbf{x}]$ and $u \in \mathbb{R}^n$. Then,*

$$p(\mathbf{x}) = p(u) + \nabla p(u)^T(\mathbf{x} - u) + (\mathbf{x} - u)^T \Big(\int_0^1 \int_0^t \nabla^2 p(u + s(\mathbf{x} - u)) dsdt\Big)(\mathbf{x} - u).$$

Proof. Fix $x, u \in \mathbb{R}^n$ and consider the univariate polynomial $q(t) := p(u + t(x - u))$ in $t \in \mathbb{R}$. Then, $q'(t) = \nabla p(u + t(x - u))^T(x - u)$, and $q''(t) = (x - u)^T \nabla^2 p(u + t(x - u))(x - u)$, so that $q(1) = p(x)$, $q(0) = p(u)$, $q'(0) = \nabla p(u)^T(x - u)$. The result now follows using Lemma 3.35 applied to q. □

Proof. (of Theorem 3.34) Assume $p \in \mathbb{R}[\mathbf{x}]$ is SOS-convex, $p(u) = 0$, and $\nabla p(u) = 0$ for some $u \in \mathbb{R}^n$. Lemma 3.37 gives the decomposition

$p(\mathbf{x}) = (\mathbf{x} - u)^T \tilde{P}(\mathbf{x})(\mathbf{x} - u)$, where $P(\mathbf{x}) := \nabla^2 p(\mathbf{x})$ and $\tilde{P}$ is defined as in Lemma 3.36. As p is SOS-convex, $P(\mathbf{x})$ is a SOS-matrix and thus $\tilde{P}(\mathbf{x})$ too (by Lemma 3.36). The above decomposition of p gives thus a sum of squares decomposition. ◻

We can now state the convex Positivstellensatz of Lasserre [89], which gives a simpler sum of squares representation for nonnegative polynomials on a convex semialgebraic set - under some convexity and regularity conditions. Consider the semialgebraic set K from (1.2), defined by polynomial inequalities $g_j \geq 0$ ($j = 1, \ldots, m$), and the following subset of the quadratic module $\mathbf{M}(g_1, \ldots, g_m)$:

$$\mathbf{M}_c(g_1, \ldots, g_m) := \{s_0 + \sum_{j=1}^m \lambda_j g_j \mid s_0 \in \Sigma,\ \lambda_1, \ldots, \lambda_m \in \mathbb{R}_+\}.$$

So the multipliers of the g_j's are now restricted to be nonnegative scalars instead of being sums of squares of polynomials.

Theorem 3.38. *[89] Let $p \in \mathbb{R}[\mathbf{x}]$. Assume that p attains its infimum on K, i.e. $p(x^*) = \inf_{x \in K} p(x)$ for some $x^* \in K$. Assume moreover that int$(K) \neq \emptyset$ (i.e. there exists $x \in K$ for which $g_j(x) > 0\ \forall j$) and that $p, -g_1, \ldots, -g_m$ are SOS-convex. Then, $p - p^{min} \in \mathbf{M}_c(g_1, \ldots, g_m)$. In particular, $p \in \mathbf{M}_c(g_1, \ldots, g_m)$ if $p \geq 0$ on K.*

Proof. As $\text{int}(K) \neq \emptyset$ and $f, -g_j$ are convex, the Karush-Kuhn-Tucker optimality conditions tell us that there exists $\lambda \in \mathbb{R}^m_+$ satisfying

$$\nabla p(x^*) - \sum_{j=1}^m \lambda_j \nabla g_j(x^*) = 0,\ \ \lambda_j g_j(x^*) = 0\ \forall j.$$

Consider the Lagrangian function $L_p(\mathbf{x}) := p(\mathbf{x}) - p^{\min} - \sum_{j=1}^m \lambda_j g_j(\mathbf{x})$. Then, $L_p(x^*) = 0$, $\nabla L_p(x^*) = 0$, and L_p is SOS-convex, since $\nabla^2 L_p = \nabla^2 p - \sum_j \lambda_j \nabla^2 g_j$ is a SOS-matrix. Applying Theorem 3.34, we deduce that L_p is a sum of squares of polynomials. Hence the decomposition $p = p^{\min} + L_p + \sum_{j=1}^m \lambda_j g_j$ shows that $p - p^{\min} \in \mathbf{M}_c(g_1, \ldots, g_m)$ and $p = (p - p^{\min}) + p^{\min} \in \mathbf{M}_c(g_1, \ldots, g_m)$ if $p^{\min} \geq 0$. ◻

Proof of Theorem 3.31. We now give a proof for Theorem 3.31, following the treatment in [21], which uses some nice tricks. Let $P(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]^{n \times n} := (a_{ij}(\mathbf{x}))_{i,j=1}^n$ be a symmetric polynomial matrix which is positive semidefinite, i.e. $f(\mathbf{x}, \mathbf{y}) := \mathbf{y}^T P(\mathbf{x})\mathbf{y}$ is nonnegative on $\mathbb{R}^{n+1}$. The proof uses induction on n. If $n = 1$, then $f(\mathbf{x}, \mathbf{y}) = y_1^2 a_{11}(\mathbf{x}) \geq 0$ on $\mathbb{R}^2$; this implies $a_{11}(\mathbf{x}) \geq 0$ on $\mathbb{R}$ and thus $P(\mathbf{x}) = a_{11}(\mathbf{x})$ is SOS. Assume now $n \geq 2$ and the theorem holds for $n - 1$. Write $f(\mathbf{x}, \mathbf{y})$ as

$$f(\mathbf{x}, \mathbf{y}) = y_1^2 a_{11}(\mathbf{x}) + 2y_1\Big(\sum_{i \geq 2} a_{1i}(\mathbf{x}) y_i\Big) + \underbrace{\sum_{i,j \geq 2} a_{ij}(\mathbf{x}) y_i y_i}_{g(\mathbf{x},\mathbf{y})}.$$

As this expression is ≥ 0 for all y_1, by computing the discrimant, we deduce that $\Delta := a_{11}(\mathbf{x})g(\mathbf{x},\mathbf{y}) - (\sum_{i\geq 2} a_{1i}(\mathbf{x})y_i)^2 \geq 0$ on $\mathbb{R}^n$. The induction assumption implies that Δ is a sum of squares. Now, as

$$a_{11}(\mathbf{x})f(\mathbf{x},\mathbf{y}) = \Big(y_1 a_{11}(\mathbf{x}) + \sum_{i\geq 2} a_{1i}(\mathbf{x})y_i\Big)^2 + \Delta,$$

we deduce that $a_{11}(\mathbf{x})f(\mathbf{x},\mathbf{y})$ is a sum of squares, i.e. $a_{11}(\mathbf{x})P(\mathbf{x})$ is a SOS-matrix. Note that $a_{11}(\mathbf{x})$ is nonnegative on $\mathbb{R}$ (as it is a diagonal entry of $P(\mathbf{x})$). Thus the proof will be completed if we can show the following result:

PROPOSITION 3.39. *Let $\mathbb{R}[\mathbf{x}]$ be the ring of univariate polynomials, $a(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ be a nonzero nonnegative polynomial and $P(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]^{n\times n}$. If $a(\mathbf{x})P(\mathbf{x})$ is a SOS-matrix, then $P(\mathbf{x})$ too is a SOS-matrix.*

The proof of this result relies on the following lemmas.

LEMMA 3.40. *Let the vectors $D_i, E_i \in \mathbb{R}^{2d}$ $(i = 1, \ldots, n)$ satisfy the condition:*

$$D_i^T E_j + D_j^T E_i = 0,\ D_i^T D_j = E_i^T E_j\ \forall i, i = 1, \ldots, n. \tag{3.20}$$

There exists an orthogonal matrix U of order $2d$ such that UD_i, UE_i have the following form:

$$UD_i = (s_{i1}, t_{i1}, \ldots, s_{id}, t_{id}),\ UE_i = (-t_{i1}, s_{i1}, \ldots, -t_{id}, s_{id}) \tag{3.21}$$

for some scalars s_{ij}, t_{ij}.

Proof. The proof is by induction on $n \geq 1$. By (3.20), D_1, E_1 are orthogonal and have the same length s. Thus, up to an orthogonal transformation, we can assume that $D_1 = (s, 0, 0, \ldots, 0)$ and $E_1 = (0, s, 0, \ldots, 0)$. Using again (3.20), we see that D_i, E_i $(i \geq 2)$ have the form: $D_i = (s_i, t_i, \tilde{D}_i)$ and $E_i = (-t_i, s_i, \tilde{E}_i)$; thus the first two coordinates of D_i, E_i have the desired shape. As $\tilde{D}_i, \tilde{E}_i$ $(i \geq 2)$ satisfy again (3.20), we get the final result applying induction. □

LEMMA 3.41. *If $a(\mathbf{x})P(\mathbf{x})$ is a SOS-matrix with $a(\mathbf{x}) = (\mathbf{x} - \alpha)^2$ $(\alpha \in \mathbb{R})$, then $P(\mathbf{x})$ is a SOS-matrix.*

Proof. By assumption, $a(\mathbf{x})\cdot\mathbf{y}^T P(\mathbf{x})\mathbf{y}$ is a SOS polynomial, of the form $\sum_k (\sum_{i=1}^n y_i b_{ki}(\mathbf{x}))^2$ for some $b_{ki}(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$. This polynomial vanishes at $x = \alpha$, which implies $b_{ki}(\alpha) = 0$ for all i, k. Thus $\mathbf{x} - \alpha$ divides each $b_{ki}(\mathbf{x})$, i.e. $b_{ki}(\mathbf{x}) = (\mathbf{x} - \alpha)c_{ki}(\mathbf{x})$ for some $c_{ki}(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$. This gives the decomposition $\mathbf{y}^T P(\mathbf{x})\mathbf{y} = \sum_k (\sum_{i=1}^n y_i c_{ki}(\mathbf{x}))^2$, thus showing that $P(\mathbf{x})$ is a SOS-matrix. □

LEMMA 3.42. *If $a(\mathbf{x})P(\mathbf{x})$ is a SOS-matrix with $a(\mathbf{x}) = (\mathbf{x} + \alpha)^2 + \beta^2$ $(\alpha, \beta \in \mathbb{R})$, then $P(\mathbf{x})$ is a SOS-matrix.*

Proof. Up to a change of variables we can assume that $a(\mathbf{x}) = \mathbf{x}^2 + 1$. By assumption, $(\mathbf{x}^2 + 1)P(\mathbf{x}) = (A_i(\mathbf{x})^T A_j(\mathbf{x}))_{i,j=1}^n$ for some vectors

$A_i(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]^{2d}$ (for some d). Taking residues modulo the ideal $(\mathbf{x}^2+1)$, we can write $A_i(\mathbf{x}) = (\mathbf{x}^2+1)C_i(\mathbf{x}) + \mathbf{x}D_i + E_i$, where $C_i(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]^{2d}$, $D_i, E_i \in \mathbb{R}^{2d}$. This implies

$$\begin{aligned}(\mathbf{x}^2+1)P(\mathbf{x}) = \;& (\mathbf{x}^2+1)^2\ C_i(\mathbf{x})^T C_j(\mathbf{x}) \\ &+ (\mathbf{x}^2+1)\ (C_i(\mathbf{x})^T E_j + C_j(\mathbf{x})^T E_i + D_i^T D_j) \\ &+ \mathbf{x}(\mathbf{x}^2+1)\ (C_i(\mathbf{x})^T D_j + C_j(\mathbf{x})^T D_i) \\ &+ \mathbf{x}\ (D_i^T E_j + D_j^T E_i) \\ &+ E_i^T E_j - D_i^T D_j.\end{aligned}$$

Therefore, the vectors D_i, E_i satisfy (3.20). Hence, by Lemma 3.40, we can assume without loss of generality that they also satisfy (3.21) (replacing accordingly $C_i(\mathbf{x})$ by $UC_i(\mathbf{x})$.) Write the components of $C_i(\mathbf{x})$ as $C_i(\mathbf{x})^T = (C_{i1}(\mathbf{x}), \ldots, C_{i,2d}(\mathbf{x}))$. Then we define the new the polynomial vectors $F_i(\mathbf{x})$ obtained by recombining pairs of coordinates of $C_i(\mathbf{x})$ in the following way:

$$\begin{aligned}F_i(\mathbf{x})^T := (\;& C_{i2}(\mathbf{x}) + \mathbf{x}C_{i1}(\mathbf{x}), \mathbf{x}C_{i2}(\mathbf{x}) - C_{i1}(\mathbf{x}), \ldots \\ & \ldots, C_{i,2d}(\mathbf{x}) + \mathbf{x}C_{i,2d-1}(\mathbf{x}), \mathbf{x}C_{i,2d}(\mathbf{x}) - C_{i,2d-1}(\mathbf{x})).\end{aligned}$$

One can verify that $F_i(\mathbf{x})^T F_j(\mathbf{x}) = (\mathbf{x}^2+1)C_i(\mathbf{x})^T C_j(\mathbf{x})$. Next define the polynomial vectors $B_i(\mathbf{x}) := F_i(\mathbf{x}) + D_i$. Then, $P(\mathbf{x}) = (B_i(\mathbf{x})^T B_j(\mathbf{x}))_{i,j=1}^n$, thus showing that $P(\mathbf{x})$ is a SOS-matrix. To see it, it suffices to verify that $F_i(\mathbf{x})^T D_j = C_i(\mathbf{x})^T(\mathbf{x}D_j + E_j)$, which can be verified using the fact that D_i, E_i have the form (3.21). □

We can now conclude the proof of Proposition 3.39 (and thus of Theorem 3.31). As $a(\mathbf{x}) \geq 0$ on $\mathbb{R}$, $a(\mathbf{x})$ is a product of powers of factors of the form $(\mathbf{x}-\alpha)^2$ or $(\mathbf{x}+\alpha)^2 + \beta^2$. So the result follows by 'peeling off the factors' of $a(\mathbf{x})$ and using Lemmas 3.41 and 3.42.

3.7. Proof of Putinar's theorem. Schweighofer [150] gave a proof for Theorem 3.20 which is more elementary than Putinar's original proof and uses only Pólya's theorem (Theorem 3.25). Later he communicated to us the following shorter and simpler proof which combines some classical ideas from real algebraic geometry with an ingeneous argument of Marshall (in Claim 3.48).

Definition 3.43. *Call a set $M \subseteq \mathbb{R}[\mathbf{x}]$ a* quadratic module *if it contains* 1 *and is closed under addition and multiplication with squares, i.e., $1 \in M$, $M + M \subseteq M$ and $\Sigma M \subseteq M$. Call a quadratic module M* proper *if $-1 \notin M$ (i.e. $M \neq \mathbb{R}[\mathbf{x}]$).*

Given $g_1, \ldots, g_m \in \mathbb{R}[\mathbf{x}]$ the set $\mathbf{M}(g_1, \ldots, g_m)$ introduced in (3.13) is obviously a quadratic module. We begin with some preliminary results about quadratic modules.

Lemma 3.44. *If $M \subseteq \mathbb{R}[\mathbf{x}]$ is a quadratic module, then $\mathcal{I} := M \cap -M$ is an ideal.*

Proof. For $f \in \mathbb{R}[\mathbf{x}]$ and $g \in \mathcal{I}$, $fg = \left(\frac{f+1}{2}\right)^2 g - \left(\frac{f-1}{2}\right)^2 g \in \mathcal{I}$. □

LEMMA 3.45. *Let $M \subseteq \mathbb{R}[\mathbf{x}]$ be a maximal proper quadratic module. Then $M \cup -M = \mathbb{R}[\mathbf{x}]$.*

Proof. Assume $f \in \mathbb{R}[\mathbf{x}] \setminus (M \cup -M)$. By maximality of M, the quadratic modules $M + f\Sigma$ and $M - f\Sigma$ are not proper, i.e., we find $g_1, g_2 \in M$ and $s_1, s_2 \in \Sigma$ such that $-1 = g_1 + s_1 f$ and $-1 = g_2 - s_2 f$. Multiplying the first equation by s_2 and the second one by s_1, we get $s_1 + s_2 + s_1 g_2 + s_2 g_1 = 0$. This implies $s_1, s_2 \in \mathcal{I} := M \cap -M$. Since $\mathcal{I}$ is an ideal, we get $s_1 f \in \mathcal{I} \subseteq M$ and therefore $-1 = g_1 + s_1 f \in M$, a contradiction. □

LEMMA 3.46. *Let $M \subseteq \mathbb{R}[\mathbf{x}]$ be a maximal proper quadratic module which is Archimedean, set $\mathcal{I} := M \cap -M$ and let $f \in \mathbb{R}[\mathbf{x}]$. Then there is exactly one $a \in \mathbb{R}$ such that $f - a \in \mathcal{I}$.*

Proof. Consider the sets

$$A := \{a \in \mathbb{R} \mid f - a \in M\} \qquad \text{and} \qquad B := \{b \in \mathbb{R} \mid b - f \in M\}.$$

As M is Archimedean, the sets A and B are not empty. We have to show that $A \cap B$ is a singleton. Since M is proper, it does not contain any negative real number. Therefore $a \leq b$ for all $a \in A$, $b \in B$. Set $a_0 := \sup A$ and $b_0 := \inf B$. Thus $a_0 \leq b_0$. Moreover, $a_0 = b_0$. Indeed if $a_0 < c < b_0$, then $f - c \notin M \cup -M$, which contradicts the fact that $\mathbb{R}[\mathbf{x}] = M \cup -M$ (by Lemma 3.45). It suffices now to show that $a_0 \in A$ and $b_0 \in B$, since this will imply that $A \cap B = \{a_0\}$ and thus conclude the proof. We show that $a_0 = \sup A \in A$. For this assume that $a_0 \notin A$, i.e., $f - a_0 \notin M$. Then $M' := M + (f - a_0)\Sigma$ is a quadratic module that cannot be proper by the maximality of M; that is, $-1 = g + (f - a_0)s$ for some $g \in M$ and $s \in \Sigma$. As M is Archimedean we can choose $N \in \mathbb{N}$ such that $N - s \in M$ and $\epsilon \in \mathbb{R}$ such that $0 < \epsilon < \frac{1}{N}$. As $a_0 - \epsilon \in A$, we have $f - (a_0 - \epsilon) \in M$. Then we have $-1 + \epsilon s = g + (f - a_0 + \epsilon)s \in M$ and $\epsilon N - \epsilon s \in M$. Adding these two equations, we get $\epsilon N - 1 \in M$ which is impossible since $\epsilon N - 1 < 0$ and M is proper. One can prove that $b_0 \in B$ in the same way. □

We now prove Theorem 3.20. Assume $p \in \mathbb{R}[\mathbf{x}]$ is positive on K; we show that $p \in \mathbf{M}(g_1, \ldots, g_m)$. We state two intermediary results.

CLAIM 3.47. *There exists $s \in \Sigma$ such that $sp \in 1 + \mathbf{M}(g_1, \ldots, g_m)$.*

Proof. We have to prove that the quadratic module $M_0 := \mathbf{M}(g_1, \ldots, g_m) - p\Sigma$ is not proper. For this assume that M_0 is proper; we show the existence of $a \in K$ for which $p(a) \leq 0$, thus contradicting the assumption $p > 0$ on K. By Zorn's lemma, we can extend M_0 to a maximal proper quadratic module $M \supseteq M_0$. As $M \supseteq \mathbf{M}(g_1, \ldots, g_m)$, M is Archimedean. Applying Lemma 3.46, there exists $a \in \mathbb{R}^n$ such that $\mathbf{x}_i - a_i \in \mathcal{I} := M \cap -M$ for all $i \in \{1, \ldots, n\}$. Since $\mathcal{I}$ is an ideal (by Lemma 3.44), $f - f(a) \in \mathcal{I}$ for any $f \in \mathbb{R}[\mathbf{x}]$. In particular, for $f = g_j$, we find that $g_j(a) = g_j - (g_j - g_j(a)) \in M$ since $g_j \in \mathbf{M}(g_1, \ldots, g_m) \subseteq M$ and $-(g_j - g_j(a)) \in M$, which implies $g_j(a) \geq 0$. Therefore, $a \in K$. Finally, $-p(a) = (p - p(a)) - p \in M$ since $p - p(a) \in \mathcal{I} \subseteq M$ and $-p \in M_0 \subseteq M$, which implies $-p(a) \geq 0$. □

Claim 3.48. *There exist $g \in \mathbf{M}(g_1, \ldots, g_m)$ and $N \in \mathbb{N}$ such that $N - g \in \Sigma$ and $gp \in 1 + \mathbf{M}(g_1, \ldots, g_m)$.*

Proof. (Marshall [106, 5.4.4]). Choose s as in Claim 3.47, i.e. $s \in \Sigma$ and $sp - 1 \in \mathbf{M}(g_1, \ldots, g_m)$. Using (3.16), there exists $k \in \mathbb{N}$ such that $2k - s,\ 2k - s^2p - 1 \in \mathbf{M}(g_1, \ldots, g_m)$. Set $g := s(2k - s)$ and $N := k^2$. Then $g \in \mathbf{M}(g_1, \ldots, g_m)$, $N - g = k^2 - 2ks + s^2 = (k - s)^2 \in \Sigma$. Moreover, $gp - 1 = s(2k - s)p - 1 = 2k(sp - 1) + (2k - s^2p - 1) \in \mathbf{M}(g_1, \ldots, g_m)$, since $sp - 1, 2k - s^2p - 1 \in \mathbf{M}(g_1, \ldots, g_m)$. □

We can now conclude the proof. Choose g, N as in Claim 3.48 and $k \in \mathbb{N}$ such that $k + p \in \mathbf{M}(g_1, \ldots, g_m)$. We may assume $N > 0$. Note that

$$(k - \frac{1}{N}) + p = \frac{1}{N}\Big((N - g)(k + p) + (gp - 1) + kg\Big) \in \mathbf{M}(g_1, \ldots, g_m).$$

Applying this iteratively we can make $k = (kN)\frac{1}{N}$ smaller and smaller until reaching 0 and thus obtain $p \in \mathbf{M}(g_1, \ldots, g_m)$. This concludes the proof of Theorem 3.20.

3.8. The cone of sums of squares is closed. As we saw earlier, for d even, the inclusion $\Sigma_{n,d} \subseteq \mathcal{P}_{n,d}$ is strict except in the three special cases $(n, d) = (1, d)$, $(n, 2)$, $(2, 4)$. One may wonder how much the two cones differ in the other cases. This question will be addressed in Section 7.1, where we will mention a result of Blekherman (Theorem 7.1) showing that, when the degree d is fixed and the number n of variables grows, there are much more nonnegative polynomials than sums of squares. On the other hand, one can show that any nonnegative polynomial is the limit (coordinate-wise) of a sequence of SOS polynomials (cf. Theorem 7.3); note that the degrees of the polynomials occurring in such a sequence cannot be bounded, since the cone $\Sigma_{n,d}$ is a closed set.

We now prove a more general result (Theorem 3.49), which implies directly that $\Sigma_{n,d}$ is closed (for the case $K = \mathbb{R}^n$). Given polynomials $g_1, \ldots, g_m \in \mathbb{R}[\mathbf{x}]$ and an integer t, set $g_0 := 1$ and define the set

$$\mathbf{M}_t(g_1, \ldots, g_m) := \{\sum_{j=0}^{m} s_j g_j \mid s_j \in \Sigma, \deg(s_j g_j) \leq t\ (0 \leq j \leq m)\}, \quad (3.22)$$

which can be seen as the "truncation at degree t" of the quadratic module $\mathbf{M}(g_1, \ldots, g_m)$. Let K be as in (1.2). Its interior $\mathrm{int}(K)$ (for the Euclidean topology) consists of the points $x \in K$ for which there exists a (full dimensional) ball centered at x and contained in K. Obviously

$$K' := \{x \in \mathbb{R}^n \mid g_j(x) > 0\ \forall j = 1, \ldots, m\} \subseteq \mathrm{int}(K).$$

The inclusion may be strict (e.g. $0 \in \text{int}(K)\backslash K'$ for $K = \{x \in \mathbb{R} \mid x^2 \geq 0\}$). However,

$$\text{int}(K) \neq \emptyset \Longleftrightarrow K' \neq \emptyset \tag{3.23}$$

assuming no g_j is the zero polynomial. Indeed if $K' = \emptyset$ and B is a ball contained in K, then the polynomial $\prod_{j=1}^m g_j$ vanishes on K and thus on B, hence it must be the zero polynomial, a contradiction. The next result will also be used later in Section 6 to show the absence of a duality gap between the moment/SOS relaxations.

Theorem 3.49. *[130, 150] If K has a nonempty interior then $\mathbf{M}_t(g_1, \ldots, g_m)$ is closed in $\mathbb{R}[\mathbf{x}]_t$ for any $t \in \mathbb{N}$.*

Proof. Note that $\deg(s_j g_j) \leq t$ is equivalent to $\deg s_j \leq 2k_j$, setting $k_j := \lfloor (t - \deg(g_j))/2 \rfloor$. Set $\Lambda_j := \dim \mathbb{R}[\mathbf{x}]_{k_j} = |\mathbb{N}^n_{k_j}|$. Then any polynomial $f \in \mathbf{M}_t(g_1, \ldots, g_m)$ is of the form $f = \sum_{j=0}^m s_j g_j$ with $s_j = \sum_{l_j=1}^{\Lambda_j} (u^{(j)}_{l_j})^2$ for some $u^{(j)}_1, \ldots, u^{(j)}_{\Lambda_j} \in \mathbb{R}[\mathbf{x}]_{k_j}$. In other words, $\mathbf{M}_t(g_1, \ldots, g_m)$ is the image of the following map

$$\begin{array}{ll} \varphi : & \mathcal{D} := (\mathbb{R}[\mathbf{x}]_{k_0})^{\Lambda_0} \times \ldots \times (\mathbb{R}[\mathbf{x}]_{k_m})^{\Lambda_m} \to \mathbb{R}[\mathbf{x}]_t \\ & u = ((u^{(0)}_{l_0})^{\Lambda_0}_{l_0=1}), \ldots, (u^{(m)}_{l_m})^{\Lambda_m}_{l_m=1})) \mapsto \varphi(u) = \sum_{j=0}^m \sum_{l_j=1}^{\Lambda_j} (u^{(j)}_{l_j})^2 g_j. \end{array}$$

We may identify the domain $\mathcal{D}$ of φ with the space $\mathbb{R}^\Lambda$ (of suitable dimension Λ); choose a norm on this space and let S denote the unit sphere in $\mathcal{D}$. Then $V := \varphi(S)$ is a compact set in the space $\mathbb{R}[\mathbf{x}]_t$, which is also equipped with a norm. Note that any $f \in \mathbf{M}_t(g_1, \ldots, g_m)$ is of the form $f = \lambda v$ for some $\lambda \in \mathbb{R}_+$ and $v \in V$. We claim that $0 \notin V$. Indeed, by assumption, $\text{int}(K) \neq \emptyset$ and thus, by (3.23), there exists a full dimensional ball $B \subseteq K$ such that each polynomial g_j ($j = 1, \ldots, m$) is positive on B. Hence, for any $u \in S$, if $\varphi(u)$ vanishes on B then each polynomial arising as component of u vanishes on B, implying $u = 0$. This shows that $\varphi(u) \neq 0$ if $u \in S$, i.e. $0 \notin V$. We now show that $\mathbf{M}_t(g_1, \ldots, g_m)$ is closed. For this consider a sequence $f_k \in \mathbf{M}_t(g_1, \ldots, g_m)$ ($k \geq 0$) converging to a polynomial f; we show that $f \in \mathbf{M}_t(g_1, \ldots, g_m)$. Write $f_k = \lambda_k v_k$ where $v_k \in V$ and $\lambda_k \in \mathbb{R}_+$. As V is compact there exists a subsequence of $(v_k)_{k \geq 0}$, again denoted as $(v_k)_{k \geq 0}$ for simplicity, converging say to $v \in V$. As $0 \notin V$, $v \neq 0$ and thus $\lambda_k = \frac{\|f_k\|}{\|v_k\|}$ converges to $\frac{\|f\|}{\|v\|}$ as $k \to \infty$. Therefore, $f_k = \lambda_k v_k$ converges to $\frac{\|f\|}{\|v\|} v$ as $k \to \infty$, which implies $f = \frac{\|f\|}{\|v\|} v \in \mathbf{M}_t(g_1, \ldots, g_m)$. □

Corollary 3.50. *The cone $\Sigma_{n,d}$ is a closed cone.*

Proof. Apply Theorem 3.49 to $\mathbf{M}_d(g_1) = \Sigma_{n,d}$ for $g_1 := 1$. □

When the set K has an empty interior one can prove an analogue of Theorem 3.49 after factoring out through the vanishing ideal of K. More precisely, let $\mathcal{I}(K) = \{f \in \mathbb{R}[\mathbf{x}] \mid f(x) = 0 \; \forall x \in K\}$ be the vanishing ideal of K, and consider the mapping $p \in \mathbb{R}[\mathbf{x}] \mapsto p' := p + \mathcal{I}(K) \in \mathbb{R}[\mathbf{x}]/\mathcal{I}(K)$

mapping any polynomial to its coset in $\mathbb{R}[\mathbf{x}]/\mathcal{I}(K)$. Define the image under this mapping of the quadratic module $\mathbf{M}_t(g_1,\dots,g_m)$ as

$$\mathbf{M}'_t(g_1,\dots,g_m) = \{p' = p + \mathcal{I}(K) \mid p \in \mathbf{M}_t(g_1,\dots,g_m)\} \subseteq \mathbb{R}[\mathbf{x}]_t/\mathcal{I}(K). \tag{3.24}$$

When K has a nonempty interior, $\mathcal{I}(K) = \{0\}$ and thus $\mathbf{M}'_t(g_1,\dots,g_m) = \mathbf{M}_t(g_1,\dots,g_m)$. Marshall [104] proved the following result, extending Theorem 3.49. (See [106, Lemma 4.1.4] for an analogous result where the ideal $\mathcal{I}(K)$ is replaced by its subideal $\sqrt{\mathbf{M}(g_1,\dots,g_m) \cap -\mathbf{M}(g_1,\dots,g_m)}$.)

Theorem 3.51. *[104] The set $\mathbf{M}'_t(g_1,\dots,g_m)$ is closed in $\mathbb{R}[\mathbf{x}]_t/\mathcal{I}(K)$.*

Proof. The proof is along the same lines as that of Theorem 3.49, except one must now factor out through the ideal $\mathcal{I}(K)$. Set $g_0 := 1$, $J := \{j \in \{0,1,\dots,m\} \mid g_j \notin \mathcal{I}(K)\}$, $\mathrm{Ann}(g) := \{p \in \mathbb{R}[\mathbf{x}] \mid pg \in \mathcal{I}(K)\}$ for $g \in \mathbb{R}[\mathbf{x}]$. For $j = 0,\dots,m$, set $k_j := \lfloor (t - \deg(g_j))/2 \rfloor$ (as in the proof of Theorem 3.49), $\mathcal{A}_j := \mathbb{R}[\mathbf{x}]_{k_j} \cap \mathrm{Ann}(g_j)$. Let $\mathcal{B}_j \subseteq \mathbb{R}[\mathbf{x}]_{k_j}$ be a set of monomials forming a basis of $\mathbb{R}[\mathbf{x}]_{k_j}/\mathcal{A}_j$; that is, any polynomial $f \in \mathbb{R}[\mathbf{x}]_{k_j}$ can be written in a unique way as $p = r + q$ where $r \in \mathrm{Span}_{\mathbb{R}}(\mathcal{B}_j)$ and $q \in \mathcal{A}_j$. Let $\Lambda_j := |\mathcal{B}_j| = \dim \mathbb{R}[\mathbf{x}]_{k_j}/\mathcal{A}_j$. Consider the mapping

$$\begin{aligned} \varphi:\ \mathcal{D} := \prod_{j\in J} (\mathbb{R}[\mathbf{x}]_{k_j}/\mathcal{A}_j)^{\Lambda_j} &\quad\rightarrow\quad \mathbb{R}[\mathbf{x}]_t/\mathcal{I}(K) \\ u = ((u_{l_j}^{(j)} + \mathcal{A}_j)_{l_j=1}^{\Lambda_j})_{j\in J} &\quad\mapsto\quad \varphi(u) = \sum_{j\in J} \sum_{l_j=1}^{\Lambda_j} (u_{l_j}^{(j)})^2 g_j + \mathcal{I}(K). \end{aligned}$$

Note first that φ is well defined; indeed, $u - v \in \mathcal{A}_j$ implies $(u-v)g_j \in \mathcal{I}(K)$ and thus $u^2 g_j - v^2 g_j = (u+v)(u-v)g_j \in \mathcal{I}(K)$. Next we claim that the image of the domain $\mathcal{D}$ under φ is precisely the set $\mathbf{M}'_t(g_1,\dots,g_m)$. That is,

$$\forall f \in \mathbf{M}_t(g_1,\dots,g_m),\ \exists u \in \mathcal{D} \ \text{ s.t. } \ f - \sum_{j\in J} \sum_{l_j=1}^{\Lambda_j} (u_{l_j}^{(j)})^2 g_j \in \mathcal{I}(K). \tag{3.25}$$

For this write $f = \sum_{j=0}^m s_j g_j$ where $s_j \in \Sigma$ and $\deg(s_j) \le t - \deg(g_j)$. Thus $f \equiv \sum_{j\in J} s_j g_j \mod \mathcal{I}(K)$. Say $s_j = \sum_{h_j} (a_{h_j}^{(j)})^2$ where $a_{h_j}^{(j)} \in \mathbb{R}[\mathbf{x}]_{k_j}$. Write $a_{h_j}^{(j)} = r_{h_j}^{(j)} + q_{h_j}^{(j)}$, where $r_{h_j}^{(j)} \in \mathrm{Span}_{\mathbb{R}}(\mathcal{B}_j)$ and $q_{h_j}^{(j)} \in \mathcal{A}_j$. Then, $s_j g_j = \sum_{j\in J} (\sum_{h_j} (r_{h_j}^{(j)})^2) g_j \mod \mathcal{I}(K)$, since $q_{h_j}^{(j)} g_j \in \mathcal{I}(K)$ as $q_{h_j}^{(j)} \in \mathcal{A}_j \subseteq \mathrm{Ann}(g_j)$. Moreover, as each $r_{h_j}^{(j)}$ lies in $\mathrm{Span}_{\mathbb{R}}(\mathcal{B}_j)$ with $|\mathcal{B}_j| = \Lambda_j$, by the Gram-matrix method (recall Section 3.3), we deduce that $\sum_{h_j} (r_{h_j}^{(j)})^2$ can be written as another sum of squares involving only Λ_j squares, i.e. $\sum_{h_j} (r_{h_j}^{(j)})^2 = \sum_{l_j=1}^{\Lambda_j} (u_{l_j}^{(j)})^2$ with $u_{l_j}^{(j)} \in \mathrm{Span}_{\mathbb{R}}(\mathcal{B}_j)$; this shows (3.25).

We now show that $\varphi^{-1}(0) = 0$. For this assume that $\varphi(u) = 0$ for some $u \in \mathcal{D}$, i.e. $f := \sum_{j\in J} \sum_{l_j=1}^{\Lambda_j} (u_{l_j}^{(j)})^2 g_j \in \mathcal{I}(K)$; we show that $u_{l_j}^{(j)} \in \mathcal{A}_j$ for all j, l_j. Fix $x \in K$. Then $f(x) = 0$ and, as $g_j(x) \geq 0 \ \forall j$, $(u_{l_j}^{(j)}(x))^2 g_j(x) = 0$, i.e. $u_{l_j}^{(j)}(x) g_j(x) = 0 \ \forall j, l_j$. This shows that each polynomial $u_{l_j}^{(j)} g_j$ lies in $\mathcal{I}(K)$, that is, $u_{l_j}^{(j)} \in \mathcal{A}_j$. We can now proceed as in the proof of Theorem 3.49 to conclude that $M'_t(g_1, \ldots, g_m)$ is a closed set in $\mathbb{R}[\mathbf{x}]_t/\mathcal{I}(K)$. □

4. Moment sequences and moment matrices.

4.1. Some basic facts.

We introduce here some basic definitions and facts about measures, moment sequences and moment matrices.

4.1.1. Measures. Let $\mathcal{X}$ be a Hausdorff locally compact topological space. Being *Hausdorff* means that for distinct points $x, x' \in \mathcal{X}$ there exist disjoint open sets U and U' with $x \in U$ and $x' \in U'$; being *locally compact* means that for each $x \in \mathcal{X}$ there is an open set U containing x whose closure $\bar{U}$ is compact. Then $\mathcal{B}(\mathcal{X})$ denotes the *Borel σ-algebra*, defined as the smallest collection of subsets of $\mathcal{X}$ which contains all open sets and is closed under taking set differences, and countable unions and intersections; sets in $\mathcal{B}(\mathcal{X})$ are called *Borel sets.* A *Borel measure* is any measure μ on $\mathcal{B}(\mathcal{X})$, i.e. a function from $\mathcal{B}(\mathcal{X})$ to $\mathbb{R}_+ \cup \{\infty\}$ satisfying: $\mu(\emptyset) = 0$, and $\mu(\cup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} \mu(A_i)$ for any pairwise disjoint $A_i \in \mathcal{B}(\mathcal{X})$. Thus measures are implicitly assumed to be positive. A probability measure is a measure with total mass $\mu(\mathcal{X}) = 1$.

Throughout we will consider Borel measures on $\mathbb{R}^n$. Thus when speaking of 'a measure' on $\mathbb{R}^n$ we mean a (positive) Borel measure on $\mathbb{R}^n$.

Given a measure μ on $\mathbb{R}^n$, its *support* $\text{supp}(\mu)$ is the smallest closed set $S \subseteq \mathbb{R}^n$ for which $\mu(\mathbb{R}^n \setminus S) = 0$. We say that μ is a measure *on K* or a measure *supported by K* $\subseteq \mathbb{R}^n$ if $\text{supp}(\mu) \subseteq K$.

Given $x \in \mathbb{R}^n$, δ_x denotes the *Dirac measure* at x, with support $\{x\}$ and having mass 1 at x and mass 0 elsewhere.

When the support of a measure μ is finite, say, $\text{supp}(\mu) = \{x_1, \ldots, x_r\}$, then μ is of the form $\mu = \sum_{i=1}^r \lambda_i \delta_{x_i}$ for some $\lambda_1, \ldots, \lambda_r > 0$; the x_i are called the *atoms* of μ and one also says that μ is a *r-atomic* measure.

4.1.2. Moment sequences. Given a measure μ on $\mathbb{R}^n$, the quantity $y_\alpha := \int x^\alpha \mu(dx)$ is called its *moment of order α*. Then, the sequence $(y_\alpha)_{\alpha\in\mathbb{N}^n}$ is called the sequence of moments of the measure μ and, given $t \in \mathbb{N}$, the (truncated) sequence $(y_\alpha)_{\alpha\in\mathbb{N}_t^n}$ is called the sequence of moments of μ up to order t. When y is the sequence of moments of a measure we also say that μ is a *representing measure* for y. The sequence of moments of the Dirac measure δ_x is the vector $\zeta_x := (x^\alpha)_{\alpha\in\mathbb{N}^n}$, called the *Zeta vector* of x (see the footnote on page 131 for a motivation). Given an integer $t \geq 1$, $\zeta_{t,x} := (x^\alpha)_{\alpha\in\mathbb{N}_t^n}$ denotes the *truncated Zeta vector.*

A basic problem in the theory of moments concerns the characterization of *(infinite or truncated) moment sequences*, i.e., the characterization of those sequences $y = (y_\alpha)_\alpha$ that are the sequences of moments of some measure. Given a subset $K \subseteq \mathbb{R}^n$, the K*-moment problem* asks for the characterization of those sequences that are sequences of moments of some measure supported by the set K. This problem has received considerable attention in the literature, especially in the case when $K = \mathbb{R}^n$ (the basic moment problem) or when K is a compact semialgebraic set, and it turns out to be related to our polynomial optimization problem, as we see below in this section. For more information on the moment problem see e.g. [11, 12, 28, 29, 30, 31, 42, 76, 77, 134, 145] and references therein.

4.1.3. Moment matrices. The following notions of moment matrix and shift operator play a central role in the moment problem. Given a sequence $y = (y_\alpha)_{\alpha\in\mathbb{N}^n} \in \mathbb{R}^{\mathbb{N}^n}$, its *moment matrix* is the (infinite) matrix $M(y)$ indexed by $\mathbb{N}^n$, with (α, β)th entry $y_{\alpha+\beta}$, for $\alpha, \beta \in \mathbb{N}^n$. Similarly, given an integer $t \geq 1$ and a (truncated) sequence $y = (y_\alpha)_{\alpha\in\mathbb{N}^n_{2t}} \in \mathbb{R}^{\mathbb{N}^n_{2t}}$, its *moment matrix of order* t is the matrix $M_t(y)$ indexed by $\mathbb{N}^n_t$, with (α, β)th entry $y_{\alpha+\beta}$, for $\alpha, \beta \in \mathbb{N}^n_t$.

Given $g \in \mathbb{R}[\mathbf{x}]$ and $y \in \mathbb{R}^{\mathbb{N}^n}$, define the new sequence

$$gy := M(y)g \in \mathbb{R}^{\mathbb{N}^n}, \tag{4.1}$$

called *shifted vector*, with αth entry $(gy)_\alpha := \sum_\beta g_\beta y_{\alpha+\beta}$ for $\alpha \in \mathbb{N}^n$. The notation gy will also be used for denoting the truncated vector $((gy)_\alpha)_{\alpha\in\mathbb{N}^n_t}$ of $\mathbb{R}^{\mathbb{N}^n_t}$ for an integer $t \geq 1$. The moment matrices of gy are also known as the *localizing matrices*, since they can be used to "localize" the support of a representing measure for y.

4.1.4. Moment matrices and (bi)linear forms on $\mathbb{R}[\mathbf{x}]$. Given $y \in \mathbb{R}^{\mathbb{N}^n}$, define the linear form $L_y \in (\mathbb{R}[\mathbf{x}])^*$ by

$$L_y(f) := y^T\text{vec}(f) = \sum_\alpha y_\alpha f_\alpha \quad \text{for } f = \sum_\alpha f_\alpha \mathbf{x}^\alpha \in \mathbb{R}[\mathbf{x}]. \tag{4.2}$$

We will often use the following simple 'calculus' involving moment matrices.

Lemma 4.1. *Let $y \in \mathbb{R}^{\mathbb{N}^n}$, $L_y \in (\mathbb{R}[\mathbf{x}])^*$ the associated linear form from (4.2), and let $f, g, h \in \mathbb{R}[\mathbf{x}]$.*

(i) $L_y(fg) = \text{vec}(f)^T M(y)\text{vec}(g)$; *in particular,* $L_y(f^2) = \text{vec}(f)^T M(y)\text{vec}(f)$, $L_y(f) = \text{vec}(1)^T M(y)\text{vec}(f)$.

(ii) $L_y(fgh) = \text{vec}(f)^T M(y)\text{vec}(gh) = \text{vec}(fg)^T M(y)\text{vec}(h) = \text{vec}(f)^T M(hy)\text{vec}(g)$.

Proof. (i) Setting $f = \sum_\alpha f_\alpha \mathbf{x}^\alpha$, $g = \sum_\beta g_\beta \mathbf{x}^\beta$, we have $fg = \sum_\gamma(\sum_{\alpha,\beta|\alpha+\beta=\gamma} f_\alpha g_\beta)\mathbf{x}^\gamma$. Then $L_y(fg) = \sum_\gamma y_\gamma(\sum_{\alpha,\beta|\alpha+\beta=\gamma} f_\alpha g_\beta)$, while $\text{vec}(f)^T M(y)\text{vec}(g) = \sum_\alpha \sum_\beta f_\alpha g_\beta y_{\alpha+\beta}$, thus equal to $L_y(fg)$. The last part of (i) follows directly.

(ii) By (i) $\mathrm{vec}(f)^T M(y)\mathrm{vec}(gh) = \mathrm{vec}(fg)^T M(y)\mathrm{vec}(h) = L_y(fgh)$, in turn equal to $\sum_{\alpha,\beta,\gamma} f_\alpha g_\beta h_\gamma y_{\alpha+\beta+\gamma}$. Finally, $\mathrm{vec}(f)^T M(hy)\mathrm{vec}(g) = \sum_\delta (hy)_\delta (fg)_\delta = \sum_\delta (\sum_\gamma h_\gamma y_{\gamma+\delta})(\sum_{\alpha,\beta|\alpha+\beta=\delta} f_\alpha g_\beta)$ which, by exchanging the summations, is equal to $\sum_{\alpha,\beta,\gamma} f_\alpha g_\beta h_\gamma y_{\alpha+\beta+\gamma} = L_y(fgh)$. □

Given $y \in \mathbb{R}^{\mathbb{N}^n}$, we can also define the bilinear form on $\mathbb{R}[\mathbf{x}]$

$$(f,g) \in \mathbb{R}[\mathbf{x}] \times \mathbb{R}[\mathbf{x}] \mapsto L_y(fg) = \mathrm{vec}(f)^T M(y)\mathrm{vec}(g),$$

whose associated quadratic form

$$f \in \mathbb{R}[\mathbf{x}] \mapsto L_y(f^2) = \mathrm{vec}(f)^T M(y)\mathrm{vec}(f)$$

is positive semidefinite, i.e. $L_y(f^2) \geq 0$ for all $f \in \mathbb{R}[\mathbf{x}]$, precisely when the moment matrix $M(y)$ is positive semidefinite.

4.1.5. Necessary conditions for moment sequences. The next lemma gives some easy well known necessary conditions for moment sequences.

LEMMA 4.2. *Let $g \in \mathbb{R}[\mathbf{x}]$ and $d_g = \lceil \deg(g)/2 \rceil$.*

(i) *If $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$ is the sequence of moments (up to order $2t$) of a measure μ, then $M_t(y) \succeq 0$ and $\mathrm{rank}\, M_t(y) \leq |\mathrm{supp}(\mu)|$. Moreover, for $p \in \mathbb{R}[\mathbf{x}]_t$, $M_t(y)p = 0$ implies $\mathrm{supp}(\mu) \subseteq V_\mathbb{R}(p) = \{x \in \mathbb{R}^n \mid p(x) = 0\}$. Therefore, $\mathrm{supp}(\mu) \subseteq V_\mathbb{R}(\mathrm{Ker}\, M_t(y))$.*

(ii) *If $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$ $(t \geq d_g)$ is the sequence of moments of a measure μ supported by the set $K := \{x \in \mathbb{R}^n \mid g(x) \geq 0\}$, then $M_{t-d_g}(gy) \succeq 0$.*

(iii) *If $y \in \mathbb{R}^{\mathbb{N}^n}$ is the sequence of moments of a measure μ, then $M(y) \succeq 0$. Moreover, if $\mathrm{supp}(\mu) \subseteq \{x \in \mathbb{R}^n \mid g(x) \geq 0\}$, then $M(gy) \succeq 0$ and, if μ is r-atomic, then $\mathrm{rank}\ M(y) = r$.*

Proof. (i) For $p \in \mathbb{R}[\mathbf{x}]_t$, we have:

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

This shows that $M_t(y) \succeq 0$. If $M_t(y)p = 0$, then $0 = p^T M_t(y) p = \int p(x)^2 \mu(dx)$. This implies that the support of μ is contained in the set $V_\mathbb{R}(p)$ of real zeros of p. [To see it, note that, as $V_\mathbb{R}(p)$ is a closed set, $\mathrm{supp}(\mu) \subseteq V_\mathbb{R}(p)$ holds if we can show that $\mu(\mathbb{R}^n \setminus V_\mathbb{R}(p)) = 0$. Indeed, $\mathbb{R}^n \setminus V_\mathbb{R}(p) = \bigcup_{k \geq 0} U_k$, setting $U_k := \{x \in \mathbb{R}^n \mid p(x)^2 \geq \frac{1}{k}\}$ for positive $k \in \mathbb{N}$. As $0 = \int p(x)^2 \mu(dx) = \int_{\mathbb{R}^n \setminus V_\mathbb{R}(p)} p(x)^2 \mu(dx) \geq \int_{U_k} p(x)^2 \mu(dx) \geq \frac{1}{k}\mu(U_k)$, this implies $\mu(U_k) = 0$ for all k and thus $\mu(\mathbb{R}^n \setminus V_\mathbb{R}(p)) = 0$.] The inequality $\mathrm{rank}\, M_t(y) \leq |\mathrm{supp}(\mu)|$ is trivial if μ has an infinite support. So assume that μ is r-atomic, say, $\mu = \sum_{i=1}^r \lambda_i \delta_{x_i}$ where $\lambda_1, \ldots, \lambda_r > 0$ and $x_1, \ldots, x_r \in \mathbb{R}^n$. Then, $M_t(y) = \sum_{i=1}^r \lambda_i \zeta_{t,x_i} (\zeta_{t,x_i})^T$, which shows that

rank $M_t(y) \leq r$.
(ii) For $p \in \mathbb{R}[\mathbf{x}]_{t-d_g}$, we have:

$$p^T M_{t-d_g}(gy)p = \sum_{\alpha,\beta\in\mathbb{N}^n_{t-d_g}} p_\alpha p_\beta (gy)_{\alpha+\beta}$$

$$= \sum_{\alpha,\beta\in\mathbb{N}^n_{t-d_g}} \sum_{\gamma\in\mathbb{N}^n} p_\alpha p_\beta g_\gamma y_{\alpha+\beta+\gamma} = \int_K g(x)p(x)^2\mu(dx) \geq 0.$$

This shows that $M_{t-d_g}(gy) \succeq 0$.
(iii) The first two claims follow directly from (i), (ii). Assume now $\mu = \sum_{i=1}^r \lambda_i \delta_{x_i}$ where $\lambda_i > 0$ and $x_1, \ldots, x_r$ are distinct points of $\mathbb{R}^n$. One can easily verify that the vectors ζ_{x_i} $(i = 1, \ldots, r)$ are linearly independent (using, e.g., the existence of interpolation polynomials at $x_1, \ldots, x_r$; see Lemma 2.3). Then, as $M(y) = \sum_{i=1}^r \lambda_i \zeta_{x_i}\zeta_{x_i}^T$, rank $M(y) = r$. □

Note that the inclusion $\text{supp}(\mu) \subseteq V_\mathbb{R}(\text{Ker}\, M_t(y))$ from Lemma 4.2 (i) may be strict in general; see Fialkow [41] for such an example. On the other hand, we will show in Theorem 5.29 that when $M_t(y) \succeq 0$ with $\text{rank}\, M_t(y) = \text{rank}\, M_{t-1}(y)$, then equality $\text{supp}(\mu) = V_\mathbb{R}(\text{Ker}\, M_t(y))$ holds. The next result follows directly from Lemma 4.2; d_K was defined in (1.10).

Corollary 4.3. *If $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$ is the sequence of moments (up to order $2t$) of a measure supported by the set K then, for any $t \geq d_K$,*

$$M_t(y) \succeq 0, \quad M_{t-d_{g_j}}(g_j y) \succeq 0 \ (j = 1, \ldots, m). \tag{4.3}$$

We will discuss in Section 5 several results of Curto and Fialkow showing that, under certain restrictions on the rank of the matrix $M_t(y)$, the condition (4.3) is sufficient for ensuring that y is the sequence of moments of a measure supported by K. Next we indicate how the above results lead to the moment relaxations for the polynomial optimization problem.

4.2. Moment relaxations for polynomial optimization. Lasserre [78] proposed the following strategy to approximate the problem (1.1). First observe that

$$p^{\min} := \inf_{x\in K} p(x) = \inf_\mu \int_K p(x)\mu(dx)$$

where the second infimum is taken over all probability measures μ on $\mathbb{R}^n$ supported by the set K. Indeed, for any $x_0 \in K$, $p(x_0) = \int p(x)\mu(dx)$ for the Dirac measure $\mu := \delta_{x_0}$, showing $p^{\min} \geq \inf_\mu \int p(x)\mu(dx)$. Conversely, as $p(x) \geq p^{\min}$ for all $x \in K$, $\int_K p(x)\mu(dx) \geq \int_K p^{\min}\mu(dx) = p^{\min}$, since μ is a probability measure. Next note that $\int p(x)\mu(dx) = \sum_\alpha p_\alpha \int x^\alpha \mu(dx) = p^T y$, where $y = (\int x^\alpha \mu(dx))_\alpha$ denotes the sequence of moments of μ. Therefore, $p^{\min}$ can be reformulated as

$$p^{\min} = \inf\ p^T y \ \text{ s.t. } y_0 = 1, \ y \text{ has a representing measure on } K. \tag{4.4}$$

Following Lemma 4.2 one may instead require in (4.4) the weaker conditions $M(y) \succeq 0$ and $M(g_j y) \succeq 0\ \forall j$, which leads to the following lower bound for $p^{\min}$

$$\begin{aligned} p^{\mathrm{mom}} &:= \inf_{y \in \mathbb{R}^{\mathbb{N}^n}} p^T y \ \text{ s.t. } y_0 = 1, M(y) \succeq 0, M(g_j y) \succeq 0\ (1 \leq j \leq m) \\ &= \inf_{L \in (\mathbb{R}[\mathbf{x}])^*} L(p) \ \text{ s.t. } L(1) = 1, L(f) \geq 0\ \forall f \in \mathbf{M}(g_1, \ldots, g_m). \end{aligned} \tag{4.5}$$

Here $\mathbf{M}(g_1, \ldots, g_m)$ is the quadratic module generated by the g_j's, introduced in in (3.13). The equivalence between the two formulations in (4.5) follows directly using the correspondence (4.2) between $\mathbb{R}^{\mathbb{N}^n}$ and linear functionals on $\mathbb{R}[\mathbf{x}]$, and Lemma 4.1 which implies

$$M(y) \succeq 0, M(g_j y) \succeq 0\ \forall j \leq m \Longleftrightarrow L_y(f) \geq 0\ \forall f \in \mathbf{M}(g_1, \ldots, g_m). \tag{4.6}$$

It is not clear how to compute the bound p^{mom} since the program (4.5) involves infinite matrices. To obtain a semidefinite program we consider instead truncated moment matrices in (4.5), which leads to the following hierarchy of lower bounds for $p^{\min}$

$$\begin{aligned} p_t^{\mathrm{mom}} &= \inf_{L \in (\mathbb{R}[\mathbf{x}]_{2t})^*} L(p) \ \text{ s.t. } && L(1) = 1, \\ & && L(f) \geq 0\ \forall f \in \mathbf{M}_{2t}(g_1, \ldots, g_m) \\ &= \inf_{y \in \mathbb{R}^{\mathbb{N}^n_{2t}}} p^T y \ \text{ s.t. } && y_0 = 1,\ M_t(y) \succeq 0, \\ & && M_{t-d_{g_j}}(g_j y) \succeq 0\ (j = 1, \ldots, m) \end{aligned} \tag{4.7}$$

for $t \geq \max(d_p, d_K)$. Here $\mathbf{M}_{2t}(g_1, \ldots, g_m)$ is the truncated quadratic module, introduced in (3.22). The equivalence between the two formulations in (4.7) follows from the truncated analogue of (4.6):

$$M_t(y) \succeq 0, M_{t-d_j}(g_j y) \succeq 0\ \forall j \leq m \Longleftrightarrow L_y(f) \geq 0\ \forall f \in \mathbf{M}_{2t}(g_1, \ldots, g_m).$$

Thus p_t^{mom} can be computed via a semidefinite program involving matrices of size $|\mathbb{N}^n_t|$. Obviously, $p_t^{\mathrm{mom}} \leq p_{t+1}^{\mathrm{mom}} \leq p^{\mathrm{mom}} \leq p^{\min}$. Moreover,

$$p_t^{\mathrm{sos}} \leq p_t^{\mathrm{mom}}; \tag{4.8}$$

indeed if $p - \rho \in \mathbf{M}_{2t}(g_1, \ldots, g_m)$ and L is feasible for (4.7) then $L(p) - \rho = L(p - \rho) \geq 0$. Therefore, $p^{\mathrm{sos}} \leq p^{\mathrm{mom}}$.

Lemma 4.4. *If the set K has a nonempty interior, then the program (4.7) is strictly feasible.*

Proof. Let μ be a measure with $\mathrm{supp}(\mu) = B$ where B is a ball contained in K. (For instance, define μ by $\mu(A) := \lambda(A \cap B)$ for any Borel set A, where $\lambda(\cdot)$ is the Lebesgue measure on $\mathbb{R}^n$.) Let y be the sequence of moments of μ. Then, $M(g_j y) \succ 0$ for all $j = 0, \dots, m$, setting $g_0 := 1$. Positive semidefiniteness is obvious. If $p \in \operatorname{Ker} M(g_j y)$ then $\int_B p(x)^2 g_j(x)\mu(dx) = 0$, which implies $B = \operatorname{supp}(\mu) \subseteq V_{\mathbb{R}}(g_j p)$ and thus $p = 0$. □

In the next section we consider the quadratic case, when all $p, g_1, \dots, g_m$ are quadratic polynomials and we show that in the convex case the moment relaxation of order 1 is exact. Then in later sections we discuss in more detail the duality relationship between sums of squares of polynomials and moment sequences. We will come back to both SOS/moment hierarchies and their application to the optimization problem (1.1) in Section 6.

4.3. Convex quadratic optimization (revisited). We consider here problem (1.1) in the case when the polynomials p, g_j are quadratic, of the form $p = \mathbf{x}^T A\mathbf{x} + 2a^T\mathbf{x} + \alpha$, $g_j = \mathbf{x}^T A_j\mathbf{x} + 2a_j^T\mathbf{x} + \alpha_j$, where A, A_j are given symmetric $n \times n$ matrices, $a, a_j \in \mathbb{R}^n$, and $\alpha, \alpha_j \in \mathbb{R}$.

Define the $(n+1)\times(n+1)$ matrices $P := \begin{pmatrix} 1 & a^T \\ a & A \end{pmatrix}$, $P_j := \begin{pmatrix} 1 & a_j^T \\ a_j & A_j \end{pmatrix}$, so that $p(x) = \left\langle P, \begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix} \right\rangle$ and $g_j(x) = \left\langle P_j, \begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix} \right\rangle$. The moment matrix $M_1(y)$ is of the form $\begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix}$ for some $x \in \mathbb{R}^n$ and some symmetric $n \times n$ matrix X, and the localizing constraint $M_0(g_j y)$ reads $\left\langle P_j, \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \right\rangle$. Therefore, the moment relaxation of order 1 can be reformulated as

$$p_1^{\mathrm{mom}} = \inf_{x,X} \left\langle P, \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \right\rangle \quad \text{s.t.} \quad \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \succeq 0$$
$$\left\langle P_j, \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \right\rangle \geq 0 \ (j = 1, \dots, m) \tag{4.9}$$

Let K_1 denote the set of all $x \in \mathbb{R}^n$ for which there exists a symmetric $n \times n$ matrix X satisfying the constraints in (4.9). Define also the set

$$K_2 := \{x \in \mathbb{R}^n \mid \sum_{j=1}^m t_j g_j(x) \geq 0 \ \text{ for all } t_j \geq 0 \text{ for which } \sum_{j=1}^m t_j A_j \preceq 0\}.$$

Obviously, $K \subseteq K_1$. The next result of Fujie and Kojima [43] shows equality $K_1 = K_2$.

Proposition 4.5. *[43] $K_1 = K_2$. In particular, if $g_1, \dots, g_m$ are concave quadratic polynomials, then $K = K_1 = K_2$.*

Proof. First we check the inclusion $K_1 \subseteq K_2$. Let $x \in K_1$ and X so that the constraints in (4.9) hold. Then, $X - xx^T \succeq 0$. Assume $\sum_j t_j A_j \preceq 0$

with all $t_j \geq 0$. Then $\langle \sum_j t_j A_j, X - xx^T \rangle \leq 0$ and thus

$$\begin{aligned} 0 \leq \sum_j t_j \Big\langle P_j, \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \Big\rangle & = \sum_j t_j (\langle A_j, X \rangle + 2a_j^T x + \alpha_j) \\ & \leq \sum_j t_j (\langle A_j, xx^T \rangle + 2a_j^T x + \alpha_j) \\ & = \sum_j t_j g_j(x). \end{aligned}$$

We now show $K_1 = K_2$. Suppose for contradiction that there exists $x \in K_2 \setminus K_1$. Then $\mathcal{A} \cap \mathcal{B} = \emptyset$, after setting

$$\mathcal{A} := \{X \mid \langle A_j, X \rangle + 2a_j^T x + \alpha_j \geq 0 \ \forall j = 1, \ldots, m\},$$

$$\mathcal{B} := \{X \mid X - xx^T \succeq 0\}, \ \mathcal{B}_0 := \{X \mid X - xx^T \succ 0\}.$$

The set $\mathcal{A}$ is an affine subspace and $\mathcal{B}$ is a convex subset of the space of symmetric $n \times n$ matrices. Therefore there exists a hyperplane $\langle U, X \rangle + \gamma = 0$ such that (i) $\langle U, X \rangle + \gamma \geq 0$ for all $X \in \mathcal{A}$, (ii) $\langle U, X \rangle + \gamma \leq 0$ for all $X \in \mathcal{B}$, and (iii) $\langle U, X \rangle + \gamma < 0$ for all $X \in \mathcal{B}_0$. (Here we use the separation theorem; cf. e.g. Barvinok [7, §II].) Using Farkas' lemma, (i) implies that $U = \sum_j t_j A_j$ for some scalars $t_j \geq 0$ satisfying $\gamma \geq \sum_j (2a_j^T x + \alpha_j)$. Using (ii) we deduce that $U \preceq 0$. As $x \in K_2$, this implies $\sum_j t_j g_j(x) \geq 0$. On the other hand, pick $X \in \mathcal{B}_0$, so that $\langle U, X \rangle + \gamma < 0$ by (iii), and $\langle U, X - xx^T \rangle \leq 0$ as $U \preceq 0$ and $X - xx^T \succeq 0$. Then,

$$\sum_j t_j g_j(x) = \langle U, xx^T - X \rangle + \langle U, X \rangle + \sum_j t_j (2a_j^T x + \alpha_j) \leq \langle U, X \rangle + \gamma < 0,$$

yielding a contradiction. When all g_j are concave, i.e. $A_j \preceq 0$ for all j, then $K_2 = K$ clearly holds. □

COROLLARY 4.6. *Consider problem (1.1) where all polynomials $p, -g_1, \ldots, -g_m$ are quadratic and convex. Then, $p^{min} = p_1^{mom}$.*

Proof. Pick (x, X) feasible for (4.9). Thus $x \in K_2 = K$ by Proposition 4.5. Therefore,

$$p^{\min} \leq p(x) = \Big\langle P, \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \Big\rangle + \langle A, xx^T - X \rangle \leq \Big\langle P, \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \Big\rangle,$$

which implies $p^{\min} \leq p_1^{\text{mom}}$ and thus equality holds. □

In the case $m = 1$ of just one quadratic constraint, the corollary remains valid without the convexity assumption, which follows from the well known S-lemma. (For a detailed treatment see e.g. [10, §4.3.5])

THEOREM 4.7. **(The S-lemma)** *Let p, g_1 be two quadratic polynomials and assume that there exists $x_0 \in \mathbb{R}^n$ with $g_1(x_0) > 0$. The following assertions are equivalent:*

(i) $g_1(x) \geq 0 \Longrightarrow p(x) \geq 0$ (that is, $p_{\min} \geq 0$).

(ii) *There exists* $\lambda \geq 0$ *satisfying* $f(x) \geq \lambda g_1(x)$ *for all* $x \in \mathbb{R}^n$.

(iii) *There exists* $\lambda \geq 0$ *satisfying* $P \succeq \lambda P_1$ *(where* P, P_1 *are the matrices as defined at the beginning of the section).*

Proposition 4.8. *Consider problem (1.1) where* $m = 1$ *and* p, g_1 *are quadratic polynomials and assume that there exists* $x_0 \in \mathbb{R}^n$ *with* $g_1(x_0) > 0$. *Then,* $p_{\min} = p_1^{mom}$.

Proof. Consider the semidefinite program (4.9) defining p_1^{mom} (with $m = 1$) which reads:

$$p_1^{\mathrm{mom}} = \min \langle P, Y\rangle \text{ s.t. } \langle P_1, Y\rangle \geq 0,\ \langle E_{00}, Y\rangle = 1,\ Y \succeq 0,$$

where E_{00} is the matrix with all zero entries except 1 at its north-west corner. The dual semidefinite program reads:

$$\mu^* := \sup \mu \text{ s.t. } P - \lambda P_1 - \mu E_{00} \succeq 0,\ \lambda \geq 0.$$

By weak duality, $p_1^{\mathrm{mom}} \geq \mu^*$. (In fact strong duality holds by the assumption on g_1.) We now verify that $\mu^* \geq p_{\min}$. Indeed, by definition of $p_{\min}$, $g(x) \geq 0$ implies $p(x) \geq p_{\min}$. Applying the S-lemma we deduce that there exists $\lambda \geq 0$ for which $\lambda g(x) \leq p(x) - p_{\min}$ for all x or, equivalently, $P - p_{\min}E_{00} - \lambda P_1 \succeq 0$. The latter shows that $(\lambda, \mu := p_{\min})$ is dual feasible and thus $p_{\min} \leq \mu^*$. Therefore, equality holds throughout: $p_{\min} = p_1^{\mathrm{mom}} = \mu^*$. □

4.4. The moment problem. The moment problem asks for the characterization of the sequences $y \in \mathbb{R}^{\mathbb{N}^n}$ having a representing measure; the analogous problem can be posed for truncated sequences $y \in \mathbb{R}^{\mathbb{N}^n_t}$ ($t \geq 1$ integer). This problem is intimately linked to the characterization of the duals of the cone $\mathcal{P}$ of nonnegative polynomials (from (3.1)) and of the cone Σ of sums of squares (from (3.2)).

4.4.1. Duality between sums of squares and moment sequences. For an $\mathbb{R}$-vector space A, A^* denotes its *dual* vector space consisting of all linear maps $L : A \to \mathbb{R}$. Any $a \in A$ induces an element $\Lambda_a \in (A^*)^*$ by setting $\Lambda_a(L) := L(a)$ for $L \in A^*$; hence there is a natural homomorphism from A to $(A^*)^*$, which is an isomorphism when A is finite dimensional. Given a cone $B \subseteq A$, its dual cone is $B^* := \{L \in A^* \mid L(b) \geq 0\ \forall b \in B\}$. There is again a natural homomorphism from B to $(B^*)^*$, which is an isomorphism when A is finite dimensional and B is a closed convex cone. Here we consider $A = \mathbb{R}[\mathbf{x}]$ and the convex cones $\mathcal{P}, \Sigma \subseteq \mathbb{R}[\mathbf{x}]$, with dual cones

$$\begin{aligned}\mathcal{P}^* &= \{L \in (\mathbb{R}[\mathbf{x}])^* \mid L(p) \geq 0\ \forall p \in \mathcal{P}\},\\ \Sigma^* &= \{L \in (\mathbb{R}[\mathbf{x}])^* \mid L(p^2) \geq 0\ \forall p \in \mathbb{R}[\mathbf{x}]\}.\end{aligned}$$

As mentioned earlier, we may identify a polynomial $p = \sum_\alpha p_\alpha \mathbf{x}^\alpha$ with its sequence of coefficients $\mathrm{vec}(p) = (p_\alpha)_\alpha \in \mathbb{R}^\infty$, the set of sequences in

$\mathbb{R}^{\mathbb{N}^n}$ with finitely many nonzero components; analogously we may identify a linear form $L \in (\mathbb{R}[\mathbf{x}])^*$ with the sequence $y := (L(\mathbf{x}^\alpha))_\alpha \in \mathbb{R}^{\mathbb{N}^n}$, so that $L = L_y$ (recall (4.2)), i.e. $L(p) = \sum_\alpha p_\alpha y_\alpha = y^T \text{vec}(p)$. In other words, we identify $\mathbb{R}[\mathbf{x}]$ with $\mathbb{R}^\infty$ via $p \mapsto \text{vec}(p)$ and $\mathbb{R}^{\mathbb{N}^n}$ with $(\mathbb{R}[\mathbf{x}])^*$ via $y \mapsto L_y$.

We now describe the duals of the cones $\mathcal{P}, \Sigma \subseteq \mathbb{R}[\mathbf{x}]$. For this consider the following cones in $\mathbb{R}^{\mathbb{N}^n}$

$$\mathcal{M} := \{y \in \mathbb{R}^{\mathbb{N}^n} \mid y \text{ has a representing measure}\}, \tag{4.10}$$

$$\mathcal{M}_{\succeq} := \{y \in \mathbb{R}^{\mathbb{N}^n} \mid M(y) \succeq 0\}. \tag{4.11}$$

PROPOSITION 4.9. *The cones $\mathcal{M}$ and $\mathcal{P}$ (resp., $\mathcal{M}_{\succeq}$ and Σ) are duals of each other. That is, $\mathcal{P} = \mathcal{M}^*$, $\mathcal{M}_{\succeq} = \Sigma^*$, $\mathcal{M} = \mathcal{P}^*$, $\Sigma = (\mathcal{M}_{\succeq})^*$.*

Proof. The first two equalities are easy. Indeed, if $p \in \mathcal{P}$ and $y \in \mathcal{M}$ has a representing measure μ, then $y^T\text{vec}(p) = \sum_\alpha p_\alpha y_\alpha = \sum_\alpha p_\alpha \int_K x^\alpha \mu(dx) = \int p(x)\mu(dx) \geq 0$, which shows the inclusions $\mathcal{P} \subseteq \mathcal{M}^*$ and $\mathcal{M} \subseteq \mathcal{P}^*$. The inclusion $\mathcal{M}^* \subseteq \mathcal{P}$ follows from the fact that, if $p \in \mathcal{M}^*$ then, for any $x \in \mathbb{R}^n$, $p(x) = \text{vec}(p)^T\zeta_x \geq 0$ (since $\zeta_x = (x^\alpha)_\alpha \in \mathcal{M}$ as it admits the Dirac measure δ_x as representing measure) and thus $p \in \mathcal{P}$. Given $y \in \mathbb{R}^{\mathbb{N}^n}$, $M(y) \succeq 0$ if and only if $\text{vec}(p)^T M(y)\text{vec}(p) = y^T\text{vec}(p^2) \geq 0$ for all $p \in \mathbb{R}[\mathbf{x}]$ (use Lemma 4.1), i.e. $y^T\text{vec}(f) \geq 0$ for all $f \in \Sigma$; this shows $\mathcal{M}_{\succeq} = \Sigma^*$ and thus the inclusion $\Sigma \subseteq (\mathcal{M}_{\succeq})^*$. The remaining two inclusions $\mathcal{P}^* \subseteq \mathcal{M}$ and $(\mathcal{M}_{\succeq})^* \subseteq \Sigma$ are proved, respectively, by Haviland [56] and by Berg, Christensen and Jensen [12]. (Cf. Section 4.6 below for a proof of Haviland's result.) □

Obviously, $\mathcal{M} \subseteq \mathcal{M}_{\succeq}$ (by Lemma 4.2) and $\Sigma \subseteq \mathcal{P}$. As we saw earlier, the inclusion $\Sigma \subseteq \mathcal{P}$ holds with equality when $n = 1$ and it is strict for $n \geq 2$. Therefore, $\mathcal{M} = \mathcal{M}_{\succeq}$ when $n = 1$; this result is due to Hamburger and is recorded below for further reference.

THEOREM 4.10. **(Hamburger)** *Let $n = 1$. Then $\mathcal{M} = \mathcal{M}_{\succeq}$. That is, for $L \in \mathbb{R}[\mathbf{x}]^*$, $L(p) \geq 0$ for all $p \in \Sigma$ if and only if there exists a measure μ on $\mathbb{R}$ such that $L(p) = \int_{\mathbb{R}} p(x)\mu(dx)$ for all $p \in \mathbb{R}[\mathbf{x}]$.*

Proof. By Lemma 3.5, $\mathcal{P} = \Sigma$ when $n = 1$, which implies $\mathcal{P}^* = \Sigma^*$ and thus $\mathcal{M} = \mathcal{M}_{\succeq}$, using Proposition 4.9. □

The inclusion $\mathcal{M} \subseteq \mathcal{M}_{\succeq}$ is strict when $n \geq 2$. There are however some classes of sequences y for which the reverse implication

$$y \in \mathcal{M}_{\succeq} \Longrightarrow y \in \mathcal{M} \tag{4.12}$$

holds. Curto and Fialkow [28] show that this is the case when the matrix $M(y)$ has finite rank.

THEOREM 4.11. [28] *If $M(y) \succeq 0$ and $M(y)$ has finite rank r, then y has a (unique) r-atomic representing measure.*

We will come back to this result in Section 5.1 (cf. Theorem 5.1) below. This result plays in fact a crucial role in the application to polynomial optimization, since it permits to give an optimality certificate for the semidefinite hierarchy based on moment matrices; see Section 6 for details. We next discuss another class of sequences for which the implication (4.12) holds, namely for bounded sequences.

4.4.2. Bounded moment sequences. Berg, Christensen, and Ressel [13] show that the implication (4.12) holds when the sequence y is bounded, i.e., when there is a constant $C > 0$ for which $|y_\alpha| \leq C$ for all $\alpha \in \mathbb{N}^n$. More generally, Berg and Maserick [14] show that (4.12) holds when y is *exponentially bounded*[1], i.e. when $|y_\alpha| \leq C_0 C^{|\alpha|}$ for all $\alpha \in \mathbb{N}^n$, for some constants $C_0, C > 0$. The next result shows that a sequence $y \in \mathbb{R}^{\mathbb{N}^n}$ has a representing measure supported by a compact set if and only if it is exponentially bounded with $M(y) \succeq 0$.

Theorem 4.12. [14] *Let $y \in \mathbb{R}^{\mathbb{N}^n}$, let $C > 0$ and let $K := [-C, C]^n$. Then y has a representing measure supported by the set K if and only if $M(y) \succeq 0$ and there is a constant $C_0 > 0$ such that $|y_\alpha| \leq C_0 C^{|\alpha|}$ for all $\alpha \in \mathbb{N}^n$.*

The proof uses the following intermediary results.

Lemma 4.13. *Assume $M(y) \succeq 0$ and $|y_\alpha| \leq C_0 C^{|\alpha|}$ for all $\alpha \in \mathbb{N}^n$, for some constants $C_0, C > 0$. Then $|y_\alpha| \leq y_0 C^{|\alpha|}$ for all $\alpha \in \mathbb{N}^n$.*

Proof. If $y_0 = 0$ then $y = 0$ since $M(y) \succeq 0$ and the lemma holds. Assume $y_0 > 0$. Rescaling y we may assume $y_0 = 1$; we show $|y_\alpha| \leq C^{|\alpha|}$ for all α. As $M(y) \succeq 0$, we have $y_\alpha^2 \leq y_{2\alpha}$ for all α. Then, $|y_\alpha| \leq (y_{2^k\alpha})^{1/2^k}$ for any integer $k \geq 1$ (easy induction) and thus $|y_\alpha| \leq (C_0 C^{2^k|\alpha|})^{1/2^k} = C_0^{1/2^k} C^{|\alpha|}$. Letting k go to ∞, we find $|y_\alpha| \leq C^{|\alpha|}$. □

Lemma 4.14. *Given $C > 0$ and $K = [-C, C]^n$, the set*

$$S := \{y \in \mathbb{R}^{\mathbb{N}^n} \mid y_0 = 1,\ M(y) \succeq 0,\ |y_\alpha| \leq C^{|\alpha|}\ \forall \alpha \in \mathbb{N}^n\}$$

is a convex set whose extreme points are the Zeta vectors $\zeta_x = (x^\alpha)_{\alpha \in \mathbb{N}^n}$ for $x \in K$.

Proof. S is obviously convex. Let y be an extreme point of S. Fix $\alpha_0 \in \mathbb{N}^n$. Our first goal is to show

$$y_{\alpha+\alpha_0} = y_\alpha y_{\alpha_0}\ \forall \alpha \in \mathbb{N}^n. \tag{4.13}$$

For this, define the sequence $y^{(\epsilon)} \in \mathbb{R}^{\mathbb{N}^n}$ by $y_\alpha^{(\epsilon)} := C^{|\alpha_0|} y_\alpha + \epsilon y_{\alpha+\alpha_0}$ for $\alpha \in \mathbb{N}^n$, for $\epsilon \in \{\pm 1\}$. Therefore, $|y_\alpha^{(\epsilon)}| \leq C^{|\alpha_0|}(1+\epsilon) C^{|\alpha|}\ \ \forall \alpha$. We now show that $M(y^{(\epsilon)}) \succeq 0$. Fix $p \in \mathbb{R}[\mathbf{x}]$; we have to show that

$$p^T M(y^{(\epsilon)}) p = \sum_{\gamma,\gamma'} p_\gamma p_{\gamma'} y_{\gamma+\gamma'}^{(\epsilon)} \geq 0. \tag{4.14}$$

[1] Our definition is equivalent to that of Berg and Maserick [14] who say that y is exponentially bounded when $|y_\alpha| \leq C_0 \sigma(\alpha)\ \forall \alpha$, for some $C_0 > 0$ and some function, called an absolute value, $\sigma : \mathbb{N}^n \to \mathbb{R}_+$ satisfying $\sigma(0) = 1$ and $\sigma(\alpha+\beta) \leq \sigma(\alpha)\sigma(\beta)$ $\forall \alpha, \beta \in \mathbb{N}^n$. Indeed, setting $C := \max_{i=1,\ldots,n} \sigma(e_i)$ we have $\sigma(\alpha) \leq C^{|\alpha|}$ and, conversely, the function $\alpha \mapsto \sigma(\alpha) := C^{|\alpha|}$ is an absolute value.

For this, define the new sequence $z := M(y)\mathrm{vec}(p^2) \in \mathbb{R}^{\mathbb{N}^n}$ with $z_\alpha = \sum_{\gamma,\gamma'} p_\gamma p_{\gamma'} y_{\alpha+\gamma+\gamma'}$ for $\alpha \in \mathbb{N}^n$. Then, $|z_\alpha| \leq (\sum_{\gamma,\gamma'} |p_\gamma p_{\gamma'}|C^{|\gamma|+|\gamma'|})C^{|\alpha|}$ $\forall\alpha$. Moreover, $M(z) \succeq 0$. Indeed, using the fact that $z = M(y)\mathrm{vec}(p^2) = gy$ for $g := p^2$ (recall (4.1)) combined with Lemma 4.1, we find that $q^T M(z)q = q^T M(gy)q = \mathrm{vec}(pq)^T M(y)\mathrm{vec}(pq) \geq 0$ for all $q \in \mathbb{R}[\mathbf{x}]$. Hence Lemma 4.13 implies $-z_0 C^{|\alpha|} \leq z_\alpha \leq z_0 C^{|\alpha|}$ $\forall\alpha$; applying this to $\alpha = \alpha_0$, we get immediately relation (4.14). Therefore, $M(y^{(\epsilon)}) \succeq 0$. Applying again Lemma 4.13, we deduce that $|y^{(\epsilon)}_\alpha| \leq y^{(\epsilon)}_0 C^{|\alpha|}$ $\forall\alpha$.

If $y^{(\epsilon)}_0 = 0$ for some $\epsilon \in \{\pm 1\}$, then $y^{(\epsilon)} = 0$, which implies directly (4.13). Assume now $y^{(\epsilon)}_0 > 0$ for both $\epsilon = 1, -1$. Then each $\frac{y^{(\epsilon)}}{y^{(\epsilon)}_0}$ belongs to S and $y = \frac{y^{(1)}_0}{2C^{|\alpha_0|}}\frac{y^{(1)}}{y^{(1)}_0} + \frac{y^{(-1)}_0}{2C^{|\alpha_0|}}\frac{y^{(-1)}}{y^{(-1)}_0}$ is a convex combination of two points of S. As y is an extreme point of S, $y \in \left\{\frac{y^{(1)}}{y^{(1)}_0}, \frac{y^{(-1)}}{y^{(-1)}_0}\right\}$, which implies again (4.13).

As relation (4.13) holds for all $\alpha_0 \in \mathbb{N}^n$, setting $x := (y_{e_i})_{i=1}^n$, we find that $x \in K$ and $y_\alpha = x^\alpha$ for all α, i.e. $y = \zeta_x$. □

Proof of Theorem 4.12. Assume that $M(y) \succeq 0$ and $|y_\alpha| \leq C_0 C^{|\alpha|}$ for all α; we show that y has a representing measure supported by K. By Lemma 4.13, $|y_\alpha| \leq y_0 C^{|\alpha|}$ $\forall\alpha$. If $y_0 = 0$, $y = 0$ and we are done. Assume $y_0 = 1$ (else rescale y). Then y belongs to the convex set S introduced in Lemma 4.14, whose set of extreme points is $S_0 := \{\zeta_x \mid x \in [-C, C]^n\}$. We can now apply the Krein-Milman theorem[2] to the convex compact set S and conclude that y belongs to the closure of the convex hull of S_0. That is, $y = \lim_{i\to\infty} y^{(i)}$ (coordinate-wise convergence), where each $y^{(i)}$ belongs to the convex hull of S_0 and thus has a (finite atomic) representing measure $\mu^{(i)}$ on $[-C, C]^n$. Using Haviland's theorem (Theorem 4.15), we can conclude that y too has a representing measure on $[-C, C]^n$.

Conversely, assume that y has a representing measure μ supported by K. Then, as $|x_i| \leq C$ for all i, $|y_\alpha| \leq \int_K |x^\alpha|\mu(dx) \leq \mu(K)C^{|\alpha|}$, which concludes the proof of Theorem 4.12. □

4.5. The K-moment problem.

We now consider the K-moment problem where, as in (1.2), $K = \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$ is a

[2] Let $\mathcal{X}$ be a locally convex topological vector space (i.e., whose topology is defined by a family of seminorms) which is Hausdorff. The Krein-Milman theorem claims that any convex compact subset S of $\mathcal{X}$ is equal to the closure of the convex hull of its set of extreme points.

semialgebraic set. Define the cones

$$\mathcal{M}_K := \{y \in \mathbb{R}^{\mathbb{N}^n} \mid y \text{ has a representing measure supported by } K\} \tag{4.15}$$

$$\mathcal{M}^{sch}_{\succeq}(g_1, \ldots, g_m) := \{y \in \mathbb{R}^{\mathbb{N}^n} \mid M(g_J y) \succeq 0 \ \forall J \subseteq \{1, \ldots, m\}\}, \tag{4.16}$$

$$\mathcal{M}^{put}_{\succeq}(g_1, \ldots, g_m) := \{y \in \mathbb{R}^{\mathbb{N}^n} \mid M(y) \succeq 0, \ M(g_j y) \succeq 0 \ (1 \leq j \leq m)\}, \tag{4.17}$$

setting $g_\emptyset := 1$, $g_J := \prod_{j \in J} g_j$ for $J \subseteq \{1, \ldots, m\}$. (The indices 'sch' and 'put' refer respectively to Schmüdgen and to Putinar; see Theorems 4.16 and 4.17 below.) Consider also the cone

$$\mathcal{P}_K = \{p \in \mathbb{R}[\mathbf{x}] \mid p(x) \geq 0 \ \forall x \in K\}$$

and recall the definition of $T(g_1, \ldots, g_m)$ from (3.12) and $\mathbf{M}(g_1, \ldots, g_m)$ from (3.13). Obviously,

$$\begin{aligned} \mathcal{M}_K \subseteq \mathcal{M}^{sch}_{\succeq}(g_1, \ldots, g_m) \subseteq \mathcal{M}^{put}_{\succeq}(g_1, \ldots, g_m), \\ \mathbf{M}(g_1, \ldots, g_m) \subseteq T(g_1, \ldots, g_m) \subseteq \mathcal{P}_K. \end{aligned}$$

One can verify that

$$\begin{aligned} \mathcal{P}_K = (\mathcal{M}_K)^*, \ \mathcal{M}^{sch}_{\succeq}(g_1, \ldots, g_m) = (T(g_1, \ldots, g_m))^*, \\ \mathcal{M}^{put}_{\succeq}(g_1, \ldots, g_m) = (\mathbf{M}(g_1, \ldots, g_m))^*. \end{aligned} \tag{4.18}$$

(The details are analogous to those for Proposition 4.9, using Lemma 4.1.)

The next result of Haviland [56] shows equality $\mathcal{M}_K = (\mathcal{P}_K)^*$; see Section 4.6 below for a proof.

Theorem 4.15. *[56]* **(Haviland's theorem)** *Let K be a closed subset in $\mathbb{R}^n$. The following assertions are equivalent for $L \in \mathbb{R}[\mathbf{x}]^*$.*

(i) $L(p) \geq 0$ for any polynomial $p \in \mathbb{R}[\mathbf{x}]$ such that $p \geq 0$ on K.

(ii) There exists a measure μ on K such that $L(p) = \int_K p(x)\mu(dx)$ for all $p \in \mathbb{R}[\mathbf{x}]$.

The following results of Schmüdgen [145] and Putinar [134] can be seen as the counterparts (on the 'moment side') of Theorems 3.16 and 3.20 (on the 'sos side'). We refer e.g. to [106, 133] for a detailed treatment and background. As we observe in Section 4.7 below, the results on the 'sos side' easily imply the results on the 'moment side', i.e., Theorems 3.16 and 3.20 imply Theorems 4.16 and 4.17, respectively. In fact the reverse implications also hold; [145, Cor. 3] shows how to derive Theorem 3.16 from Theorem 4.16[3]. We also present a direct proof of Theorem 4.16 in Section 4.7.

[3] Cf. the discussion of Marshall [106, §6.1] regarding the correctness of this proof. We will sketch an alternative argument of Marshall in Proposition 4.23 below.

THEOREM 4.16. *[145]* **(Schmüdgen's theorem)** *If K is compact, then $\mathcal{M}_K = \mathcal{M}_{\succeq}^{sch}(g_1, \ldots, g_m)$.*

THEOREM 4.17. *[134]* **(Putinar's theorem)** *Assume $\mathbf{M}(g_1, \ldots, g_m)$ is Archimedean, i.e. (3.16) holds. Then $\mathcal{M}_K = \mathcal{M}_{\succeq}^{put}(g_1, \ldots, g_m)$.*

In the univariate case we have the following classical results for the moment problem on an interval.

THEOREM 4.18. **(Stieltjes)** *The following assertions are equivalent for $L \in \mathbb{R}[\mathbf{x}]^*$.*

(i) There exists a measure μ on $\mathbb{R}_+$ such that $L(p) = \int p(x)\mu(dx)$ for all $p \in \mathbb{R}[\mathbf{x}]$.

(ii) $L(p), L(\mathbf{x}p) \geq 0$ for all $p \in \Sigma$, i.e., $M(y), M(\mathbf{x}y) \succeq 0$, where $y = (L(\mathbf{x}^i))_{i\geq 0}$ and $\mathbf{x}y = (L(\mathbf{x}^{i+1}))_{i\geq 0}$.

Proof. Directly from Theorem 3.21 combined with Theorem 4.15. □

THEOREM 4.19. **(Hausdorff)** *(cf. [73]) The following assertions are equivalent for $L \in \mathbb{R}[\mathbf{x}]^*$.*

(i) There exists a measure μ on $[a, b]$ such that $L(p) = \int p(x)\mu(dx)$ for all $p \in \mathbb{R}[\mathbf{x}]$.

(ii) $L((\mathbf{x} - a)p), L((b - \mathbf{x})p) \geq 0$ for all $p \in \Sigma$, i.e., $M((\mathbf{x} - a)y) \succeq 0$ and $M((b - \mathbf{x})y) \succeq 0$.

(iii) $L(p), L((\mathbf{x} - a)(b - \mathbf{x})p) \geq 0$ for all $p \in \Sigma$, i.e., $M(y) \succeq 0$ and $M((\mathbf{x} - a)(b - \mathbf{x})y) \succeq 0$.

Here, $y = (L(\mathbf{x}^i))_{i\geq 0}$, and recall that $(\mathbf{x} - a)y = (L(\mathbf{x}^{i+1}) - aL(\mathbf{x}^i))_{i\geq 0}$, $(b - \mathbf{x})y = (L(b\mathbf{x}^i) - L(\mathbf{x}^{i+1}))_{i\geq 0}$, and $(\mathbf{x} - a)(b - \mathbf{x})y = (-L(\mathbf{x}^{i+2}) + (a + b)L(\mathbf{x}^{i+1}) - abL(\mathbf{x}^i))_{i\geq 0}$.

Proof. Directly from Theorem 3.23 combined with Theorem 4.15. □

4.6. Proof of Haviland's theorem. We give here a proof of Haviland's theorem (Theorem 4.15), following the treatment in [106]. The proof uses the following version of the Riesz representation theorem.

THEOREM 4.20. **(Riesz representation theorem)** *Let $\mathcal{X}$ be a Hausdorff locally compact space and let $\mathcal{C}_c(\mathcal{X}, \mathbb{R})$ be the space of continuous functions $f : \mathcal{X} \to \mathbb{R}$ with compact support, i.e. for which the closure of the set $\{x \in \mathcal{X} \mid f(x) \neq 0\}$ is compact. Let $L \in (\mathcal{C}_c(\mathcal{X}, \mathbb{R}))^*$ be a positive linear function, i.e. $L(f) \geq 0$ for all $f \in \mathcal{C}_c(\mathcal{X}, \mathbb{R})$ such that $f \geq 0$ on $\mathcal{X}$. Then there exists a Borel measure μ on $\mathcal{X}$ such that $L(f) = \int_K f(x)\mu(dx)$ for all $f \in \mathcal{C}_c(\mathcal{X}, \mathbb{R})$.*

Here we consider the topological space $\mathcal{X} = K$ (obviously locally compact). Let $\mathcal{C}'(K, \mathbb{R})$ denote the set of continuous functions $f : K \to \mathbb{R}$ for which there exists $a \in \mathbb{R}[\mathbf{x}]$ such that $|f| \leq |a|$ on K. Obviously, $\mathbb{R}[\mathbf{x}]_{|K} \subseteq \mathcal{C}'(K, \mathbb{R})$ and $\mathcal{C}_c(K, \mathbb{R}) \subseteq \mathcal{C}'(K, \mathbb{R})$; here $\mathbb{R}[\mathbf{x}]_{|K}$ denotes the set of polynomial functions restricted to K. Assume $L \in \mathbb{R}[\mathbf{x}]^*$ satisfies (i), i.e. $L(p) \geq 0$ for all $p \in \mathcal{P}_K$. First we extend L to a positive linear function L' on $\mathcal{C}'(K, \mathbb{R})$.

Lemma 4.21. *There exists a linear map $L' : \mathcal{C}'(K, \mathbb{R}) \to \mathbb{R}$ extending L and which is positive, i.e. $L'(f) \geq 0$ for all $f \in \mathcal{C}'(K, \mathbb{R})$ such that $f \geq 0$ on K.*

Proof. Using Zorn's lemma[4], one can prove the existence of a pair (V, L'), where V is a vector space such that $\mathbb{R}[\mathbf{x}]_{|K} \subseteq V \subseteq \mathcal{C}'(K, \mathbb{R})$ and $L' \in V^*$ is positive (i.e. $L'(f) \geq 0$ for all $f \in V$ such that $f \geq 0$ on K), which is maximal with respect to the following partial ordering:

$$(V_1, L_1) \leq (V_2, L_2) \text{ if } V_1 \subseteq V_2 \text{ and } (L_2)_{|V_1} = L_1.$$

It suffices now to verify that $V = \mathcal{C}'(K, \mathbb{R})$. Suppose for contradiction that $g \in \mathcal{C}'(K, \mathbb{R}) \setminus V$. Let $p \in \mathbb{R}[\mathbf{x}]$ such that $|g| \leq |p|$ on K. As $|p| \leq q := \frac{1}{2}(p^2 + 1)$, we have $-q \leq g \leq q$ on K, where $-q, q \in V$. The quantities

$$\rho_1 := \sup L'(f_1) \text{ s.t. } f_1 \in V,\ f_1 \leq g,\ \rho_2 := \inf L'(f_2) \text{ s.t. } f_2 \in V,\ f_2 \geq g$$

are well defined (as $L(f_1) \leq L(q)$ and $L(f_2) \geq L(-q)$ for any feasible f_1, f_2) and $\rho_1 \leq \rho_2$. Fix $c \in \mathbb{R}$ such that $\rho_1 \leq c \leq \rho_2$ and define the linear extension L'' of L' to $V'' := V + \mathbb{R}g$ by setting $L''(g) := c$. Then, L'' is positive. Indeed, assume $f + \lambda g \geq 0$ on K, where $f \in V$ and $\lambda \in \mathbb{R}$. If $\lambda = 0$ then $L''(f + \lambda g) = L'(f) \geq 0$. If $\lambda > 0$, then $-\frac{1}{\lambda} f \leq g$, implying $L'(-\frac{1}{\lambda} f) \leq \rho_1 \leq c = L''(g)$ and thus $L''(f + \lambda g) \geq 0$. If $\lambda < 0$, then $-\frac{1}{\lambda} f \geq g$, implying $L'(-\frac{1}{\lambda} f) \geq \rho_2 \geq c = L''(g)$ and thus $L''(f + \lambda g) \geq 0$. As $(V, L') < (V'', L'')$, this contradicts the maximality assumption on (V, L). □

In particular, L' is a positive linear map on $\mathcal{C}_c(K, \mathbb{R})$ and thus, applying Theorem 4.20, we can find a Borel measure μ on K such that

$$L'(f) = \int_K f(x)\mu(dx) \quad \forall f \in \mathcal{C}_c(K, \mathbb{R}).$$

To conclude the proof, it remains to show that this also holds for any $f \in \mathcal{C}'(K, \mathbb{R})$. Fix $f \in \mathcal{C}'(K, \mathbb{R})$. Without loss of generality we can suppose that $f \geq 0$ on K (else write $f = f_+ - f_-$ where $f_+ := \max\{f, 0\}$ and $f_- := -\min\{f, 0\}$). We now show how to approach f via a sequence of functions $f_i \in \mathcal{C}_c(K, \mathbb{R})$. For this, set $g := f + \sum_{j=1}^n x_j^2$, and

$$K_i := \{x \in K \mid g(x) \leq i\}, \quad H_i := \Big\{x \in K_{i+1} \mid g(x) \geq i + \frac{1}{2}\Big\}.$$

Then, K_i is compact, $K_i \subseteq K_{i+1}$, and $K = \cup_{i \geq 0} K_i$. Using Urysohn's lemma[5], we can construct continuous functions $h_i : K_{i+1} \to [0, 1]$ such

[4] Zorn's lemma states that every partially ordered set in which every chain has an upper bound contains at least one maximal element.

[5] Urysohn's lemma states that, if A and B are disjoint closed subsets of a compact Hausdorff (more generally, normal) space C, then there is a continuous function h on C taking value 1 on A and 0 on B.

that $h_i = 1$ on K_i and $h_i = 0$ on H_i. Extend h_i to K by setting $h_i = 0$ on $K \setminus K_{i+1}$. Define $f_i := fh_i$. Then, $0 \leq f_i \leq f$, $f_i = f$ on K_i, and $f_i = 0$ outside of K_{i+1}. In particular, $f_i \in \mathcal{C}_c(K, \mathbb{R})$, so that $L'(f_i) = \int_K f_i(x)\mu(dx)$. Now, $0 \leq f - f_i \leq \frac{1}{i}g^2$ on K; indeed, this is obvious on K_i and, on $K \setminus K_i$, $g > i$ implies $g^2 \geq ig \geq if \geq i(f - f_i)$. Thus, $0 \leq \int_K (f(x) - f_i(x))\mu(dx) \leq \frac{1}{i}\int_K g^2(x)\mu(dx)$, which shows that $L'(f) = lim_{i\to\infty}L'(f_i) = \lim_{i\to\infty}\int_K f_i(x)\mu(dx) = \int_K f(x)\mu(dx)$. This concludes the proof of Theorem 4.15

4.7. Proof of Schmüdgen's theorem. We give here a proof of Theorem 4.16, following the treatment in [106, §6.1]. This proof is due to Wörmann [177], who proved the following characterization for K compact. To simplify the notation we set $T := T(g_1, \ldots, g_m)$ and $T^{**} := (T^*)^*$ stands for the double dual of T, thus consisting of the polynomials p satisfying $L(p) \geq 0$ for all $L \in T^*$.

THEOREM 4.22. *[177] The set K is compact if and only if T is Archimedean, i.e., for all $p \in \mathbb{R}[\mathbf{x}]$ there exists $N \in \mathbb{N}$ for which $N \pm p \in T$.*

The following result relates the SOS/moment assertions in Schmüdgen's theorem.

PROPOSITION 4.23. *Consider the following assertions:*
(i) For $f \in \mathbb{R}[\mathbf{x}]$, $f > 0$ on $K \Longrightarrow f \in T$.
(ii) $\mathcal{P}_K = T^{**}$.
(iii) $\mathcal{M}_K = \mathcal{M}_{\succeq}^{sch}$.
Then, (i) $\Longrightarrow$ (ii) $\Longleftrightarrow$ (iii). Moreover, (i) $\Longleftrightarrow$ (ii) $\Longleftrightarrow$ (iii) when K is compact.

Proof. (i) $\Longrightarrow$ (ii): Let $f \in \mathcal{P}_K$ and $L \in T^*$; we show that $L(f) \geq 0$. As $f \in \mathcal{P}_K$, for any $\epsilon > 0$, $f + \epsilon > 0$ on K, and thus $f + \epsilon \in T$. Hence, $L(f + \epsilon) \geq 0$, implying $L(f) \geq -\epsilon L(1)$ for all $\epsilon > 0$ and thus $L(f) \geq 0$. This shows $f \in T^{**}$.
(ii) $\Longrightarrow$ (iii): $\mathcal{P}_K = T^{**}$ implies $\mathcal{P}_K^* = T^{***} = T^*$ and thus (iii), since $\mathcal{P}_K^* = \mathcal{M}_K$ (by Haviland's theorem) and $\mathcal{M}_{\succeq}^{sch} = T^*$ (by (4.18)).
(iii) $\Longrightarrow$ (ii): $\mathcal{M}_K = \mathcal{M}_{\succeq}^{sch}$ implies $\mathcal{P}_K = (\mathcal{M}_K)^* = (\mathcal{M}_{\succeq}^{sch})^* = T^{**}$.
Assuming K compact, (iii) $\Longrightarrow$ (i): First, note that 1 lies in the interior of T (because, for any $p \in \mathbb{R}[\mathbf{x}]$, there exists $N > 0$ for which $1 + \frac{1}{N}p \in T$, since T is Archimedean by Theorem 4.22). Let U denote the interior of T; thus U is a non-empty open set in $\mathbb{R}[\mathbf{x}]$. Assume $f \in \mathbb{R}[\mathbf{x}]$ is positive on K and $f \notin T$. Consider the cone $C := \mathbb{R}_+ f$ in $\mathbb{R}[\mathbf{x}]$. Thus $U \cap C = \emptyset$. We can apply the separation theorem [106, Thm 3.6.2] and deduce the existence of $L \in \mathbb{R}[\mathbf{x}]^*$ satisfying $L(f) \leq 0$, $L > 0$ on U, and thus $L \geq 0$ on T. By (iii), L has a representing measure μ on K. As K is compact, there exists $c > 0$ such that $f \geq c$ on K. Thus $L(f) = \int_K f(x)\mu(dx) \geq cL(1) > 0$, and we reach a contradiction. □

As Corollary 4.26 below shows (combined with Proposition 4.23), the following result implies directly Theorem 4.16.

Theorem 4.24. *Given $f \in \mathbb{R}[\mathbf{x}]$ and $k \in \mathbb{R}_+$, if $-k \leq f \leq k$ on K, then $k^2 - f^2 \in T^{**}$.*

Corollary 4.25. *For $f \in \mathbb{R}[\mathbf{x}]$ and $b \in \mathbb{R}$, if $0 \leq f \leq b$ on K, then $f \in T^{**}$.*

Proof. Setting $g := f - \frac{b}{2}$ and $k := \frac{b}{2}$, we have $-k \leq g \leq k$ on K. Thus $k^2 - g^2 \in T^{**}$ by Theorem 4.24. If $k = 0$, for any $\epsilon > 0$, $\epsilon^2 + 2\epsilon g = (g + \epsilon)^2 + (-g^2) \in T^{**}$; for all $L \in T^*$, this implies $\epsilon L(1) + 2L(g) \geq 0$ and thus $L(g) \geq 0$ by letting $\epsilon \to 0$, thus showing $g \in T^{**}$. Suppose now $k > 0$. Then, for $\epsilon = \pm 1$, $k + \epsilon g = \frac{1}{2k}((k + \epsilon g)^2 + (k^2 - g^2)) \in T^{**}$. In both cases we find that $f = g + k \in T^{**}$. □

Corollary 4.26. *If K is compact, then $\mathcal{P}_K = T^{**}$.*

Proof. The inclusion $\mathcal{P}_K \subseteq T^{**}$ follows directly from Corollary 4.25, since any polynomial is upper bounded on the compact set K. As the reverse inclusion is obvious, we have equality $\mathcal{P}_K = T^{**}$. □

We now proceed to prove Theorem 4.24, which uses the Positivstellensatz in a crucial way. Fix $l > k$ so that $l^2 - f^2 > 0$ on K. Thus, by Theorem 3.12 (i), $(l^2 - f^2)p = 1 + q$, for some $p, q \in T$. We first claim:

$$(l^{2i} - f^{2i})p \in T \quad \forall i \geq 1. \tag{4.19}$$

If $i = 1$, $(l^2 - f^2)p = 1 + q \in T$. If $i \geq 2$, then $(l^{2i+2} - f^{2i+2})p = l^2(l^{2i} - f^{2i})p - f^{2i}(l^2 - f^2)p \in T$, since $(l^{2i} - f^{2i})p \in T$ using induction. Next we claim:

$$l^{2i+2}p - f^{2i} \in T \quad \forall i \geq 1. \tag{4.20}$$

Indeed, $l^{2i+2}p - f^{2i} = l^2(l^{2i} - f^{2i})p + f^{2i}(l^2p - 1) \in T$, since $(l^{2i} - f^{2i})p \in T$ by (4.19) and $l^2p - 1 = f^2p + q \in T$.

Fix $L \in T^*$ and consider the linear map L_1 on the univariate polynomial ring $\mathbb{R}[\mathbf{t}]$ defined by $L_1(r(\mathbf{t})) := L(r(f))$, where we substitute variable $\mathbf{t}$ by f in the polynomial $r \in \mathbb{R}[\mathbf{t}]$. Obviously, L_1 is positive on squares and thus, by Hamburger' theorem (Theorem 4.10), there exists a Borel measure ν on $\mathbb{R}$ such that $L_1(r) = \int_{\mathbb{R}} r(t)\nu(dt)$ for $r \in \mathbb{R}[\mathbf{t}]$. Fix $\lambda > l$ and let χ_λ denote the charactersitic function of the subset $(-\infty, -\lambda) \cup (\lambda, \infty)$ of $\mathbb{R}$. As $\lambda^{2i}\chi_\lambda(t) \leq t^{2i}$ for all $t \in \mathbb{R}$, we obtain:

$$\lambda^{2i} \int_{\mathbb{R}} \chi_\lambda(t)\nu(dt) \leq \int_{\mathbb{R}} t^{2i}\nu(dt) = L_1(\mathbf{t}^{2i}) = L(f^{2i}) \leq l^{2i+2}L(p),$$

using (4.20) for the latter inequality. As $l < \lambda$, letting $i \to \infty$, we obtain $\int_{\mathbb{R}} \chi_\lambda(t)\nu(dt) = 0$. In turn, letting $\lambda \to l$, this implies $\int_{\mathbb{R}} \chi_l(t)\nu(dt) = 0$ and thus the support of ν is contained in $[-l, l]$. As $t^2 \leq l^2$ on $[-l, l]$, we obtain:

$$L(f^2) = L_1(\mathbf{t}^2) = \int_{\mathbb{R}} t^2\nu(dt) = \int_{[-l,l]} t^2\nu(dt) \leq \int_{[-l,l]} l^2\nu(dt) = L(l^2).$$

This proves $L(l^2-f^2) \geq 0$ for all $l > k$ and thus, letting $l \to k$, $L(k^2-f^2) \geq 0$. That is, $k^2 - f^2 \in T^{**}$, which concludes the proof of Theorem 4.24.

Finally, observe that Theorem 4.17 follows easily from Theorem 3.20. Indeed, using similar arguments as for Proposition 4.23, one can verify that Theorem 4.17 implies $\mathcal{P}_K = \mathbf{M}(g_1, \ldots, g_m)^{**}$, which in turn implies $\mathcal{M}_K = \mathcal{M}_{\succeq}^{put}$.

5. More about moment matrices.

5. More about moment matrices. We group here several results about moment matrices, mostly from Curto and Fialkow [28, 29, 30], which will have important applications to the optimization problem (1.1).

5.1. Finite rank moment matrices. We have seen in Lemma 4.2 (iii) that, if a sequence $y \in \mathbb{R}^{\mathbb{N}^n}$ has a r-atomic representing measure, then its moment matrix $M(y)$ is positive semidefinite and its rank is equal to r. Curto and Fialkow [28] show that the reverse implication holds. More precisely, they show the following result, thus implying Theorem 4.11.

THEOREM 5.1. [28] *Let* $y \in \mathbb{R}^{\mathbb{N}^n}$.

(i) *If* $M(y) \succeq 0$ *and* $M(y)$ *has finite rank* r, *then* y *has a unique representing measure* μ. *Moreover,* μ *is* r*-atomic and* $\mathrm{supp}(\mu) = V_{\mathbb{C}}(\mathrm{Ker}\, M(y))$ $(\subseteq \mathbb{R}^n)$.

(ii) *If* y *has a* r*-atomic representing measure, then* $M(y) \succeq 0$ *and* $M(y)$ *has rank* r.

Assertion (ii) is just Lemma 4.2 (iii). We now give a simple proof for Theorem 5.1 (i) (taken from [95]), which uses an algebraic tool (the Real Nullstellensatz) in place of the tools from functional analysis (the spectral theorem and the Riesz representation theorem) used in the original proof of Curto and Fialkow [28].

Recall that one says that 'a polynomial f lies in the kernel of $M(y)$' when $M(y)f := M(y)\mathrm{vec}(f) = 0$, which permits to identify the kernel of $M(y)$ with a subset of $\mathbb{R}[\mathbf{x}]$. Making this identification enables us to claim that 'the kernel of $M(y)$ is an ideal in $\mathbb{R}[\mathbf{x}]$' (as observed by Curto and Fialkow [28]) or, when $M(y) \succeq 0$, that 'the kernel is a radical ideal' (as observed by Laurent [95]) or even 'a real radical ideal' (as observed by Möller [107], or Scheiderer [143]). Moreover, linearly independent sets of columns of $M(y)$ correspond to linearly independent sets in the quotient vector space $\mathbb{R}[\mathbf{x}]/\,\mathrm{Ker}\, M(y)$. These properties, which play a crucial role in the proof, are reported in the next two lemmas.

LEMMA 5.2. *The kernel* $\mathcal{I} := \{p \in \mathbb{R}[\mathbf{x}] \mid M(y)p = 0\}$ *of a moment matrix* $M(y)$ *is an ideal in* $\mathbb{R}[\mathbf{x}]$. *Moreover, if* $M(y) \succeq 0$, *then* $\mathcal{I}$ *is a real radical ideal.*

Proof. We apply Lemma 4.1. Assume $f \in \mathcal{I}$ and let $g \in \mathbb{R}[\mathbf{x}]$. For any $h \in \mathbb{R}[\mathbf{x}]$, $\mathrm{vec}(h)^T M(y)\mathrm{vec}(fg) = \mathrm{vec}(hg)^T M(y)\mathrm{vec}(f) = 0$, implying that $fg \in \mathcal{I}$. Assume now $M(y) \succeq 0$. We show that $\mathcal{I}$ is real radical. In view of

Lemma 2.2, it suffices to show that, for any $g_1, \dots, g_m \in \mathbb{R}[\mathbf{x}]$,

$$\sum_{j=1}^{m} g_j^2 \in \mathcal{I} \Longrightarrow g_1, \dots, g_m \in \mathcal{I}.$$

Indeed, if $\sum_{j=1}^m g_j^2 \in \mathcal{I}$ then $0 = \text{vec}(1)^T M(y) \text{vec}(\sum_{j=1}^m g_j^2) = \sum_{j=1}^m g_j^T M(y) g_j$. As $g_j^T M(y) g_j \geq 0$ since $M(y) \succeq 0$, this implies $0 = g_j^T M(y) g_j$ and thus $g_j \in \mathcal{I}$ for all j. □

Lemma 5.3. *Let $\mathcal{B} \subseteq \mathbb{T}_n$. Then, $\mathcal{B}$ indexes a (maximum) linearly independent set of columns of $M(y)$ if and only if $\mathcal{B}$ is a (maximum) linearly independent subset of the quotient vector space $\mathbb{R}[\mathbf{x}]/\operatorname{Ker} M(y)$.*

Proof. Immediate verification. □

Proof of Theorem 5.1(i). Assume $M(y) \succeq 0$ and $r := \text{rank } M(y) < \infty$. By Lemmas 5.2 and 5.3, the set $\mathcal{I} := \operatorname{Ker} M(y)$ is a real radical zero-dimensional ideal in $\mathbb{R}[\mathbf{x}]$. Hence, using (2.1) and Theorem 2.6, $V_\mathbb{C}(\mathcal{I}) \subseteq \mathbb{R}^n$ and $|V_\mathbb{C}(\mathcal{I})| = \dim \mathbb{R}[\mathbf{x}]/\mathcal{I} = r$. Let $p_v \in \mathbb{R}[\mathbf{x}]$ $(v \in V_\mathbb{C}(\mathcal{I}))$ be interpolation polynomials at the points of $V_\mathbb{C}(\mathcal{I})$. Setting $\lambda_v := p_v^T M(y) p_v$, we now claim that the measure $\mu := \sum_{v \in V_\mathbb{C}(\mathcal{I})} \lambda_v \delta_v$ is the unique representing measure for y.

Lemma 5.4. *$M(y) = \sum_{v \in V_\mathbb{C}(\mathcal{I})} \lambda_v \zeta_v \zeta_v^T$.*

Proof. Set $N := \sum_{v \in V_\mathbb{C}(\mathcal{I})} \lambda_v \zeta_v \zeta_v^T$. We first show that $p_u^T M(y) p_v = p_u^T N p_v$ for all $u, v \in V_\mathbb{C}(\mathcal{I})$. This identity is obvious if $u = v$. If $u \neq v$ then $p_u^T N p_v = 0$; on the other hand, $p_u^T M(y) p_v = \text{vec}(1)^T M(y) \text{vec}(p_u p_v) = 0$ where we use Lemma 4.1 for the first equality and the fact that $p_u p_v \in \mathcal{I}(V_\mathbb{C}(\mathcal{I})) = \mathcal{I}$ for the second equality. As the set $\{p_v \mid v \in V_\mathbb{C}(\mathcal{I})\}$ is a basis of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ (by Lemma 2.5), we deduce that $f^T M(y) g = f^T N g$ for all $f, g \in \mathbb{R}[\mathbf{x}]$, implying $M(y) = N$. □

Lemma 5.5. *The measure $\mu = \sum_{v \in V_\mathbb{C}(\mathcal{I})} \lambda_v \delta_v$ is r-atomic and it is the unique representing measure for y.*

Proof. μ is a representing measure for y by Lemma 5.4 and μ is r-atomic since $p_v^T M(y) p_v > 0$ as $p_v \notin \mathcal{I}$ for $v \in V_\mathbb{C}(\mathcal{I})$. We now verify the unicity of such measure. Say, μ' is another representing measure for y. By Lemma 4.2, $r = \operatorname{rank} M(y) \leq r' := |\operatorname{supp}(\mu')|$; moreover, $\operatorname{supp}(\mu') \subseteq V_\mathbb{C}(\mathcal{I})$, implying $r' \leq |V_\mathbb{C}(\mathcal{I})| = r$. Thus, $r = r'$, $\operatorname{supp}(\mu') = \operatorname{supp}(\mu) = V_\mathbb{C}(\mathcal{I})$ and $\mu = \mu'$. □

This concludes the proof of Theorem 5.1. □

We now make a simple observation about the degrees of interpolation polynomials at the points of $V_\mathbb{C}(\operatorname{Ker} M(y))$, which will be useful for the proof of Theorem 5.33 below.

Lemma 5.6. *Assume $M(y) \succeq 0$ and $r :=$ rank $M(y) < \infty$. Set $\mathcal{I} := \operatorname{Ker} M(y)$. If, for some integer $t \geq 1$, $\operatorname{rank} M_t(y) = r$, then there exist interpolation polynomials p_v $(v \in V_\mathbb{C}(\mathcal{I}))$ having degree at most t.*

Proof. As rank $M_t(y)$ = rank $M(y)$, one can choose a basis $\mathcal{B}$ of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ where $\mathcal{B} \subseteq \mathbb{T}^n_t$. (Recall Lemma 5.3.) Let q_v ($v \in V_{\mathbb{C}}(\mathcal{I})$) be interpolation polynomials at the points of $V_{\mathbb{C}}(\mathcal{I})$. Replacing each q_v by its residue p_v modulo $\mathcal{I}$ w.r.t. the basis $\mathcal{B}$, we obtain a new set of interpolation polynomials p_v ($v \in V_{\mathbb{C}}(\mathcal{I})$) with $\deg p_v \leq t$. □

As we saw in Lemma 5.2, the kernel of an *infinite* moment matrix is an ideal in $\mathbb{R}[\mathbf{x}]$. We now observe that, although the kernel of a *truncated* moment matrix cannot be claimed to be an ideal, it yet enjoys some 'truncated ideal like' properties. We will use the notion of *flat extension* of a matrix, introduced earlier in Definition 1.1, as well as Lemma 1.2.

Lemma 5.7. *Let $f, g \in \mathbb{R}[\mathbf{x}]$.*
(i) *If $\deg(fg) \leq t-1$ and $M_t(y) \succeq 0$, then*

$$f \in \operatorname{Ker} M_t(y) \Longrightarrow fg \in \operatorname{Ker} M_t(y). \tag{5.1}$$

(ii) *If $\deg(fg) \leq t$ and $\operatorname{rank} M_t(y) = \operatorname{rank} M_{t-1}(y)$, then (5.1) holds.*

Proof. It suffices to show the result for $g = \mathbf{x}_i$ since the general result follows from repeated applications of this special case. Then, $h := f\mathbf{x}_i = \sum_\alpha f_\alpha \mathbf{x}^{\alpha+e_i} = \sum_{\alpha|\alpha\geq e_i} f_{\alpha-e_i}\mathbf{x}^\alpha$. For $\alpha \in \mathbb{N}^n_{t-1}$, we have:

$$\begin{aligned}(M_t(y)h)_\alpha &= \sum_\gamma h_\gamma y_{\alpha+\gamma} = \sum_{\gamma|\gamma\geq e_i} f_{\gamma-e_i} y_{\alpha+\gamma} \\ &= \sum_\gamma f_\gamma y_{\alpha+\gamma+e_i} = (M_t(y)f)_{\alpha+e_i} = 0.\end{aligned}$$

In view of Lemma 1.2, this implies $M_t(y)h = 0$ in both cases (i), (ii). □

Example 5.8. *Here is an example showing that Lemma 5.7 (i) cannot be extended to the case when $\deg(fg) = t$. For this, consider the sequence $y := (1,1,1,1,2) \in \mathbb{R}^{\mathbb{N}^1_4}$ (here $n = 1$). Thus,*

$$M_2(y) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} \succeq 0. \tag{5.2}$$

Then the polynomial $1-\mathbf{x}$ belongs to the kernel of $M_2(y)$, but the polynomial $\mathbf{x}(1-\mathbf{x}) = \mathbf{x} - \mathbf{x}^2$ does not belong to the kernel of $M_2(y)$.

5.2. Finite atomic measures for truncated moment sequences. Theorem 5.1 characterizes the infinite sequences having a finite atomic representing measure. The next question is to characterize the *truncated* sequences $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$ having a finite atomic representing measure μ. It turns out that, for a truncated sequence, the existence of a representing measure implies the existence of another one with a *finite* support (this is not true for infinite sequences). This result, due to Bayer and Teichmann [9], strengthens an earlier result of Putinar [135] which assumed the existence of a measure with a compact support.

Theorem 5.9. [9] *If a truncated sequence $y \in \mathbb{R}^{\mathbb{N}^n_t}$ has a representing measure μ, then it has another representing measure ν which is finitely atomic with* $\text{supp}(\nu) \subseteq \text{supp}(\mu)$ *and* $|\text{supp}(\nu)| \le \binom{n+t}{t}$.

Proof. Let $S := \text{supp}(\mu)$, i.e., S is the smallest closed set for which $\mu(\mathbb{R}^n \setminus S) = 0$. Let $\mathcal{C} \subseteq \mathbb{R}^{\mathbb{N}^n_t}$ denote the convex cone generated by the vectors $\zeta_{t,x} = (x^\alpha)_{\alpha \in \mathbb{N}^n_t}$ for $x \in S$. Then its closure $\overline{\mathcal{C}}$ is a closed convex cone in $\mathbb{R}^{\mathbb{N}^n_t}$ and therefore it is equal to the intersection of its supporting halfspaces. That is,

$$\overline{\mathcal{C}} = \{z \in \mathbb{R}^{\mathbb{N}^n_t} \mid c^T z \ge 0 \ \forall c \in H\}$$

for some $H \subseteq \mathbb{R}^{\mathbb{N}^n_t}$. Obviously, $y \in \overline{\mathcal{C}}$ since, for any $c \in H$,

$$c^T y = \sum_\alpha c_\alpha y_\alpha = \int_S (\sum_\alpha c_\alpha x^\alpha) \mu(dx) \ge 0$$

as $\sum_\alpha c_\alpha x^\alpha = c^T \zeta_{t,x} \ge 0$ for all $x \in S$. Moreover,

$$y \text{ belongs to the relative interior of } \overline{\mathcal{C}}. \tag{5.3}$$

To see it, consider a supporting hyperplane $\{z \mid c^T z = 0\}$ $(c \in H)$ that does not contain $\overline{\mathcal{C}}$. We show that $c^T y > 0$. For this, assume $c^T y = 0$ and set $X := \{x \in S \mid c^T \zeta_{t,x} > 0\}$, $X_k := \{x \in S \mid c^T \zeta_{t,x} \ge \frac{1}{k}\}$ for $k \ge 1$ integer. Then, $X \ne \emptyset$ and $X = \bigcup_{k \ge 1} X_k$. We have

$$0 = c^T y = \int_X c^T \zeta_{t,x} \mu(dx) \ge \int_{X_k} c^T \zeta_{t,x} \mu(dx) \ge \frac{1}{k} \mu(X_k) \ge 0,$$

implying $\mu(X_k) = 0$. This shows that $\mu(X) = 0$. Now, the set $S \setminus X = \{x \in S \mid c^T \zeta_{t,x} = 0\}$ is closed with $\mu(\mathbb{R}^n \setminus (S \setminus X)) = \mu(X) + \mu(\mathbb{R}^n \setminus S) = 0$. We reach a contradiction since $S \setminus X$ is a strict closed subset of S.

Therefore, (5.3) holds and thus y belongs to the cone $\mathcal{C}$, since the two cones $\mathcal{C}$ and its closure $\overline{\mathcal{C}}$ have the same relative interior. Using Carathéodory's theorem, we deduce that y can be written as a conic combination of at most $|\mathbb{N}^n_t| = \binom{n+t}{t}$ vectors $\zeta_{t,x}$ $(x \in S)$; that is, y has an atomic representing measure on S with at most $\binom{n+t}{t}$ atoms. □

Remark 5.10. *As pointed out by M. Schweighofer, the above proof can be adapted to show the following stronger result: If $y \in \mathbb{R}^{\mathbb{N}^n_t}$ has a representing measure μ and $S \subseteq \mathbb{R}^n$ is a measurable set with $\mu(\mathbb{R}^n \setminus S) = 0$, then y has another representing measure ν which is finite atomic with* $\text{supp}(\nu) \subseteq S$ *and* $|\text{supp}(\nu)| \le \binom{n+t}{t}$.

To see it, let $\mathcal{I} \subseteq \mathbb{R}[\mathbf{x}]$ denote an ideal which is maximal with respect to the property that $\mu(\mathbb{R}^n \setminus (V_{\mathbb{R}}(\mathcal{I}) \cap S)) = 0$ (such an ideal exists by assumption *since $\mathbb{R}[\mathbf{x}]$ is Noetherian). Set $S' := V_{\mathbb{R}}(\mathcal{I}) \cap S$; thus $\mu(\mathbb{R}^n \setminus S') = 0$. We now construct a measure ν with* $\operatorname{supp}(\nu) \subseteq S'$. *For this consider the cone $\mathcal{C}'$, defined as the conic hull of the vectors $\zeta_{t,x}$ ($x \in S'$). Again one can show that y lies in the relative interior of the closure $\overline{\mathcal{C}'}$; the proof is analogous but needs some more argument. Namely, suppose $c^T z = 0$ is a supporting hyperplane for $\overline{\mathcal{C}'}$ that does not contain $\overline{\mathcal{C}'}$. As above, the set $X := \{x \in S' \mid c^T\zeta_{t,x} > 0\}$ has measure $\mu(X) = 0$. Consider the polynomial $f := \sum_\alpha c_\alpha \mathbf{x}^\alpha \in \mathbb{R}[\mathbf{x}]_t$ and the ideal $\mathcal{J} := \mathcal{I} + (f) \subseteq \mathbb{R}[\mathbf{x}]$. Then, $V_{\mathbb{R}}(\mathcal{J}) = V_{\mathbb{R}}(\mathcal{I}) \cap V_{\mathbb{R}}(f)$, $V_{\mathbb{R}}(\mathcal{J}) \cap S = \{x \in S' \mid c^T\zeta_{t,x} = 0\} = S' \setminus X$, and thus $\mu(\mathbb{R}^n \setminus (V_{\mathbb{R}}(\mathcal{J}) \cap S)) = \mu(X) + \mu(\mathbb{R}^n \setminus S') = 0$. This implies $\mathcal{J} = \mathcal{I}$ (by our maximality assumption on $\mathcal{I}$) and thus $f \in \mathcal{I}$. Hence f vanishes on $S' \subseteq V_{\mathbb{R}}(\mathcal{I})$ which implies $X = \emptyset$, yielding a contradiction. The rest of the proof is analogous: y lies in the relative interior of $\overline{\mathcal{C}'}$, thus of $\mathcal{C}'$; hence we can apply Caratheodory's theorem and conclude that y can be written as a conic combination of at most $\binom{n+t}{t}$ vectors $\zeta_{t,x}$ with $x \in S'$.*

EXAMPLE 5.11. *As an illustration, consider the case $n = t = 1$ and the uniform measure μ on $[-1, 1]$ whose sequence of moments is $y = (2, 0) \in \mathbb{R}^{\mathbb{N}^1_1}$. Theorem 5.9 tells us that there is another representing measure ν for y with at most two atoms. Indeed, the Dirac measure $\delta_{\{0\}}$ at the origin represents y, but if we exclude the origin then we need two atoms to represent y; namely $\nu = \delta_{\{\epsilon\}} + \delta_{\{-\epsilon\}}$ represents y for any $\epsilon > 0$.*

Finding alternative measures with a small number of atoms is also known as the problem of finding cubature (or quadrature) rules for measures. The next result is a direct consequence of Theorem 5.9.

COROLLARY 5.12. *For a measurable set $K \subseteq \mathbb{R}^n$ and $y \in \mathbb{R}^{\mathbb{N}^n_t}$, the following assertions (i)-(iii) are equivalent: (i) y has a representing measure on K; (ii) y has an atomic representing measure on K; (iii) $y = \sum_{i=1}^N \lambda_i \delta_{x_i}$ for some $\lambda_i > 0$, $x_i \in K$.*

As we saw earlier, Haviland's theorem claims that an infinite sequence $y \in \mathbb{R}^{\mathbb{N}^n}$ has a representing measure on a closed subset $K \subseteq \mathbb{R}^n$ if and only if $y^T p \geq 0$ for all polynomials p nonnegative on K; that is, $\mathcal{M}_K = (\mathcal{P}_K)^*$ in terms of conic duality. One may naturally wonder whether there is an analogue of this result for the truncated moment problem. For this, define

$$\mathcal{P}_{K,t} := \{p \in \mathbb{R}[\mathbf{x}]_t \mid p \geq 0 \text{ on } K\},$$
$$\mathcal{M}_{K,t} := \{y \in \mathbb{R}^{\mathbb{N}^n_t} \mid y \text{ has a representing measure on } K\}.$$

Obviously, $\mathcal{M}_{K,t} \subseteq (\mathcal{P}_{K,t})^*$; Tchakaloff [162] proved that equality holds when K is compact.

THEOREM 5.13. *[162] If $K \subseteq \mathbb{R}^n$ is compact, then $\mathcal{M}_{K,t} = (\mathcal{P}_{K,t})^*$.*

Proof. Set $\Delta_N := \{(\lambda_1, \ldots, \lambda_N) \in \mathbb{R}^N_+ \mid \sum_{i=1}^N \lambda_i = 1\}$ and $N := \binom{n+t}{t}$.

Consider the mapping

$$\begin{array}{lll}\varphi: & \Delta_N \times K^N & \to \mathcal{M}_{K,t} \\ & (\lambda_1, \ldots, \lambda_N, v_1, \ldots, v_N) & \mapsto \sum_{i=1}^N \lambda_i \zeta_{t,v_i},\end{array}$$

whose image is $\mathcal{M}_{K,t} \cap \{y \mid y_0 = 1\}$ in view of Theorem 5.9. As φ is continuous and K is compact, $\mathcal{M}_{K,t} \cap \{y \mid y_0 = 1\}$ is the continuous image of a compact set, and thus it is a compact set. This implies that $\mathcal{M}_{K,t}$ is a closed set. Indeed suppose $y^{(i)} \in \mathcal{M}_{K,t}$ converges to $y \in \mathbb{R}^{\mathbb{N}^n_t}$; we show that $y \in \mathcal{M}_{K,t}$. If $y_0 > 0$ then $y_0^{(i)} > 0$ for i large enough; the limit of $\frac{y^{(i)}}{y_0^{(i)}}$ belongs to $\mathcal{M}_{K,t} \cap \{z \mid z_0 = 1\}$ and is equal to $\frac{y}{y_0}$, which implies $y \in \mathcal{M}_{K,t}$. It suffices now to show that $y_0 = 0$ implies $y = 0$. For this, let $\mu^{(i)}$ be a representing measure on K for $y^{(i)}$. As K is compact, there exists a constant $C > 0$ such that $|x^\alpha - 1| \leq C$ for all $x \in K$ and $\alpha \in \mathbb{N}^n_t$. Then, $|y_\alpha^{(i)} - y_0^{(i)}| = |\int_K (x^\alpha - 1) d\mu^{(i)}(x)| \leq \int_K |x^\alpha - 1| d\mu^{(i)}(x) \leq C \int_K d\mu^{(i)}(x) = Cy_0^{(i)}$. Taking limits as $i \to \infty$, this implies $|y_\alpha - y_0| \leq Cy_0$ and thus $y_0 = 0$ implies $y = 0$.

Now suppose $y \in (\mathcal{P}_{K,t})^* \setminus \mathcal{M}_{K,t}$. Thus there is a hyperplane separating y from the closed convex set $\mathcal{M}_{K,t}$; that is, there exists $p \in \mathbb{R}[\mathbf{x}]_t$ for which $y^T p < 0$ and $z^T p \geq 0$ for all $z \in \mathcal{M}_{K,t}$. Hence $p \notin \mathcal{P}_{K,t}$ and thus $p(a) < 0$ for some $a \in K$. Now, $z := \zeta_{t,a} \in \mathcal{M}_{K,t}$ implies $z^T p = p(a) \geq 0$, yielding a contradiction. Therefore, $\mathcal{M}_{K,t} = (\mathcal{P}_{K,t})^*$, which concludes the proof. □

Here is an example (taken from [31]) showing that the inclusion $\mathcal{M}_{K,t} \subseteq (\mathcal{P}_{K,t})^*$ can be strict for general K.

Example 5.14. *Consider again the sequence* $y := (1,1,1,1,2) \in \mathbb{R}^{\mathbb{N}^1_4}$ *(with* $n = 1$*) from Example 5.8, whose moment matrix of order 2 is shown in (5.2). Then* $y \in (\mathcal{P}_{\mathbb{R},4})^*$*, since* $M_2(y) \succeq 0$ *and any univariate nonnegative polynomial is a sum of squares. However* y *does not have a representing measure. Indeed, if* μ *is a representing measure for* y*, then its support is contained in* $V_{\mathbb{C}}(\operatorname{Ker} M_2(y)) \subseteq \{1\}$ *since the polynomial* $1 - \mathbf{x}$ *lies in* $\operatorname{Ker} M_2(y)$*. But then* μ *would be the Dirac measure* $\delta_{\{1\}}$ *which would imply* $y_4 = 1$*, a contradiction. On the other hand, note that the subsequence* $(1,1,1,1)$ *of* y *containing its components up to order 3 does admit the Dirac measure* δ_1 *as representing measure (cf. Theorem 5.15 below for a general statement). Moreover some small perturbation of* y *does have a representing measure (cf. Example 5.18).*

Although the inclusion $\mathcal{M}_{K,t} \subseteq (\mathcal{P}_{K,t})^*$ is strict in general, Curto and Fialkow [31] can prove the following results. We omit the proofs, which use the Riesz representation theorem and the technical result from Theorem 5.17 about limits of measures.

Theorem 5.15. [31, Th. 2.4] *Let* $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$ *and let* K *be a closed subset of* $\mathbb{R}^n$*. If* $y \in (\mathcal{P}_{K,2t})^*$*, then the subsequence* $(y_\alpha)_{\alpha \in \mathbb{N}^n_{2t-1}}$ *has a representing measure on* K*.*

THEOREM 5.16. [31, Th. 2.2] *Let $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$ and let K be a closed subset of $\mathbb{R}^n$. Then y has a representing measure on K if and only if y admits an extension $\tilde{y} \in \mathbb{R}^{\mathbb{N}^n_{2t+2}}$ such that $\tilde{y} \in (\mathcal{P}_{K,2t+2})^*$.*

The following result of Stochel [160] shows that the truncated moment problem is in fact more general than the (infinite) moment problem. Moreover, it can be used to derive Haviland's theorem from Theorem 5.16.

THEOREM 5.17. [160, Th. 4] *Let $y \in \mathbb{R}^{\mathbb{N}^n}$ and let $K \subseteq \mathbb{R}^n$ be a closed set. Then y has a representing measure on K if and only if, for each integer $t \geq 1$, the subsequence $(y_\alpha)_{\alpha \in \mathbb{N}^n_t}$ has a representing measure on K.*

The inclusion $\mathcal{M}_{K,t} \subseteq (\mathcal{P}_{K,t})^*$ is strict in general. However, in the case when $K = \mathbb{R}^n$, one can show that equality $\text{cl}(\mathcal{M}_{K,t}) = (\mathcal{P}_{K,t})^*$ holds in the three exceptional cases of Hilbert's theorem characterizing when all nonnegative polynomials are sums of squares of polynomials. Note that we have to take the closure of the cone $\mathcal{M}_{K,t}$ since this cone is not closed in general when K is not compact. As an illustration consider again the sequence y from Example 5.14, which in fact lies in $\text{cl}(\mathcal{M}_{K,t}) \setminus \mathcal{M}_{K,t}$.

EXAMPLE 5.18. *Consider again the sequence $y = (1, 1, 1, 1, 2) \in \mathbb{R}^{\mathbb{N}^1_4}$ from Example 5.14. We saw there that $y \in (\mathcal{P}_{K,t})^*$ but $y \notin \mathcal{M}_{K,t}$ ($K = \mathbb{R}$, $n = 1$, $t = 4$). However, $y \in cl(\mathcal{M}_{K,t})$. To see it consider the sequence $y_\epsilon = (1+\epsilon^4, 1+\epsilon^3, 1+\epsilon^2, 1+\epsilon, 2)$ for $\epsilon > 0$. Then the sequence y_ϵ admits the atomic measure $\delta_{\{1\}} + \epsilon^4 \delta_{\{1/\epsilon\}}$ as representing measure and $y = \lim_{\epsilon \to 0} y_\epsilon$, which shows that y belongs to the closure of the convex hull of the moment curve $\{\zeta_{4,a} \mid a \in \mathbb{R}\}$.*

THEOREM 5.19. *[60] Let $K = \mathbb{R}^n$ and assume $n = 1$, or $t = 2$, or $(n = 2, t = 4)$. Then $cl(\mathcal{M}_{K,t}) = (\mathcal{P}_{K,t})^* = \{y \in \mathbb{R}^{\mathbb{N}^n_t} \mid M_t(y) \succeq 0\}$.*

Proof. By Hilbert's theorem (Theorem 3.4), we know that any polynomial in $\mathcal{P}_{K,t}$ is a sum of squares and thus $(\mathcal{P}_{K,t})^* = \{y \in \mathbb{R}^{\mathbb{N}^n_t} \mid M_t(y) \succeq 0\}$. Suppose that y satisfies $M_t(y) \succeq 0$ but $y \notin \text{cl}(\mathcal{M}_{K,t})$. Then there exists a hyperplane separating the (closed convex) cone $\text{cl}(\mathcal{M}_{K,t})$ from y. That is, there exists a polynomial $p \in \mathbb{R}[\mathbf{x}]_t$ for which $p^T y < 0$ and $p \geq 0$ on $\mathbb{R}^n$. Then p is a sum of squares (since we are in one of the three exceptional cases of Hilbert's theorem) which together with $M_t(y) \succeq 0$ implies $p^T y \geq 0$, yielding a contradiction. □

We refer to Henrion [59, 60] for concrete applications of this result to get explicit semidefinite representations of some rational varieties.

5.3. Flat extensions of moment matrices. The main result in this section is Theorem 5.20 below, which provides a key result about flat extensions of moment matrices. Indeed it permits to extend a truncated sequence $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$, satisfying some 'flat extension' assumption, to an infinite sequence $\tilde{y} \in \mathbb{R}^{\mathbb{N}^n}$, satisfying $\text{rank}\, M(\tilde{y}) = \text{rank}\, M_t(y)$. In this way one can then apply the tools developed for infinite moment sequences (e.g., Theorem 5.1) to truncated sequences. Recall the notion of 'flat extension' from Definition 1.1. In particular, $M_t(y)$ is a *flat extension* of $M_{t-1}(y)$ if $\operatorname{rank} M_t(y) = \operatorname{rank} M_{t-1}(y)$.

Theorem 5.20. (**Flat extension theorem** [28]) *Let $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$. If $M_t(y)$ is a flat extension of $M_{t-1}(y)$, then one can extend y to a (unique) vector $\tilde{y} \in \mathbb{R}^{\mathbb{N}^n_{2t+2}}$ in such a way that $M_{t+1}(\tilde{y})$ is a flat extension of $M_t(y)$.*

We will now give two proofs for this result. The first proof (which follows the treatment in [28]) is completely elementary but with technical details. The second proof (which follows the treatment in [98]) is less technical; it uses some results of commutative algebra and permits to show a more general result (cf. Theorem 5.24).

5.3.1. First proof of the flat extension theorem. We begin with a characterization of moment matrices, which we will use in the proof. By definition, a matrix M (indexed by $\mathbb{N}^n_t$) is a moment matrix if $M_{\alpha,\beta} = M_{\alpha',\beta'}$ whenever $\alpha+\beta = \alpha'+\beta'$. The next lemma shows that it suffices to check this for 'adjacent' pairs (α,β) and (α',β'), namely for the pairs for which the Hamming distance between α and α' is at most 2.

Lemma 5.21. *Let M be a symmetric matrix indexed by $\mathbb{N}^n_t$. Then, M is a moment matrix, i.e., $M = M_t(y)$ for some sequence $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$, if and only if the following holds:*

(i) $M_{\alpha,\beta} = M_{\alpha-e_i,\beta+e_i}$ *for all* $\alpha,\beta \in \mathbb{N}^n_t$, $i \in \{1,\ldots,n\}$ *such that* $\alpha_i \geq 1$, $|\beta| \leq t-1$.

(ii) $M_{\alpha,\beta} = M_{\alpha-e_i+e_j,\beta+e_i-e_j}$ *for all* $\alpha,\beta \in \mathbb{N}^n_t$, $i,j \in \{1,\ldots,n\}$ *such that* $\alpha_i,\beta_j \geq 1$, $|\alpha| = |\beta| = t$.

Proof. The 'if part' being obvious, we show the 'only if' part. That is, we assume that (i), (ii) hold and we show that $M(\alpha,\beta) = M(\alpha',\beta')$ whenever $\alpha+\beta = \alpha'+\beta'$. For this we use induction on the parameter $\delta_{\alpha\beta,\alpha'\beta'} := \min(\|\alpha-\alpha'\|_1, \|\alpha-\beta'\|_1)$. If $\delta_{\alpha\beta,\alpha'\beta'} = 0$, then $(\alpha,\beta) = (\alpha',\beta')$ and there is nothing to prove. If $\delta_{\alpha\beta,\alpha'\beta'} = 1$, then the result holds by assumption (i). Assume now that $\delta_{\alpha\beta,\alpha'\beta'} \geq 2$.

Consider first the case when $|\alpha|+|\beta| \leq 2t-1$. As $\alpha \neq \alpha'$ we may assume without loss of generality that $\alpha'_i \geq \alpha_i + 1$ for some i, implying $\beta'_i \leq \beta_i - 1$. Define $(\alpha'',\beta'') := (\alpha-e_i,\beta+e_i)$. Then, $\delta_{\alpha\beta,\alpha''\beta''} = \delta_{\alpha\beta,\alpha'\beta'} - 1$. If $|\beta'| \leq t-1$, then $M_{\alpha,\beta} = M_{\alpha'',\beta''}$ by the induction assumption and $M_{\alpha'',\beta''} = M_{\alpha',\beta'}$ by (i), implying the desired result. Assume now that $|\beta'| = t$ and thus $|\alpha'| \leq t-1$. Then, $|\alpha|-|\alpha'| = t-|\beta| \geq 0$ and thus $\alpha_i \geq \alpha'_i + 1$ for some i, yielding $\beta'_i \geq \beta_i + 1$. Define $(\alpha'',\beta'') := (\alpha'+e_i,\beta'-e_i)$. Then $\delta_{\alpha\beta,\alpha''\beta''} = \delta_{\alpha\beta,\alpha'\beta'} - 1$. Therefore, $M_{\alpha,\beta} = M_{\alpha'',\beta''}$ by the induction assumption and $M_{\alpha'',\beta''} = M_{\alpha',\beta'}$ by (i), implying the desired result.

We can now suppose that $|\alpha| = |\beta| = |\alpha'| = |\beta'| = t$. Hence, $\alpha'_i \geq \alpha_i + 1$ for some i and $\beta'_j \geq \beta_j + 1$ for some j; moreover $i \neq j$. Define $(\alpha'',\beta'') := (\alpha'-e_i+e_j,\beta'+e_i-e_j)$. Then, $\delta_{\alpha\beta,\alpha''\beta''} = \delta_{\alpha\beta,\alpha'\beta'} - 2$. Therefore, $M_{\alpha,\beta} = M_{\alpha'',\beta''}$ by the induction assumption and $M_{\alpha'',\beta''} = M_{\alpha',\beta'}$ by (ii), implying the desired result. ∎

Set $M_t := M_t(y) = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$, where $A := M_{t-1}(y)$. By assumption, M_t is a flat extension of A. Our objective is to construct a flat extension $N := \begin{pmatrix} M_t & D \\ D^T & E \end{pmatrix}$ of M_t, which is a moment matrix. As M_t is a flat extension of A, we can choose a subset $\mathcal{B} \subseteq \mathbb{N}^n_{t-1}$ indexing a maximum set of linearly independent columns of M_t. Then any column of M_t can be expressed (in a unique way) as a linear combination of columns indexed by $\mathcal{B}$. In other words, for any polynomial $p \in \mathbb{R}[\mathbf{x}]_t$, there exists a unique polynomial $r \in \mathrm{Span}_{\mathbb{R}}(\mathcal{B})$ for which $p - r \in \operatorname{Ker} M_t$.

Lemma 5.7 (ii) plays a central role in the construction of the matrix N, i.e., of the matrices D and E. Take $\gamma \in \mathbb{N}^n_{t+1}$ with $|\gamma| = t+1$. Say, $\gamma_i \geq 1$ for some $i = 1, \ldots, n$ and $\mathbf{x}^{\gamma - e_i} - r \in \operatorname{Ker} M_t$, where $r \in \mathrm{Span}_{\mathbb{R}}(\mathcal{B})$. Then it follows from Lemma 5.7 (ii) that $\mathbf{x}_i(\mathbf{x}^{\gamma-e_i} - r)$ belongs to the kernel of N, the desired flat extension of M_t. In other words, $N\mathrm{vec}(\mathbf{x}^\gamma) = N\mathrm{vec}(\mathbf{x}_i r)$, which tells us how to define the γth column of N, namely, by $D\mathrm{vec}(\mathbf{x}^\gamma) = M_t\mathrm{vec}(\mathbf{x}_i r)$ and $E\mathrm{vec}(\mathbf{x}^\gamma) = D^T\mathrm{vec}(\mathbf{x}_i r)$. We now verify that these definitions are *good*, i.e., that they do not depend on the choice of the index i for which $\gamma_i \geq 1$.

Lemma 5.22. *Let $\gamma \in \mathbb{N}^n$ with $|\gamma| = t+1$, $\gamma_i, \gamma_j \geq 1$ and let $r, s \in$ Span$_{\mathbb{R}}(\mathcal{B})$ for which $\mathbf{x}^{\gamma-e_i} - r$, $\mathbf{x}^{\gamma-e_j} - s \in \operatorname{Ker} M_t$. Then we have $M_t\mathrm{vec}(\mathbf{x}_i r - \mathbf{x}_j s) = 0$ (implying that D is well defined) and $D^T\mathrm{vec}(\mathbf{x}_i r - \mathbf{x}_j s) = 0$ (implying that E is well defined).*

Proof. We first show that $M_t\mathrm{vec}(\mathbf{x}_i r - \mathbf{x}_j s) = 0$. In view of Lemma 1.2 (ii), it suffices to show that $\mathrm{vec}(\mathbf{x}^\alpha)^T M_t\mathrm{vec}(\mathbf{x}_i r - \mathbf{x}_j s) = 0$ for all $\alpha \in \mathbb{N}^n_{t-1}$. Fix $\alpha \in \mathbb{N}^n_{t-1}$. Then,

$$\begin{aligned}\mathrm{vec}(\mathbf{x}^\alpha)^T M_t\mathrm{vec}(\mathbf{x}_i r - \mathbf{x}_j s) &= \mathrm{vec}(\mathbf{x}_i\mathbf{x}^\alpha)^T M_t r - \mathrm{vec}(\mathbf{x}_j\mathbf{x}^\alpha)^T M_t s \\ &= \mathrm{vec}(\mathbf{x}_i\mathbf{x}^\alpha)^T M_t\mathrm{vec}(\mathbf{x}^{\gamma-e_i}) - \mathrm{vec}(\mathbf{x}_j\mathbf{x}^\alpha)^T M_t\mathrm{vec}(\mathbf{x}^{\gamma-e_j}) \\ &= y^T\mathrm{vec}(\mathbf{x}_i\mathbf{x}^\alpha\mathbf{x}^{\gamma-e_i}) - y^T\mathrm{vec}(\mathbf{x}_j\mathbf{x}^\alpha\mathbf{x}^{\gamma-e_j}) = 0,\end{aligned}$$

where we have used the fact that $r - \mathbf{x}^{\gamma-e_i}, s - \mathbf{x}^{\gamma-e_j} \in \operatorname{Ker} M_t$ for the second equality, and Lemma 4.1 for the third equality. We now show that $D^T\mathrm{vec}(\mathbf{x}_i r - \mathbf{x}_j s) = 0$, i.e., $\mathrm{vec}(\mathbf{x}_i r - \mathbf{x}_j s)^T D\mathrm{vec}(\mathbf{x}^\delta) = 0$ for all $|\delta| = t+1$. Fix $\delta \in \mathbb{N}^n_{t+1}$. Say, $\delta_k \geq 1$ and $\mathbf{x}^{\delta-e_k} - u \in \operatorname{Ker} M_t$, where $u \in \mathrm{Span}_{\mathbb{R}}(\mathcal{B})$. Then, $D\mathrm{vec}(\mathbf{x}^\delta) = M_t\mathrm{vec}(\mathbf{x}_k u)$ by construction. Using the above, this implies $\mathrm{vec}(\mathbf{x}_i r - \mathbf{x}_j s)^T D\mathrm{vec}(\mathbf{x}^\delta) = \mathrm{vec}(\mathbf{x}_i r - \mathbf{x}_j s)^T M_t\mathrm{vec}(\mathbf{x}_k u) = 0$. □

We now verify that the matrix N is a moment matrix, i.e., that N satisfies the conditions (i), (ii) from Lemma 5.21.

Lemma 5.23.

(i) *$N_{\gamma,\delta} = N_{\gamma+e_i,\delta-e_i}$ for $\gamma, \delta \in \mathbb{N}^n_{t+1}$ with $\delta_i \geq 1$ and $|\gamma| \leq t$.*

(ii) *$N_{\gamma,\delta} = N_{\gamma-e_j+e_i,\delta+e_j-e_i}$ for $\gamma, \delta \in \mathbb{N}^n_{t+1}$ with $\gamma_j \geq 1$, $\delta_i \geq 1$, $|\gamma| = |\delta| = t+1$.*

Proof. (i) Assume $\mathbf{x}^{\delta-e_i}-r$, $\mathbf{x}^{\gamma}-s \in \operatorname{Ker} M_t$ for some $r, s \in \operatorname{Span}_{\mathbb{R}}(\mathcal{B})$; then $\mathbf{x}^{\delta}-\mathbf{x}_i r$, $\mathbf{x}_i\mathbf{x}^{\gamma}-\mathbf{x}_i s \in \operatorname{Ker} N$, by construction. We have

$$\begin{aligned}
\operatorname{vec}(\mathbf{x}^{\gamma})^T N \operatorname{vec}(\mathbf{x}^{\delta}) &= \operatorname{vec}(\mathbf{x}^{\gamma})^T N \operatorname{vec}(\mathbf{x}_i r) = \operatorname{vec}(\mathbf{x}^{\gamma})^T M_t \operatorname{vec}(\mathbf{x}_i r)\\
&= s^T M_t \operatorname{vec}(\mathbf{x}_i r) = \operatorname{vec}(\mathbf{x}_i s)^T M_t r = \operatorname{vec}(\mathbf{x}_i s)^T M_t \operatorname{vec}(\mathbf{x}^{\delta-e_i})\\
&= \operatorname{vec}(\mathbf{x}_i s)^T N \operatorname{vec}(\mathbf{x}^{\delta-e_i}) = \operatorname{vec}(\mathbf{x}_i\mathbf{x}^{\gamma})^T N \operatorname{vec}(\mathbf{x}^{\delta-e_i}).
\end{aligned}$$

This shows $N_{\gamma,\delta} = N_{\gamma+e_i,\delta-e_i}$.
(ii) Let $r, s \in \operatorname{Span}_{\mathbb{R}}(\mathcal{B})$ for which $\mathbf{x}^{\delta-e_i}-r, \mathbf{x}^{\gamma-e_j}-s \in \operatorname{Ker} M_t$. Then, $\mathbf{x}^{\delta}-\mathbf{x}_i r,\ \mathbf{x}_j\mathbf{x}^{\delta-e_i}-\mathbf{x}_j r,\ \mathbf{x}^{\gamma}-\mathbf{x}_j s,\ \mathbf{x}_i\mathbf{x}^{\gamma-e_j}-\mathbf{x}_i s \in \operatorname{Ker} N$ by construction. We have

$$\begin{aligned}
\operatorname{vec}(\mathbf{x}^{\gamma})^T N \operatorname{vec}(\mathbf{x}^{\delta}) &= \operatorname{vec}(\mathbf{x}_j s)^T N \operatorname{vec}(\mathbf{x}_i r) = \operatorname{vec}(\mathbf{x}_j s)^T M_t \operatorname{vec}(\mathbf{x}_i r)\\
&= \operatorname{vec}(\mathbf{x}_i s)^T M_t \operatorname{vec}(\mathbf{x}_j r) = \operatorname{vec}(\mathbf{x}_i s)^T N \operatorname{vec}(\mathbf{x}_j r)\\
&= \operatorname{vec}(\mathbf{x}^{\gamma-e_j+e_i})^T N \operatorname{vec}(\mathbf{x}^{\delta-e_i+e_j}),
\end{aligned}$$

which shows $N_{\gamma,\delta} = N_{\gamma-e_j+e_i,\delta+e_j-e_i}$. □

This concludes the proof of Theorem 5.20.

5.3.2. A generalized flat extension theorem. Theorem 5.20 considers moment matrices indexed by $\mathbb{N}^n_t$, corresponding to the full set of monomials of degree at most t. We now state a more general result which applies to matrices indexed by an arbitrary set of monomials (assumed to be connected to 1). We need some definitions.

Given a set of monomials $\mathcal{C} \subseteq \mathbb{T}^n$, set

$$\mathcal{C}^+ := \mathcal{C} \cup \bigcup_{i=1}^{n} \mathbf{x}_i\mathcal{C} = \{m, \mathbf{x}_1 m, \ldots, \mathbf{x}_n m \mid m \in \mathcal{C}\},\ \partial\mathcal{C} := \mathcal{C}^+ \setminus \mathcal{C},$$

called respectively the *closure* and the *border* of $\mathcal{C}$. Set also

$$\mathcal{C}\cdot\mathcal{C} := \{ab \mid a, b \in \mathcal{C}\}.$$

Recall that $\mathcal{C} \subseteq \mathbb{T}_n$ is *connected to 1* if $1 \in \mathcal{C}$ and every non-constant monomial $m \in \mathcal{C}$ can be written as $m = \mathbf{x}_{i_1}\cdots\mathbf{x}_{i_k}$ with $\mathbf{x}_{i_1}, \mathbf{x}_{i_1}\mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_1}\cdots\mathbf{x}_{i_k} \in \mathcal{C}$. For instance, $\mathcal{C}$ is connected to 1 if $\mathcal{C}$ is closed under taking divisions. For example, $\{1, \mathbf{x}_2, \mathbf{x}_1\mathbf{x}_2\}$ is connected to 1 but $\{1, \mathbf{x}_1\mathbf{x}_2\}$ is not.

Here it will turn out to be more convenient to index matrices by monomials $a = \mathbf{x}^{\alpha} \in \mathbb{T}^n$ rather than by integer sequences $\alpha \in \mathbb{N}^n$. So we now consider matrices M indexed by an arbitrary monomial set $\mathcal{C} \subseteq \mathbb{T}^n$; then M is said to be a *moment matrix* if $M_{a,b} = M_{a',b'}$ for all $a, b, a', b' \in \mathcal{C}$ with $ab = a'b'$. Thus the entries of M are given by a sequence y indexed by the set $\mathcal{C}\cdot\mathcal{C}$. For convenience we also denote M by $M_{\mathcal{C}}(y)$. When $\mathcal{C} = \mathbb{T}^n_t$, $M_{\mathcal{C}}(y) = M_t(y)$, i.e. we find the moment matrix $M_t(y)$ which we have considered so far.

We can now formulate the generalized flat extension theorem.

THEOREM 5.24. *(***Generalized flat extension theorem** *[98]) Consider a sequence $y = (y_a)_{a \in \mathcal{C}^+ \cdot \mathcal{C}^+}$, where $\mathcal{C} \subseteq \mathbb{T}^n$ is a finite set of monomials assumed to be connected to 1. If $M_{\mathcal{C}^+}(y)$ is a flat extension of $M_{\mathcal{C}}(y)$, then there exists a (unique) sequence $\tilde{y} = (\tilde{y})_{a \in \mathbb{T}^n}$ for which $M(\tilde{y})$ is a flat extension of $M_{\mathcal{C}^+}(y)$.*

We first give an example (taken from [98]) showing that the assumption that $\mathcal{C}$ be connected to 1 cannot be removed in the above theorem.

EXAMPLE 5.25. *In the univariate case, consider the set of monomials $\mathcal{C} := \{1, \mathbf{x}^3\}$, with border $\partial\mathcal{C} = \{\mathbf{x}, \mathbf{x}^4\}$; $\mathcal{C}$ is not connected to 1 and $\mathcal{C}^+ \cdot \mathcal{C}^+$ consists of all monomials of degree at most 8. Consider the sequence $y \in \mathbb{R}^8$ with entries $y_m = 1$ for $m \in \{1, \mathbf{x}, \mathbf{x}^2\}$, $y_m = b$ for $m \in \{\mathbf{x}^3, \mathbf{x}^4, \mathbf{x}^5\}$, and $y_m = c$ for $m \in \{\mathbf{x}^6, \mathbf{x}^7, \mathbf{x}^8\}$, where $b, c \in \mathbb{R}$ satisfy $c \neq b^2$. Then $M_{\mathcal{C}^+}(y)$ is a flat extension of $M_{\mathcal{C}}(y)$, but there is no flat extension of $M_{\mathcal{C}^+}(y)$. Indeed, such an extension has the form:*

$$M_{\mathcal{C}^+ \cup \{\mathbf{x}^2\}}(y) = \begin{array}{c} \\ 1 \\ \mathbf{x}^3 \\ \mathbf{x} \\ \mathbf{x}^4 \\ \mathbf{x}^2 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & \mathbf{x}^3 & \mathbf{x} & \mathbf{x}^4 & \mathbf{x}^2 \end{array} \\ \begin{pmatrix} 1 & b & 1 & c & 1 \\ b & c & b & c & b \\ 1 & b & 1 & b & b \\ b & c & b & c & c \\ 1 & b & b & c & b \end{pmatrix} \end{array}.$$

As the polynomials $1 - \mathbf{x}$ and $\mathbf{x}^3 - \mathbf{x}^4$ belong to $\operatorname{Ker} M_{\mathcal{C}^+}(y)$, they also belong to $\operatorname{Ker} M_{\mathcal{C}^+ \cup \{\mathbf{x}^2\}}(y)$, which implies $b = 1$ and $b = c$, contradicting our choice $c \neq b^2$.

When $\mathcal{C} = \mathbb{T}^n_t$, Theorem 5.24 coincides with the flat extension theorem from Theorem 5.20. Note that, in this case, the following additionnal condition holds:

$$\exists \mathcal{B} \subseteq \mathcal{C} \text{ s.t. } \mathcal{B} \text{ is connected to 1 and indexes a column basis of } M_{\mathcal{C}}(y). \tag{5.4}$$

It is indeed not difficult to show that a moment matrix $M_t(y)$ satisfies this condition (5.4); for instance such $\mathcal{B}$ can be constructed by selecting the linearly independent columns of $M_t(y)$ in a greedy way after ordering the monomials of $\mathbb{T}^n_t$ according to a total degree ordering.

Although Theorem 5.24 does not need the condition (5.4), its proof turns out to be a bit simpler when this additionnal condition holds, one reason for this being that we can then use the result of Theorem 2.16 about commuting multiplication matrices. We now give a proof of Theorem 5.24 and we will point out where the simplification occurs when (5.4) holds.

Proof of Theorem 5.24. For the proof it is convenient to use the language of linear forms. Namely, to the sequence $y \in \mathbb{R}^{\mathcal{C}^+ \cdot \mathcal{C}^+}$ corresponds the linear form $L \in (\text{Span}(\mathcal{C}^+ \cdot \mathcal{C}^+))^*$ defined by $L(p) := \sum_{a \in \mathcal{C}^+ \cdot \mathcal{C}^+} p_a y_a$ for

$p = \sum_{a\in\mathcal{C}^+\cdot\mathcal{C}^+} p_a a \in \mathrm{Span}(\mathcal{C}^+\cdot\mathcal{C}^+)$. Let us set for short $M_{\mathcal{C}^+} := M_{\mathcal{C}^+}(y) = (L(cc'))_{c,c'\in\mathcal{C}^+}$ and $M_{\mathcal{C}} := M_{\mathcal{C}}(y) = (L(cc')_{c,c'\in\mathcal{C}}$. Our assumption is that $\mathrm{rank}\, M_{\mathcal{C}^+} = \mathrm{rank}\, M_{\mathcal{C}}$. This directly implies the following observations which we will often use in the proof: For $p \in \mathrm{Span}(\mathcal{C}^+)$,

$$\begin{array}{c} p \in \mathrm{Ker}\, M_{\mathcal{C}^+} \overset{\text{def.}}{\Longleftrightarrow} L(ap) = 0 \ \forall a \in \mathcal{C}^+ \Longleftrightarrow L(ap) = 0 \ \forall a \in \mathcal{C}, \\ p \in \mathrm{Ker}\, M_{\mathcal{C}^+} \ \text{ and } \ \mathbf{x}_i p \in \mathrm{Span}(\mathcal{C}^+) \implies \mathbf{x}_i p \in \mathrm{Ker}\, M_{\mathcal{C}^+}. \end{array} \tag{5.5}$$

Our objective is to construct a linear form $\tilde{L} \in \mathbb{R}[\mathbf{x}]^*$ whose moment matrix $\tilde{M} := (L(a\cdot b))_{a,b\in\mathbb{T}^n}$ is a flat extension of $M_{\mathcal{C}^+}$. Let $\mathcal{B} \subseteq \mathcal{C}$ index a maximum set of linearly independent columns of $M_{\mathcal{C}}$. We may assume that $1 \in \mathcal{B}$, as L would be zero if no such basis exists. Then we have the following direct sum decomposition:

$$\mathrm{Span}(\mathcal{C}^+) = \mathrm{Span}(\mathcal{B}) \oplus \mathrm{Ker}\, M_{\mathcal{C}^+}.$$

Let π denote the projection from $\mathrm{Span}(\mathcal{C}^+)$ on $\mathrm{Span}(\mathcal{B})$ along $\mathrm{Ker}\, M_{\mathcal{C}^+}$, so that $f(p) := p - \pi(p) \in \mathrm{Ker}\, M_{\mathcal{C}^+}$ for all $p \in \mathrm{Span}(\mathcal{C}^+)$. Then the set

$$F := \{f(m) = m - \pi(m) \mid m \in \partial\mathcal{B}\} \subseteq \mathrm{Ker}\, M_{\mathcal{C}^+}$$

is a rewriting family for $\mathcal{B}$, i.e. all polynomials can be expressed modulo the ideal (F) in the linear span of $\mathcal{B}$ (recall the definition from Section 2.5). Thus,

$$\dim \mathbb{R}[\mathbf{x}]/(F) \leq |\mathcal{B}|. \tag{5.6}$$

As in (2.17), define for each $i = 1, \ldots, n$ the linear operator

$$\begin{array}{rccc} \chi_i : & \mathrm{Span}(\mathcal{B}) & \rightarrow & \mathrm{Span}(\mathcal{B}) \\ & p & \mapsto & \pi(\mathbf{x}_i p) \end{array}$$

which can be viewed as an 'abstract' multiplication operator. First we show that $\chi_1, \ldots, \chi_n$ commute pairwise.

Lemma 5.26. $\chi_i \circ \chi_j = \chi_j \circ \chi_i$.

Proof. Let $m \in \mathcal{B}$. Write $\pi(\mathbf{x}_i m) := \sum_{b\in\mathcal{B}} \lambda_b^i b$ ($\lambda_b^i \in \mathbb{R}$). We have:

$$\begin{array}{rl} \chi_j \circ \chi_i(m) & = \chi_j(\sum_{b\in\mathcal{B}} \lambda_b^i b) = \sum_{b\in\mathcal{B}} \lambda_b^i \chi_j(b) \\ & = \sum_{b\in\mathcal{B}} \lambda_b^i (\mathbf{x}_j b - f(\mathbf{x}_j b)) \\ & = \mathbf{x}_j(\sum_{b\in\mathcal{B}} \lambda_b^i b) - \sum_{b\in\mathcal{B}} \lambda_b^i f(\mathbf{x}_j b) \\ & = \mathbf{x}_j(\mathbf{x}_i m - f(\mathbf{x}_i m)) - \sum_{b\in\mathcal{B}} \lambda_b^i f(\mathbf{x}_j b). \end{array}$$

Therefore,

$$\begin{array}{rl} p & := \chi_j \circ \chi_i(m) - \chi_i \circ \chi_j(m) \\ & = \underbrace{\mathbf{x}_i f(\mathbf{x}_j m) - \mathbf{x}_j f(\mathbf{x}_i m)}_{p_1} + \underbrace{\sum_{b\in\mathcal{B}} \lambda_b^j f(\mathbf{x}_j b) - \lambda_b^i f(\mathbf{x}_i b)}_{p_2}. \end{array}$$

We show that $p_1 \in \operatorname{Ker} M_{\mathcal{C}^+}$. Indeed, for any $a \in \mathcal{C}$, $L(ap_1) = L(a\mathbf{x}_i f(\mathbf{x}_j m)) - L(a\mathbf{x}_j f(\mathbf{x}_i m)) = 0$ since $a\mathbf{x}_i, a\mathbf{x}_j \in \mathcal{C}^+$ and $f(\mathbf{x}_i m)$, $f(\mathbf{x}_j m) \in \operatorname{Ker} M_{\mathcal{C}^+}$; in view of (5.5), this shows $p_1 \in \operatorname{Ker} M_{\mathcal{C}^+}$. As $p_2 \in \operatorname{Ker} M_{\mathcal{C}^+}$ too, this implies $p \in \operatorname{Ker} M_{\mathcal{C}^+}$ and thus $p = 0$, because $p \in \operatorname{Span}(\mathcal{B})$. □

If the set $\mathcal{B}$ is connected to 1, then we can conclude using Theorem 2.16 that $\mathcal{B}$ is a linear basis of $\mathbb{R}[\mathbf{x}]/(F)$. Thus we have the direct sum decomposition: $\mathbb{R}[\mathbf{x}] = \operatorname{Span}(\mathcal{B}) \oplus (F)$. Letting $\tilde{\pi}$ denote the projection onto $\operatorname{Span}(\mathcal{B})$ along (F), we can define $\tilde{L} \in \mathbb{R}[\mathbf{x}]^*$ by $\tilde{L}(p) = L(\tilde{\pi}(p))$ for all $p \in \mathbb{R}[\mathbf{x}]$. Remains to verify that $\tilde{L}$ is the desired extension of L.

We now proceed with the proof without the assumption that $\mathcal{B}$ is connected to 1. We cannot use Theorem 2.16, but we will use some ideas of its proof. Namely, as the χ_i's commute, we can define $f(\chi) := f(\chi_1, \ldots, \chi_n)$ for any polynomial f, and thus the mapping

$$\begin{array}{rccc} \varphi: & \mathbb{R}[\mathbf{x}] & \rightarrow & \operatorname{Span}(\mathcal{B}) \\ & f & \mapsto & f(\chi)(1) \end{array}$$

(recall that $1 \in \mathcal{B}$). Note that $\varphi(fg) = f(\chi)(\varphi(g))$ for any $f, g \in \mathbb{R}[\mathbf{x}]$. From this follows that $\operatorname{Ker}\varphi$ is an ideal in $\mathbb{R}[\mathbf{x}]$.

LEMMA 5.27. *$\varphi = \pi$ on Span($\mathcal{C}^+$).*

Proof. We show that $\varphi(m) = \pi(m)$ for all $m \in \mathcal{C}^+$ by induction on the degree of m. If $m = 1$ the result is obvious. Let $m \in \mathcal{C}^+$. As $\mathcal{C}$ is connected to 1, we can write $m = \mathbf{x}_i m_1$. Thus $\varphi(m_1) = \pi(m_1)$ by the induction assumption. Then, $\varphi(m) = \chi_i(\varphi(m_1)) = \chi_i(\pi(m_1)) = \pi(\mathbf{x}_i \pi(m_1))$ is thus equal to $\mathbf{x}_i \pi(m_1) + \kappa$ for some $\kappa \in \operatorname{Ker} M_{\mathcal{C}^+}$. On the other hand, $m = \mathbf{x}_i m_1 = \mathbf{x}_i(\pi(m_1) + \kappa_1)$ for some $\kappa_1 \in \operatorname{Ker} M_{\mathcal{C}^+}$. Thus, $m = \varphi(m) - \kappa + \mathbf{x}_i \kappa_1$. Hence, $\mathbf{x}_i \kappa_1 \in \operatorname{Span}(\mathcal{C}^+)$. As $\kappa_1 \in \operatorname{Ker} M_{\mathcal{C}^+}$, this implies $\mathbf{x}_i \kappa_1 \in \operatorname{Ker} M_{\mathcal{C}^+}(y)$ and thus $\varphi(m) = \pi(m)$. □

This implies directly:

$$L(pq) = L(p\ \varphi(q)) \quad \text{for all } p, q \in \operatorname{Span}(\mathcal{C}^+). \tag{5.7}$$

Moreover, $\varphi(b) = b$ for all $b \in \mathcal{B}$, so that φ is onto. Hence φ is the projection onto $\operatorname{Span}(\mathcal{B})$ and we have the direct sum decomposition:

$$\mathbb{R}[\mathbf{x}] = \operatorname{Span}(\mathcal{B}) \oplus \operatorname{Ker}\varphi.$$

We now define $\tilde{L} \in \mathbb{R}[\mathbf{x}]^*$ by

$$\tilde{L}(f) = L(\varphi(f)) \quad \text{for all } f \in \mathbb{R}[\mathbf{x}].$$

Note that $\operatorname{Ker}\varphi \subseteq \operatorname{Ker} M(\tilde{L})$. Indeed, if $p \in \operatorname{Ker}\varphi$ and $q \in \mathbb{R}[\mathbf{x}]$, $\tilde{L}(pq) = L(\varphi(pq)) = L(q(\chi)(\varphi(p))) = 0$. We now verify that $\tilde{L}$ is the desired flat extension of L. The next lemma implies that $\tilde{L}$ is an extension of L.

Lemma 5.28. *$L(pq) = L(\varphi(pq))$ for all $p, q \in$ Span$(\mathcal{C}^+)$.*

Proof. First we show that $L(mb) = L(\varphi(mb))$ for all $b \in \mathcal{B}$, $m \in \mathcal{C}^+$, by induction on the degree of m. The result is obvious if $m = 1$. Else, write $m = \mathbf{x}_i m_1$ where $m_1 \in \mathcal{C}^+$. By the induction assumption, we know that $L(m_1 q) = L(\varphi(m_1 q))$ for all $q \in \mathrm{Span}(\mathcal{B})$. Let $b \in \mathcal{B}$. We have $L(mb) = L(m_1 \mathbf{x}_i b) = L(m_1 \varphi(\mathbf{x}_i b))$ (by 5.7), which in turn is equal to $L(\varphi(m_1 \varphi(\mathbf{x}_i b)))$ (using the induction assumption), and thus to $L(\varphi(mb))$, since $\varphi(\mathbf{x}_i b) = \chi_i(b)$ giving $\varphi(m_1 \varphi(\mathbf{x}_i b)) = m_1(\chi)(\varphi(\mathbf{x}_i b)) = m_1(\chi)(\chi_i(b)) = m(\chi)(b) = \varphi(mb)$. Finally, for $p, q \in \mathrm{Span}(\mathcal{C}^+)$, $L(pq) = L(p\varphi(q))$ (by 5.7), which in turn is equal to $L(\varphi(p\varphi(q)))$ (by the above) and thus to $L(p(\chi)(\varphi(q))) = L(\varphi(pq))$, which concludes the proof. ◻

Thus $\tilde{L}$ is an extension of L. Indeed, for $p, q \in \mathrm{Span}(\mathcal{C}^+)$, $\tilde{L}(pq) = L(\varphi(pq))$ is equal to $L(pq)$ by Lemma 5.28. Moreover, $M(\tilde{L})$ is a *flat* extension of $M_{\mathcal{C}^+}(L)$. Indeed, on the one hand, we have $\mathrm{rank}\, M(\tilde{L}) \geq \mathrm{rank}\, M_{\mathcal{C}^+}(L) = |\mathcal{B}|$. On the other hand, as $(F) \subseteq (\mathrm{Ker}\, M_{\mathcal{C}^+}) \subseteq \mathrm{Ker}\, \varphi \subseteq \mathrm{Ker}\, M(\tilde{L})$, we have

$$\begin{aligned}\mathrm{rank}\, M(\tilde{L}) = \dim \mathbb{R}[\mathbf{x}]/\,\mathrm{Ker}\, M(\tilde{L}) \leq \dim \mathbb{R}[\mathbf{x}]/\,\mathrm{Ker}\, \varphi \\ \leq \dim \mathbb{R}[\mathbf{x}]/(\mathrm{Ker}\, M_{\mathcal{C}^+}) \leq \dim \mathbb{R}[\mathbf{x}]/(F) \leq |\mathcal{B}|\end{aligned}$$

(recall (5.6)). Thus equality holds throughout, $\mathrm{rank}\, M(\tilde{L}) = \mathrm{rank}\, M_{\mathcal{C}^+}(L)$, and $\mathrm{Ker}\, M(\tilde{L}) = \mathrm{Ker}\, \varphi = (\mathrm{Ker}\, M_{\mathcal{C}^+})$.

Finally, suppose that $L' \in \mathbb{R}[\mathbf{x}]^*$ is another linear form whose moment matrix M' is a flat extension of $M_{\mathcal{C}^+}$, then $\mathrm{Ker}\, \varphi = (\mathrm{Ker}\, M_{\mathcal{C}^+}) \subseteq \mathrm{Ker}\, M'$. This implies that $L'(p) = L'(\varphi(p)) = L(\varphi(p)) = \tilde{L}(p)$ for all $p \in \mathbb{R}[\mathbf{x}]$. This shows the unicity of the flat extension of $M_{\mathcal{C}^+}$, which concludes the proof of Theorem 5.24.

5.4. Flat extensions and representing measures. We group here several results about the truncated moment problem. The first result from Theorem 5.29 essentially follows from the flat extension theorem (Theorem 5.20) combined with Theorem 5.1 about finite rank (infinite) moment matrices. This result is in fact the main ingredient that will be used for the extraction procedure of global minimizers in the polynomial optimization problem (see Section 6.7).

Theorem 5.29. *Let $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$ for which $M_t(y) \succeq 0$ and $\mathrm{rank}\, M_t(y) = \mathrm{rank}\, M_{t-1}(y)$. Then y can be extended to a (unique) vector $\tilde{y} \in \mathbb{R}^{\mathbb{N}^n}$ satisfying $M(\tilde{y}) \succeq 0$, $\mathrm{rank}\, M(\tilde{y}) = \mathrm{rank}\, M_t(y)$, and $\mathrm{Ker}\, M(\tilde{y}) = (\mathrm{Ker}\, M_t(y))$, the ideal generated by $\mathrm{Ker}\, M_t(y)$. Moreover, any set $\mathcal{B} \subseteq \mathbb{T}^n_{t-1}$ indexing a maximum nonsingular principal submatrix of $M_{t-1}(y)$ is a basis of $\mathbb{R}[\mathbf{x}]/(\mathrm{Ker}\, M_t(y))$. Finally, $\tilde{y}$, and thus y, has a (unique) representing measure μ, which is r-atomic with $\mathrm{supp}(\mu) = V_{\mathbb{C}}(\mathrm{Ker}\, M_t(y))$.*

Proof. Applying iteratively Theorem 5.20 we find an extension $\tilde{y} \in \mathbb{R}^{\mathbb{N}^n}$ of y for which $M(\tilde{y})$ is a flat extension of $M_t(y)$; thus $\mathrm{rank}\, M(\tilde{y}) =$

$\operatorname{rank} M_t(y) =: r$ and $M(\tilde{y}) \succeq 0$. By Theorem 5.1, $\tilde{y}$ has a (unique) representing measure μ, which is r-atomic and satisfies $\operatorname{supp}(\mu) = V_{\mathbb{C}}(\operatorname{Ker} M(\tilde{y}))$. To conclude the proof, it suffices to verify that $(\operatorname{Ker} M_t(y)) = \operatorname{Ker} M(\tilde{y})$, as this implies directly $\operatorname{supp}(\mu) = V_{\mathbb{C}}(\operatorname{Ker} M(\tilde{y})) = V_{\mathbb{C}}(\operatorname{Ker} M_t(y))$. Obviously, $\operatorname{Ker} M_t(y) \subseteq \operatorname{Ker} M(\tilde{y})$, implying $(\operatorname{Ker} M_t(y)) \subseteq \operatorname{Ker} M(\tilde{y})$. We now show the reverse inclusion. Let $\mathcal{B} \subseteq \mathbb{T}^n_{t-1}$ index a maximum nonsingular principal submatrix of $M_{t-1}(y)$. Thus $|\mathcal{B}| = r$ and $\mathcal{B}$ also indexes a maximum nonsingular principal submatrix of $M(\tilde{y})$. Hence, by Lemma 5.3, $\mathcal{B}$ is a basis of $\mathbb{R}[\mathbf{x}]/\operatorname{Ker} M(\tilde{y})$. We show that $\mathcal{B}$ is a generating set in $\mathbb{R}[\mathbf{x}]/(\operatorname{Ker} M_t(y))$; that is, for all $\beta \in \mathbb{N}^n$,

$$\mathbf{x}^\beta \in \operatorname{Span}_{\mathbb{R}}(\mathcal{B}) + (\operatorname{Ker} M_t(y)). \tag{5.8}$$

We prove (5.8) using induction on $|\beta|$. If $|\beta| \leq t$, (5.8) holds since $\mathcal{B}$ indexes a basis of the column space of $M_t(y)$. Assume $|\beta| \geq t+1$. Write $\mathbf{x}^\beta = \mathbf{x}_i \mathbf{x}^\gamma$ where $|\gamma| = |\beta| - 1$. By the induction assumption, $\mathbf{x}^\gamma = \sum_{\mathbf{x}^\alpha \in \mathcal{B}} \lambda_\alpha \mathbf{x}^\alpha + q$, where $\lambda_\alpha \in \mathbb{R}$ and $q \in (\operatorname{Ker} M_t(y))$. Then, $\mathbf{x}^\beta = \mathbf{x}_i \mathbf{x}^\gamma = \sum_{\mathbf{x}^\alpha \in \mathcal{B}} \lambda_\alpha \mathbf{x}_i \mathbf{x}^\alpha + \mathbf{x}_i q$. Obviously, $\mathbf{x}_i q \in (\operatorname{Ker} M_t(y))$. For $\mathbf{x}^\alpha \in \mathcal{B}$, $\deg(\mathbf{x}_i \mathbf{x}^\alpha) \leq t$ and, therefore, $\mathbf{x}_i \mathbf{x}^\alpha \in \operatorname{Span}_{\mathbb{R}}(\mathcal{B}) + (\operatorname{Ker} M_t(y))$. From this follows that $\mathbf{x}^\beta \in \operatorname{Span}_{\mathbb{R}}(\mathcal{B}) + (\operatorname{Ker} M_t(y))$. Thus (5.8) holds for all $\beta \in \mathbb{N}^n$. Take $p \in \operatorname{Ker} M(\tilde{y})$. In view of (5.8), we can write $p = p_0 + q$, where $p_0 \in \operatorname{Span}_{\mathbb{R}}(\mathcal{B})$ and $q \in (\operatorname{Ker} M_t(y))$. Hence, $p - q \in \operatorname{Ker} M(\tilde{y}) \cap \operatorname{Span}_{\mathbb{R}}(\mathcal{B})$, which implies $p - q = 0$, since $\mathcal{B}$ is a basis of $\mathbb{R}[\mathbf{x}]/\operatorname{Ker} M(\tilde{y})$. Therefore, $p = q \in (\operatorname{Ker} M_t(y))$, which concludes the proof for equality $\operatorname{Ker} M(\tilde{y}) = (\operatorname{Ker} M_t(y))$. □

We now give several results characterizing existence of a finite atomic measure for truncated sequences. By Lemma 4.2 (i), a necessary condition for the existence of a finite atomic reprenting measure μ for a sequence $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$ is that its moment matrix $M_t(y)$ has rank at most $|\operatorname{supp}(\mu)|$. Theorem 5.30 below gives a characterization for the existence of a *minimum* atomic measure, i.e., satisfying $|\operatorname{supp}(\mu)| = \operatorname{rank}\ M_t(y)$. Then Theorem 5.31 deals with the general case of existence of a finite atomic representing measure and Theorems 5.33 and 5.34 give the analogous results for existence of a measure supported by a prescribed semialgebraic set. In these results, the notion of *flat extension* studied in the preceding section plays a central role.

THEOREM 5.30. [28] *The following assertions are equivalent for $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$.*

(i) *y has a $(\operatorname{rank} M_t(y))$-atomic representing measure.*

(ii) *$M_t(y) \succeq 0$ and one can extend y to a vector $\tilde{y} \in \mathbb{R}^{\mathbb{N}^n_{2t+2}}$ in such a way that $M_{t+1}(\tilde{y})$ is a flat extension of $M_t(y)$.*

Proof. Directly from Theorems 5.1 and 5.20. □

THEOREM 5.31. [29, 41] *Let $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$, $r := \operatorname{rank} M_t(y)$ and $v := |V_{\mathbb{R}}(\operatorname{Ker} M_t(y))| \leq \infty$. Consider the following assertions:*

(i) *y has a representing measure.*

(ii) *y has a $\binom{n+2t}{2t}$-atomic representing measure.*
(iii) *$M_t(y) \succeq 0$ and there exists an integer $k \geq 0$ for which y can be extended to a vector $\tilde{y} \in \mathbb{R}^{\mathbb{N}^n_{2(t+k+1)}}$ in such a way that $M_{t+k}(\tilde{y}) \succeq 0$ and $M_{t+k+1}(\tilde{y})$ is a flat extension of $M_{t+k}(\tilde{y})$.*
(iv) *$r \leq v < \infty$ and y can be extended to a vector $\tilde{y} \in \mathbb{R}^{\mathbb{N}^n_{2(t+v-r+1)}}$ in such a way that $M_{t+v-r+1}(\tilde{y}) \succeq 0$ and $\operatorname{rank} M_{t+v-r+1}(\tilde{y}) \leq |V_{\mathbb{R}}(\operatorname{Ker} M_{t+v-r+1}(\tilde{y}))|$.*

Then, (i) $\Longleftrightarrow$ (ii) $\Longleftrightarrow$ (iii) and, when $v < \infty$, (i) $\Longleftrightarrow$ (iv). Moreover one can assume in (iii) that $k \leq \binom{n+2t}{2t} - r$ and, when $v < \infty$, that $k \leq v - r$.

Proof. The equivalence of (i) and (ii) follows from Theorem 5.9 and the implication (iii) $\Longrightarrow$ (i) follows from Theorem 5.29.

Assume now that (ii) holds; i.e., y has a finite atomic representing measure μ with $|\operatorname{supp}(\mu)| \leq \binom{n+2t}{2t}$. First we show that (iii) holds with $k \leq \binom{n+2t}{2t} - r$. Indeed, y can be extended to the sequence $\tilde{y} \in \mathbb{R}^{\mathbb{N}^n}$ consisting of all the moments of the measure μ. Then $M(\tilde{y}) \succeq 0$ and rank $M(\tilde{y}) = |\operatorname{supp}(\mu)| = |V_{\mathbb{R}}(\operatorname{Ker} M(\tilde{y}))|$ (by Theorem 5.1 (ii)). As $\operatorname{rank} M_{t+k}(\tilde{y}) \leq \operatorname{rank} M_{t+k+1}(\tilde{y}) \leq \operatorname{rank} M(\tilde{y}) \leq \binom{n+2t}{2t}$ for all $k \in \mathbb{N}$, there exists a smallest integer k for which $\operatorname{rank} M_{t+k}(\tilde{y}) = \operatorname{rank} M_{t+k+1}(\tilde{y})$. Then, $\operatorname{rank} M_{t+k}(\tilde{y}) > \operatorname{rank} M_{t+k-1}(\tilde{y}) > \ldots > \operatorname{rank} M_t(\tilde{y}) = \operatorname{rank} M_t(y) = r$, which implies $\operatorname{rank} M_{t+k}(\tilde{y}) \geq r + k$ and thus $k \leq \binom{n+2t}{2t} - r$. This concludes the proof of equivalence of (i), (ii) and (iii).

Assume now that (ii) holds and $v := |V_{\mathbb{R}}(\operatorname{Ker} M_t(y))| < \infty$; we show that (iv) holds and $k \leq v - r$. We use the fact that, for any $s \geq t$, $\operatorname{Ker} M_t(y) \subseteq \operatorname{Ker} M_s(\tilde{y}) \subseteq \operatorname{Ker} M(\tilde{y})$, which implies $|V_{\mathbb{R}}(\operatorname{Ker} M(\tilde{y}))| \leq |V_{\mathbb{R}}(\operatorname{Ker} M_s(\tilde{y}))| \leq |V_{\mathbb{R}}(\operatorname{Ker} M_t(y))|$. By the above we have: $r + k \leq \operatorname{rank} M_{t+k}(\tilde{y}) \leq \operatorname{rank} M(\tilde{y}) = |V_{\mathbb{R}}(\operatorname{Ker} M(\tilde{y}))| \leq |V_{\mathbb{R}}(\operatorname{Ker} M_t(y))| = v$, giving $k \leq v - r$ (and thus $r \leq v$). Moreover, $\operatorname{rank} M_{t+v-r+1}(\tilde{y}) \leq \operatorname{rank} M(\tilde{y}) \leq |V_{\mathbb{R}}(\operatorname{Ker} M_{t+v-r+1}(\tilde{y}))|$, showing (iv).

Finally assume (iv) holds; we show (iii). Indeed we have $r = \operatorname{rank} M_t(y) \leq \operatorname{rank} M_{t+v-r+1}(\tilde{y}) \leq |V_{\mathbb{R}}(\operatorname{Ker} M_{t+v-r+1}(\tilde{y}))| \leq |V_{\mathbb{R}}(\operatorname{Ker} M_t(y))| = v$. Hence, there exists $k \in \{0, \ldots, v - r\}$ for which $\operatorname{rank} M_{t+k+1}(\tilde{y}) = \operatorname{rank} M_{t+k}(\tilde{y})$ for, if not, we would have $\operatorname{rank} M_{t+v-r+1}(\tilde{y}) \geq \operatorname{rank} M_t(y) + v - r + 1 = v + 1$, contradicting $\operatorname{rank} M_{t+v-r+1}(\tilde{y}) \leq v$. Thus (iii) holds, moreover with $k \leq v - r$. □

Remark 5.32. *Theorem 5.31 provides conditions characterizing the existence of a representing measure for a truncated sequence. It is however not clear how to check these conditions and the smallest integer k for which (iii) holds as well as the gap $v - r$ may be large. We refer to Fialkow [41] for a detailed treatment of such issues.*

Let us observe here that in some instances the bound $v - r$ is better than the bound $\binom{n+2t}{2t} - r$. For instance, as observed in [41], in the 2-dimensional case ($n = 2$), $v \leq t^2$ by Bezout theorem, implying $\binom{2t+2}{2t} - v \geq \binom{2t+2}{2t} - t^2 = t^2 + 3t + 1$. Moreover, Fialkow [41] constructs an instance with large gap $v - r \geq \binom{t-1}{2}$. For this choose two polynomi-

als $p, q \in \mathbb{R}[x_1, x_2]_t$ having t^2 common zeros in $\mathbb{R}^2$, i.e., $|V_{\mathbb{R}}(p,q)| = t^2$. Let μ be a measure on $\mathbb{R}^2$ with support $V_{\mathbb{R}}(p,q)$ and let y be its sequence of moments. Then, $V_{\mathbb{R}}(\operatorname{Ker} M_t(y)) = V_{\mathbb{R}}(p,q)$ and thus $v = t^2$. Indeed, $t^2 = |\operatorname{supp}(\mu)| \leq |V_{\mathbb{R}}(\operatorname{Ker} M_t(y))|$ and $V_{\mathbb{R}}(\operatorname{Ker} M_t(y)) \subseteq V_{\mathbb{R}}(p,q)$ since $p, q \in \operatorname{Ker} M_t(y)$. Moreover, $r = \operatorname{rank} M_t(y) \leq |\mathbb{N}^2_t| - 2 = \binom{t+2}{2} - 2$ which implies $v - r \geq t^2 - \binom{t+2}{2} + 2 = \binom{t-1}{2}$.

The next two theorems (from Curto and Fialkow [30]) extend the results from Theorems 5.30 and 5.31 to truncated sequences having a finite atomic representing measure whose support is contained in a prescribed semialgebraic set K. As indicated in [95], they can be derived easily from Theorems 5.30 and 5.31 using Lemma 5.6. In what follows K is as in (1.2) and d_K as in (1.10). One may assume w.l.o.g. that the polynomials g_j defining K are not constant; thus $d_{g_j} \geq 1$. For convenience we set $d_K := 1$ if $m = 0$, i.e., if there are no constraints defining the set K, in which case $K = \mathbb{R}^n$.

Theorem 5.33. [30] *Let K be the set from (1.2) and $d_K = \max_{j=1,\ldots,m} d_{g_j}$. The following assertions are equivalent for $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$.*

(i) *y has a* (rank $M_t(y)$)*-atomic representing measure μ whose support is contained in K.*

(ii) *$M_t(y) \succeq 0$ and y can be extended to a vector $\tilde{y} \in \mathbb{R}^{\mathbb{N}^n_{2(t+d_K)}}$ in such a way that $M_{t+d_K}(\tilde{y})$ is a flat extension of $M_t(y)$ and $M_t(g_j\tilde{y}) \succeq 0$ for $j = 1, \ldots, m$.*

Then, setting $r_j :=$ rank $M_t(g_j\tilde{y})$, exactly $r - r_j$ of the atoms in the support of μ belong to the set of roots of the polynomial $g_j(x)$. Moreover μ is a representing measure for $\tilde{y}$.

Proof. The implication (i) $\Longrightarrow$ (ii) follows from Theorem 5.1 together with Lemma 4.2 (ii). Conversely, assume that (ii) holds and set $r := \operatorname{rank} M_t(y)$. By Theorem 5.30 ((ii) $\Longrightarrow$ (i)), y has a r-atomic representing measure μ; say, $\mu = \sum_{v \in S} \lambda_v \delta_v$ where $\lambda_v > 0$, $|S| = r$. We prove that $S \subseteq K$; that is, $g_j(v) \geq 0$ for all $v \in S$. By Lemma 5.6, there exist interpolation polynomials p_v ($v \in S$) having degree at most t. Then, $p_v^T M_t(g_j y) p_v = \sum_{u \in S} (p_v(u))^2 g_j(u) \lambda_u = g_j(v)\lambda_v \geq 0$, since $M_t(g_j y) \succeq 0$. This implies that $g_j(v) \geq 0$ for all $j = 1, \ldots, m$ and $v \in S$, and thus $S \subseteq K$. That is, the measure μ is supported by the set K.

We now verify that $r - r_j$ of the points of S are zeros of the polynomial g_j. Denote by $\tilde{y} \in \mathbb{R}^{\mathbb{N}^n}$ the (infinite) sequence of moments of the measure μ; then $g_j\tilde{y}$ is the (infinite) sequence of moments of the measure $\mu_j := \sum_{v \in S} \lambda_v g_j(v) \delta_v$. Thus, $\tilde{y}$ (resp., $g_j\tilde{y}$) is an extension of y (resp., $g_j y$). Moreover, rank $M(g_j\tilde{y}) = |\{v \in S \mid g_j(v) > 0\}|$. We now verify that $M(g_j\tilde{y})$ is a flat extension of $M_t(g_j\tilde{y})$, which implies that $r_j = |\{v \in S \mid g_j(v) > 0\}|$, giving the desired result. For this we note that $\operatorname{Ker} M(\tilde{y}) \subseteq \operatorname{Ker} M(g_j\tilde{y})$. Indeed, if $p \in \operatorname{Ker} M(\tilde{y})$ then, using Lemma 4.1, $p^T M(g_j\tilde{y}) p = \operatorname{vec}(pg_j)^T M(\tilde{y}) p = 0$. Now, as $M(\tilde{y})$ is a flat extension of $M_t(y)$, it follows that $M(g_j\tilde{y})$ is a flat extension of $M_t(g_j\tilde{y})$ too. □

Theorem 5.34. [30] *Let K be the set from (1.2) and $d_K = \max_{j=1,\ldots,m} d_{g_j}$. The following assertions are equivalent for $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$.*

(i) *y has a (finite atomic) representing measure whose support is contained in K.*

(ii) *$M_t(y) \succeq 0$ and there exists an integer $k \geq 0$ for which y can be extended to a vector $\tilde{y} \in \mathbb{R}^{\mathbb{N}^n_{2(t+k+d_K)}}$ in such a way that $M_{t+k+d_K}(\tilde{y}) \succeq 0$, $M_{t+k+d_K}(\tilde{y})$ is a flat extension of $M_{t+k}(\tilde{y})$, and $M_{t+k}(g_j\tilde{y}) \succeq 0$ for $j = 1, \ldots, m$.*

Proof. Analogously using Theorems 5.31 and 5.33. □

5.5. The truncated moment problem in the univariate case. We review here some of the main results about the truncated moment problem in the univariate case. Namely we give characterizations for the truncated sequences $y \in \mathbb{R}^{\mathbb{N}^1_t}$ having a representing measure on an interval $K = [0, \infty)$ and $K = [-1, 1]$. We follow the treatment in [27] but simplify some arguments. We need a definition.

Given a sequence $y \in \mathbb{R}^{\mathbb{N}^1_{2t}}$, consider its moment matrix $M_t(y)$, indexed by $\{1, \mathbf{x}, \ldots, \mathbf{x}^t\}$. Then, $\text{rk}(y)$ is defined as the smallest integer r for which the column of $M_t(y)$ indexed by $\mathbf{x}^r$ lies in the linear span of the columns indexed by $\{1, \mathbf{x}, \ldots, \mathbf{x}^{r-1}\}$ if such integer exists, and $\text{rk}(y) := t + 1$ otherwise. When $M_t(y) \succeq 0$, $\text{rk}(y)$ is also equal to the smallest integer r for which $M_r(y)$ is singular, with $\text{rk}(y) = t + 1$ if $M_t(y) \succ 0$ (easy verification using Lemma 1.2 (iii)).

Assume $M_t(y) \succeq 0$ and let $r := \text{rk}(y)$. If $r = t + 1$, then $M_t(y) \succ 0$ and $\text{rank}\, M_t(y) = t + 1 = r$. Assume now $r \leq t$; that is, $M_{r-1}(y) \succ 0$ and $\text{rank}\, M_r(y) = \text{rank}\, M_{r-1}(y) = r$. We now observe that $\text{rank}\, M_t(y) = r$ or $r + 1$. Indeed there exists a unique polynomial f of degree r, belonging to the kernel of $M_t(y)$. By the 'ideal-like' property of $\text{Ker}\, M_t(y)$ (Lemma 5.7), $\mathbf{x}^p f \in \text{Ker}\, M_t(y)$ for all $p \leq t - r - 1$, and thus $\text{rank}\, M_{t-1}(y) = \text{rank}\, M_{r-1}(y) = r$. Hence there are two cases:

(a) Either $\mathbf{x}^{t-r} f \in \text{Ker}\, M_t(y)$; then $\text{rank}\, M_t(y) = r$ (i.e. $M_t(y)$ is a flat extension of $M_{r-1}(y)$).

(b) Or $\mathbf{x}^{t-r} f \notin \text{Ker}\, M_t(y)$; then $\text{rank}\, M_t(y) = r + 1$.

The next lemma gives more details.

Lemma 5.35. *Given $y = (y_i)_{i=0}^{2t}$, let $r := rk(y)$ and assume $M_t(y) \succeq 0$. The following assertions are equivalent.*

(i) $\text{rank}\, M_t(y) = r$.

(ii) There exists scalars $y_{2t+1}, y_{2t+2} \in \mathbb{R}$ for which the extended sequence $\tilde{y}^T := (y^T, y_{2t+1}, y_{2t+2})$ satisfies $M_{t+1}(\tilde{y})$ is a flat extension of $M_t(y)$.

(iii) There exists scalars $y_{2t+1}, y_{2t+2} \in \mathbb{R}$ for which the extended sequence $\tilde{y}^T := (y^T, y_{2t+1}, y_{2t+2})$ satisfies $M_{t+1}(\tilde{y}) \succeq 0$.

Proof. (ii) $\Longrightarrow$ (iii) is obvious. We now show (iii) $\Longrightarrow$ (i). Assume (iii) holds. If $M_t(y) \succ 0$, then $\text{rank}\, M_t(y) = t + 1 = r$ and (i) holds. Otherwise, let f be the unique polynomial of degree r in $\text{Ker}\, M_t(y)$ (as defined above). Lemma 5.7 applied to $M_{t+1}(\tilde{y})$ implies $\mathbf{x}^{t-r} f \in \text{Ker}\, M_{t+1}(\tilde{y})$ and thus

$\mathbf{x}^{t-r}f \in \operatorname{Ker} M_t(y)$. Hence we are in the above case (a), showing (i).
Remains to show (i) $\Longrightarrow$ (ii). When $M_t(y) \succ 0$, set $y_{2t+1} := 0$, $v^T := (y_{t+1}, \dots, y_{2t}, y_{2t+1})$, $\lambda := (M_t(y))^{-1}v$, and $y_{2t+2} := v^T\lambda$. Then $M_{t+1}(\tilde{y})$ is a flat extension of $M_t(y)$, showing (ii). Assume now $\operatorname{rank} M_t(y) = r \le t$. Then $M_t(y)$ is a flat extension of $M_{t-1}(y)$ and thus the existence of a flat extension $M_{t+1}(\tilde{y})$ of $M_t(y)$ follows from the flat extension theorem (Theorem 5.20). We now give the explicit argument, which is simple in the univariate case.

For an integer $0 \le i \le t$, let $v[i]$ denote the column of $M_t(y)$ indexed by $\mathbf{x}^i$, and set $v[t+1]^T := (y_{t+1}, \dots, y_{2t}, y_{2t+1})$ where y_{2t+1} has to be determined. Thus $v[i]_j = y_{i+j}$. Write the unique polynomial $f \in \operatorname{Ker} M_t(y)$ of degree r as $f = \mathbf{x}^r - \sum_{j=0}^{r-1} \lambda_j \mathbf{x}^j$. By assumption,

$$v[r+p] = \sum_{j=0}^{r-1} \lambda_j v[j+p] \quad \forall p = 0, 1, \dots, t-r. \tag{5.9}$$

Set $y_{2t+1} := \sum_{j=0}^{r-1} \lambda_j y_{j-r+2t+1}$. It suffices now to verify that $v[t+1] = \sum_{j=0}^{r-1} \lambda_j v[j+t-r+1]$, i.e. $v[t+1]_s = \sum_{j=0}^{r-1} \lambda_j v[j+t-r+1]_s$ for all $s = 0, \dots, t$. This is true for $s = t$ by the definition of y_{2t+1}. For $s \le t-1$, $v[t+1]_s = v[t]_{s+1} = \sum_{j=0}^{r-1} \lambda_j v[j+t-r]_{s+1} = \sum_{j=0}^{r-1} \lambda_j v[j+t-r+1]_s$, where we have used the case $p = t-r$ of (5.9) for the second equation. □

We now present characterizations for the Hamburger, Stieltjes and Hausdorff truncated moment problems. We will use the results from the preceding sections about flat extensions of moment matrices; namely, the following result from Theorem 5.30: If $M_t(y) \succeq 0$ has a flat extension $M_{t+1}(\tilde{y})$ then y has an atomic representing measure. For a sequence $y = (y_i)_{i=0}^t$ and $0 \le i \le j \le t$, we set $y[i;j]^T = (y_i, y_{i+1}, \dots, y_j)$.

THEOREM 5.36. **(Hamburger truncated moment problem)**

(i) **(odd case)** $y = (y_i)_{i=0}^{2t+1}$ *has a representing measure on* $\mathbb{R} \Longleftrightarrow M_t(y) \succeq 0$ *and* $y[t+1; 2t+1] \in \mathcal{R}(M_t(y))$.

(ii) **(even case)** $y = (y_i)_{i=0}^{2t}$ *has a representing measure on* $\mathbb{R} \Longleftrightarrow M_t(y) \succeq 0$ *and* $\operatorname{rank} M_t(y) = rk(y)$.

Proof. (i) If y has a measure, then y has an extension $\tilde{y}$ for which $M_{t+1}(\tilde{y}) \succeq 0$; the condition $y[t+1; 2t+1] \in \mathcal{R}(M_t(y))$ follows using Lemma 1.2 (iv) applied to $M_{t+1}(\tilde{y})$. Conversely, $y[t+1; 2t+1] \in \mathcal{R}(M_t(y))$ implies $y[t+1; 2t+1] = M_t(y)\lambda$ for some $\lambda \in \mathbb{R}^{t+1}$. Set $y_{2t+2} := y[t+1; 2t+1]^T\lambda$, $\tilde{y}^T := (y^T, y_{2t+2})$; then we obtain a flat extension $M_{t+1}(\tilde{y})$ of $M_t(y)$, thus showing that y has a representing measure.
(ii) follows using Lemma 5.35. □

EXAMPLE 5.37. *As an illustration consider again the sequence* $y = (1, 1, 1, 1, 2) \in \mathbb{R}^{\mathbb{N}_4^1}$ *from Example 5.14. Then* $M_2(y) \succeq 0$ *but* y *has no representing measure, which is consistent with the fact that* $rk(y) = 1 < \operatorname{rank} M_2(y) = 2$. *However, recall from Example 5.18 that a small perturbation of* y *does have a representing measure.*

Theorem 5.38. **(Stieltjes truncated moment problem)**

(i) **(odd case)** $y = (y_i)_{i=0}^{2t+1}$ *has a representing measure on* $\mathbb{R}_+ \iff M_t(y) \succeq 0$, $M_t(\mathbf{x}y) \succeq 0$, *and* $y[t+1; 2t+1] \in \mathcal{R}(M_t(y))$.

(ii) **(even case)** $y = (y_i)_{i=0}^{2t}$ *has a representing measure on* $\mathbb{R}_+ \iff M_t(y) \succeq 0$, $M_{t-1}(\mathbf{x}y) \succeq 0$, *and* $y[t+1; 2t] \in \mathcal{R}(M_{t-1}(\mathbf{x}y))$.

Proof. We only show the implications '$\Longleftarrow$' as the reverse implications can be shown with similar arguments as in the previous theorem.
(i) Let $y = (y_i)_{i=0}^{2t+1}$ such that $M_t(y) \succeq 0$, $M_t(\mathbf{x}y) \succeq 0$, and $y[t+1; 2t+1] \in \mathcal{R}(M_t(y))$. Then, $y[t+1; 2t+1] = M_t(y)\lambda$ for some $\lambda \in \mathbb{R}^{t+1}$. Set $y_{2t+2} := y[t+1; 2t+1]^T\lambda$ and $\tilde{y}^T := (y^T, y_{2t+2})$. Then $M_{t+1}(\tilde{y})$ is a flat extension of $M_t(y)$. Thus $\tilde{y}$ has an atomic measure $\mu = \sum_{i=1}^r a_i\delta_{x_i}$ (with $a_i > 0$ and $x_i \in \mathbb{R}$). Using Lemma 5.6 we can find interpolation polynomials p_{x_i} at the x_i's having degree at most t. The condition $M_t(\mathbf{x}y) \succeq 0$ can then be used to derive $x_i \geq 0$ from $p_{x_i}^T M_t(\mathbf{x}y)p_{x_i} \geq 0$. Thus μ is supported by $\mathbb{R}_+$, showing (i).
(ii) Let $y = (y_i)_{i=0}^{2t}$ such that $M_t(y) \succeq 0$, $M_{t-1}(\mathbf{x}y) \succeq 0$, and $y[t+1; 2t] \in \mathcal{R}(M_{t-1}(\mathbf{x}y))$. As $y[t+1; 2t] \in \mathcal{R}(M_{t-1}(\mathbf{x}y))$, we can find y_{2t+1} for which $M_t(\mathbf{x}\tilde{y})$ is a flat extension of $M_{t-1}(\mathbf{x}y)$, after setting $\tilde{y}^T := (y^T, y_{2t+1})$. As the matrix $Z := \begin{pmatrix} M_{t-1}(\mathbf{x}y) \\ y[t+1; 2t]^T \end{pmatrix}$ coincides with the submatrix of $M_t(y)$ consisting of its columns indexed by $\{\mathbf{x}, \ldots, \mathbf{x}^{t-1}\}$, $\tilde{y}[t+1; 2t+1]$ belongs to $\mathcal{R}(Z) \subseteq \mathcal{R}(M_t(y))$. Together with $M_t(\tilde{y}) = M_t(y) \succeq 0$ and $M_t(\mathbf{x}\tilde{y}) \succeq 0$, we deduce from the preceding case (i) that $\tilde{y}$, and thus y, has a representing measure on $\mathbb{R}_+$. □

Theorem 5.39. **(Hausdorff truncated moment problem)** *Let $a < b$ be real numbers.*

(i) **(odd case)** $y = (y_i)_{i=0}^{2t+1}$ *has a representing measure on* $[a, b] \iff bM_t(y) - M_t(\mathbf{x}y) \succeq 0$ *and* $M_t(\mathbf{x}y) - aM_t(y) \succeq 0$.

(ii) **(even case)** $y = (y_i)_{i=0}^{2t}$ *has a representing measure on* $[a, b] \iff M_t(y) \succeq 0$ *and* $-M_{t-1}(\mathbf{x}^2y) + (a+b)M_{t-1}(\mathbf{x}y) - abM_{t-1}(y) \succeq 0$.

Proof. To show (i) (resp., (ii)), apply Theorem 3.23 (ii) (resp., the even case of (i)) combined with Theorem 5.13.
We now give another direct proof for (i). First we show that $y[t+1; 2t+1] \in \mathcal{R}(M_t(y))$. Indeed,

$$(b-a)M_t(y) = \underbrace{(bM_t(y) - M_t(\mathbf{x}y))}_{Z_1} + \underbrace{(M_t(\mathbf{x}y) - aM_t(y))}_{Z_2},$$

where $Z_1, Z_2 \succeq 0$. This implies $\operatorname{Ker} M_t(y) \subseteq \operatorname{Ker} Z_1$, and thus $\operatorname{Ker} M_t(y) \subseteq \operatorname{Ker} M_t(\mathbf{x}y)$. As $y[t+1; 2t+1]$ is a row of $M_t(\mathbf{x}y)$ (namely, its row indexed by $\mathbf{x}^t$), $y[t+1; 2t+1] \in (\operatorname{Ker} M_t(\mathbf{x}y))^\perp \subseteq (\operatorname{Ker} M_t(y))^\perp = \mathcal{R}(M_t(y))$. The rest of the proof now goes along the same lines as for Theorem 5.38 (i). Namely, define y_{2t+2} and $\tilde{y}^T := (y^T, y_{2t+2})$ such that $M_{t+1}(\tilde{y})$ is a flat extension of $M_t(y)$. Thus $\tilde{y}$ has an atomic representing measure $\mu = \sum_{i=1}^r a_i\delta_{x_i}$ and it suffices to verify that all $x_i \in [a, b]$. This can be verified as in the previous theorem using interpolation polynomials at the x_i's of degree at most t. $\square$

Part 2: Application to Optimization

6. Back to the polynomial optimization problem.

6.1. Hierarchies of relaxations. We consider again the optimization problem (1.1). Following Lasserre [78] and as explained earlier, hierarchies of semidefinite programming relaxations can be constructed for (1.1); namely, the SOS relaxations (3.8) (introduced in Section 3.4), that are based on relaxing polynomial positivity by sums of squares representations, and the moment relaxations (4.7) (introduced in Section 4.2), that are based on relaxing existence of a representing measure by positive semidefiniteness of moment matrices. For convenience we repeat the formulation of the bounds p_t^{sos} from (3.8) and of the bounds p_t^{mom} from (4.7). Recall

$$d_p = \lceil \deg(p)/2 \rceil,\ d_{g_j} = \lceil \deg(g_j)/2 \rceil,\ d_K = \begin{cases} \max(d_{g_1}, \dots, d_{g_m}) \\ 1 \text{ if } m = 0 \end{cases} \tag{6.1}$$

Then, for any integer $t \geq \max(d_p, d_K)$,

$$\begin{aligned} p_t^{\text{sos}} &= \sup \rho \ \text{ s.t. } \ p - \rho \in \mathbf{M}_{2t}(g_1, \dots, g_m) \\ &= \sup \rho \text{ s.t. } \ p - \rho = s_0 + \textstyle\sum_{j=1}^m s_j g_j \ \text{ for some } s_0, s_j \in \Sigma \\ &\qquad\qquad \text{with } \deg(s_0), \deg(s_j g_j) \leq 2t. \end{aligned} \tag{6.2}$$

$$\begin{aligned} p_t^{\text{mom}} &= \inf_{L \in (\mathbb{R}[\mathbf{x}]_{2t})^*} L(p) \ \text{ s.t. } \ L(1) = 1, \\ &\qquad\qquad L(f) \geq 0 \ \forall f \in \mathbf{M}_{2t}(g_1, \dots, g_m) \\ &= \inf_{y \in \mathbb{R}^{\mathbb{N}^n_{2t}}} p^T y \ \text{ s.t. } \ y_0 = 1,\ M_t(y) \succeq 0, \\ &\qquad\qquad M_{t-d_{g_j}}(g_j y) \succeq 0 \ (j = 1, \dots, m). \end{aligned} \tag{6.3}$$

We refer to program (6.2) as the *SOS relaxation* of order t, and to program (6.3) as the *moment relaxation* of order t. The programs (6.2) and (6.3) are semidefinite programs involving matrices of size $\binom{n+t}{t} = O(n^t)$ and $O(n^{2t})$ variables. Hence, for any *fixed* t, p_t^{mom} and p_t^{sos} can be computed in polynomial time (to any precision). In the remaining of Section 6 we study in detail some properties of these bounds. In particular,

(i) ***Duality:*** $p_t^{\text{sos}} \leq p_t^{\text{mom}}$ and, under some condition on the set K, the two bounds p_t^{mom} and p_t^{sos} coincide.

(ii) ***Convergence:*** Under certain conditions on the set K, there is **asymptotic** (sometimes even **finite**) **convergence** of the bounds p_t^{mom} and p_t^{sos} to p^{min}.

(iii) ***Optimality certificate:*** Under some conditions, the relaxations are exact, i.e. $p_t^{\text{sos}} = p_t^{\text{mom}} = p^{\text{min}}$ (or at least $p_t^{\text{mom}} = p^{\text{min}}$).

(iv) ***Finding global minimizers:*** Under some conditions, one is able to extract some global minimizers for the original problem (1.1) from an optimum solution to the moment relaxation (6.3).

6.2. Duality. One can verify that the two programs (6.3) and (6.2) are dual semidefinite programs (cf. [78]), which implies $p_t^{\mathrm{sos}} \leq p_t^{\mathrm{mom}}$ by weak duality; this inequality also follows directly as noted earlier in (4.8). We now give a condition ensuring that strong duality holds, i.e. there is no duality gap between (6.3) and (6.2).

THEOREM 6.1. *[78, 150] If K has a nonempty interior (i.e. there exists a full dimensional ball contained in K), then $p_t^{mom} = p_t^{sos}$ for all $t \geq \max(d_p, d_K)$. Moreover, if (6.2) is feasible then it attains its supremum.*

Proof. We give two arguments. The first argument comes from [150] and relies on Theorem 3.49. Let $\rho > p_t^{\mathrm{sos}}$, i.e. $p - \rho \notin \mathbf{M}_{2t}(g_1, \ldots, g_m)$. As $\mathbf{M}_{2t}(g_1, \ldots, g_m)$ is a closed convex cone (by Theorem 3.49), there exists a hyperplane strictly separating $p - \rho$ from $\mathbf{M}_{2t}(g_1, \ldots, g_m)$; that is, there exists $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$ with

$$y^T \mathrm{vec}(p - \rho) < 0 \quad \text{and } y^T \mathrm{vec}(f) \geq 0 \; \forall f \in \mathbf{M}_{2t}(g_1, \ldots, g_m). \tag{6.4}$$

If $y_0 > 0$ then we may assume $y_0 = 1$ by rescaling. Then y is feasible for (6.3), which implies $p_t^{\mathrm{mom}} \leq y^T \mathrm{vec}(p) < \rho$. As this is true for all $\rho > p_t^{\mathrm{sos}}$, we deduce that $p_t^{\mathrm{mom}} \leq p_t^{\mathrm{sos}}$ and thus equality holds. Assume now $y_0 = 0$. Pick $x \in K$ and set $z := y + \epsilon \zeta_{2t,x}$ where $\zeta_{2t,x} = (x^\alpha)_{|\alpha| \leq 2t}$. Then, $z^T \mathrm{vec}(p - \rho) < 0$ if we choose $\epsilon > 0$ small enough and $z^T \mathrm{vec}(f) \geq 0$ for all $f \in \mathbf{M}_{2t}(g_1, \ldots, g_m)$, that is, z satisfies (6.4). As $z_0 = \epsilon > 0$, the previous argument (applied to z in place of y) yields again $p_t^{\mathrm{mom}} = p_t^{\mathrm{sos}}$. Finally, if (6.2) is feasible then it attains its supremum since $\mathbf{M}_{2t}(g_1, \ldots, g_m)$ is closed and one can bound the variable ρ.

The second argument, taken from [78], works under the assumption that (6.2) is feasible and uses the strong duality theorem for semidefinite programming. Indeed, by Lemma 4.4, the program (6.3) is strictly feasible and thus, by Theorem 1.3, there is no duality gap and (6.2) attains its supremum. □

PROPOSITION 6.2.

(i) If $\mathbf{M}(g_1, \ldots, g_m)$ is Archimedean, then the SOS relaxation (6.2) is feasible for t large enough.

(ii) If the ball constraint $R^2 - \sum_{i=1}^n \mathbf{x}_i^2 \geq 0$ is present in the description of K, then the feasible region of the moment relaxation (6.3) is bounded and the infimum is attained in (6.3).

Proof. (i) Using (3.16), $p + N \in \mathbf{M}(g_1, \ldots, g_m)$ for some N and thus $-N$ is feasible for (6.2) for t large enough.
(ii) Let y be feasible for (6.3). With $g := R^2 - \sum_{i=1} \mathbf{x}_i^2$, $(gy)_{2\beta} = R^2 y_{2\beta} - \sum_{i=1}^n y_{2\beta+2e_i}$. Thus the constraint $M_{t-1}(gy) \succeq 0$ implies $y_{2\beta+2e_i} \leq R^2 y_{2\beta}$ for all $|\beta| \leq t-1$ and $i = 1, \ldots, n$. One can easily derive (using induction on $|\beta|$) that $y_{2\beta} \leq R^{2|\beta|}$ for $|\beta| \leq t$. This in turn implies $|y_\gamma| \leq R^{|\gamma|}$ for $|\gamma| \leq 2t$. Indeed, write $\gamma = \alpha + \beta$ with $|\alpha|, |\beta| \leq t$; then as $M_t(y) \succeq 0$,

$y_{\alpha+\beta}^2 \leq y_{2\alpha}y_{2\beta} \leq R^{2|\alpha|}R^{2|\beta|}$, giving $|y_\gamma| \leq R^{|\gamma|}$. This shows that the feasible region to (6.3) is bounded and thus compact (as it is closed). Thus (6.3) attains its infimum. ∎

The next example (taken from [150]) shows that the infimum may not be attained in (6.3) even when K has a nonempty interior.

Example 6.3. *Consider the problem* $p^{min} := \inf_{x\in K} x_1^2$, *where* $K \subseteq \mathbb{R}^2$ *is defined by the polynomial* $g_1 = \mathbf{x}_1\mathbf{x}_2 - 1 \geq 0$. *Then* $p^{min} = p_t^{mom} = 0$ *for any* $t \geq 1$, *but these optimum values are not attained. Indeed, for small* $\epsilon > 0$, *the point* $x := (\epsilon, 1/\epsilon)$ *lies in* K, *which implies* $p^{min} \leq \epsilon^2$. *As* $p_t^{mom} \geq 0$ *(since* $y_{20} \geq 0$ *for any* y *feasible for (6.3)), this gives* $p_t^{mom} = p^{min} = 0$. *On the other hand* $y_{20} > 0$ *for any feasible* y *for (6.3); indeed* $M_0(g_1y) \succeq 0$ *implies* $y_{11} \geq 1$, *and* $y_{20} = 0$ *would imply* $y_{11} = 0$ *since* $M_1(y) \succeq 0$. *Thus the infimum is not attained in (6.3) in this example. Note that the above still holds if we add the constraints* $-2 \leq \mathbf{x}_1 \leq 2$ *and* $2 \leq \mathbf{x}_2 \leq 2$ *to the description of* K *to make it compact.*

On the other hand, when K has an empty interior, the duality gap may be infinite. We now give such an instance (taken from [150]) where $-\infty = p_t^{\text{sos}} < p_t^{\text{mom}} = p^{\text{min}}$.

Example 6.4. *Consider the problem* $p^{min} := \min_{x\in K} x_1x_2$, *where* $K := \{x \in \mathbb{R}^2 \mid g_1(x), g_2(x), g_3(x) \geq 0\}$ *with* $g_1 := -\mathbf{x}_2^2$, $g_2 := 1 + \mathbf{x}_1$, $g_3 := 1 - \mathbf{x}_1$. *Thus* $K = [-1, 1] \times \{0\}$. *Obviously,* $p^{min} = 0$. *We verify that* $p_1^{mom} = 0$, $p_1^{sos} = -\infty$. *For this let* y *be feasible for the program (6.3) for order* $t = 1$; *we show that* $y_{e_1+e_2} = 0$. *Indeed,* $(M_1(y))_{e_2,e_2} = y_{2e_2} \geq 0$ *and* $(M_0(g_1y))_{0,0} = -y_{2e_2} \geq 0$ *imply* $y_{2e_2} = 0$. *Thus the* e_2*th column of* $M_1(y)$ *is zero, which gives* $y_{e_1+e_2} = (M_1(y))_{e_1,e_2} = 0$. *Assume now that* ρ *is feasible for the program (6.2) at order* $t = 1$. *That is,* $\mathbf{x}_1\mathbf{x}_2 - \rho = \sum_i (a_i + b_i\mathbf{x}_1 + c_i\mathbf{x}_2)^2 - e_1\mathbf{x}_2^2 + e_2(1 + \mathbf{x}_1) + e_3(1 - \mathbf{x}_1)$ *for some* $a_i, b_i, c_i \in \mathbb{R}$ *and* $e_1, e_2, e_3 \in \mathbb{R}_+$. *Looking at the coefficient of* $\mathbf{x}_1^2$ *we find* $0 = \sum_i b_i^2$ *and thus* $b_i = 0$ *for all* i. *Looking at the coefficient of* $\mathbf{x}_1\mathbf{x}_2$ *we find* $1 = 0$, *a contradiction. Therefore there is no feasible solution, i.e.,* $p_1^{sos} = -\infty$. *On the other hand,* $p_2^{sos} = 0$ *since, for all* $\epsilon > 0$, $p_2^{sos} \geq -\epsilon$ *as* $\mathbf{x}_1\mathbf{x}_2 + \epsilon = \frac{(\mathbf{x}_2+2\epsilon)^2}{8\epsilon}(\mathbf{x}_1 + 1) + \frac{(\mathbf{x}_2-2\epsilon)^2}{8\epsilon}(-\mathbf{x}_1 + 1) - \frac{1}{4\epsilon}\mathbf{x}_2^2$.

What if K has an empty interior? When K has a nonempty interior the moment/SOS relaxations behave nicely; indeed there is no duality gap (Theorem 6.1) and the optimum value is attained under some conditions (cf. Proposition 6.2). Marshall [104] has studied in detail the case when K has an empty interior. He proposes to exploit the presence of equations to sharpen the SOS/moment bounds, in such a way that there is no duality gap between the sharpened bounds. Consider an ideal $\mathcal{J} \subseteq \mathcal{I}(K)$, where $\mathcal{I}(K) = \{f \in \mathbb{R}[\mathbf{x}] \mid f(x) = 0 \ \forall x \in K\}$ is the vanishing ideal of K; thus $\mathcal{I}(K) = \{0\}$ if K has a nonempty interior. Marshall makes the following assumption:

$$\mathcal{J} \subseteq \mathbf{M}(g_1, \ldots, g_m). \tag{6.5}$$

If this assumption does not hold and $\{h_1, \dots, h_{m_0}\}$ is a system of generators of the ideal $\mathcal{J}$, it suffices to add the polynomials $\pm h_1, \dots, \pm h_{m_0}$ in order to obtain a representation of K that fulfills (6.5). Now one may work with polynomials modulo the ideal $\mathcal{J}$. Let

$$\mathbf{M}'_{2t}(g_1, \dots, g_m) := \{p' \mid p \in \mathbf{M}_{2t}(g_1, \dots, g_m)\} \subseteq \mathbb{R}[\mathbf{x}]_{2t}/\mathcal{J}$$

be the image of $\mathbf{M}_{2t}(g_1, \dots, g_m)$ under the map $p \mapsto p' := p \mod \mathcal{J}$ from $\mathbb{R}[\mathbf{x}]$ to $\mathbb{R}[\mathbf{x}]/\mathcal{J}$. (This set was introduced in (3.24) for the ideal $\mathcal{J} = \mathcal{I}(K)$.) Consider the following refinement of the SOS relaxation (6.2)

$$\begin{aligned} p_t^{\text{sos,eq}} &:= \sup \rho \ \text{ s.t. } (p - \rho)' \in \mathbf{M}'_{2t}(g_1, \dots, g_m) \\ &= \sup \rho \ \text{ s.t. } p - \rho \in \mathbf{M}_{2t}(g_1, \dots, g_m) + \mathcal{J}. \end{aligned} \tag{6.6}$$

For the analogue of (6.3), we now consider linear functionals on $\mathbb{R}[\mathbf{x}]_{2t}/\mathcal{J}$, i.e. linear functionals on $\mathbb{R}[\mathbf{x}]_{2t}$ vanishing on $\mathcal{J} \cap \mathbb{R}[\mathbf{x}]_{2t}$, and define

$$\begin{aligned} p_t^{\text{mom,eq}} := & \inf_{L \in (\mathbb{R}[\mathbf{x}]_{2t}/\mathcal{J})^*} L(p) \\ & \text{s.t. } L(1) = 1, L(f) \geq 0 \ \forall f \in \mathbf{M}'_{2t}(g_1, \dots, g_m). \end{aligned} \tag{6.7}$$

Then, $p_t^{\text{sos}} \leq p_t^{\text{sos,eq}} \leq p^{\text{sos}}$, where the last inequality follows using (6.5); $p_t^{\text{sos,eq}} \leq p_t^{\text{mom,eq}}$, $p_t^{\text{mom}} \leq p_t^{\text{mom,eq}} \leq p^{\min}$. Moreover, $p_t^{\text{mom}} = p_t^{\text{mom,eq}}$, $p_t^{\text{sos}} = p_t^{\text{sos,eq}}$ if K has a nonempty interior since then $\mathcal{J} = \mathcal{I}(K) = \{0\}$. Marshall [104] shows the following extension of Theorem 6.1, which relies on Theorem 3.51 showing that $\mathbf{M}'_{2t}(g_1, \dots, g_m)$ is closed when $\mathcal{J} = \mathcal{I}(K)$. We omit the details of the proof which are similar to those for Theorem 6.1.

THEOREM 6.5. *[104] When $\mathcal{J} = \mathcal{I}(K)$ satisfies (6.5), $p_t^{sos,eq} = p_t^{mom,eq}$ for all $t \geq \max(d_p, d_K)$.*

As a consequence,

$$\sup_t p_t^{\text{mom}} = p^{\text{sos}} \ \text{ if } \ \mathcal{J} = \mathcal{I}(K) \subseteq \mathbf{M}(g_1, \dots, g_m).$$

Indeed, $p_t^{\text{mom}} \leq p_t^{\text{mom,eq}} = p_t^{\text{sos,eq}} \leq p_t^{\text{sos}} \leq p^{\text{sos}}$ implies $\sup_t p_t^{\text{mom}} \leq p^{\text{sos}}$. On the other hand, $p_t^{\text{mom}} \geq p_t^{\text{sos}}$ implies $\sup_t p_t^{\text{mom}} \geq \sup_t p_t^{\text{sos}} = p^{\text{sos}}$, which gives equality $\sup_t p_t^{\text{mom}} = p^{\text{sos}}$.

If we know a basis $\{h_1, \dots, h_{m_0}\}$ of $\mathcal{J}$ then we can add the equations $h_j = 0$ $(j \leq m_0)$, leading to an enriched representation for the set K of the form (2.5). Assuming $\mathcal{J} = \mathcal{I}(K)$, the SOS/moment bounds with respect to the description (2.5) of K are related to the bounds (6.6), (6.7) by

$$p_t^{\text{sos}} \leq p_t^{\text{mom}} \leq p_t^{\text{mom,eq}} = p_t^{\text{sos,eq}}. \tag{6.8}$$

LEMMA 6.6. *Assume that $\mathcal{J} = \mathcal{I}(K)$, $\{h_1, \dots, h_{m_0}\}$ is a Gröbner basis of $\mathcal{J}$ for a total degree ordering, and* $\deg(h_j)$ *is even* $\forall j \leq m_0$. *Then equality holds throughout in (6.8).*

Proof. Let ρ be feasible for (6.6); we show that ρ is feasible for (6.2), implying $p_t^{\text{sos,eq}} = p_t^{\text{sos}}$. We have $p - \rho = \sum_{j=0}^m s_j g_j + q$ where $s_j \in \Sigma$, $\deg(s_j g_j) \leq 2t$ and $q \in \mathcal{J}$. Then $q = \sum_{j=1}^{m_0} u_j h_j$ with $\deg(u_j h_j) \leq 2t$ (since the h_j's form a Gröbner basis for a total degree ordering) and thus $\deg(u_j) \leq 2(t - d_{h_j})$ (since $\deg(h_j) = 2d_{h_j}$ is even), i.e. ρ is feasible for (6.2). □

Remark 6.7. *As each equation $h_j = 0$ is treated like two inequalities $\pm h_j \geq 0$, we have $f \in \mathbf{M}_{2t}(g_1, \dots, g_m, \pm h_1, \dots, \pm h_{m_0})$ if and only if $f = \sum_{j=0}^m s_j g_j + \sum_{j=1}^{m_0}(u_j' - u_j'')h_j$ for some $s_j, u_j', u_j'' \in \Sigma$ with* $\deg(s_j g_j), \deg(u_j' h_j), \deg(u_j'' h_j) \leq 2t$. *As* $\deg(u_j' h_j), \deg(u_j'' h_j) \leq 2t$ *is equivalent to* $\deg(u_j'), \deg(u_j'') \leq 2(t - d_{h_j})$*, one may equivalently write $\sum_{j=1}^{m_0}(u_j' - u_j'')h_j = \sum_{j=1}^{m_0} u_j h_j$ where $u_j \in \mathbb{R}[\mathbf{x}]_{2(t-d_{h_j})}$. Note that* $\deg(u_j) \leq 2(t - d_{h_j})$ *implies* $\deg(u_j h_j) \leq 2t$*, but the reverse does not hold, except if at least one of* $\deg(u_j)$, $\deg(h_j)$ *is even. This is why we assume in Lemma 6.6 that* $\deg(h_j)$ *is even. As an illustration, consider again Example 6.4, where $\mathcal{I}(K) = (\mathbf{x}_2)$. If we add the equation $\mathbf{x}_2 = 0$ to the description of K, we still get $p_1^{sos} = -\infty$ (since the multiplier of $\mathbf{x}_2$ in a decomposition of $\mathbf{x}_1\mathbf{x}_2 - \rho \in \mathbf{M}_2(\mathbf{x}_1 + 1, 1 - \mathbf{x}_1, \pm\mathbf{x}_2)$ should be a scalar), while $p_1^{sos,eq} = 0$ (since $\mathbf{x}_1$ is now allowed as multiplier of $\mathbf{x}_2$).*

6.3. Asymptotic convergence. The asymptotic convergence of the SOS/moment bounds to $p^{\min}$ follows directly from Putinar's theorem (Theorem 3.20); recall Definition 3.18 for an Archimedean quadratic module.

Theorem 6.8. [78] *If $\mathbf{M}(g_1, \dots, g_m)$ is Archimedean, then $p^{sos} = p^{mom} = p^{min}$, i.e.* $\lim_{t\to\infty} p_t^{sos} = \lim_{t\to\infty} p_t^{mom} = p^{min}$.

Proof. Given $\epsilon > 0$, the polynomial $p - p^{\min} + \epsilon$ is positive on K. By Theorem 3.20, it belongs to $\mathbf{M}(g_1, \dots, g_m)$ and thus the scalar $p^{\min} - \epsilon$ is feasible for the program (6.2) for some t. Therefore, there exists t for which $p_t^{\text{sos}} \geq p^{\min} - \epsilon$. Letting ϵ go to 0, we find that $p^{\text{sos}} = \lim_{t\to\infty} p_t^{\text{sos}} \geq p^{\min}$, implying $p^{\text{sos}} = p^{\text{mom}} = p^{\min}$. □

Note that if we would have a representation result valid for *nonnegative* (instead of *positive*) polynomials, this would immediately imply the *finite convergence* of the bounds $p_t^{\text{sos}}, p_t^{\text{mom}}$ to $p^{\min}$. For instance, Theorem 2.4 in Section 2.1 gives such a reprentation result in the case when the description of K involves a set of polynomial equations generating a zero-dimensional radical ideal. Thus we have the following result.

Corollary 6.9. *Assume K is as in (2.5) and $h_1, \dots, h_{m_0}$ generate a zero-dimensional radical ideal. Then, $p_t^{sos} = p_t^{mom} = p^{min}$ for t large enough.*

Proof. Directly from Theorem 2.4, as in the proof of Theorem 6.8. □

In the non-compact case, convergence to $p^{\min}$ may fail. For instance, Marshall [104] shows that when K contains a full dimensional cone then, for all $t \geq \max(d_p, d_K)$, $p_t^{\text{sos}} = p_t^{\text{mom}}$, which can be strictly smaller than $p^{\min}$. This applies in particular to the case $K = \mathbb{R}^n$.

6.4. Approximating the unique global minimizer via the moment relaxations. Here we prove that when (1.1) has a *unique* global minimizer, then this minimizer can be approximated from the optimum solutions to the moment relaxations (6.3). We show in fact a stronger result (Theorem 6.11); this result is taken from Schweighofer [150] (who however formulates it in a slightly more general form in [150]). Recall the definition of the quadratic module $\mathbf{M}(g_1,\ldots,g_m)$ from (3.13) and of its truncation $\mathbf{M}_t(g_1,\ldots,g_m)$ from (3.22). Define the set of global minimizers of (1.1)

$$K_p^{\min} := \{x \in K \mid p(x) = p^{\min}\}. \tag{6.9}$$

DEFINITION 6.10. *Given $y^{(t)} \in \mathbb{R}^{\mathbb{N}^n_{2t}}$, $y^{(t)}$ is* nearly optimal *for (6.3) if $y^{(t)}$ is feasible for (6.3) and $\lim_{t\to\infty} p^T y^{(t)} = \lim_{t\to\infty} p_t^{mom}$.*

THEOREM 6.11. *[150] Assume $\mathbf{M}(g_1,\ldots,g_m)$ is Archimedian, $K_p^{\min} \neq \emptyset$, and let $y^{(t)}$ be nearly optimal solutions to (6.3). Then, $\forall \epsilon > 0 \;\; \exists t_0 \geq \max(d_p, d_K) \;\; \forall t \geq t_0 \;\; \exists \mu$ probability measure on $K_p^{\min}$ such that $\max_{i=1,\ldots,n} |y_{e_i}^{(t)} - \int x_i \mu(dx)| \leq \epsilon$.*

Proof. As $\mathbf{M}(g_1,\ldots,g_m)$ is Archimedian, we deduce from (3.16) that

$$\forall k \in \mathbb{N} \;\; \exists N_k \in \mathbb{N} \;\; \forall \alpha \in \mathbb{N}^n_k \;\; N_k \pm \mathbf{x}^\alpha \in M_{N_k}(g_1,\ldots,g_m). \tag{6.10}$$

Define the sets $Z := \prod_{\alpha\in\mathbb{N}^n}[-N_{|\alpha|}, N_{|\alpha|}]$, $C_0 := \{z \in Z \mid z_0 = 1\}$, $C_f := \{z \in Z \mid z^T f \geq 0\}$ for $f \in \mathbf{M}(g_1,\ldots,g_m)$, $C_\delta := \{z \in Z \mid |z^T p - p^{\min}| \leq \delta\}$ for $\delta > 0$, and

$$C := \{z \in Z \mid \begin{array}{l} \max_{i=1,\ldots,n} |z_{e_i} - \int x_i \mu(dx)| > \epsilon \\ \forall \mu \text{ probability measure on } K_p^{\min}\}. \end{array}$$

CLAIM 6.12. $\bigcap_{f\in\mathbf{M}(g_1,\ldots,g_m)} C_f \cap \bigcap_{\delta>0} C_\delta \bigcap C_0 \bigcap C = \emptyset$.

Proof. Assume z lies in the intersection. As $z \in C_0 \cap \bigcap_{f\in\mathbf{M}(g_1,\ldots,g_m)} C_f$, we deduce using (4.6) that $z \in \mathcal{M}^{put}_{\succeq}(g_1,\ldots,g_m)$ (recall (4.17)). Hence, by Theorem 4.17, $z \in \mathcal{M}_K$, i.e. z has a representing measure μ which is a probability measure on the set K. As $z \in \cap_{\delta>0} C_\delta$, we have $p^T z = p^{\min}$, i.e. $\int (p(x) - p^{\min})\mu(dx) = 0$, which implies that the support of μ is contained in the set $K_p^{\min}$, thus contradicting the fact that $z \in C$. □

As Z is a compact set (by Tychonoff's theorem) and all the sets C_f, C_δ, C_0, C are closed subsets of Z, there exists a finite collection of those sets having an empty intersection. That is, there exist $f_1,\ldots,f_s \in \mathbf{M}(g_1,\ldots,g_m)$, $\delta > 0$ such that

$$C_{f_1} \cap \ldots \cap C_{f_s} \cap C_\delta \cap C_0 \cap C = \emptyset. \tag{6.11}$$

Choose an integer $t_1 \geq \max(d_p, d_K)$ such that $f_1,\ldots,f_s \in M_{2t_1}(g_1,\ldots,g_m)$. Then choose an integer t_0 such that $t_0 \geq t_1$, $2t_0 \geq$

$\max(N_k \mid k \leq 2t_1)$ (recall (6.10)) and $|p^T y^{(t)} - p^{\min}| \leq \delta$ for all $t \geq t_0$. We now verify that this t_0 satisfies the conclusion of the theorem. For this fix $t \geq t_0$. Consider the vector $z \in \mathbb{R}^{\mathbb{N}^n}$ defined by $z_\alpha := y_\alpha^{(t)}$ if $|\alpha| \leq 2t_1$, and $z_\alpha := 0$ otherwise.

Claim 6.13. *$z \in Z$.*

Proof. Let $\alpha \in \mathbb{N}^n$ with $|\alpha| =: k \leq 2t_1$. Then $N_k \pm \mathbf{x}^\alpha \in M_{N_k}(g_1, \ldots, g_m) \subseteq M_{2t_0}(g_1, \ldots, g_m) \subseteq \mathbf{M}_{2t}(g_1, \ldots, g_m)$. As $y^{(t)}$ is feasible for (6.3) we deduce that $(y^{(t)})^T \mathrm{vec}(N_k \pm \mathbf{x}^\alpha) \geq 0$, implying $|y_\alpha^{(t)}| \leq N_k = N_{|\alpha|}$. □

Obviously $z \in C_0$. Next $z \in C_\delta$ since $|z^T p - p^{\min}| = |(y^{(t)})^T p - p^{\min}| \leq \delta$ as $2t_1 \geq \deg(p)$. Finally, for any $r = 1, \ldots, s$, $z \in C_{f_r}$ since $z^T f_r = (y^{(t)})^T f_r \geq 0$ as $f_r \in M_{2t_1}(g_1, \ldots, g_m) \subseteq \mathbf{M}_{2t}(g_1, \ldots, g_m)$. As the set in (6.11) is empty, we deduce that $z \notin C$. Therefore, there exists a probability measure μ on $K_p^{\min}$ for which $\max_i |y_{e_i}^{(t)} - \int x_i \mu(dx)| = \max_i |z_{e_i} - \int x_i \mu(dx)| \leq \epsilon$. This concludes the proof of Theorem 6.11. □

Corollary 6.14. *Assume $\mathbf{M}(g_1, \ldots, g_m)$ is Archimedian and problem (1.1) has a unique minimizer x^*. Let $y^{(t)}$ be nearly optimal solutions to (6.3). Then $\lim_{t \to \infty} y_{e_i}^{(t)} = x_i^*$ for each $i = 1, \ldots, n$.*

Proof. Directly from Theorem 6.11 since the Dirac measure δ_{x^*} at x^* is the unique probability measure on $K_p^{\min}$. □

6.5. Finite convergence. Here we show finite convergence for the moment/SOS relaxations, when the description of the semialgebraic set K contains a set of polynomial equations $h_1 = 0, \ldots, h_{m_0} = 0$ generating a zero-dimensional ideal. The case when the polynomial equations $h_1, \ldots, h_{m_0}$ generate a *radical* zero-dimensional ideal was considered earlier in Corollary 6.9. We now consider the non-radical case. Theorem 6.15 below extends a result of Laurent [96] and uses ideas from Lasserre et al. [90].

Theorem 6.15. *Consider the problem (1.1) of minimizing $p \in \mathbb{R}[\mathbf{x}]$ over the set $K = \{x \in \mathbb{R}^n \mid h_j(x) = 0 \ (j = 1, \ldots, m_0),\ g_j(x) \geq 0 \ (j = 1, \ldots, m)\}$ (as in (2.5)). Set $\mathcal{J} := (h_1, \ldots, h_{m_0})$.*

(i) *If $|V_\mathbb{C}(\mathcal{J})| < \infty$, then $p^{min} = p_t^{mom} = p_t^{sos}$ for t large enough.*

(ii) *If $|V_\mathbb{R}(\mathcal{J})| < \infty$, then $p^{min} = p_t^{mom}$ for t large enough.*

Proof. Fix $\epsilon > 0$. The polynomial $p - p^{\min} + \epsilon$ is positive on K. For the polynomial $u := -\sum_{j=1}^{m_0} h_j^2$, the set $\{x \in \mathbb{R}^n \mid u(x) \geq 0\} = V_\mathbb{R}(\mathcal{J})$ is compact (in fact, finite under (i) or (ii)) and u belongs to the quadratic module generated by the polynomials $\pm h_1, \ldots, \pm h_{m_0}$. Hence we are in the Archimedean case and we can apply Theorem 3.20. Therefore, there is a decomposition

$$p - p^{\min} + \epsilon = s_0 + \sum_{j=1}^{m} s_j g_j + q, \tag{6.12}$$

where s_0, s_j are sums of squares and $q \in \mathcal{J}$. To finish the proof we distinguish the two cases (i), (ii).

Consider first the case (i) when $|V_{\mathbb{C}}(\mathcal{J})| < \infty$. Let $\{f_1, \ldots, f_L\}$ be a Gröbner basis of $\mathcal{J}$ for a total degree monomial ordering, let $\mathcal{B}$ be a basis of $\mathbb{R}[\mathbf{x}]/\mathcal{J}$, and set $d_{\mathcal{B}} := \max_{b \in \mathcal{B}} \deg(b)$ (which is well defined as $|\mathcal{B}| < \infty$ since $\mathcal{J}$ is zero-dimensional). Consider the decomposition (6.12). Say, $s_j = \sum_i s_{i,j}^2$ and write $s_{i,j} = r_{i,j} + q_{i,j}$, where $r_{i,j}$ is a linear combination of members of $\mathcal{B}$ and $q_{i,j} \in \mathcal{J}$; thus $\deg(r_{i,j}) \leq d_{\mathcal{B}}$. In this way we obtain another decomposition:

$$p - p^{\min} + \epsilon = s_0' + \sum_{j=1}^{m} s_j' g_j + q', \tag{6.13}$$

where s_0', s_j' are sums of squares, $q' \in \mathcal{J}$ and $\deg(s_0'), \deg(s_j') \leq 2d_{\mathcal{B}}$. Set

$$T_0 := \max(2d_p, 2d_{\mathcal{B}}, 2d_{\mathcal{B}} + 2d_{g_1}, \ldots, 2d_{\mathcal{B}} + 2d_{g_m}). \tag{6.14}$$

Then, $\deg(s_0'), \deg(s_j'g_j), \deg(p - p^{\min} + \epsilon) \leq T_0$ and thus $\deg(q') \leq T_0$. Therefore, q' has a decomposition $q' = \sum_{l=1}^{L} u_l f_l$ with $\deg(u_l f_l) \leq \deg(q') \leq T_0$ (because we use a total degree monomial ordering). We need to find a decomposition of q' with bounded degrees in the original basis $\{h_1, \ldots, h_{m_0}\}$ of $\mathcal{J}$. For this, write $f_l = \sum_{j=1}^{m_0} a_{l,j} h_j$ where $a_{l,j} \in \mathbb{R}[\mathbf{x}]$. Then, $q' = \sum_{l=1}^{L} u_l (\sum_{j=1}^{m_0} a_{l,j} h_j) = \sum_{j=1}^{m_0} (\sum_{l=1}^{L} a_{l,j} u_l) h_j =: \sum_{j=1}^{m_0} b_j h_j$, setting $b_j := \sum_{l=1}^{L} a_{l,j} u_l$. As $\deg(u_l) \leq T_0$, we have $\deg(b_j h_j) \leq 2d_{h_j} + T_0 + \max_{l=1}^{L} \deg(a_{l,j})$. Thus, $\deg(b_j h_j) \leq T_g$ after setting $T_g := T_0 + \max_{l,j}(\deg(a_{l,j}) + 2d_{h_j})$, which is a constant not depending on ϵ. Therefore we can conclude that $p^{\min} - \epsilon$ is feasible for the program (6.2) for all $t \geq T_1 := \lceil T_g/2 \rceil$. This implies $p_t^{\text{sos}} \geq p^{\min} - \epsilon$ for all $t \geq T_1$. Letting ϵ go to zero, we find $p_t^{\text{sos}} \geq p^{\min}$ and thus $p_t^{\text{sos}} = p^{\min}$ for $t \geq T_1$, which concludes the proof in case (i).

Consider now the case (ii) when $|V_{\mathbb{R}}(\mathcal{J})| < \infty$. Let y be feasible for the program (6.3); that is, $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$ satisfies $y_0 = 1$, $M_t(y) \succeq 0$, $M_{t-d_{h_j}}(h_j y) = 0$ $(j = 1, \ldots, m_0)$, $M_{t-d_{g_j}}(g_j y) \succeq 0$ $(j = 1, \ldots, m)$. We show $p^T y \geq p^{\min}$ for t large enough. We need the following observations about the kernel of $M_t(y)$. First, for $j = 1, \ldots, m_0$, $h_j \in \operatorname{Ker} M_t(y)$ for $t \geq 2d_{h_j}$ (directly, from the fact that $M_{t-d_{h_j}}(h_j y) = 0$). Moreover, for t large enough, $\operatorname{Ker} M_t(y)$ contains any given finite set of polynomials of $\mathcal{I}(V_{\mathbb{R}}(\mathcal{J}))$.

CLAIM 6.16. *Let $f_1, \ldots, f_L \in \mathcal{I}(V_{\mathbb{R}}(\mathcal{J}))$. There exists $t_1 \in \mathbb{N}$ such that $f_1, \ldots, f_L \in \operatorname{Ker} M_t(y)$ for all $t \geq t_1$.*

Proof. Fix $l = 1, \ldots, L$. As $f_l \in \mathcal{I}(V_{\mathbb{R}}(\mathcal{J}))$, by the Real Nullstellensatz (Theorem 2.1), $f_l^{2m_l} + \sum_i p_{l,i}^2 = \sum_{j=1}^{m} u_{l,j} h_j$ for some $p_{l,i}, u_{l,j} \in \mathbb{R}[\mathbf{x}]$ and $m_l \in \mathbb{N} \setminus \{0\}$. Set $t_1 := \max(\max_{j=1}^{m_0} 2d_{h_j}, 1 + \max_{l \leq L, j \leq m_0} \deg(u_{l,j} h_j))$ and let $t \geq t_1$. Then, each $u_{l,j} h_j$ lies in $\operatorname{Ker} M_t(y)$ by Lemma 5.7. Therefore,

$f_l^{2m_l} + \sum_i p_{l,i}^2 \in \operatorname{Ker} M_t(y)$, which implies $f_l^{m_l}, p_{l,i} \in \operatorname{Ker} M_t(y)$. An easy induction permits to conclude that $f_l \in \operatorname{Ker} M_t(y)$. □

Let $\{f_1, \ldots, f_L\}$ be a Gröbner basis of $\mathcal{I}(V_{\mathbb{R}}(\mathcal{J}))$ for a total degree monomial ordering, let $\mathcal{B}$ be a basis of $\mathbb{R}[\mathbf{x}]/\mathcal{I}(V_{\mathbb{R}}(\mathcal{J}))$, and set $d_{\mathcal{B}} := \max_{b \in \mathcal{B}} \deg(b)$ (which is well defined since $|\mathcal{B}| < \infty$ as $\mathcal{I}(V_{\mathbb{R}}(\mathcal{J}))$ is zero-dimensional). Given $\epsilon > 0$, consider the decomposition (6.12) where s_0, s_j are sums of squares and $q \in \mathcal{J}$. As in case (i), we can derive another decomposition (6.13) where s'_0, s'_j are s.o.s., $q' \in \mathcal{I}(V_{\mathbb{R}}(\mathcal{J}))$, and $\deg(s'_0), \deg(s'_j) \leq 2d_{\mathcal{B}}$. Then, $\deg(s'_0), \deg(s'_j g_j), \deg q' \leq T_0$ with T_0 being defined as in (6.14) and we can write $q' = \sum_{l=1}^{L} u_l f_l$ with $\deg(u_l f_l) \leq T_0$. Fix $t \geq \max(T_0 + 1, t_1)$. Then, $u_l f_l \in \operatorname{Ker} M_t(y)$ (by Lemma 5.7 and Claim 6.16) and thus $q' \in \operatorname{Ker} M_t(y)$. Moreover, $\operatorname{vec}(1)^T M_t(y) \operatorname{vec}(s'_j g_j) \geq 0$; to see it, write $s'_j = \sum_i a_{i,j}^2$ and note that $\operatorname{vec}(1)^T M_t(y) \operatorname{vec}(s'_j g_j) = \sum_i a_{i,j}^T M_{t-d_{g_j}}(g_j y) a_{i,j} \geq 0$ since $M_{t-d_{g_j}}(g_j y) \succeq 0$. Therefore, we deduce from (6.13) that $\operatorname{vec}(1)^T M_t(y) \operatorname{vec}(p - p^{\min} + \epsilon) \geq 0$, which gives $p^T y = 1^T M_t(y) p \geq p^{\min} - \epsilon$ and thus $p_t^{\text{mom}} \geq p^{\min} - \epsilon$. Letting ϵ go to 0, we obtain $p_t^{\text{mom}} \geq p^{\min}$ and thus $p_t^{\text{mom}} = p^{\min}$. □

Question 6.17. *Does there exist an example with $|V_{\mathbb{C}}(\mathcal{J})| = \infty$, $|V_{\mathbb{R}}(\mathcal{J})| < \infty$ and where $p_t^{sos} < p^{min}$ for all t ?*

The finite convergence result from Theorem 6.15 applies, in particular, to the case when K is contained in a discrete grid $K_1 \times \ldots \times K_n$ with $K_i \subseteq \mathbb{R}$ finite, considered by Lasserre [80], and by Lasserre [79] in the special case $K \subseteq \{0,1\}^n$. We will come back to the topic of exploiting equations in Section 8.2.

6.6. Optimality certificate. We now formulate some stopping criterion for the moment hierarchy (6.3), i.e. some condition permitting to claim that the moment relaxation (6.3) is in fact exact, i.e. $p_t^{\text{mom}} = p^{\min}$, and to extract some global minimizer for (1.1).

A first easy such condition is as follows. Let y be an optimum solution to (6.3) and $x^* := (y_{10\ldots0}, \ldots, y_{0\ldots01})$ the point in $\mathbb{R}^n$ consisting of the coordinates of y indexed by $\alpha \in \mathbb{N}^n$ with $|\alpha| = 1$. Then

$$\begin{array}{c} x^* \in K \text{ and } p_t^{\text{mom}} = p(x^*) \Longrightarrow p_t^{\text{mom}} = p^{\min} \text{ and} \\ x^* \text{ is a global minimizer of } p \text{ over } K. \end{array} \tag{6.15}$$

Indeed $p^{\min} \leq p(x^*)$ as $x \in K$, which together with $p(x^*) = p_t^{\text{mom}} \leq p^{\min}$ implies equality $p_t^{\text{mom}} = p^{\min}$ and x^* is a minimizer of p over K. Note that $p_t^{\text{mom}} = p(x^*)$ automatically holds if p is linear. According to Theorem 6.11 this condition has a good chance to be satisfied (approximatively) when problem (1.1) has a unique minimizer. See Examples 6.24, 6.25 for instances where the criterion (6.15) is satisfied.

We now see another stopping criterion, which may work when problem (1.1) has a *finite* number of global minimizers. This stopping criterion, which has been formulated by Henrion and Lasserre [62], deals with the rank of the moment matrix of an optimal solution to (6.3) and is based on the result of Curto and Fialkow from Theorem 5.33. As in (6.9), $K_p^{\min}$ denotes the set of global minimizers of p over the set K. Thus $K_p^{\min} \neq \emptyset$, e.g., when K is compact. The next result is based on [62] combined with ideas from [90].

THEOREM 6.18. *Let $t \geq \max(d_p, d_K)$ and let $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$ be an optimal solution to the program (6.3). Assume that the following rank condition holds:*

$$\exists s \;\; s.t. \;\; \max(d_p, d_K) \leq s \leq t \;\; and \;\; \operatorname{rank} M_s(y) = \operatorname{rank} M_{s-d_K}(y). \quad (6.16)$$

Then $p_t^{mom} = p^{min}$ and $V_{\mathbb{C}}(\operatorname{Ker} M_s(y)) \subseteq K_p^{\min}$. Moreover, equality $V_{\mathbb{C}}(\operatorname{Ker} M_s(y)) = K_p^{\min}$ holds if $\operatorname{rank} M_t(y)$ is maximum among all optimal solutions to (6.3).

Proof. By assumption, $p_t^{\text{mom}} = p^T y$, $M_t(y) \succeq 0$, rank $M_s(y) =$ rank $M_{s-d_K}(y) =: r$ and $M_{s-d_K}(g_j y) \succeq 0$ for $j = 1, \ldots, m$, where $\max(d_p, d_K) \leq s \leq t$. As $s \geq d_K$, we can apply Theorem 5.33 and conclude that the sequence $(y_\alpha)_{\alpha \in \mathbb{N}^n_{2s}}$ has a r-atomic representing measure $\mu = \sum_{i=1}^r \lambda_i \delta_{v_i}$, where $v_i \in K$, $\lambda_i > 0$ and $\sum_{i=1}^r \lambda_i = 1$ (since $y_0 = 1$). As $s \geq d_p$, $p_t^{\text{mom}} = p^T y = \sum_{|\alpha| \leq 2s} p_\alpha y_\alpha = \sum_{i=1}^r \lambda_i p(v_i) \geq p^{\min}$, since $p(v_i) \geq p^{\min}$ for all i. On the other hand, $p^{\min} \geq p_t^{\text{mom}}$. This implies that $p^{\min} = p_t^{\text{mom}}$ and that each v_i is a minimizer of p over the set K, i.e., $\operatorname{supp}(\mu) = \{v_1, \ldots, v_r\} \subseteq K_p^{\min}$. As $\operatorname{supp}(\mu) = V_{\mathbb{C}}(\operatorname{Ker} M_s(y))$ by Theorem 5.29, we obtain $V_{\mathbb{C}}(\operatorname{Ker} M_s(y)) \subseteq K_p^{\min}$.

Assume now that $\operatorname{rank} M_t(y)$ is maximum among all optimal solutions to (6.3). By Lemma 1.4, $\operatorname{Ker} M_t(y) \subseteq \operatorname{Ker} M_t(y')$ for any other optimal solution y' to (6.3). For any $v \in K_p^{\min}$, $y' := \zeta_{2t,v}$ is feasible for (6.3) with objective value $p^T y' = p(v) = p^{\min}$; thus y' is an optimal solution and thus $\operatorname{Ker} M_t(y) \subseteq \operatorname{Ker} M_t(\zeta_{2t,v})$. This implies $\operatorname{Ker} M_t(y) \subseteq \bigcap_{v \in K_p^{\min}} \operatorname{Ker} M_t(\zeta_{2t,v}) \subseteq \mathcal{I}(K_p^{\min})$. Therefore, $\operatorname{Ker} M_s(y) \subseteq \operatorname{Ker} M_t(y) \subseteq \mathcal{I}(K_p^{\min})$, which implies $K_p^{\min} \subseteq V_{\mathbb{C}}(\operatorname{Ker} M_s(y))$ and thus equality $V_{\mathbb{C}}(\operatorname{Ker} M_s(y)) = K_p^{\min}$ holds. □

Hence, if at some order $t \geq \max(d_p, d_K)$ one can find a maximum rank optimal solution to the moment relaxation (6.3) which satisfies the rank condition (6.16), then one can find *all* the global minimizers of p over the set K, by computing the common zeros to the polynomials in $\operatorname{Ker} M_s(y)$. In view of Theorem 5.29 and Lemma 5.2, the ideal $(\operatorname{Ker} M_s(y))$ is (real) radical and zero-dimensional. Hence its variety $V_{\mathbb{C}}(\operatorname{Ker} M_s(y))$ is finite. Moreover one can determine this variety with the eigenvalue method, described in Section 2.4. This extraction procedure is presented in Henrion and Lasserre [62] and is implemented in their optimization software GloptiPoly.

The second part of Theorem 6.18, asserting that all global minimizers are found when having a maximum rank solution, relies on ideas from [90].

When p is the constant polynomial 1 and K is defined by the equations $h_1 = 0, \ldots, h_{m_0} = 0$, then $K_p^{\min}$ is the set of all common real roots of the h_j's. The paper [90] explains in detail how the moment methodology applies to finding real roots, and [91] extends this to complex roots.

As we just proved, if (6.16) holds for a maximum rank optimal solution y to (6.3), then $K_p^{\min} = V_{\mathbb{C}}(\operatorname{Ker} M_s(y))$ is finite. Hence the conditions of Theorem 6.18 can apply *only* when p has finitely many global minimizers over the set K. We will give in Example 6.24 an instance with infinitely many global minimizers and thus, as predicted, the rank condition (6.16) is not satisfied. We now see an example where the set $K_p^{\min}$ of global minimizers is finite but yet the conditions of Theorem 6.18 are never met.

Example 6.19. *We give here an example where $|K_p^{\min}| < \infty$ and $p_t^{mom} = p_t^{sos} < p^{min}$; hence condition (6.16) does not hold. Namely consider the problem*

$$p^{min} = \min_{x \in K} p(x) \quad \textit{where } K := \{x \in \mathbb{R}^n \mid g_1(x) := 1 - \sum_{i=1}^{n} x_i^2 \geq 0\}.$$

Assume that p is a homogeneous polynomial which is positive (i.e., $p(x) > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$), but not a sum of squares. Then, $p^{min} = 0$ and the origin is the unique minimizer, i.e., $K_p^{\min} = \{0\}$. Consider the moment relaxation (6.3) and the dual SOS relaxation (6.2) for $t \geq d_p$. As $\mathbf{M}(g_1)$ is Archimedean, the SOS relaxation (6.2) is feasible for t large enough. Moreover, as K has a nonempty interior, there is no duality gap, i.e. $p_t^{mom} = p_t^{sos}$, and the supremum is attained in (6.2) (apply Theorem 6.1). We now verify that $p_t^{sos} = p_t^{mom} < p^{min} = 0$. Indeed, if $p_t^{sos} = 0$, then $p = s_0 + s_1(1 - \sum_{i=1}^n \mathbf{x}_i^2)$ where $s_0, s_1 \in \mathbb{R}[\mathbf{x}]$ are sums of squares. It is not difficult to verify that this implies that p must be a sum of squares (see [35, Prop. 4]), yielding a contradiction. Therefore, on this example, $p_t^{mom} = p_t^{sos} < p^{min}$ and thus the rank condition (6.16) cannot be satisfied. This situation is illustrated in Example 6.25. There we choose $p = M + \epsilon(\mathbf{x}_1^6 + \mathbf{x}_2^6 + \mathbf{x}_3^6)$ where M is the Motzkin form (introduced in Example 3.7). Thus p is a homogeneous positive polynomial and there exists $\epsilon > 0$ for which p_ϵ is not SOS (for if not $M = \lim_{\epsilon \to 0} p_\epsilon$ would be SOS since the cone $\Sigma_{3,6}$ is closed).

On the other hand, we now show that the rank condition (6.16) in Theorem 6.18 holds for t large enough when the description of the set K comprises a system of equations $h_1 = 0, \ldots, h_{m_0} = 0$ having finitely many real zeros. Note that the next result also provides an alternative proof for Theorem 6.15 (ii), which does not use Putinar's theorem but results about moment matrices instead.

Theorem 6.20. *[90, Prop. 4.6] Let K be as in (2.5), let $\mathcal{J}$ be the ideal generated by $h_1, \ldots, h_{m_0}$ and assume that $|V_{\mathbb{R}}(\mathcal{J})| < \infty$. For t large enough, there exists an integer s,* $\max(d_K, d_p) \leq s \leq t$, *such that* $\operatorname{rank} M_s(y) = \operatorname{rank} M_{s-d_K}(y)$ *for any feasible solution y to (6.3).*

Proof. As in the proof of Theorem 6.15 (ii), let $\{f_1, \ldots, f_L\}$ be a Gröbner basis of $\mathcal{I}(V_{\mathbb{R}}(\mathcal{J}))$ for a total degree monomial ordering. By Claim 6.16, there exists $t_1 \in \mathbb{N}$ such that $f_1, \ldots, f_L \in \operatorname{Ker} M_t(y)$ for all $t \geq t_1$. Let $\mathcal{B}$ be a basis of $\mathbb{R}[\mathbf{x}]/\mathcal{I}(V_{\mathbb{R}}(\mathcal{J}))$ and $d_{\mathcal{B}} := \max_{b \in \mathcal{B}} \deg(b)$. Write any monomial as $x^\alpha = r^{(\alpha)} + \sum_{l=1}^{L} p_l^{(\alpha)} f_l$, where $r^{(\alpha)} \in \operatorname{Span}_{\mathbb{R}}(\mathcal{B})$, $p_l^{(\alpha)} \in \mathbb{R}[\mathbf{x}]$. Set $t_2 := \max(t_1, d_{\mathcal{B}} + d_K, d_p)$ and $t_3 := 1 + \max(t_2, \deg(p_l^{(\alpha)} f_l)$ for $l \leq L, |\alpha| \leq t_2)$. Fix $t \geq t_3$ and let y be feasible for (6.3). We claim that $\operatorname{rank} M_{t_2}(y) = \operatorname{rank} M_{t_2 - d_K}(y)$. Indeed, consider $\alpha \in \mathbb{N}^n_{t_2}$. As $\deg(p_l^{(\alpha)} f_l) \leq t - 1$ and $f_l \in \operatorname{Ker} M_t(y)$, we deduce (using Lemma 5.7) that $p_l^{(\alpha)} f_l \in \operatorname{Ker} M_t(y)$ and thus $\mathbf{x}^\alpha - r^{(\alpha)} \in \operatorname{Ker} M_t(y)$. As $\deg(r^{(\alpha)}) \leq d_{\mathcal{B}} \leq t_2 - d_K$, this shows that the αth column of $M_t(y)$ can be written as a linear combination of columns of $M_{t_2 - d}(y)$; that is, $\operatorname{rank} M_{t_2}(y) = \operatorname{rank} M_{t_2 - d_K}(y)$. □

Let us conclude this section with a brief discussion about the assumptions made in Theorem 6.18. A first basic assumption we made there is that the moment relaxation (6.3) attains its minimum. This is the case, e.g., when the feasible region of (6.3) is bounded (which happens e.g. when a ball constraint is present in the description of K, cf. Proposition 6.2), or when program (6.2) is strictly feasible (recall Theorem 1.3). A second basic question is to find conditions ensuring that there is no duality gap, i.e. $p_t^{\text{mom}} = p_t^{\text{sos}}$, since this is needed if one wants to solve the semidefinite programs using a primal-dual interior point algorithm. This is the case, e.g. when K has a nonempty interior (by Theorem 6.1) or when any of the programs (6.3) or (6.2) is strictly feasible (recall Theorem 1.3).

Another question raised in Theorem 6.18 is to find an optimum solution to a semidefinite program with maximum rank. It is in fact a property of most interior-point algorithms that they return a maximum rank optimal solution. This is indeed the case for the SDP solver SeDuMi used within GloptiPoly. More precisely, when both primal and dual problems (6.3) and (6.2) are strictly feasible, then the interior-point algorithm SeDuMi constructs a sequence of points on the so-called central path, which has the property of converging to an optimal solution of maximum rank. SeDuMi also finds a maximum rank optimal solution under the weaker assumption that (6.3) is feasible and attains its minimum, (6.2) is feasible, and $p_t^{\text{mom}} = p_t^{\text{sos}} < \infty$. Indeed SeDuMi applies the so-called extended self-dual embedding technique, which consists of embedding the given program into a new program satisfying the required strict feasibility property; a maximum rank optimal solution for the original problem can then be derived from a maximum rank optimal solution to the embedded problem. See [33, Ch. 4], [176, Ch. 5] for details. (This issue is also discussed in [90] in the context of solving systems of polynomial equations.)

There are many further numerical issues arising for the practical implementation of the SOS/moment method. Just to name a few, the numerical instability of linear algebra dealing with matrices with a Hankel type structure, or the numerically sensitive issue of computing ranks, etc. To address the first issue, Löfberg and Parrilo [100] suggest to use sampling to represent polynomials and other non-monomial bases of the polynomial ring $\mathbb{R}[\mathbf{x}]$; see also [34] where promising numerical results are reported for the univariate case, and [141].

Another issue, which is discussed in detail by Waki et al. [173], is that numerical errors in the computations may lead the SDP solver to return the wrong value. As an example they consider the problem

$$p^{\min} := \min_{x \in K} x \;\text{ where } K := \{x \in \mathbb{R} \mid x \geq 0, x^2 - 1 \geq 0\}. \tag{6.17}$$

Obviously, $p^{\min} = 1$ and it is not difficult to check that $p_t^{\text{mom}} = p_t^{\text{sos}} = 0$ for all $t \geq 1$. However, as noted in [173], the values returned e.g. by the SDP solver SeDuMi when solving the SOS/moment relaxations of order $t = 5, 6$ are approximatively equal to 1. So in practice the relaxations converge to $p_{\min}$, while this is not true in theory! A fact which might help explain this strange behaviour is that, while $\mathbf{x} - 1$ does not have any decomposition $s_0 + \mathbf{x}s_1 + (\mathbf{x}^2 - 1)s_2$ with $s_0, s_1, s_2 \in \Sigma$, a small perturbation of it does have such a decomposition. Namely, for any $\epsilon > 0$ there exists $r \in \mathbb{N}$ for which the perturbed polynomial $\mathbf{x} - 1 + \epsilon\mathbf{x}^{2r}$ has a decomposition $s_0 + \mathbf{x}s_1 + (\mathbf{x}^2 - 1)s_2$ with $s_0, s_1, s_2 \in \Sigma$. So it seems that, due to numerical errors introduced in the course of computation, the SDP solver is not solving the original problem (6.17) but some perturbation of it which behaves better. As a remedy to this Waki et al. [173] propose to use a solver using multiple-precision computations like SDPA-GMP [44]; applied to the above problem this solver indeed returns the correct value 0 for the moment/SOS relaxations of order 5. Henrion and Lasserre [62] had also reported similar findings about minimizing the Motzkin polynomial.

6.7. Extracting global minimizers. We explain here how to extract global minimizers to the problem (1.1) assuming we are in the situation of Theorem 6.18. That is, $y \in \mathbb{R}^{\mathbb{N}^n_{2t}}$ is an optimal solution to the program (6.3) satisfying the rank condition (6.16). Then, as claimed in Theorem 6.18 (and its proof), $p_t^{\text{mom}} = p^{\min}$, y has a r-atomic representing measure $\mu = \sum_{i=1}^r \lambda_i \delta_{v_i}$, where $\lambda_i > 0$, $\sum_{i=1}^r \lambda_i = 1$, $r = \operatorname{rank} M_s(y)$, and $V_{\mathbb{C}}(\operatorname{Ker} M_s(y)) = \{v_1, \ldots, v_r\} \subseteq K_p^{\min}$, the set of optimal solutions to (1.1). The question we now address is how to find the v_i's from the moment matrix $M_s(y)$. We present the method proposed by Henrion and Lasserre [62], although our description differs in some steps and follows the implementation proposed by Jibetean and Laurent [69] and presented in detail in Lasserre et al. [90].

Denote by $\tilde{y}$ the (infinite) sequence of moments of the measure μ. Then, $M(\tilde{y})$ is a flat extension of $M_s(y)$. Hence, by Theorem 5.29, $\mathcal{I} := \operatorname{Ker} M(\tilde{y}) = (\operatorname{Ker} M_s(y))$ and any set $\mathcal{B} \subseteq \mathbb{T}^n_{s-1}$ indexing a maximum nonsingular principal submatrix of $M_{s-1}(y)$ is a basis of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. One can now determine $V_{\mathbb{C}}(\operatorname{Ker} M_s(y)) = V_{\mathbb{C}}(\mathcal{I})$ with the eigenvalue method presented in Section 2.4.

In a first step we determine a subset $\mathcal{B} \subseteq \mathbb{T}^n_{s-d_K}$ indexing a maximum nonsingular principal submatrix of $M_s(y)$. We can find such set $\mathcal{B}$ in a 'greedy manner', by computing the successive ranks of the north-east corner principal submatrices of $M_{s-d_K}(y)$. Starting from the constant monomial 1, we insert in $\mathcal{B}$ as many low degree monomials as possible.

In a second step, for each $i = 1, \dots, n$, we construct the multiplication matrix $M_{\mathbf{x}_i}$ of the 'multiplication by $\mathbf{x}_i$' operator $m_{\mathbf{x}_i}$ (recall (2.6)) with respect to the basis $\mathcal{B}$ of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. By definition, for $\mathbf{x}^\beta \in \mathcal{B}$, the $\mathbf{x}^\beta$th column of $M_{\mathbf{x}_i}$ contains the residue of the monomial $\mathbf{x}_i\mathbf{x}^\beta$ modulo $\mathcal{I}$ w.r.t. the basis $\mathcal{B}$. That is, setting $M_{\mathbf{x}_i} := (a^{(i)}_{\alpha,\beta})_{\mathbf{x}^\alpha,\mathbf{x}^\beta \in \mathcal{B}}$, the polynomial $\mathbf{x}_i\mathbf{x}^\beta - \sum_{\mathbf{x}^\alpha \in \mathcal{B}} a^{(i)}_{\alpha,\beta}\mathbf{x}^\alpha$ belongs to $\mathcal{I}$ and thus to $\operatorname{Ker} M_s(y)$. From this we immediately derive the following explicit characterization for $M_{\mathbf{x}_i}$ from the moment matrix $M_s(y)$.

LEMMA 6.21. *Let M_0 denote the principal submatrix of $M_s(y)$ indexed by $\mathcal{B}$ and let U_i denote the submatrix of $M_s(y)$ whose rows are indexed by $\mathcal{B}$ and whose columns are indexed by the set $\mathbf{x}_i\mathcal{B} := \{\mathbf{x}_i\mathbf{x}^\alpha \mid \mathbf{x}^\alpha \in \mathcal{B}\}$. Then, $M_{\mathbf{x}_i} = M_0^{-1}U_i$.*

Given a polynomial $h \in \mathbb{R}[\mathbf{x}]$, the multiplication matrix of the 'multiplication by h' operator w.r.t. the basis $\mathcal{B}$ is then given by $M_h = h(M_{\mathbf{x}_1}, \dots, M_{\mathbf{x}_n})$. In view of Theorem 2.9, the eigenvectors of M_h^T are the vectors $\zeta_{\mathcal{B},v} = (v^\alpha)_{\mathbf{x}^\alpha \in \mathcal{B}}$ with respective eigenvalues $h(v)$ for $v \in V_\mathbb{C}(\mathcal{I})$. A nice feature of the ideal $\mathcal{I} = \operatorname{Ker} M(\tilde{y}) = (\operatorname{Ker} M_s(y))$ is that it is (real) radical. Hence, if the values $h(v)$ for $v \in V_\mathbb{C}(\mathcal{I})$ are all distinct, then the matrix M_h^T is non-derogatory, i.e., its eigenspaces are 1-dimensional and spanned by the vectors $\zeta_{\mathcal{B},v}$ (for $v \in V_\mathbb{C}(\mathcal{I})$) (recall Lemma 2.12). In that case, one can recover the vectors $\zeta_{\mathcal{B},v}$ directly from the right eigenvectors of M_h. Then it is easy - in fact, immediate if $\mathcal{B}$ contains the monomials $\mathbf{x}_1, \dots, \mathbf{x}_n$ - to recover the components of v from the vector $\zeta_{\mathcal{B},v}$. According to [24], if we choose h as a random linear combination of the monomials $\mathbf{x}_1, \dots, \mathbf{x}_n$ then, with high probability, the values $h(v)$ at the distinct points of $V_\mathbb{C}(\mathcal{I})$ are all distinct.

6.8. Software and examples. Several software packages have been developed for computing sums of squares of polynomials and optimizing polynomials over semialgebraic sets.

- **GloptiPoly**, developed by Henrion and Lasserre [61], implements the moment/SOS hierarchies (6.3), (6.2), and the techniques described in this section for testing optimality and extracting global optimizers. See `http://www.laas.fr/~henrion/software/gloptipoly/` The software has been recently updated to treat more general moment problems; cf. [63].
- **SOSTOOLS**, developed by Prajna, Papachristodoulou, Seiler and Parrilo [132], is dedicated to formulate and compute sums of squares optimization programs. See `http://www.mit.edu/~parrilo/sostools/`

order t	rank sequence	bound p_t^{mom}	solution extracted
1	(1,7)	unbounded	none
2	(**1**,**1**, 21)	-310	(5,1,5,0,5,10)

Table 1
Moment relaxations for Example 6.22

• **SparsePOP**, developed by Waki, Kim, Kojima and Muramatsu [172], implements sparse moment/SOS relaxations for polynomial optimization problems having some sparsity pattern (see Section 8.1). See `http://www.is.titech.ac.jp/~kojima/SparsePOP/`
• **Yalmip**, developed by Löfberg, is a MATLAB toolbox for rapid prototyping of optimization problems, which implements in particular the sum-of-squares and moment based approaches. See `http://control.ee.ethz.ch/~joloef/yalmip.php`

We conclude with some small examples. See e.g. [78, 62, 90] for more examples.

Example 6.22. *Consider the problem:*

$$\begin{array}{ll} \min & p(x) = -25(x_1-2)^2 - (x_2-2)^2 - (x_3-1)^2 \\ & \qquad -(x_4-4)^2 - (x_5-1)^2 - (x_6-4)^2 \\ s.t. & (x_3-3)^2 + x_4 \geq 4, \;\; (x_5-3)^2 + x_6 \geq 4 \\ & x_1 - 3x_2 \leq 2, \;\; -x_1 + x_2 \leq 2, \; x_1 + x_2 \leq 6, \\ & x_1 + x_2 \geq 2, \; 1 \leq x_3 \leq 5, \; 0 \leq x_4 \leq 6, \\ & 1 \leq x_5 \leq 5, \; 0 \leq x_6 \leq 10, \; x_1, x_2 \geq 0. \end{array}$$

As shown in Table 1, GloptiPoly finds the optimum value -310 *and a global minimizer* $(5,1,5,0,5,10)$ *at the relaxation of order* $t=2$. *This involves then the computation of a SDP with 209 variables, one semidefinite constraint involving a matrix of size 28 (namley,* $M_2(y) \succeq 0$*) and 16 semidefinite constraints involving matrices size 7 (namely,* $M_1(g_j y) \succeq 0$*, corresponding to the 16 constraints* $g_j \geq 0$ *of degree 1 or 2).*

Example 6.23. *Consider the problem*

$$\begin{array}{ll} \min & p(x) = -x_1 - x_2 \\ s.t. & x_2 \leq 2x_1^4 - 8x_1^3 + 8x_1^2 + 2 \\ & x_2 \leq 4x_1^4 - 32x_1^3 + 88x_1^2 - 96x_1 + 36 \\ & 0 \leq x_1 \leq 3, \; 0 \leq x_2 \leq 4. \end{array}$$

As shown in Table 2, GloptiPoly solves the problem at optimality at the relaxation of order $t = 4$.

Example 6.24. *Consider the problem:*

$$\begin{array}{ll} \min & p(x) = x_1^2 x_2^2 (x_1^2 + x_2^2 - 3x_3^2) + x_3^6 \\ s.t. & x_1^2 + x_2^2 + x_3^2 \leq 1, \end{array}$$

order t	rank sequence	bound p_t^{mom}	solution extracted
2	(1,1,4)	-7	none
3	(1,2,2,4)	-6.6667	none
4	(**1,1,1,1**,6)	-5.5080	(2.3295,3.1785)

Table 2
Moment relaxations for Example 6.23

order t	rank sequence	bound p_t^{mom}	value reached by moment vector
3	(1, 4, 9, 13)	-0.0045964	$7\ 10^{-26}$
4	(1, 4, 10, 20, 29)	-0.00020329	$3\ 10^{-30}$
5	(1, 4, 10, 20, 34, 44)	$-2.8976\ 10^{-5}$	$3\ 10^{-36}$
6	(1, 4, 10, 20, 34, 56, 84)	$-6.8376\ 10^{-6}$	$6\ 10^{-42}$
7	(1, 4, 10, 20, 35, 56, 84, 120)	$-2.1569\ 10^{-6}$	$4\ 10^{-43}$

Table 3
Moment relaxations for Example 6.24

of minimizing the Motzkin form over the unit ball. As we see in Table 3, the moment bounds p_t^{mom} converge to $p^{min} = 0$, but optimality cannot be detected via the rank condition (6.16) since it is never satisfied. This is to be expected since p has infinitely many global minimizers over the unit ball. However the criterion (6.15) applies here; indeed GloptiPoly returns that the relaxed vector $x^* := (y_{e_i})_{i=1}^3$ (where y is the optimum solution to the moment relaxation) is feasible (i.e. lies in the unit ball) and reaches the objective value which is mentioned in the last column of Table 3; here $x^* \sim 0$.

Example 6.25. *Consider the problem*

$$\begin{array}{ll} \min & p(x) = x_1^2x_2^2(x_1^2+x_2^2-3x_3^2)+x_3^6+\epsilon(x_1^6+x_2^6+x_3^6) \\ s.t. & x_1^2+x_2^2+x_3^2 \leq 1, \end{array}$$

of minimizing the perturbed Motzkin form over the unit ball. For any $\epsilon > 0$, $p^{min} = 0$ and p is positive, i.e. the origin is the unique global minimizer. Moreover, p_ϵ is SOS if and only if $\epsilon \geq \epsilon^* \sim 0.01006$ [171]. Hence, as explained in Example 6.19, it is to be expected that for $\epsilon < \epsilon^*$, the rank condition (6.16) does not hold. This is confirmed in Table 4 which gives results for $\epsilon = 0.01$. Again the criterion (6.15) applies, i.e. the moment vector y yields the global minimizer $x^* = (y_{e_i})_{i=1}^3$, $x^* \sim 0$, and the last column gives the value of p_ϵ evaluated at x^*.

The following example (taken from [90]) illustrates how the method can be used to compute the real roots to a system of polynomial equations: $h_1(x) = 0, \dots, h_m(x) = 0$. Indeed this problem can be cast as finding the

t	rank sequence	bound p_t^{mom}	value reached by moment vector
3	(1, 4, 9, 13)	$-2.11\ 10^{-5}$	$1.67\ 10^{-44}$
4	(1, 4, 10, 20, 35)	$-1.92\ 10^{-9}$	$4.47\ 10^{-60}$
5	(1, 4, 10, 20, 35, 56)	$2.94\ 10^{-12}$	$1.26\ 10^{-44}$
6	(1, 4, 10, 20, 35, 56, 84)	$3.54\ 10^{-12}$	$1.5\ 10^{-44}$
7	(1, 4, 10, 20, 35, 56, 84, 120)	$4.09\ 10^{-12}$	$2.83\ 10^{-43}$
8	(1, 4, 10, 20, 35, 56, 84, 120, 165)	$4.75\ 10^{-12}$	$5.24\ 10^{-44}$

Table 4

Moment relaxations for Example 6.25

global minimizers of the minimization problem:

$$\min\ p(x) := 1 \text{ s.t. } h_1(x) = 0, \ldots, h_m(x) = 0.$$

Thus when the system has finitely many real roots one can find them by solving a sufficiently large semidefinite moment relaxation.

Example 6.26. *Consider the polynomials*

$$\begin{aligned} h_1 &= 5\mathbf{x}_1^9 - 6\mathbf{x}_1^5\mathbf{x}_2 + \mathbf{x}_1\mathbf{x}_2^4 + 2\mathbf{x}_1\mathbf{x}_3, \\ h_2 &= -2\mathbf{x}_1^6\mathbf{x}_2 + 2\mathbf{x}_1^2\mathbf{x}_2^3 + 2\mathbf{x}_2\mathbf{x}_3, \\ h_3 &= \mathbf{x}_1^2 + \mathbf{x}_2^2 - 0.265625. \end{aligned}$$

Table 5 shows the rank sequences for the moment relaxations of order $t = 5, \ldots, 8$. Here we have $d_K = 5$, where $K := \{x \in \mathbb{R}^3 \mid h_1(x) = 0, \ldots, h_3(x) = 0\}$. At order $t = 8$, the flatness condition holds, but we have only $\operatorname{rank} M_4(y) = \operatorname{rank} M_2(y)$, while the condition (6.16) would require $\operatorname{rank} M_s(y) = \operatorname{rank} M_{s-5}(y)$ for some $s \geq 5$. Thus at order $t = 8$, we can extract 8 real points, but we cannot claim that they lie in the set K. However it suffices to check afterwards that the extracted points lie in K, i.e. satisfy the equations $h_1 = 0, \ldots, h_3 = 0$. The column ’accuracy’ shows the quantity $\max_{j,v} |h_j(v)|$, where v runs over the extracted points and $j = 1, 2, 3$. The extracted real roots are:

$$\begin{aligned} x_1 &= (-0.515, -0.000153, -0.0124)\,, \\ x_2 &= (-0.502, 0.119, 0.0124)\,, \\ x_3 &= (0.502, 0.119, 0.0124)\,, \\ x_4 &= (0.515, -0.000185, -0.0125)\,, \\ x_5 &= (0.262, 0.444, -0.0132)\,, \\ x_6 &= (-2.07e\text{-}5, 0.515, -1.27e\text{-}6)\,, \\ x_7 &= (-0.262, 0.444, -0.0132)\,, \\ x_8 &= (-1.05e\text{-}5, -0.515, -7.56e\text{-}7)\,. \end{aligned}$$

order t	rank sequence	accuracy
5	1 4 8 16 25 34	—
6	1 3 9 15 22 26 32	—
7	1 3 8 10 12 16 20 24	—
8	1 4 8 8 8 12 16 20 24	4.6789e-5

TABLE 5
Moment relaxations for Example 6.26

Note that there are 20 complex roots in total.

7. Application to optimization - Some further selected topics.

7.1. Approximating positive polynomials by sums of squares.

We now come back to the comparison between nonnegative polynomials and sums of squares of polynomials. As we saw earlier, the parameters (n, d) for which every nonnegative polynomial of degree d in n variables is a sum of squares have been characterized by D. Hilbert; namely, they are ($n = 1, d$ even), ($n \geq 1, d = 2$), and ($n = 2, d = 4$). Thus for any other pair (n, d) (d even) there exist nonnegative polynomials that cannot be written as a sum of squares. A natural question is whether there are many or few such polynomials. Several answers may be given depending whether the degree and the number of variables are fixed or not. First, on the negative side, Blekherman [16] has shown that when the degree d is fixed but the number n of variables grows, then there are significantly more positive polynomials than sums of squares. More precisely, for $d \in \mathbb{N}$ even, consider the cone $\mathbf{H}_d$ (resp., $\mathbf{\Sigma}_d$) of homogeneous polynomials $p \in \mathbb{R}[\mathbf{x}]$ of degree d that are nonnegative on $\mathbb{R}^n$ (resp., a sum of squares). In order to compare the two cones, Blekherman considers their sections $\widehat{\mathbf{H}}_d := \mathbf{H}_d \cap H$ and $\widehat{\mathbf{\Sigma}}_d := \mathbf{\Sigma}_d \cap H$ by the hyperplane $H := \{p \mid \int_{S^{n-1}} p(x)\mu(dx) = 1\}$, where μ is the rotation invariant probability measure on the unit sphere S^{n-1}.

THEOREM 7.1. *[16] There exist constants $C_1, C_2 > 0$ depending on d only such that for any n large enough,*

$$C_1 n^{(d/2-1)/2} \leq \left(\frac{vol\ \widehat{\mathbf{H}}_d}{vol\ \widehat{\mathbf{\Sigma}}_d} \right)^{1/D} \leq C_2 n^{(d/2-1)/2},$$

where $D := \binom{n+d-1}{d} - 1$.

However, on the positive side, Berg [11] has shown that, when the number of variables is fixed but the degree is variable, then the cone of sums of squares is dense in the cone of polynomials nonnegative on $[-1, 1]^n$. While Berg's result is existential, Lasserre and Netzer [92] have provided an explicit and very simple sum of squares approximation, which we present in Theorem 7.2 below. Previously, Lasserre [84] had given an analogous result for polynomials nonnegative on the whole space $\mathbb{R}^n$, presented in Theorem 7.3 below. To state the results we need the following polynomials:

$$\theta_t := \sum_{k=0}^{t} \sum_{i=1}^{n} \frac{\mathbf{x}_i^{2k}}{k!}, \quad \Theta_t := 1 + \sum_{i=1}^{n} \mathbf{x}_i^{2t}, \tag{7.1}$$

defined for any $t \in \mathbb{N}$.

Theorem 7.2. *[92] Let $f \in \mathbb{R}[\mathbf{x}]$ be a polynomial nonnegative on $[-1,1]^n$ and let Θ_t be as in (7.1). For any $\epsilon > 0$, there exists $t_0 \in \mathbb{N}$ such that the polynomial $f + \epsilon\Theta_t$ is a sum of squares for all $t \geq t_0$.*

Theorem 7.3. *[84] Let $f \in \mathbb{R}[\mathbf{x}]$ be a polynomial nonnegative on $\mathbb{R}^n$ and let θ_t be as in (7.1). For any $\epsilon > 0$, there exists $t_0 \in \mathbb{N}$ such that $f + \epsilon\theta_t$ is a sum of squares for all $t \geq t_0$.*

In both cases the proof relies on a result about existence of a representing measure, combined with some elementary bounds on the entries of positive semidefinite moment matrices. For Theorem 7.2 we use the result from Theorem 4.12 about existence of a representing measure for bounded sequences. On the other hand, for Theorem 7.3, we need the following (more difficult) result of Carleman (for the case $n = 1$) and Nussbaum (for $n \geq 1$). Recall that $e_1, \ldots, e_n$ denote the standard unit vectors in $\mathbb{R}^n$. Thus, for $y \in \mathbb{R}^{\mathbb{N}^n}$, y_{2ke_i} is its entry indexed by $2ke_i$, i.e. $y_{2ke_i} = y_{(0,\ldots,0,2k,0,\ldots,0)}$ where $2k$ is at the ith position.

Theorem 7.4. *[119] Given $y \in \mathbb{R}^{\mathbb{N}^n}$, if $M(y) \succeq 0$ and*

$$\sum_{k=0}^{\infty} y_{2ke_i}^{-1/2k} = \infty \quad (i = 1, \ldots, n) \tag{7.2}$$

then y has a (unique) representing measure.

Observe that Theorem 7.2 in fact implies Theorem 7.3. Indeed, if $f \geq 0$ on $\mathbb{R}^n$, then $f \geq 0$ on $[-1,1]^n$ and thus, by Theorem 7.2, there exist $\epsilon > 0$ and $t_0 \in \mathbb{N}$ for which $f + \epsilon\Theta_t$ is SOS for all $t \geq t_0$. It suffices now to observe that $\theta_t = s + \frac{1}{t!}\Theta_t$, where $s = \sum_{k=1}^{t-1} \sum_{i=1}^{n} \frac{\mathbf{x}_i^{2k}}{k!} + \frac{-1+nt!}{t!}$ is SOS, which shows that $f + \epsilon t!\theta_t$ is SOS.

In what follows we first give the proof of Theorem 7.2 and, although it implies Theorem 7.3, we also give a direct proof for Theorem 7.3 since it might be instructive to see how Carleman's result (Theorem 7.4) is used in the proof.

7.1.1. Bounds on entries of positive semidefinite moment matrices. We begin with some elementary bounds from [84, 92] on the entries of $M_t(y)$, which will be used in the proof, and might be of independent interest. As we now see, when $M_t(y) \succeq 0$, all entries y_α can be bounded in terms of y_0 and $y_{(2t,0,\dots,0)}$, $\dots$, $y_{(0,\dots,0,2t)}$, the entries indexed by the constant monomial 1 and the highest order monomials $\mathbf{x}_1^{2t},\dots,\mathbf{x}_n^{2t}$. For $0 \le k \le t$, set

$$\tau_k := \max(y_{(2k,0,\dots,0)},\dots,y_{(0,\dots,0,2k)}) = \max_{i=1,\dots,n} y_{2ke_i};$$

thus $\tau_0 = y_0$. We will use the inequality $y^2_{\alpha+\beta} \le y_{2\alpha}y_{2\beta}$ (for $\alpha,\beta \in \mathbb{N}^n_t$), which follows from the fact that the submatrix of $M_t(y)$ indexed by $\{\alpha,\beta\}$ is positive semidefinite.

LEMMA 7.5. *Assume $M_t(y) \succeq 0$ and $n = 1$. Then $y_{2k} \le \max(\tau_0,\tau_t)$ for $0 \le k \le t$.*

Proof. The proof is by induction on $t \ge 0$. If $t = 0, 1$, the result is obvious. Assume $t \ge 1$ and the result holds for $t-1$, i.e. $y_0,\dots,y_{2t-2} \le \max(y_0, y_{2t-2})$; we show that $y_0,\dots,y_{2t} \le \max(y_0,y_{2t})$. This is obvious if $y_0 \ge y_{2t-2}$. Assume now $y_0 \le y_{2t-2}$. As $y^2_{2t-2} \le y_{2t-4}y_{2t} \le y_{2t-2}y_{2t}$, we deduce $y_{2t-2} \le y_{2t}$ and thus $y_0,\dots,y_{2t} \le y_{2t} = \max(y_0,y_{2t})$. □

LEMMA 7.6. *Assume $M_t(y) \succeq 0$. Then $y_{2\alpha} \le \tau_k$ for all $|\alpha| = k \le t$.*

Proof. The case $n = 1$ being obvious, we first consider the case $n = 2$. Say $s := \max_{|\alpha|=k} y_{2\alpha}$ is attained at $y_{2\alpha^*}$. As $2\alpha^*_1 \ge k \iff 2\alpha^*_2 \le k$, we may assume w.l.o.g. $2\alpha^*_1 \ge k$. Write $2\alpha^* = (k,0) + (2\alpha^*_1 - k, 2\alpha^*_2) = (k,0) + (k - 2\alpha^*_2, 2\alpha^*_2)$. Then $y^2_{2\alpha^*} \le y_{(2k,0)}y_{(2k-4\alpha^*_2,4\alpha^*_2)}$. Now $y^2_{2\alpha^*} = s^2$, $y_{(2k-4\alpha^*_2,4\alpha^*_2)} \le s$, $y_{(2k,0)} \le \tau_k$, which implies $s \le \tau_k$. This shows the result in the case $n = 2$.

Assume now $n \ge 3$ and the result holds for $n-1$. Thus $y_{2\alpha} \le \tau_k$ if $|\alpha| = k$ and $\alpha_i = 0$ for some i. Assume now $1 \le \alpha_1 \le \dots \le \alpha_n$. Consider the sequences $\gamma := (2\alpha_1, 0, \alpha_3 + \alpha_2 - \alpha_1, \alpha_4,\dots,\alpha_n)$ and $\gamma' := (0, 2\alpha_2, \alpha_3 + \alpha_1 - \alpha_2, \alpha_4, \dots, \alpha_n)$. Thus $|\gamma| = |\gamma'| = |\alpha| = k$, $\gamma + \gamma' = 2\alpha$. As $\gamma_2 = \gamma'_1 = 0$, we have $y_{2\gamma}, y_{2\gamma'} \le \tau_k$. Hence $y^2_{2\alpha} = y^2_{\gamma+\gamma'} \le y_{2\gamma}y_{2\gamma'} \le \tau^2_k$, implying $y_{2\alpha} \le \tau_k$. □

COROLLARY 7.7. *Assume $M_t(y) \succeq 0$. Then $|y_\alpha| \le \max_{0\le k\le t} \tau_k = \max(\tau_0,\tau_t)$ for all $|\alpha| \le 2t$.*

Proof. Using Lemma 7.5, we see that $y_{(2k,0,\dots,0)} \le \max(y_0, y_{2t,0,\dots,0)}) \le \max(\tau_0,\tau_t)$, implying $\tau_k \le \max(\tau_0,\tau_t)$ and thus $\max_{0\le k\le t}\tau_k = \max(\tau_0,\tau_t) =: \tau$. By Lemma 7.6 we deduce $y_{2\alpha} \le \tau$ for $|\alpha| \le t$. Consider now $|\gamma| \le 2t$. Write $\gamma = \alpha + \beta$ with $|\alpha|, |\beta| \le t$. Then $y^2_\gamma \le y_{2\alpha}y_{2\beta} \le \tau^2$, giving $|y_\gamma| \le \tau$. □

We mention for completeness another result for bounding entries of a positive semidefinite moment matrix. This result is used in [86] for giving an explicit set of conditions ensuring that a polynomial p is a sum of squares, the conditions depending only on the coefficients of p.

LEMMA 7.8. *[86] If $M_t(y) \succeq 0$ and $y_0 = 1$, then $|y_\alpha|^{1/|\alpha|} \le \tau_t^{1/2t}$ for all $|\alpha| \le 2t$.*

Proof. Use induction on $t \geq 1$. The result holds obviously for $t = 1$. Assume the result holds for $t \geq 1$ and let $M_{t+1}(y) \succeq 0$, $y_0 = 1$. By the induction assumption, $|y_\alpha|^{1/|\alpha|} \leq \tau_t^{1/2t}$ for $|\alpha| \leq 2t$. By Lemma 7.6, $|y_\alpha| \leq \tau_{t+1}$ for $|\alpha| = 2t+2$. We first show $\tau_t^{1/t} \leq \tau_{t+1}^{1/(t+1)}$. For this, say $\tau_t = y_{2te_1}$; then $\tau_t^2 = y_{2te_1}^2 \leq y_{2(t+1)e_1} y_{2(t-1)e_1} \leq \tau_{t+1}\tau_t^{(2t-2)/2t}$, which gives $\tau_t^{1/t} \leq \tau_{t+1}^{1/(t+1)}$. Remains only to show that $|y_\alpha|^{1/|\alpha|} \leq \tau_{t+1}^{1/t+1}$ for $|\alpha| = 2t+1$ (as the case $|\alpha| \leq 2t$ follows using the induction assumption, and $|\alpha| = 2t+2$ has been settled above). Say, $|\alpha| = 2t+1$ and $\alpha = \beta + \gamma$ with $|\beta| = t$, $|\gamma| = t+1$. Then $y_\alpha^2 \leq y_{2\beta} y_{2\gamma} \leq \tau_t \tau_{t+1} \leq \tau_{t+1}^{t/(t+1)} \tau_{t+1} = \tau_{t+1}^{(2t+1)/(t+1)}$, giving the desired result. □

7.1.2. Proof of Theorem 7.2. The following result is crucial for the proof of Theorem 7.2.

PROPOSITION 7.9. *Given a polynomial $f \in \mathbb{R}[\mathbf{x}]$ consider the program*

$$\epsilon_t^* := \inf f^T y \text{ s.t. } M_t(y) \succeq 0,\ y^T \Theta_t \leq 1 \tag{7.3}$$

for any integer $t \geq d_f = \lceil \deg(f)/2 \rceil$. Recall the polynomial Θ_t from (7.1). Then,

(i) $-\infty < \epsilon_t^* \leq 0$ *and the infimum is attained in (7.3).*

(ii) *For $\epsilon \geq 0$, the polynomial $f + \epsilon\Theta_t$ is a sum of squares if and only if $\epsilon \geq -\epsilon_t^*$. In particular, f is a sum of squares if and only if $\epsilon_t^* = 0$.*

(iii) *If the polynomial $f \in \mathbb{R}[\mathbf{x}]$ is nonnegative on $[-1,1]^n$, then $\lim_{t\to\infty} \epsilon_t^* = 0$.*

Proof. Let y be feasible for the program (7.3). Then $0 \leq y_0, y_{(2t,0,\ldots,0)}, \cdots, y_{(0,\ldots,0,2t)} \leq 1$ (from the linear constraint $y^T\Theta_t \leq 1$) which, using Corollary 7.7, implies $|y_\alpha| \leq 1$ for all α. Hence the feasible region of (7.3) is bounded and nonempty (as $y = 0$ is feasible). Therefore the infimum is attained in program (7.3) and $-\infty < \epsilon_t^* \leq 0$, showing (i). One can verify that the dual semidefinite program of (7.3) reads

$$d_t^* := \sup_{\lambda \geq 0} -\lambda \ \text{ s.t. } f + \lambda\Theta_t \text{ is a sum of squares.}$$

As the program (7.3) is strictly feasible (choose for y the sequence of moments of a measure with positive density on $\mathbb{R}^n$, with finite moments up to order $2t$, rescaled so as to satisfy $y^T\Theta_t \leq 1$), its dual semidefinite program attains it supremum and there is no duality gap, i.e. $\epsilon_t^* = d_t^*$. Thus $f + \epsilon\Theta_t$ is a sum of squares if and only if $-\epsilon \leq d_t^* = \epsilon_t^*$, i.e. $\epsilon \geq -\epsilon_t^*$, showing (ii).

We now show (iii). Say $\epsilon_t^* = f^T y^{(t)}$, where $y^{(t)}$ is an optimum solution to (7.3) with, as we saw above, $y^{(t)} \in [-1,1]^{\mathbb{N}^n_{2t}}$. Complete $y^{(t)}$ to a sequence $\tilde{y}^{(t)} = (y^{(t)}, 0, \ldots, 0) \in [-1,1]^{\mathbb{N}^n}$. As $[-1,1]^{\mathbb{N}^n}$ is compact, there exists a converging subsequence $(y^{(t_l)})_{l\geq 0}$, converging to $y^* \in [-1,1]^{\mathbb{N}^n}$ in the product topology. In particular, there is coordinate-wise convergence, i.e. $(y_\alpha^{(t_l)})_{l\geq 0}$ converges to y_α^*, for all α. Therefore $M(y^*) \succeq 0$.

As $y^* \in [-1,1]^{\mathbb{N}^n}$, we deduce using Theorem 4.12 that y^* has a representing measure μ on $[-1,1]^n$. In particular, $\epsilon^*_{t_l} = f^T y^{(t_l)}$ converges to $f^T y^* = \int_{[-1,1]^n} f(x)\mu(dx)$. By assumption, $f \geq 0$ on $[-1,1]^n$ and thus $f^T y^* \geq 0$. On the other hand, $f^T y^* \leq 0$ since $\epsilon^*_t \leq 0$ for all t. Thus $f^T y^* = 0$. This shows that the only accumulation point of the sequence ϵ_t is 0 and thus ϵ_t converges to 0. □

We can now conclude the proof of Theorem 7.2. Let $\epsilon > 0$. By Proposition 7.9 (iii), $\lim_{t\to\infty} \epsilon^*_t = 0$. Hence there exists $t_0 \in \mathbb{N}$ such that $\epsilon^*_t \geq -\epsilon$ for all $t \geq t_0$. Applying Proposition 7.9 (ii), we deduce that $f + \epsilon\Theta_t$ is a sum of squares.

As an example, consider the univariate polynomial $f = 1 - \mathbf{x}^2$, obviously nonnegative on $[-1,1]$. Then, for $\epsilon \geq (t-1)^{t-1}/t^t$, the polynomial $f + \epsilon\mathbf{x}^{2t}$ is nonnegative on $\mathbb{R}$ and thus a sum of squares (see [92] for details).

7.1.3. Proof of Theorem 7.3. We now turn to the proof of Theorem 7.3, whose details are a bit more technical. Given an integer $M > 0$, consider the program

$$\mu^*_M := \inf_{\mu} \int f(x)\mu(dx) \text{ s.t. } \int \sum_{i=1}^n e^{x_i^2}\mu(dx) \leq ne^{M^2}, \tag{7.4}$$

where the infimum is taken over all probability measures μ on $\mathbb{R}^n$.

LEMMA 7.10. *Let $f \in \mathbb{R}[\mathbf{x}]$ and assume $f^{min} := \inf_{x\in\mathbb{R}^n} f(x) > -\infty$. Then $\mu^*_M \downarrow f^{min}$ as $M \to \infty$.*

Proof. Obviously, the sequence $(\mu^*_M)_M$ is monotonically non-increasing and $\mu^*_M \geq f^{\min}$. Next observe that $\mu^*_M \leq \inf_{\|x\|_\infty \leq M} f(x)$ since the Dirac measure $\mu = \delta_x$ at any point x with $\|x\|_\infty \leq M$ is feasible for (7.4) with objective value $f(x)$. Now $\inf_{\|x\|_\infty \leq M} f(x)$ converges to $f^{\min}$ as $M \to \infty$, which implies that $\mu^*_M \downarrow f^{\min}$ as $M \to \infty$. □

The idea is now to approach the optimum value of (7.4) via a sequence of moment relaxations. Namely, for any integer $t \geq d_f = \lceil \deg(f)/2 \rceil$, consider the semidefinite program

$$\epsilon^*_{t,M} := \inf f^T y \text{ s.t. } M_t(y) \succeq 0,\ y_0 = 1,\ y^T\theta_t \leq ne^{M^2} \tag{7.5}$$

whose dual reads

$$d^*_{t,M} := \sup_{\gamma,\lambda} \gamma - \lambda ne^{M^2} \text{ s.t. } \lambda \geq 0,\ \gamma + \lambda\theta_r \text{ is a sum of squares.} \tag{7.6}$$

The next result is crucial for the proof of Theorem 7.3.

PROPOSITION 7.11. *Fix $M > 0$, $t \geq d_f$, and assume $f^{min} > -\infty$. The following holds for the programs (7.5) and (7.6).*

(i) *The optimum is attained in both programs (7.5) and (7.6) and there is no duality gap, i.e. $\epsilon^*_{t,M} = d^*_{t,M}$.*

(ii) $\epsilon^*_{t,M} \uparrow \mu^*_M$ *as* $t \to \infty$.

Proof. (i) As (7.5) is strictly feasible, its dual (7.6) attains its optimum and there is no duality gap. The infimum is attained in program (7.5) since the feasible region is bounded (directly using the constraint $y^T\theta_t \leq ne^{M^2}$ together with Corollary 7.7) and nonempty (as $y = (1, 0, \dots, 0)$ is feasible for (7.5)).

(ii) We begin with observing that $(\epsilon^*_{t,M})_t$ is monotonically nondecreasing; hence $\lim_{t\to\infty} \epsilon^*_{t,M} = \sup_t \epsilon^*_{t,M}$. Let μ be feasible for (7.4) and let y be its sequence of moments. Then, for any integer $t \geq d_f$, $\int f(x)\mu(dx) = f^T y$, $M_t(y) \succeq 0$, $y_0 = 1$ and, as $\sum_{k=0}^\infty x_i^{2k}/k! = e^{x_i^2}$, the constraint $\int \sum_{i=1}^n e^{x_i^2}\mu(dx) \leq ne^{M^2}$ implies $y^T\theta_t \leq ne^{M^2}$. That is, y is feasible for (7.5) and thus $\mu^*_M \geq \epsilon^*_{t,M}$. This shows $\mu^*_M \geq \lim_{t\to\infty} \epsilon^*_{t,M}$.

We now show the reverse inequality. For this we first note that if y is feasible for (7.5), then $\max_{i\leq n, k\leq t} y_{2ke_i} \leq t!ne^{M^2} =: \sigma_t$ and thus $\max_{|\alpha|\leq 2t} |y_\alpha| \leq \sigma_t$ (by Corollary 7.7). Moreover, for any $s \leq t$, $|y_\alpha| \leq \sigma_s$ for $|\alpha| \leq 2s$ (since the restriction of y to $\mathbb{R}^{\mathbb{N}^n_{2s}}$ is feasible for the program (7.5) with s in place of t).

Say $\epsilon^*_{t,M} = f^T y^{(t)}$, where $y^{(t)}$ is an optimum solution to (7.5) (which is attained by (i)). Define $\tilde{y}^{(t)} = (y^{(t)}, 0 \dots 0) \in \mathbb{R}^{\mathbb{N}^n}$ and $\hat{y}^{(t)} \in \mathbb{R}^{\mathbb{N}^n}$ by $\hat{y}^{(t)}_\alpha := \tilde{y}^{(t)}_\alpha/\sigma_s$ if $2s - 1 \leq |\alpha| \leq 2s$, $s \geq 0$. Thus each $\hat{y}^{(t)}$ lies in the compact set $[-1, 1]^{\mathbb{N}^n}$. Hence there is a converging subsequence $(\hat{y}^{(t_l)})_{l\geq 0}$, converging say to $\hat{y} \in [-1, 1]^{\mathbb{N}^n}$. In particular, $\lim_{l\to\infty} \hat{y}^{(t_l)}_\alpha = \hat{y}_\alpha$ for all α. Define $y^* \in \mathbb{R}^{\mathbb{N}^n}$ by $y^*_\alpha := \sigma_s \hat{y}_\alpha$ for $2s - 1 \leq |\alpha| \leq 2s$, $s \geq 0$. Then $\lim_{l\to\infty} \tilde{y}^{(t_l)}_\alpha = y^*_\alpha$ for all α and $\lim_{l\to\infty} y^{(t_l)}_\alpha = y^*_\alpha$ for all $|\alpha| \leq 2t_l$. From this follows that $M(y^*) \succeq 0$, $y^*_0 = 1$, and $(y^*)^T\theta_r \leq ne^{M^2}$ for any $r \geq 0$. In particular, $\sum_{k=0}^\infty \sum_{i=1}^n \frac{y^*_{2ke_i}}{k!} \leq ne^{M^2}$, which implies[6] $\sum_{k=0}^\infty (y_{2ke_i})^{-1/2k} = \infty$ for all i. That is, condition (7.2) holds and thus, by Theorem 7.4, y^* has a representing measure μ. As μ is feasible for (7.4), this implies $\mu^*_M \leq \int f(x)\mu(dx) = f^T y^* = \lim_{l\to\infty} f^T y^{(t_l)} = \lim_{l\to\infty} \epsilon^*_{t_l,M}$. Hence we find $\mu^*_M \leq \lim_{l\to\infty} \epsilon^*_{t_l,M} \leq \lim_{t\to\infty} \epsilon^*_{t,M} \leq \mu^*_M$ and thus equality holds throughout, which shows (ii). □

We can now conclude the proof of Theorem 7.3. We begin with two easy observations. First it suffices to show the existence of $t_0 \in \mathbb{N}$ for which $f + \epsilon\theta_{t_0}$ is a sum of squares (since this obviously implies that $f + \epsilon\theta_t$ is a sum of squares for all $t \geq t_0$). Second we note that it suffices to show the result for the case $f^{\min} > 0$. Indeed, if $f^{\min} = 0$, consider the polynomial $g := f + n\epsilon/2$ with $g^{\min} = n\epsilon/2 > 0$. Hence, for some $t_0 \in \mathbb{N}$, $g + (\epsilon/2)\theta_{t_0}$ is a sum of squares. As $(\epsilon/2)(\theta_{t_0} - n)$ is a sum of squares, we find that $f + \epsilon\theta_{t_0} = g + (\epsilon/2)\theta_{t_0} + (\epsilon/2)(\theta_{t_0} - n)$ is a sum of squares.

[6] Indeed if $a_k > 0$, $C \geq 1$ satisfy $a_k \leq Ck!$ for all $k \geq 1$, then $a_k \leq Ck^k$, implying $a_k^{-1/2k} \geq C^{-1/2k}/\sqrt{k}$ and thus $\sum_{k\geq 1} a_k^{-1/2k} = \infty$.

So assume $f^{\min} > 0$ and $f^{\min} > 1/M$, where M is a positive integer. By Proposition 7.11 (ii), $\epsilon^*_{t_M,M} \geq \mu^*_M - 1/M \geq f^{\min} - 1/M > 0$ for some integer t_M. By Proposition 7.11 (i), we have $\epsilon^*_{t_M,M} = \gamma_M - \lambda_M n e^{M^2}$, where $\lambda_M \geq 0$ and $f - \gamma_M + \lambda_M \theta_{t_M} =: q$ is a sum of squares. Hence $f + \lambda_M \theta_{t_M} = q + \gamma_M$ is a sum of squares, since $\gamma_M = n\lambda_M e^{M^2} + \epsilon^*_{t_M,M} \geq 0$. Moreover, evaluating at the point 0, we find $f(0) - \gamma_M + \lambda_M n = q(0) \geq 0$, i.e. $f(0) - f^{\min} + f^{\min} - \epsilon^*_{t_M,M} - \lambda_M n e^{M^2} + \lambda_M n \geq 0$. As $f^{\min} - \epsilon^*_{t_M,M} \leq 1/M$, this implies $\lambda_M \leq \frac{1/M + f(0) - f^{\min}}{n(e^{M^2}-1)}$. Therefore, $\lim_{M\to\infty} \lambda_M = 0$. We can now conclude the proof: Given $\epsilon > 0$, choose M in such a way that $f^{\min} > 1/M$ and $\lambda_M < \epsilon$. Then $f + \epsilon\theta_{t_M}$ is a sum of squares.

We refer to [82], [92] for further approximation results by sums of squares for polynomials nonnegative on a basic closed semialgebraic set.

7.2. Unconstrained polynomial optimization. In this section we come back to the unconstrained minimization problem (1.3) which, given a polynomial $p \in \mathbb{R}[\mathbf{x}]$, asks for its infimum $p^{\min} = \inf_{x\in\mathbb{R}^n} p(x)$. There is quite a large literature on this problem; we sketch here only some of the methods that are most relevant to the topic of this survey. We first make some general easy observations. To begin with, we may assume that $\deg(p) =: 2d$ is even, since otherwise $p^{\min} = -\infty$. Probably the most natural idea is to search for global minimizers of p within the critical points of p. One should be careful however. Indeed p may have infinitely many global minimizers, or p may have none! The latter happens, for instance, for the polynomial $p = \mathbf{x}_1^2 + (\mathbf{x}_1\mathbf{x}_2 - 1)^2$; then for $\epsilon > 0$, $p(\epsilon, 1/\epsilon) = \epsilon^2$ converges to 0 as $\epsilon \to 0$, showing $p^{\min} = 0$ but the infimum is *not* attained. Next, how can one recognize whether p has a global minimizer? As observed by Marshall [104], the highest degree homogeneous component of p plays an important role.

LEMMA 7.12. *[104] For a polynomial $p \in \mathbb{R}[\mathbf{x}]$, let $\tilde{p}$ be its highest degree component, consisting of the sum of the terms of p with maximum degree, and let $\tilde{p}^{\min}_S$ denote the minimum of $\tilde{p}$ over the unit sphere.*

(i) *If $\tilde{p}^{\min}_S < 0$ then $p^{min} = -\infty$.*

(ii) *If $\tilde{p}^{\min}_S > 0$ then p has a global minimizer. Moreover any global minimizer x satisfies $\|x\| \leq \max\left(1, \frac{1}{\tilde{p}^{\min}_S} \sum_{1\leq|\alpha|<\deg(p)} |p_\alpha|\right)$.*

Proof. (i) is obvious. (ii) Set $\deg(p) =: d$, $p = \tilde{p} + g$, where all terms of g have degree $\leq d-1$. If $p^{\min} = p(0)$, 0 is a global minimizer and we are done. Otherwise let $x \in \mathbb{R}^n$ with $p(x) \leq p(0)$ and $x \neq 0$. Then, $\tilde{p}(x) \leq p(0) - g(x) \leq \sum_{1\leq|\alpha|\leq d-1} |p_\alpha||x^\alpha|$. Combined with $\tilde{p}(x) = \tilde{p}(x/\|x\|)\|x\|^d \geq \tilde{p}^{\min}_S \|x\|^d$, and $|x^\alpha| \leq \|x\|^{|\alpha|}$ if $\|x\| \geq 1$, this gives $\|x\| \leq R := \max\left(\frac{1}{\tilde{p}^{\min}_S} \sum_{|\alpha|<\deg(p)} |p_\alpha|, 1\right)$. Hence, $p^{\min}$ is equal to the minimum of p over the ball of radius R, which shows existence of global minimizers and that they are all located within this ball. □

No conclusion can be drawn when $\tilde{p}_S^{\min} = 0$; indeed p may have a minimum (e.g. for $p = \mathbf{x}_1^2\mathbf{x}_2^2$), or a finite infimum (e.g. for $p = \mathbf{x}_1^2 + (\mathbf{x}_1\mathbf{x}_2 - 1)^2$), or an infinite infimum (e.g. for $p = \mathbf{x}_1^2 + \mathbf{x}_2$).

We now see how we can apply the general relaxation scheme from Section 6 to the unconstrained polynomial optimization problem (1.3). As there are no constraints, we find just one lower bound for $p^{\min}$:

$$p_t^{\rm sos} = p_t^{\rm mom} = p_d^{\rm sos} = p_d^{\rm mom} \leq p^{\min} \text{ for all } t \geq d,$$

with equality $p_d^{\rm sos} = p^{\min}$ if and only if $p - p^{\min}$ is a sum of squares. Indeed, by Theorem 6.1, there is no duality gap, i.e., $p_t^{\rm sos} = p_t^{\rm mom}$ for $t \geq d$. Moreover, $p_t^{\rm sos} = p_d^{\rm sos}$ since the degree of a sum of squares decomposition of $p - \rho$ ($\rho \in \mathbb{R}$) is bounded by $2d$. Finally, as (6.3) is strictly feasible, the supremum is attained in (6.2) when it is feasible. Thus, when $p_{\min} > -\infty$, $p_d^{\rm sos} = p^{\min}$ if and only if $p - p^{\min}$ is a sum of squares. Summarizing, if $p - p^{\min}$ is a sum of squares, the infimum $p^{\min}$ of p can be found via the semidefinite program (6.3) at order $t = d$. Otherwise, we just find one lower bound for $p^{\min}$. One may wonder when is this lower bound nontrivial, i.e., when is $p_d^{\rm sos} \neq -\infty$, or in other words when does there exist a scalar ρ for which $p - \rho$ is a sum of squares. Marshall [105] gives an answer which involves again the highest degree component of p.

Proposition 7.13. *[105] Let $p \in \mathbb{R}[\mathbf{x}]_{2d}$, $\tilde{p}$ its highest degree component, and $\mathbf{\Sigma}_{n,2d}$ the cone of homogeneous polynomials of degree $2d$ that are sums of squares.*

(i) *If $p_d^{sos} \neq -\infty$ then $\tilde{p}$ is a sum of squares, i.e. $\tilde{p} \in \mathbf{\Sigma}_{n,2d}$.*
(ii) *If $\tilde{p}$ is an interior point of $\mathbf{\Sigma}_{n,2d}$ then $p_d^{sos} \neq -\infty$.*

For instance, $\sum_{i=1}^n \mathbf{x}_i^{2d}$, $(\sum_{i=1}^d \mathbf{x}_i^2)^d$ are interior points of $\mathbf{\Sigma}_{n,2d}$. See [105] for details.

Example 7.14. *Here are some examples taken from [105]. For the Motzkin polynomial $p = p_M := \mathbf{x}^4\mathbf{y}^2 + \mathbf{x}^2\mathbf{y}^4 - 3\mathbf{x}^2\mathbf{y}^2 + 1$, $p^{min} = 0$, $\tilde{p} = \mathbf{x}^4\mathbf{y}^2 + \mathbf{x}^2\mathbf{y}^4$ is a sum of squares, and $p_3^{sos} = -\infty$. Thus the necessary condition from Proposition 7.13 is not sufficient.*

For $p = (\mathbf{x} - \mathbf{y})^2$, $p^{min} = p_1^{sos} = 0$, and $\tilde{p} = p$ lies on the boundary of $\mathbf{\Sigma}_{2,2}$. Thus the sufficient condition of Proposition 7.13 is not necessary.

For $p = p_M + \epsilon(\mathbf{x}^6 + \mathbf{y}^6)$, where p_M is the Motzkin polynomial, $p^{min} = \epsilon/(1+\epsilon)$, $\tilde{p}_\epsilon = \mathbf{x}^4\mathbf{y}^2 + \mathbf{x}^2\mathbf{y}^4 + \epsilon(\mathbf{x}^6 + \mathbf{y}^6)$ is an interior point of $\Sigma_{3,6}$. Thus $p_{\epsilon,3}^{sos} \neq -\infty$. Yet $\lim_{\epsilon \to 0} p_{\epsilon,3}^{sos} = -\infty$ for otherwise $p_M + \rho$ would be a sum of squares for some ρ (which is not possible, as observed in Example 3.7).

Thus arises naturally the question of designing alternative relaxation schemes to get better approximations for $p^{\min}$. A natural idea is to try to transform the unconstrained problem (1.3) into a *constrained* problem. We start with the most favourable situation when p has a minimum and moreover some information is known about the position of a global minimizer.

7.2.1. Case 1: p attains its minimum and a ball is known containing a minimizer. If p attains its minimum and if some bound R is known on the Euclidian norm of a global minimizer, then (1.3) can be reformulated as the constrained minimization problem over the ball

$$p^{\rm ball} := \min\ p(x) \text{ s.t. } \sum_{i=1}^n x_i^2 \leq R^2. \tag{7.7}$$

We can now apply the relaxation scheme from Section 6 to the semialgebraic set $K = \{x \mid \sum_{i=1}^n x_i^2 \leq R^2\}$ which obviously satisfies Putinar's assumption (3.14); thus the moment/SOS bounds converge to $p^{\rm ball} = p^{\rm min}$. This approach seems to work well if the radius R is not too large.

7.2.2. Case 2: p attains its minimum, but no information about minimizers is known. Nie, Demmel and Sturmfels [116] propose an alternative way of transforming (1.3) into a constrained problem when p attains its infimum. Define the *gradient ideal* of p

$$\mathcal{I}_p^{\rm grad} := \Big(\frac{\partial p}{\partial x_i}\ (i = 1, \ldots, n)\Big), \tag{7.8}$$

as the ideal generated by the partial derivatives of p. Since all global minimizers of p are critical points, i.e. they lie in $V_{\mathbb{R}}(\mathcal{I}_p^{\rm grad})$, the (real) gradient variety of p, the unconstrained minimization problem (1.3) can be reformulated as the constrained minimization problem over the gradient varietyi; that is,

$$p^{\rm min} = p^{\rm grad} := \min_{x \in V_{\mathbb{R}}(\mathcal{I}_p^{\rm grad})} p(x). \tag{7.9}$$

Note that the equality $p^{\rm min} = p^{\rm grad}$ may not hold if p has no minimum. E.g. for $p = \mathbf{x}_1^2 + (1 - \mathbf{x}_1\mathbf{x}_2)^2$, $p^{\rm min} = 0$ while $p^{\rm grad} = 1$ as $V_{\mathbb{C}}(\mathcal{I}_p^{\rm grad}) = \{0\}$.

We can compute the moments/SOS bounds obtained by applying the relaxation scheme from Section 6 to the semialgebraic set $V_{\mathbb{R}}(\mathcal{I}_p^{\rm grad})$. However in general this set does not satisfy the assumption (3.14), i.e. we are not in the Archimedean situation, and thus we cannot apply Theorem 6.8 (which relies on Theorem 3.20) to show the asymptotic convergence of the moment/SOS bounds to $p^{\rm grad}$. Yet asymptotic convergence *does hold* and sometimes even *finite* convergence. Nie et al. [116] show the representation results from Theorems 7.15-7.16 below, for positive (nonnegative) polynomials on their gradient variety as sums of squares modulo their gradient ideal. As an immediate application, there is asymptotic convergence of the moment/SOS bounds from the programs (6.3), (6.2) (applied to the polynomial constraints $\partial p/\partial x_i = 0$ $(i = 1, \ldots, n)$) to the parameter $p^{\rm grad}$, and thus to $p^{\rm min}$ when p attains its minimum; moreover there is finite convergence when $\mathcal{I}_p^{\rm grad}$ is radical.

Theorem 7.15. *[116] If $p(x) > 0$ for all $x \in V_{\mathbb{R}}(\mathcal{I}_p^{\text{grad}})$, then p is a sum of squares modulo its gradient ideal $\mathcal{I}_p^{\text{grad}}$, i.e., $p = s_0 + \sum_{i=1}^n s_i \partial p/\partial x_i$, where $s_i \in \mathbb{R}[\mathbf{x}]$ and s_0 is a sum of squares.*

Theorem 7.16. *[116] Assume $\mathcal{I}_p^{\text{grad}}$ is a radical ideal and $p(x) \geq 0$ for all $x \in V_{\mathbb{R}}(\mathcal{I}_p^{\text{grad}})$. Then p is a sum of squares modulo its gradient ideal $\mathcal{I}_p^{\text{grad}}$, i.e., $p = s_0 + \sum_{i=1}^n s_i \partial p/\partial x_i$, where $s_i \in \mathbb{R}[\mathbf{x}]$ and s_0 is a sum of squares.*

We postpone the proofs of these two results, which need some algebraic tools, till Section 7.3. The following example of C. Scheiderer shows that the assumption that $\mathcal{I}_p^{\text{grad}}$ is radical cannot be removed in Theorem 7.16.

Example 7.17. *Consider the polynomial $p = \mathbf{x}^8 + \mathbf{y}^8 + \mathbf{z}^8 + M$, where $M = \mathbf{x}^4\mathbf{y}^2 + \mathbf{x}^2\mathbf{y}^4 + \mathbf{z}^6 - 3\mathbf{x}^2\mathbf{y}^2\mathbf{z}^2$ is the Motzkin form. As observed earlier, M is nonnegative on $\mathbb{R}^3$ but not a sum of squares. The polynomial p is nonnegative over $\mathbb{R}^3$, thus over $V_{\mathbb{R}}(\mathcal{I}_p^{\text{grad}})$, but it is not a sum of squares modulo $\mathcal{I}_p^{\text{grad}}$. Indeed one can verify that $p - M/4 \in \mathcal{I}_p^{\text{grad}}$ and that M is not a sum of squares modulo $\mathcal{I}_p^{\text{grad}}$ (see [116] for details); thus $\mathcal{I}_p^{\text{grad}}$ is not radical.*

Let us mention (without proof) a related result of Marshall [105] which shows a representation result related to that of Theorem 7.16 but under a different assumption.

Theorem 7.18. *[105] Assume p attains its minimum and the matrix $\left(\frac{\partial^2 p}{\partial x_i \partial x_j}(x)\right)_{i,j=1}^n$ is positive definite at every global minimizer x of p. Then $p - p^{min}$ is a sum of squares modulo $\mathcal{I}_p^{\text{grad}}$.*

Summarizing, the above results of Nie et al. [116] show that the parameter p^{grad} can be approximated via converging moment/SOS bounds; when p has a minimum, then $p^{\min} = p^{\text{grad}}$ and thus $p^{\min}$ too can be approximated.

7.2.3. Case 3: p does not attain its minimum. When it is not known whether p attains its minimum, we cannot apply the previous approaches, based on minimizing over a ball or on minimizing over the gradient variety. Several strategies have been proposed in the literature which we now discuss; namely, one may perturb the polynomial in order to get a polynomial with a minimum, or one may minimize p over a suitable semialgebraic set larger than the gradient variety (the so-called gradient tentacle set, or the tangency variety).

Strategy 1: A first strategy is to perturb the polynomial in such a way that the perturbed polynomial has a minimum. For instance, Hanzon and Jibetean [55], Jibetean [68] propose the following perturbation

$$p_\epsilon := p + \epsilon\Big(\sum_{i=1}^n \mathbf{x}_i^{2d+2}\Big)$$

if p has degree $2d$, where $\epsilon > 0$. Then the perturbed polynomial p_ϵ has a minimum (e.g. because the minimum of $\sum_i \mathbf{x}_i^{2d+2}$ over the unit sphere is equal to $1/n^d > 0$; recall Lemma 7.12) and $\lim_{\epsilon\to 0} p_\epsilon^{\min} = p^{\min}$.

For fixed $\epsilon > 0$, $p_\epsilon^{\min} = p_\epsilon^{\text{grad}}$ can be obtained by minimizing p_ϵ over its gradient variety and the asymptotic convergence of the moment/SOS bounds to p_ϵ^{grad} follows from the above results of Nie et al. [116]. Alternatively we may observe that the gradient variety of p_ϵ is finite. Indeed, $\partial p_\epsilon/\partial x_i = (2d+2)\mathbf{x}_i^{2d+1} + \partial p/\partial x_i$, where $\deg(\partial p/\partial \mathbf{x}_i) < 2d$. Hence, $|V_{\mathbb{C}}(\mathcal{I}_{p_\epsilon}^{\text{grad}})| \le \dim \mathbb{R}[\mathbf{x}]/\mathcal{I}_{p_\epsilon}^{\text{grad}} \le (2d+1)^n$. By Theorem 6.15, we can conclude to the *finite* convergence of the moment/SOS bounds to $p_\epsilon^{\text{grad}} = p_\epsilon^{\min}$. Jibetean and Laurent [69] have investigated this approach and present numerical results. Moreover they propose to exploit the equations defining the gradient variety to reduce the number of variables in the moment relaxations.

Hanzon and Jibetean [55] and Jibetean [68] propose in fact an exact algorithm for computing $p^{\min}$. Roughly speaking they exploit the fact (recall Theorem 2.9) that the points of the gradient variety of p_ϵ can be obtained as eigenvalues of the multiplication matrices in the quotient space $\mathbb{R}[\mathbf{x}]/\mathcal{I}_{p_\epsilon}^{\text{grad}}$ and they study the behaviour of the limits as $\epsilon \to 0$. In particular they show that when p has a minimum, the limit set as $\epsilon \to 0$ of the set of global minimizers of p_ϵ is contained in the set of global minimizers of p, and each connected component of the set of global minimizers of p contains a point which is the limit of a branch of minimizers of p_ϵ. Their method however has a high computational cost and is thus not practical.

Strategy 2: Schweighofer [151] proposes a different strategy for dealing with the case when p has no minimum. Namely he proposes to minimize p over the following semialgebraic set, called the *principle gradient tentacle* of p,

$$K_{\nabla p} := \Big\{ x \in \mathbb{R}^n \mid \Big(\sum_{i=1}^n \Big(\frac{\partial p}{\partial x_i}(x) \Big)^2 \Big) \Big(\sum_{i=1}^n x_i^2 \Big) \le 1 \Big\},$$

which contains the gradient variety. Under some technical condition on p, Schweighofer [151] shows that $p^{\min} = \inf_{x \in K_{\nabla p}} p(x)$, and he proves a SOS representation result for positive polynomials on $K_{\nabla p}$; these two results together thus show that $p^{\min}$ can be computed via the SOS approach. We sketch the results (without proofs). We also refer to [169] for related work.

Write p as $p = p_d + p_{d-1} + \ldots + p_0$, where p_k consists of the terms of p of degree k, and $d = \deg(p)$. One says that p *has no isolated singularities at infinity* if $d = 0$, or if $d \ge 1$ and the system of equations:

$$\frac{\partial p_d}{\partial x_1} = \ldots = \frac{\partial p_d}{\partial x_n} = p_{d-1} = 0$$

has only finitely many projective zeros. Note that this is always true when $n = 2$.

Schweighofer [151] shows the following representation theorem,

Theorem 7.19. *[151] Assume $p^{min} > -\infty$. Furthermore assume that, either p has only isolated singularities at infinity (which is always true if $n = 2$), or $K_{\nabla p}$ is compact. Then the following assertions are equivalent.*

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

(ii) $p \geq 0$ *on* $K_{\nabla p}$*;*

(iii) $\forall \epsilon > 0 \ \exists s_0, s_1 \in \Sigma \ \ p + \epsilon = s_0 + s_1\Big(1 - (\sum_{i=1}^n (\partial p/\partial x_i)^2)(\sum_{i=1}^n \mathbf{x}_i^2)\Big)$.

Thus, under the conditions of the theorem, $p^{\min} = \inf_{x \in K_{\nabla p}} p(x)$ can thus be approximated using the SOS hierarchy (applied to $K_{\nabla p}$). When nothing is known about the singularities at infinity, Schweighofer [151] proposes to use modified gradient tentacles of the form

$$\Big\{x \in \mathbb{R}^n \mid \Big(\sum_{i=1}^n \Big(\frac{\partial p}{\partial x_i}(x)\Big)^{2N}\Big)\Big(\sum_{i=1}^n x_i^2\Big)^{N+1} \leq 1\Big\},$$

for various $N \in \mathbb{N}$. He shows some SOS representation results for positive polynomials over such sets, but these SOS approximations are too costly to compute to be useful practically.

Theorem 7.19 relies in particular on Theorem 7.20 below, a non trivial generalization of Schmüdgen's theorem (Theorem 3.16), which will also play a central role in the approach of Vui and Son described below. We need a definition. Given a polynomial p and a subset $S \subseteq \mathbb{R}^n$, a scalar $y \in \mathbb{R}$ is an *asymptotic value* of p on S if there exists a sequence $(x_k)_k$ of points of S for which $\lim_{k\to\infty} \|x_k\| = \infty$ and $\lim_{k\to\infty} f(x_k) = y$.

Theorem 7.20. *[151] Let $p \in \mathbb{R}[\mathbf{x}]$, let K be a semi-algebraic set defined by polynomial inequalities $g_j \geq 0$ $(j = 1, \ldots, m)$ (as in (1.2)), and let $T(g_1, \ldots, g_m)$ be the associated preordering (as in (3.12)). Assume that (i) p is bounded on K, (ii) $p > 0$ on K, and (iii) p has finitely many asymptotic values on K and all of them are positive. Then $p \in T(g_1, \ldots, g_m)$.*

Strategy 3: Vui and Son [170] propose yet another strategy for computing $p^{\min}$, which applies to *arbitrary polynomials* p. It is based on considering the *tangency variety*

$$\Gamma_p := \Big\{x \in \mathbb{R}^n \mid \text{rank}\begin{pmatrix} \frac{\partial p}{\partial x_1} & \cdots & \frac{\partial p}{\partial x_n} \\ x_1 & \ldots & x_n \end{pmatrix} \leq 1\Big\},$$

and the *truncated tangency variety*

$$\Gamma_p^0 := \{x \in \Gamma_p \mid p(x) \leq p(0)\}$$

of the polynomial p. A first basic observation is as follows.

Lemma 7.21. *For $p \in \mathbb{R}[\mathbf{x}]$, $p^{min} = \inf_{x \in \Gamma_p} p(x) = \inf_{x \in \Gamma_p^0} p(x)$.*

Proof. It suffices to show $\inf_{x\in\Gamma_p^0} p(x) \leq p^{\min}$. As $0 \in \Gamma_p^0$, this is obvious if $p^{\min} = p(0)$. Suppose now $p^{\min} < p(0)$. Pick ϵ with $0 < \epsilon < p(0) - p^{\min}$. There exists $x_\epsilon \in \mathbb{R}^n$ for which $p(x_\epsilon) \leq p^{\min} + \epsilon$. Without loss of generality we can assume that $p(x_\epsilon) = \min_{x \mid \|x\|=\|x_\epsilon\|} p(x)$. Setting $h := \sum_{i=1}^n x_i^2 - \|x_\epsilon\|^2$ and applying the first order optimality conditions, we deduce that $\lambda\nabla p(x_\epsilon) + \mu\nabla h(x_\epsilon) = 0$ for some scalars $(\lambda,\mu) \neq (0,0)$. As $\nabla h(x_\epsilon) = 2x_\epsilon$, this shows that $x_\epsilon \in \Gamma_p^0$. Therefore, $\inf_{x\in\Gamma_p^0} p(x) \leq p^{\min} + \epsilon$ which, letting $\epsilon \to 0$, implies $\inf_{x\in\Gamma_p^0} p(x) \leq p^{\min}$. □

Next note that Γ_p is a variety, as Γ_p can be defined by the polynomial equations: $g_{ij} := x_j \frac{\partial p}{\partial x_i} - x_i \frac{\partial p}{\partial x_j} = 0$ for $1 \leq i < j \leq n$. Vui and Son [170] show the following representation result for positive polynomials on the semi-algebraic set Γ_p^0.

Theorem 7.22. *For $p \in \mathbb{R}[\mathbf{x}]$, the following assertions are equivalent.*

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

(ii) $p \geq 0$ *on* Γ_p^0.

(iii) $\forall \epsilon > 0 \;\; \exists s,t \in \Sigma \;\; \exists u_{ij} \in \mathbb{R}[\mathbf{x}]$ *for* $1 \leq i < j \leq n$ *such that*

$$p + \epsilon = s + t(p(0) - p) + \sum_{1\leq i<j\leq n} u_{ij} g_{ij}.$$

Proof. (i) $\Longrightarrow$ (ii) and (iii) $\Longrightarrow$ (i) are obvious. We show (ii) $\Longrightarrow$ (iii). For this, assume $p > 0$ on Γ_p^0; we show the existence of a decomposition as in (iii) with $\epsilon = 0$. We can apply Theorem 7.20 applied to Γ_p^0, since [170] shows that p has finitely many asymptotic values on Γ_p and that all are positive. Thus p lies in the preordering defined by the polynomials $p(0)-p$, g_{ij} and $-g_{ij}$, which immediately gives the desired decomposition for p. □

Therefore, in view of Lemma 7.21 and Theorem 7.22, the infimum $p^{\min}$ of an arbitrary polynomial p over $\mathbb{R}^n$ can be approximated using the SOS hierarchy applied to the semi-algebraic set Γ_p^0.

7.3. Positive polynomials over the gradient ideal. We give here the proofs for Theorems 7.15 and 7.16 which give sums of squares representations modulo the gradient ideal for positive polynomials on the gradient variety; we mostly follow Nie et al. [116] although our proof slightly differs at some places. We begin with the following lemma which can be seen as an extension of Lemma 2.3 about existence of interpolation polynomials. Recall that a set $V \subseteq \mathbb{C}^n$ is a variety if $V = V_{\mathbb{C}}(\{p_1,\ldots,p_s\})$ for some polynomials $p_i \in \mathbb{C}[\mathbf{x}]$. When all p_i's are real polynomials, i.e. $p_i \in \mathbb{R}[\mathbf{x}]$, then $V = \overline{V} := \{\overline{v} \mid v \in V\}$, i.e. $v \in V \Leftrightarrow \overline{v} \in V$.

Lemma 7.23. *Let $V_1,\ldots,V_r$ be pairwise disjoint varieties in $\mathbb{C}^n$ such that $V_i = \overline{V}_i := \{\overline{v} \mid v \in V_i\}$ for all i. There exist polynomials $p_1,\ldots,p_r \in \mathbb{R}[\mathbf{x}]$ such that $p_i(V_j) = \delta_{i,j}$ for $i,j = 1,\ldots r$; that is, $p_i(v) = 1$ if $v \in V_i$ and $p_i(v) = 0$ if $v \in V_j$ $(j \neq i)$.*

Proof. The ideal $\mathcal{I}_i := \mathcal{I}(V_i) \subseteq \mathbb{C}[\mathbf{x}]$ is radical with $V_{\mathbb{C}}(\mathcal{I}_i) = V_i$. We have $V_{\mathbb{C}}(\mathcal{I}_i + \bigcap_{j\neq i} \mathcal{I}_j) = V_{\mathbb{C}}(\mathcal{I}_i) \cap V_{\mathbb{C}}(\bigcap_{j\neq i} \mathcal{I}_j) = V_{\mathbb{C}}(\mathcal{I}_i) \cap (\bigcup_{j\neq i} V_{\mathbb{C}}(\mathcal{I}_j)) =$

$V_i \cap (\bigcup_{j\neq i} V_j) = \emptyset$. Hence, by Hilbert's Nullstellensatz (Theorem 2.1 (i)), $1 \in \mathcal{I}_i + \bigcap_{j\neq i} \mathcal{I}_j$; say $1 = q_i + p_i$, where $q_i \in \mathcal{I}_i$ and $p_i \in \bigcap_{j\neq i} \mathcal{I}_j$. Hence $p_i(V_j) = \delta_{i,j}$ (since q_i vanishes on V_i and p_i vanishes on V_j for $j \neq i$). As $V_i = \overline{V}_i$ for all i, we can replace p_i by its real part to obtain polynomials satisfying the properties of the lemma. □

A variety $V \subseteq \mathbb{C}^n$ is irreducible if any decomposition $V = V_1 \cup V_2$, where V_1, V_2 are varieties, satisfies $V_1 = V$ or $V_2 = V$. It is a known fact that any variety can be written (in a unique way) as a finite union of irreducible varieties (known as its irreducible components) (see e.g. [25, Chap. 4]). Let $V_{\mathbb{C}}(\mathcal{I}_p^{\text{grad}}) = \bigcup_{l=1}^L V_l$ be the decomposition of the gradient variety into irreducible varieties. The following fact is crucial for the proof.

Lemma 7.24. *The polynomial p is constant on each irreducible component of its gradient variety $V_{\mathbb{C}}(\mathcal{I}_p^{\text{grad}})$.*

Proof. Fix an irreducible component V_l. We use the fact[7] that V_l is connected by finitely many differentiable paths. Given $x, y \in V_l$, assume that there exists a continuous differentiable function $\varphi : [0,1] \to V_l$ with $\varphi(0) = x$ and $\varphi(1) = y$; we show that $p(x) = p(y)$, which will imply that p is constant on V_l. Applying the mean value theorem to the function $t \mapsto g(t) := p(\varphi(t))$, we find that $g(1) - g(0) = g'(t^*)$ for some $t^* \in (0,1)$. Now $g(t) = \sum_\alpha p_\alpha \varphi(t)^\alpha$, $g'(t) = \sum_\alpha p_\alpha (\sum_{i=1}^n \alpha_i \varphi_i'(t) \frac{\varphi(t)^\alpha}{\varphi_i(t)}) = \sum_{i=1}^n \frac{\partial p}{\partial x_i}(\varphi(t))\varphi_i'(t)$, which implies $g'(t^*) = 0$ as $\varphi(t^*) \in V_l \subseteq V_{\mathbb{C}}(\mathcal{I}_p^{\text{grad}})$. Therefore, $0 = g(1) - g(0) = p(y) - p(x)$. □

We now group the irreducible components of $V_{\mathbb{C}}(\mathcal{I}_p^{\text{grad}})$ in the following way:

$$V_{\mathbb{C}}(\mathcal{I}_p^{\text{grad}}) = W_0 \cup W_1 \cup \ldots \cup W_r,$$

where the W_i's are defined as follows. Denoting by p_l the common value taken by p on the irreducible component V_l (which is well defined by Lemma 7.24), then W_0 is defined as the union of all the components V_l for which there does not exist another component $V_{l'}$ with $p_l = p_{l'}$ and such that $V_{l'}$ contains at least one real point. Next, p takes a constant value a_i on each W_i $(i = 1, \ldots, r)$, and $a_1, \ldots, a_r$ are all distinct. Thus W_0 contains no real point and any other W_i $(i = 1, \ldots, r)$ contains some real point. Moreover, $W_0, W_1, \ldots, W_r$ are pairwise disjoint, $a_1, \ldots, a_r \in \mathbb{R}$, and $\overline{W}_i = W_i$ for $0 \leq i \leq r$. Hence we can apply Lemma 7.23 and deduce the existence of polynomials $p_0, p_1, \ldots, p_r \in \mathbb{R}[\mathbf{x}]$ satisfying $p_i(W_j) = \delta_{i,j}$ for $i, j = 0, \ldots, r$.

Lemma 7.25. *$p = s_0$ modulo $\mathcal{I}(W_0)$, where s_0 is a sum of squares.*

[7] This is a nontrivial result of algebraic geometry; we thank M. Schweighofer for communicating us the following sketch of proof. Let V be an irreducible variety in $\mathbb{C}^n$. Then V is connected with respect to the usual norm topology of $\mathbb{C}^n$ (see e.g. [153]). Viewing V as a connected semialgebraic set in $\mathbb{R}^{2n}$, it follows that V is connected by a semialgebraic continuous path (see e.g. [17]). Finally, use the fact that a semialgebraic continuous path is piecewise differentiable (see [168, Chap. 7, 2, Prop. 2.5.]).

Proof. We apply the Real Nullstellensatz (Theorem 2.1 (ii)) to the ideal $\mathcal{I} := \mathcal{I}(W_0) \subseteq \mathbb{R}[\mathbf{x}]$. As $V_\mathbb{R}(\mathcal{I}) = W_0 \cap \mathbb{R}^n = \emptyset$, we have $\sqrt[\mathbb{R}]{\mathcal{I}} = \mathcal{I}(V_\mathbb{R}(\mathcal{I})) = \mathbb{R}[\mathbf{x}]$. Hence, $-1 \in \sqrt[\mathbb{R}]{\mathcal{I}}$; that is, $-1 = s + q$, where s is a sum of squares and $q \in \mathcal{I}$. Writing $p = p_1 - p_2$ with p_1, p_2 sums of squares, we find $p = p_1 + sp_2 + p_2 q$, where $s_0 := p_1 + sp_2$ is a sum of squares and $p_2 q \in \mathcal{I} = \mathcal{I}(W_0)$. □

We can now conclude the proof of Theorem 7.16. By assumption, p is nonnegative on $V_\mathbb{R}(\mathcal{I}_p^{\text{grad}})$. Hence, the values $a_1, \ldots, a_r$ taken by p on $W_1, \ldots, W_r$ are nonnegative numbers. Consider the polynomial $q := s_0 p_0^2 + \sum_{i=1}^r a_i p_i^2$, where $p_0, p_1, \ldots, p_r$ are derived from Lemma 7.23 as indicated above and s_0 is as in Lemma 7.25. By construction, q is a sum of squares. Moreover, $p - q$ vanishes on $V_\mathbb{C}(\mathcal{I}_p^{\text{grad}}) = W_0 \cup W_1 \cup \ldots \cup W_r$, since $q(x) = s_0(x) = p(x)$ for $x \in W_0$ (by Lemma 7.25) and $q(x) = a_i = p(x)$ for $x \in W_i$ $(i = 1, \ldots, r)$. As $\mathcal{I}_p^{\text{grad}}$ is radical, we deduce that $p - q \in \mathcal{I}(V_\mathbb{C}(\mathcal{I}_p^{\text{grad}})) = \mathcal{I}_p^{\text{grad}}$, which shows that p is a sum of squares modulo $\mathcal{I}_p^{\text{grad}}$ and thus concludes the proof of Theorem 7.16.

We now turn to the proof of Theorem 7.15. Our assumption now is that p is positive on $V_\mathbb{R}(\mathcal{I}_p^{\text{grad}})$; that is, $a_1, \ldots, a_r > 0$. Consider a primary decomposition of the ideal $\mathcal{I}_p^{\text{grad}}$ (see [25, Chap. 4]) as $\mathcal{I}_p^{\text{grad}} = \bigcap_{h=1}^k \mathcal{I}_h$. Then each variety $V_\mathbb{C}(\mathcal{I}_h)$ is irreducible and thus contained in W_i for some $i = 0, \ldots, r$. For $i = 0, \ldots, r$, set $\mathcal{J}_i := \bigcap_{h | V_\mathbb{C}(\mathcal{I}_h) \subseteq W_i} \mathcal{I}_h$. Then, $\mathcal{I}_p^{\text{grad}} = \mathcal{J}_0 \cap \mathcal{J}_1 \cap \ldots \cap \mathcal{J}_r$, with $V_\mathbb{C}(\mathcal{J}_i) = W_i$ for $0 \leq i \leq r$. As $V_\mathbb{C}(\mathcal{J}_i + \mathcal{J}_j) = V_\mathbb{C}(\mathcal{J}_i) \cap V_\mathbb{C}(\mathcal{J}_j) = W_i \cap W_j = \emptyset$, we have $\mathcal{J}_i + \mathcal{J}_j = \mathbb{R}[\mathbf{x}]$ for $i \neq j$. The next result follows from the Chinese reminder theorem, but we give the proof for completeness.

LEMMA 7.26. *Given $s_0, \ldots, s_r \in \mathbb{R}[\mathbf{x}]$, there exists $s \in \mathbb{R}[\mathbf{x}]$ satisfying $s - s_i \in \mathcal{J}_i$ $(i = 0, \ldots, r)$. Moreover, if each s_i is a sum of squares then s too can be chosen to be a sum of squares.*

Proof. The proof is by induction on $r \geq 1$. Assume first $r = 1$. As $\mathcal{J}_0 + \mathcal{J}_1 = \mathbb{R}[\mathbf{x}]$, $1 = u_0 + u_1$ for some $u_0 \in \mathcal{J}_0$, $u_1 \in \mathcal{J}_1$. Set $s := u_0^2 s_1 + u_1^2 s_0$; thus s is a sum of squares if s_0, s_1 are sums of squares. Moreover, $s - s_0 = u_0^2 s_1 + s_0(u_1^2 - 1) = u_0^2 s_1 - u_0(u_1 + 1)s_0 \in \mathcal{J}_0$. Analogously, $s - s_1 \in \mathcal{J}_1$.

Let s be the polynomial just constructed, satisfying $s - s_0 \in \mathcal{J}_0$ and $s - s_1 \in \mathcal{J}_1$. Consider now the ideals $\mathcal{J}_0 \cap \mathcal{J}_1, \mathcal{J}_2, \ldots, \mathcal{J}_r$. As $(\mathcal{J}_0 \cap \mathcal{J}_1) + \mathcal{J}_i = \mathbb{R}[\mathbf{x}]$ $(i \geq 2)$, we can apply the induction assumption and deduce the existence of $t \in \mathbb{R}[\mathbf{x}]$ for which $t - s \in \mathcal{J}_0 \cap \mathcal{J}_1$, $t - s_i \in \mathcal{J}_i$ $(i \geq 2)$. Moreover, t is a sum of squares if $s, s_2, \ldots, s_r$ are sums of squares, which concludes the proof. □

The above lemma shows that the mapping

$$\begin{array}{ccc} \mathbb{R}[\mathbf{x}]/\mathcal{I}_p^{\text{grad}} = \mathbb{R}[\mathbf{x}]/\cap_{i=0}^r \mathcal{J}_i & \rightarrow & \prod_{i=0}^r \mathbb{R}[\mathbf{x}]/\mathcal{J}_i \\ s \mod \mathcal{I}_p^{\text{grad}} & \mapsto & (s_i \mod \mathcal{J}_i | i = 0, \ldots, r) \end{array}$$

is a bijection. Moreover if, for all $i = 0, \ldots, r$, $p - s_i \in \mathcal{J}_i$ with s_i sum of squares, then there exists a sum of squares s for which $p - s \in \mathcal{I}_p^{\text{grad}}$. Therefore, to conclude the proof of Theorem 7.15, it suffices to show that p is a sum of squares modulo each ideal $\mathcal{J}_i$. For $i = 0$, as $V_{\mathbb{R}}(\mathcal{J}_0) = \emptyset$, this follows from the Real Nullstellensatz (same argument as for Lemma 7.25). The next lemma settles the case $i \geq 1$ and thus the proof of Theorem 7.15.

Lemma 7.27. *p is a sum of squares modulo $\mathcal{J}_i$, for $i = 1, \ldots, r$.*

Proof. By assumption, $p(x) = a_i > 0$ for all $x \in V_{\mathbb{C}}(\mathcal{J}_i) = W_i$. Hence the polynomial $u := p/a_i - 1$ vanishes on $V_{\mathbb{C}}(\mathcal{J}_i)$ and thus $u \in \mathcal{I}(V_{\mathbb{C}}(\mathcal{J}_i)) = \sqrt{\mathcal{J}_i}$; that is, using Hilbert's Nullstellensatz (Theorem 2.1 (i)), $u^m \in \mathcal{J}_i$ for some integer $m \geq 1$. The identity

$$1 + u = \left(\sum_{k=0}^{m-1} \binom{1/2}{k} u^k \right)^2 + q u^m \tag{7.10}$$

(where $q \in \text{Span}_{\mathbb{R}}(u^i \mid i \geq 0)$) gives directly that $p/a_i = 1 + u$ is a sum of squares modulo $\mathcal{J}_i$. To show (7.10), write $\left(\sum_{k=0}^{m-1} \binom{1/2}{k} u^k\right)^2 = \sum_{j=0}^{2m-2} c_j u^j$, where $c_j := \sum_k \binom{1/2}{k}\binom{1/2}{j-k}$ with the summation over k satisfying $\max(0, j - m + 1) \leq k \leq \min(j, m - 1)$. We now verify that $c_j = 1$ for $j = 0, 1$ and $c_j = 0$ for $j = 2, \ldots, m - 1$, which implies (7.10). For this fix $0 \leq j \leq m - 1$ and consider the univariate polynomial $g_j := \sum_{h=0}^{j} \binom{\mathbf{t}}{h}\binom{\mathbf{t}}{j-h} - \binom{2\mathbf{t}}{j} \in \mathbb{R}[\mathbf{t}]$; as g_j vanishes at all $t \in \mathbb{N}$, g_j is identically zero and thus $g_j(1/2) = 0$, which gives $c_j = \binom{1}{j}$ for $j \leq m - 1$, i.e. $c_0 = c_1 = 1$ and $c_j = 0$ for $2 \leq j \leq m - 1$. □

8. Exploiting algebraic structure to reduce the problem size.

In the previous sections we have seen how to construct moment/SOS approximations for the infimum of a polynomial over a semialgebraic set. The simplest instance is the unconstrained minimization problem (1.3) of computing $p^{\min}$ ($= \inf_{x \in \mathbb{R}^n} p(x)$) where p is a polynomial of degree $2d$, its moment relaxation p_d^{mom} ($= \inf p^T y$ s.t. $M_d(y) \succeq 0$, $y_0 = 1$), and its SOS relaxation p_d^{sos} ($= \sup \rho$ s.t. $p - \rho$ is a sum of squares). Recall that $p_d^{\text{mom}} = p_d^{\text{sos}}$. To compute $p_d^{\text{mom}} = p_d^{\text{sos}}$ one needs to solve a semidefinite program involving a matrix indexed by $\mathbb{N}_d^n$, thus of size $\binom{n+d}{d}$. This size becomes prohibitively large as soon as n or d is too large. It is thus of crucial importance to have methods permitting to reduce the size of this semidefinite program. For this one can exploit the specific structure of the problem at hand. For instance, the problem may have some symmetry, or may have some sparsity pattern, or may contain equations, all features which can be used to reduce the number of variables and sometimes the size of the matrices involved. See e.g. Parrilo [124] for an overview about exploiting algebraic structure in SOS programs. Much research has been done in the recent years about such issues, which we cannot cover in detail in this survey. We will only treat certain chosen topics.

8.1. Exploiting sparsity.

8.1.1. Using the Newton polynomial. Probably one of the first results about exploiting sparsity is a result of Reznick [137] about Newton polytopes of polynomials. For a polynomial $p = \sum_{|\alpha|\le d} p_\alpha \mathbf{x}^\alpha$, its *Newton polytope* is defined as

$$N(p) := \mathrm{conv}(\alpha \in \mathbb{N}^n_d \mid p_\alpha \neq 0).$$

Reznick [137] shows the following properties for Newton polytopes.

THEOREM 8.1. *[137] Given $p, q, f_1, \ldots, f_m \in \mathbb{R}[\mathbf{x}]$.*

(i) *$N(pq) = N(p) + N(q)$ and, if p, q are nonnegative on $\mathbb{R}^n$ then $N(p) \subseteq N(p+q)$.*

(ii) *If $p = \sum_{j=1}^m f_j^2$, then $N(f_j) \subseteq \frac{1}{2}N(p)$ for all j.*

(iii) *$N(p) \subseteq conv(2\alpha \mid p_{2\alpha} > 0)$.*

We illustrate the result on the following example taken from [124].

EXAMPLE 8.2. *Consider the polynomial $p = (\mathbf{x}_1^4+1)(\mathbf{x}_2^4+1)(\mathbf{x}_3^4+1)(\mathbf{x}_4^4+1) + 2\mathbf{x}_1 + 3\mathbf{x}_2 + 4\mathbf{x}_3 + 5\mathbf{x}_4$ of degree $2d = 16$ in $n = 4$ variables. Suppose we wish to find a sum of squares decomposition $p = \sum_j f_j^2$. A priori, each f_j has degree at most 8 and thus may involve the $495 = \binom{4+8}{4}$ monomials $\mathbf{x}^\alpha$ with $|\alpha| \in \mathbb{N}^4_8$. The polynomial p is however very sparse; it has only 20 terms, thus much less than the total number $4845 = \binom{4+16}{16}$ of possible terms. As a matter of fact, using the above result of Reznick, one can restrict the support of f_j to the 81 monomials $\mathbf{x}^\alpha$ with $\alpha \in \{0,1,2\}^4$. Indeed the Newton polytope of p is the cube $[0,4]^4$, thus $\frac{1}{2}N(p) = [0,2]^4$ and $\mathbb{N}^4 \cap \frac{1}{2}N(p) = \{0,1,2\}^4$.*

Kojima, Kim and Waki [71] further investigate effective methods for reducing the support of polynomials entering the sum of square decomposition of a sparse polynomial, which are based on Theorem 8.1 and further refinements.

8.1.2. Structured sparsity on the constraint and objective polynomials. We now consider the polynomial optimization problem (1.1) where some sparsity structure is assumed on the polynomials $p, g_1, \ldots, g_m$. Roughly speaking we assume that each g_j uses only a small set of variables and that p can be separated into polynomials using only these small specified sets of variables. Then under some assumption on these specified sets, when searching for a decomposition $p = s_0 + \sum_{j=1}^m s_j g_j$ with all s_j sums of squares, we may restrict our search to polynomials s_j using again the specified sets of variables. We now give the precise definitions.

For a set $I \subseteq \{1, \ldots, n\}$, let $\mathbf{x}_I$ denote the set of variables $\{\mathbf{x}_i \mid i \in I\}$ and $\mathbb{R}[\mathbf{x}_I]$ the polynomial ring in those variables. Assume $\{1, \ldots, n\} = I_1 \cup \ldots \cup I_k$ where the I_h's satisfy the property

$$\forall h \in \{1, \ldots, k-1\} \;\; \exists r \in \{1, \ldots, h\} \;\; I_{h+1} \cap (I_1 \cup \ldots \cup I_h) \subseteq I_r. \quad (8.1)$$

Note that (8.1) holds automatically for $k \leq 2$. We make the following assumptions on the polynomials $p, g_1, \ldots, g_m$:

$$p = \sum_{h=1}^{k} p_h \quad \text{where } p_h \in \mathbb{R}[\mathbf{x}_{I_h}] \tag{8.2}$$

$$\{1, \ldots, m\} = J_1 \cup \ldots \cup J_k \quad \text{and } g_j \in \mathbb{R}[\mathbf{x}_{I_h}] \text{ for } j \in J_h,\ 1 \leq h \leq k. \tag{8.3}$$

Remark 8.3. *If $I_1, \ldots, I_k$ are the maximal cliques of a chordal graph, then $k \leq n$ and (8.1) is satisfied (after possibly reordering the I_h's) and is known as the* running intersection property*. Cf. e.g. [15] for details about chordal graphs. The following strategy is proposed in [172] for identifying a sparsity structure like (8.2)-(8.3). Define the (correlative sparsity) graph $G = (V, E)$ where $V := \{1, \ldots, n\}$ and there is an edge $ij \in E$ if some term of p uses both variables $\mathbf{x}_i, \mathbf{x}_j$, or if both variables $\mathbf{x}_i, \mathbf{x}_j$ are used by some g_l $(l = 1, \ldots, m)$. Then find a chordal extension G' of G and choose the maximal cliques of G' as $I_1, \ldots, I_k$.*

Example 8.4. *For instance, the polynomials $p = \mathbf{x}_1^2\mathbf{x}_2\mathbf{x}_3 + \mathbf{x}_3\mathbf{x}_4^2 + \mathbf{x}_3\mathbf{x}_5 + \mathbf{x}_6$, $g_1 = \mathbf{x}_1\mathbf{x}_2 - 1$, $g_2 = \mathbf{x}_1^2 + \mathbf{x}_2\mathbf{x}_3 - 1$, $g_3 = \mathbf{x}_2 + \mathbf{x}_3^2\mathbf{x}_4$, $g_4 = \mathbf{x}_3 + \mathbf{x}_5$, $g_5 = \mathbf{x}_3\mathbf{x}_6$, $g_6 = \mathbf{x}_2\mathbf{x}_3$ satisfy conditions (8.2), (8.3) after setting $I_1 = \{1, 2, 3\}$, $I_2 = \{2, 3, 4\}$, $I_3 = \{3, 5\}$, $I_4 = \{3, 6\}$.*

Example 8.5. *The so-called chained singular function: $p = \sum_{i=1}^{n-3}(\mathbf{x}_i + 10\mathbf{x}_{i+1})^2 + 5(\mathbf{x}_{i+2} - \mathbf{x}_{i+3})^2 + (\mathbf{x}_{i+1} - 2\mathbf{x}_{i+2})^4 + 10(\mathbf{x}_i - 10\mathbf{x}_{i+3})^4$ satisfies (8.2) with $I_h = \{h, h+1, h+2, h+3\}$ ($h = 1, \ldots, n-3$). Cf. [172] for computational results.*

Let us now formulate the sparse moment and SOS relaxations for problem (1.1) for any order $t \geq \max(d_p, d_{g_1}, \ldots, d_{g_m})$. For $\alpha \in \mathbb{N}^n$, set $\text{supp}(\alpha) = \{i \in \{1, \ldots, n\} \mid \alpha_i \geq 1\}$. For $t \in \mathbb{N}$ and a subset $I \subseteq \{1, \ldots, n\}$ set $\Lambda_t^I := \{\alpha \in \mathbb{N}_t^n \mid \text{supp}(\alpha) \subseteq I\}$. Finally set $\Lambda_t := \cup_{h=1}^k \Lambda_t^{I_h}$. The sparse moment relaxation of order t involves a variable $y \in \mathbb{R}^{\Lambda_{2t}}$, thus having entries y_α only for $\alpha \in \mathbb{N}_{2t}^n$ with $\text{supp}(\alpha)$ contained in some I_h; moreover, it involves the matrices $M_t(y, I_h)$, where $M_t(y, I_h)$ is the submatrix of $M_t(y)$ indexed by $\Lambda_t^{I_h}$. The sparse moment relaxation of order t reads as follows

$$\begin{aligned} \widehat{p_t^{\text{mom}}} := \inf\ p^T y \text{ s.t.} \quad & y_0 = 1,\ M_t(y, I_h) \succeq 0\ (h = 1, \ldots, k) \\ & M_{t-d_{g_j}}(g_j y, I_h) \succeq 0\ (j \in J_h, h = 1, \ldots, k) \end{aligned} \tag{8.4}$$

where the variable y lies in $\mathbb{R}^{\Lambda_{2t}}$. The corresponding sparse SOS relaxation of order t reads

$$\begin{aligned} \widehat{p_t^{\text{sos}}} := \sup \rho \text{ s.t.} \quad & p - \rho = \textstyle\sum_{h=1}^k \Big(u_h + \sum_{j \in J_h} u_{jh} g_j\Big) \\ & u_h, u_{jh}\ (j \in J_h) \text{ sums of squares in } \mathbb{R}[\mathbf{x}_{Ih}] \\ & \deg(u_h), \deg(u_{jh} g_j) \leq 2t\ (h = 1, \ldots, k). \end{aligned} \tag{8.5}$$

Obviously,

$$\widehat{p_t^{\text{sos}}} \leq \widehat{p_t^{\text{mom}}} \leq p^{\min},\ \widehat{p_t^{\text{mom}}} \leq p_t^{\text{mom}},\ \widehat{p_t^{\text{sos}}} \leq p_t^{\text{sos}}.$$

The sparse relaxation is in general weaker than the dense relaxation. However when all polynomials p_h, g_j are quadratic then the sparse and dense relaxations are equivalent (cf. [172, §4.5], also [115, Th. 3.6]). We sketch the details below.

LEMMA 8.6. *Assume $p = \sum_{h=1}^k p_h$ where $p_h \in \mathbb{R}[\mathbf{x}_{I_h}]$ and the sets I_h satisfy (8.1). If $\deg(p_h) \leq 2$ for all h and p is a sum of squares, then p has a sparse sum of squares decomposition, i.e. of the form*

$$p = \sum_{h=1}^{k} s_h \quad \textit{where } s_h \in \mathbb{R}[\mathbf{x}_{I_h}] \textit{ and } s_h \textit{ is SOS.} \tag{8.6}$$

Proof. Consider the dense/sparse SOS/moment relaxations of order 1 of the problem $\min_{x\in\mathbb{R}^n} p(x)$, with optimum values p_1^{sos}, p_1^{mom}, $\widehat{p_1^{\text{sos}}}$, $\widehat{p_1^{\text{mom}}}$. The strict feasibility of the moment relaxations implies that $p_1^{\text{sos}} = p_1^{\text{mom}}$, $\widehat{p_1^{\text{sos}}} = \widehat{p_1^{\text{mom}}}$, the optimum is attained in the dense/sparse SOS relaxations, p SOS $\Longleftrightarrow$ $p_1^{\text{sos}} \geq 0$, and p has a sparse SOS decomposition (8.6) $\Longleftrightarrow$ $\widehat{p_1^{\text{sos}}} \geq 0$. Thus it suffices to show that $p_1^{\text{mom}} \leq \widehat{p_1^{\text{mom}}}$. For this let y be feasible for the program defining $\widehat{p_1^{\text{mom}}}$, i.e. $y_0 = 1$, $M_1(y, I_h) \succeq 0$ for all $h = 1, \ldots, k$. Using a result of Grone et al. [49] (which claims that any partial positive semidefinite matrix whose specified entries form a chordal graph can be completed to a fully specified positive semidefinite matrix), we can complete y to a vector $\tilde{y} \in \mathbb{R}^{\mathbb{N}_2^n}$ satisfying $M_1(\tilde{y}) \succeq 0$. Thus $\tilde{y}$ is feasible for the program defining p_1^{mom}, which shows $p_1^{\text{mom}} \leq \widehat{p_1^{\text{mom}}}$. □

COROLLARY 8.7. *Consider the problem (1.1) and assume that (8.1), (8.2), (8.3) hold. If all p_h, g_j are quadratic, then $p_1^{mom} = \widehat{p_1^{mom}}$ and $p_1^{sos} = \widehat{p_1^{sos}}$.*

Proof. Assume y is feasible for the program defining $\widehat{p_1^{\text{mom}}}$; that is, $y_0 = 1$, $M_1(y, I_h) \succeq 0$ $(h = 1, \ldots, k)$ and $(g_j y)_0 (= \sum_\alpha (g_j)_\alpha y_\alpha) \geq 0$ $(j = 1, \ldots, m)$. Using the same argument as in the proof of Lemma 8.6 we can complete y to $\tilde{y} \in \mathbb{R}^{\mathbb{N}_2^n}$ such that $M_1(\tilde{y}) \succeq 0$ and thus $\tilde{y}$ is feasible for the program defining p_1^{mom}, which shows $p_1^{\text{mom}} \leq \widehat{p_1^{\text{mom}}}$. Assume now $\rho \in \mathbb{R}$ is feasible for the program defining p_1^{sos}; that is, $p - \rho = s_0 + \sum_{j=1}^m s_j g_j$ where s is a sum of squares in $\mathbb{R}[\mathbf{x}]$ and $s_j \in \mathbb{R}_+$. Now the polynomial $p - \rho - \sum_{j=1}^m s_j g_j$ is separable (i.e. can be written as a sum of polynomials in $\mathbb{R}[\mathbf{x}_{I_h}]$); hence, by Lemma 8.6, it has a sparse sum of squares decomposition, of the form $\sum_{h=1}^k s_h$ with $s_h \in \mathbb{R}[\mathbf{x}_{I_h}]$ SOS. This shows that ρ is feasible for the program defining $\widehat{p_1^{\text{sos}}}$, giving the desired inequality $p_1^{\text{sos}} \leq \widehat{p_1^{\text{sos}}}$. □

EXAMPLE 8.8. *We give an example (mentioned in [115, Ex. 3.5]) showing that the result of Lemma 8.6 does not hold for polynomials of* *degree more than 2. Consider the polynomial* $p = p_1 + p_2$, *where* $p_1 = \mathbf{x}_1^4 + (\mathbf{x}_1\mathbf{x}_2 - 1)^2$ *and* $p_2 = \mathbf{x}_2^2\mathbf{x}_3^2 + (\mathbf{x}_3^2 - 1)^2$. *Waki [171] verified that* $0 = \widehat{p_2^{sos}} < p_2^{sos} = p^{min} \sim 0.84986$.

Waki et al. [172] have implemented the above sparse SDP relaxations. Their numerical results show that they can be solved much faster than the dense relaxations and yet they give very good approximations of $p^{\min}$. Lasserre [85, 87] proved the theoretical convergence, i.e. $\lim_{t\to\infty} \widehat{p_t^{\text{sos}}} = \lim_{t\to\infty} \widehat{p_t^{\text{mom}}} = p^{\min}$, under the assumption that K has a nonempty interior and that a ball constraint $R_h^2 - \sum_{i\in J_h} \mathbf{x}_i \geq 0$ is present in the description of K for each $h = 1, \dots, k$. Kojima and Muramatsu [72] proved the result for compact K with possibly empty interior. Grimm, Netzer and Schweighofer [48] give a simpler proof, which does not need the presence of ball constraints in the description of K but instead assumes that each set of polynomials g_j $(j \in J_h)$ generates an Archimedean module.

Theorem 8.9. *[48] Assume that, for each* $h = 1, \dots, k$, *the quadratic module* $\mathbf{M}_h := \mathbf{M}(g_j \mid j \in J_h)$ *generated by* g_j $(j \in J_h)$ *is Archimedean. Assume that (8.1) holds and that* $p, g_1, \dots, g_m$ *satisfy (8.2), (8.3). If* p *is positive on the set* $K = \{x \in \mathbb{R}^n \mid g_j(x) \geq 0 \ (j = 1, \dots, m)\}$, *then* $p \in \mathbf{M}_1 + \ldots + \mathbf{M}_k$; *that is,* $p = \sum_{h=1}^k \left(u_h + \sum_{j\in J_h} u_{jh} g_j\right)$, *where* u_h, u_{jh} *are sums of squares in* $\mathbb{R}[\mathbf{x}_{I_h}]$.

Before proving the theorem we state the application to asymptotic convergence.

Corollary 8.10. *Under the assumptions of Theorem 8.9, we have* $\lim_{t\to\infty} \widehat{p_t^{sos}} = \lim_{t\to\infty} \widehat{p_t^{mom}} = p^{min}$.

Proof. Fix $\epsilon > 0$. As $p - p^{\min} + \epsilon$ is positive on K and satisfies (8.2), we deduce from Theorem 8.9 that $p - p^{\min} + \epsilon \in \sum_{h=1}^k \mathbf{M}_h$. Thus $p^{\min} - \epsilon$ is feasible for (8.5) for some t. Hence, for every $\epsilon > 0$, there exists $t \in \mathbb{N}$ with $p^{\min} - \epsilon \leq \widehat{p_t^{\text{sos}}} \leq p^{\min}$. This shows that $\lim_{t\to\infty} \widehat{p_t^{\text{sos}}} = p^{\text{mom}}$. □

8.1.3. Proof of Theorem 8.9. We give the proof of [48] which is elementary except it uses the following special case of Putinar's theorem (Theorem 3.20): For $p \in \mathbb{R}[\mathbf{x}]$,

$$p > 0 \text{ on } \{x \mid R^2 - \sum_{i=1}^n x_i^2 \geq 0\} \Longrightarrow \exists s_0, s_1 \in \Sigma \quad p = s_0 + s_1(R^2 - \sum_{i=1}^n \mathbf{x}_i^2). \tag{8.7}$$

We start with some preliminary results.

Lemma 8.11. *Let* $C \subseteq \mathbb{R}$ *be compact. Assume* $p = p_1 + \ldots + p_k$ *where* $p_h \in \mathbb{R}[\mathbf{x}_{I_h}]$ $(h = 1, \dots, k)$ *and* $p > 0$ *on* C^n. *Then* $p = f_1 + \ldots + f_k$ *where* $f_h \in \mathbb{R}[\mathbf{x}_{I_h}]$ *and* $f_h > 0$ *on* C^{I_h} $(h = 1, \dots, k)$.

Proof. We use induction on $k \geq 2$. Assume first $k = 2$. Let $\epsilon > 0$ such that $p = p_1 + p_2 \geq \epsilon$ on C^n. Define the function F on $\mathbb{R}^{I_1 \cap I_2}$ by

$$F(y) := \min_{x \in C^{I_1 \setminus I_2}} p_1(x, y) - \frac{\epsilon}{2} \quad \text{for } y \in \mathbb{R}^{I_1 \cap I_2}.$$

The function F is continuous on $C^{I_1 \cap I_2}$. Indeed for $y, y' \in C^{I_1 \cap I_2}$ and $x, x' \in C^{I_1 \setminus I_2}$ minimizing respectively $p_1(x, y)$ and $p_1(x', y')$, we have

$$|F(y) - F(y')| \leq \max(|p_1(x, y) - p_1(x, y')|, |p_1(x', y) - p_1(x', y')|),$$

implying the uniform continuity of F on $C^{I_1 \cap I_2}$ since p_1 is uniform continuous on C^{I_1}. Next we claim that

$$p_1(x,y) - F(y) \geq \frac{\epsilon}{2}, \; p_2(y,z) + F(y) \geq \frac{\epsilon}{2} \; \forall x \in \mathbb{R}^{I_1 \setminus I_2}, y \in \mathbb{R}^{I_1 \cap I_2}, z \in \mathbb{R}^{I_2 \setminus I_1}.$$

The first follows from the definition of F. For the second note that $p_2(y, z) + F(y) = p_2(y, z) + p_1(x, y) - \frac{\epsilon}{2}$ (for some $x \in C^{I_1 \setminus I_2}$), which in turn is equal to $p(x, y, z) - \frac{\epsilon}{2} \geq \epsilon - \frac{\epsilon}{2} = \frac{\epsilon}{2}$. By the Stone-Weierstrass theorem, F can be uniformly approximated by a polynomial $f \in \mathbb{R}[\mathbf{x}_{I_1 \cap I_2}]$ satisfying $|F(y) - f(y)| \leq \frac{\epsilon}{4}$ for all $y \in C^{I_1 \cap I_2}$. Set $f_1 := p_1 - f$ and $f_2 := p_2 + f$. Thus $p = f_1 + f_2$; $f_1 > 0$ on C^{I_1} since $f_1(x, y) = p_1(x, y) - f(y) = p_1(x, y) - F(y) + F(y) - f(y) \geq \frac{\epsilon}{2} - \frac{\epsilon}{4} = \frac{\epsilon}{4}$; $f_2 > 0$ on C^{I_2} since $f_2(y, z) = p_2(y, z) + f(y) = p_2(y, z) + F(y) + f(y) - F(y) \geq \frac{\epsilon}{2} - \frac{\epsilon}{4} = \frac{\epsilon}{4}$. Thus the lemma holds in the case $k = 2$.

Assume now $k \geq 3$. Write $\tilde{I} := \cup_{h=1}^{k-1} I_h$, $\tilde{p} := p_1 + \ldots + p_{k-1} \in \mathbb{R}[\mathbf{x}_{\tilde{I}}]$, so that $p = \tilde{p} + f_k$. By the above proof, there exists $f \in \mathbb{R}[\mathbf{x}_{\tilde{I} \cap I_k}]$ such that $\tilde{p} - f > 0$ on $C^{\tilde{I}}$ and $p_k + f > 0$ on C^{I_k}. Using (8.1), it follows that $\tilde{I} \cap I_k \subseteq I_{h_0}$ for some $h_0 \leq k - 1$. Hence $f \in \mathbb{R}[\mathbf{x}_{I_{h_0}}] \cap \mathbb{R}[\mathbf{x}_{I_k}]$ and $\tilde{p} - f$ is a sum of polynomials in $\mathbb{R}[\mathbf{x}_{I_h}]$ ($h = 1, \ldots, k - 1$). Using the induction assumption for the case $k - 1$, we deduce that $\tilde{p} - f = f_1 + \ldots + f_{k-1}$ where $f_h \in \mathbb{R}[\mathbf{x}_{I_h}]$ and $f_h > 0$ on C^{I_h} for each $h \leq k - 1$. This gives $p = \tilde{p} + p_k = \tilde{p} - f + f + p_k = f_1 + \ldots + f_{k-1} + f + p_k$ which is the desired conclusion since $f + p_k \in \mathbb{R}[\mathbf{x}_{I_k}]$ and $f + p_k > 0$ on C^{I_k}. □

LEMMA 8.12. *Assume $p = p_1 + \ldots + p_k$ where $p_h \in \mathbb{R}[\mathbf{x}_{I_h}]$ and $p > 0$ on the set K. Let B be a bounded set in $\mathbb{R}^n$. There exist $t \in \mathbb{N}$, $\lambda \in \mathbb{R}$ with $0 < \lambda \leq 1$, and polynomials $f_h \in \mathbb{R}[\mathbf{x}_{I_h}]$ such that $f_h > 0$ on B and*

$$p = \sum_{j=1}^{m} (1 - \lambda g_j)^{2t} g_j + f_1 + \ldots + f_k. \tag{8.8}$$

Proof. Choose a compact set $C \subseteq \mathbb{R}$ such that $B \subseteq C^n$ and choose $\lambda \in \mathbb{R}$ such that $0 < \lambda \leq 1$ and $\lambda g_j(x) \leq 1$ for all $x \in C^n$ and $j = 1, \ldots, m$. For $t \in \mathbb{N}$ set

$$F_t := p - \sum_{j=1}^{m} (1 - \lambda g_j)^{2t} g_j.$$

Obviously $F_t \leq F_{t+1}$ on C^n. First we claim

$$\forall x \in C^n \ \exists t \in \mathbb{N}^n \ F_t(x) > 0. \tag{8.9}$$

We use the fact that $(1-\lambda g_j(x))^{2t} g_j(x)$ goes to 0 as t goes to ∞ if $g_j(x) \geq 0$, and to ∞ otherwise. If $x \in K$ then $\lim_{t\to\infty} F_t(x) = p(x)$ and thus $F_t(x) > 0$ for t large enough. If $x \in C^n \setminus K$ then $\lim_{t\to\infty} F_t(x) = \infty$ and thus $F_t(x) > 0$ again for t large enough. This shows (8.9). Next we claim

$$\exists t \in \mathbb{N} \ \forall x \in C^n \ F_t(x) > 0. \tag{8.10}$$

By (8.9), for each $x \in C^n$ there exists an open ball B_x containing x and $t_x \in \mathbb{N}$ such that $F_{t_x} > 0$ on B_x. Thus $C^n \subseteq \cup_{x\in C^n} B_x$. As C^n is compact, we must have $C^n \subseteq B_{x_1} \cup \ldots \cup B_{x_N}$ for finitely many x_i. As $F_t > 0$ on B_{x_i} for all $t \geq t_{x_i}$, we deduce that $F_t > 0$ on C^n for all $t \geq \max_{i=1,\ldots,N} t_{x_i}$, which shows (8.10). Hence we have found the decomposition $p = \sum_{j=1}^m (1-\lambda g_j)^{2t} g_j + F_t$ where $F_t > 0$ on C^n. As F_t is a sum of polynomials in $\mathbb{R}[\mathbf{x}_{I_h}]$ and $F_t > 0$ on C^n, we can apply Lemma 8.11 and deduce that $F_t = f_1 + \ldots + f_k$ where $f_h \in \mathbb{R}[\mathbf{x}_{I_h}]$ and $f_h > 0$ on C^{I_h} and thus on B. Thus (8.8) holds. □

We can now conclude the proof of Theorem 8.9. As each module $\mathbf{M}_h$ is Archimedean, we can find $R > 0$ for which $R^2 - \sum_{i\in I_h} \mathbf{x}_i^2 \in \mathbf{M}_h$ for each $h = 1, \ldots, k$. By assumption, $p > 0$ on K. We apply Lemma 8.12 to the closed ball B in $\mathbb{R}^n$ of radius R. Thus we find a decomposition as in (8.8). As $f_h > 0$ on B we deduce that $f_h \in \mathbf{M}_h$ using (8.7). Finally observe that $\sum_{j=1}^m (1-\lambda g_j)^{2t} g_j = \sum_{h=1}^k u_h$ where $u_h := \sum_{j\in J_h} (1-\lambda g_j)^{2t} g_j \in \mathbf{M}_h$. This concludes the proof of Theorem 8.9.

8.1.4. Extracting global minimizers. In some cases one is also able to extract global minimizers for the original problem (1.1) from the sparse SDP relaxation (8.4). Namely assume y is an optimum solution to the sparse moment ralaxation (8.4) and that the following rank conditions hold:

$$\operatorname{rank} M_s(y, I_h) = \operatorname{rank} M_{s-a_h}(y, I_h) \ \forall h = 1, \ldots, k, \tag{8.11}$$

$$\operatorname{rank} M_s(y, I_h \cap I_{h'}) = 1 \ \forall h \neq h' = 1, \ldots, k \ \text{ with } I_h \cap I_{h'} \neq \emptyset, \tag{8.12}$$

setting $a_h := \max_{j\in J_h} d_{g_j}$. Then we can apply the results from Sections 5.2, 6.6 to extract solutions. Namely for each $h \leq k$, by (8.11), the restriction of y to $\mathbb{R}^{\Lambda_{2t}^{I_h}}$ has a unique representing measure with support $\Delta^h \subseteq \mathbb{R}^{I_h}$. Moreover, by (8.12), if $I_h \cap I_{h'} \neq \emptyset$, then the restriction of y to $\mathbb{R}^{\Lambda_{2t}^{I_h\cap I_{h'}}}$ has a unique representing measure which is a Dirac measure at a point $x^{(hh')} \in \mathbb{R}^{I_h\cap I_{h'}}$. Therefore, any $x^{(h)} \in \Delta^h$, $x^{(h')} \in \Delta^{h'}$ coincide on $I_h \cap I_{h'}$, i.e. $x_i^{(h)} = x_i^{(h')} = x_i^{(hh')}$ for $i \in I_h \cap I_{h'}$. Therefore any point $x^* \in \mathbb{R}^n$ obtained by setting $x_i^* := x_i^{(h)}$ $(i \in I_h)$ for some $x^{(h)} \in \Delta^h$, is an optimum solution to the original problem (1.1). The rank conditions (8.11)-(8.12) are however quite restrictive.

Here is another situation when one can extract a global minimizer; namely when (1.1) has a unique global minimizer. Assume that for all t large enough we have a near optimal solution $y^{(t)}$ to the sparse moment relaxation of order t; that is, $y^{(t)}$ is feasible for (8.4) and $p^T y^{(t)} \leq \widehat{p^{\text{mom}}}_t + 1/t$. Lasserre [85] shows that, if problem (1.1) has a unique global minimizer x^*, then the vectors $(y^{(t)}_{e_i})_{i=1}^n$ converge to the global minimizer x^* as t goes to ∞.

SparsePOP software. Waki, Kim, Kojima, Muramatsu, and Sugimoto have developed the software SparsePOP, which implements the sparse moment and SOS relaxations (8.4)-(8.5) proposed in [172] for the problem (1.1). The software can be downloaded from the website `http://www.is.titech.ac.jp/~kojima/SparsePOP/`.

We also refer to [172] where another technique is proposed, based on perturbing the objective function in (1.1) which, under some conditions, permits the extraction of an approximate global minimizer.

For a detailed presentation of several examples together with computational numerical results, see in particular [172]; see also [115], and [114] for instances arising from sensor network localization (which is an instance of the distance realization problem described in Section 1).

8.2. Exploiting equations. Here we come back to the case when the semialgebraic set K is as in (2.5), i.e. there are explicit polynomial equations $h_1 = 0, \ldots, h_{m_0} = 0$ present in its decription. Let $\mathcal{J} := (h_1, \ldots, h_{m_0})$ be the ideal generated by these polynomials. As noted in Section 6.2 one can formulate SOS/moment bounds by working in the quotient ring $\mathbb{R}[\mathbf{x}]/\mathcal{J}$, which leads to a saving in the number of variables and thus in the complexity of the SDP's to be solved. Indeed suppose we know a (linear) basis $\mathcal{B}$ of $\mathbb{R}[\mathbf{x}]/\mathcal{J}$, so that $\mathbb{R}[\mathbf{x}] = \text{Span}_{\mathbb{R}}(\mathcal{B}) \oplus \mathcal{J}$. Then, for $p \in \mathbb{R}[\mathbf{x}]$,

$$p \text{ SOS} \mod \mathcal{J} \Longleftrightarrow p = \textstyle\sum_l u_l^2 + q \text{ with } u_l \in \text{Span}_{\mathbb{R}}(\mathcal{B}), q \in \mathcal{J}. \tag{8.13}$$

(This is obvious: If $p = \sum_l f_l^2 + g$ with $f_l \in \mathbb{R}[\mathbf{x}]$, $g \in \mathcal{J}$, write $f_l = u_l + v_l$ with $u_l \in \text{Span}_{\mathbb{R}}(\mathcal{B})$ and $v_l \in \mathcal{J}$, so that $p = \sum_l u_l^2 + q$ after setting $q := g + \sum_l v_l^2 + 2u_l v_l \in \mathcal{J}$.) Hence to check the existence of a SOS decomposition modulo $\mathcal{J}$, we can apply the Gram-matrix method from Section 3.3 working with matrices indexed by $\mathcal{B}$ (or a subset of it) instead of the full set of monomials. Moreover, when formulating the moment relaxations, one can use the equations $h_j = 0$ to eliminate some variables within $y = (y_\alpha)_\alpha$. Let us illustrate this on an example (taken from [124]).

EXAMPLE 8.13. *Suppose we want to minimize the polynomial* $p = 10 - \mathbf{x}_1^2 - \mathbf{x}_2$ *over* $\{(x, y) \in \mathbb{R}^2 \mid g_1 := x_1^2 + x_2^2 - 1 = 0\}$ *(the unit circle). To get a lower bound on* p^{min}*, one can compute the largest* ρ *for which* $p - \rho$ *is SOS modulo the ideal* $\mathcal{J} = (\mathbf{x}_1^2 + \mathbf{x}_2^2 - 1)$*. As* $\mathcal{B} := \{\mathbf{x}_1^i, \mathbf{x}_2\mathbf{x}_1^i \mid i \geq 0\}$

is a basis of $\mathbb{R}[\mathbf{x}]/\mathcal{J}$ *(it is the set of standard monomials w.r.t. a graded lex monomial ordering), one can first try to find a decomposition as in (8.13) using only monomials in the subset* $\{1, \mathbf{x}_1, \mathbf{x}_2\} \subseteq \mathcal{B}$*. Namely, find the largest scalar* ρ *for which*

$$\begin{aligned} 10 - \mathbf{x}_1^2 - \mathbf{x}_2 - \rho &= \begin{pmatrix} 1 \\ \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}^T \underbrace{\begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}}_{X \succeq 0} \begin{pmatrix} 1 \\ \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \mod \mathcal{J} \\ &= a + f + (d-f)\mathbf{x}_1^2 + 2b\mathbf{x}_1 + 2c\mathbf{x}_2 + 2e\mathbf{x}_1\mathbf{x}_2 \mod \mathcal{J} \end{aligned}$$

giving $10 - \rho - \mathbf{x}_1^2 - \mathbf{x}_2 = a + f + (d-f)\mathbf{x}_1^2 + 2b\mathbf{x}_1 + 2c\mathbf{x}_2 + 2e\mathbf{x}_1\mathbf{x}_2$*. Equating coefficients in both sides, we find*

$$X = \begin{pmatrix} 10 - f - \rho & 0 & -1/2 \\ 0 & f - 1 & 0 \\ -1/2 & 0 & f \end{pmatrix}.$$

One can easily verify that the largest ρ *for which* $X \succeq 0$ *is* $\rho = 35/4$*, obtained for* $f = 1$*, in which case* $X = L^T L$ *with* $L = \begin{pmatrix} -1/2 & 0 & 1 \end{pmatrix}$*, giving* $p - 35/4 = (\mathbf{x}_2 - 1/2)^2 \mod \mathcal{J}$*. This shows* $p^{min} \geq 35/4$*. Equality holds since* $p(x_1, x_2) = 35/4$ *for* $(x_1, x_2) = (\pm\sqrt{7}/2, -1/2)$*.*

On the moment side, the following program

$$\inf\ 10 - y_{20} - y_{01} \text{ s.t. } \begin{pmatrix} 1 & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & 1 - y_{20} \end{pmatrix} \succeq 0$$

gives a lower bound for p^{min}*. Here we have used the condition* $0 = (g_1 y)_{00} = y_{20} + y_{02} - y_{00}$ *stemming from the equation* $\mathbf{x}_1^2 + \mathbf{x}_2^2 - 1 = 0$*, which thus permits to eliminate the variable* y_{02}*. One can easily check that the optimum of this program is again* $35/4$*, obtained for* $y_{10} = y_{11} = 0$*,* $y_{01} = 1/2$*,* $y_{20} = 3/4$*.*

8.2.1. The zero-dimensional case. When $\mathcal{J}$ is zero-dimensional, $\mathcal{B}$ is a finite set; say $\mathcal{B} = \{b_1, \ldots, b_N\}$ where $N := \dim \mathbb{R}[\mathbf{x}]/\mathcal{J} \geq |V_{\mathbb{C}}(\mathcal{J})|$. For convenience assume $\mathcal{B}$ contains the constant monomial 1, say $b_1 = 1$. By Theorem 6.15, there is finite convergence of the SOS/moment hierarchies and thus problem (1.1) can be reformulated as the semidefinite program (6.2) or (6.3) for t large enough. Moreover the SOS bound

$$\begin{aligned} p^{\text{sos}} &= \sup \rho \text{ s.t. } p - \rho \in \mathbf{M}(g_1, \ldots, g_m, \pm h_1, \ldots, \pm h_{m_0}) \\ &= \sup \rho \text{ s.t. } p - \rho = \textstyle\sum_{j=0}^m s_j g_j \mod \mathcal{J} \text{ for some } s_j \in \Sigma \end{aligned}$$

can be computed via a semidefinite program involving $N \times N$ matrices in view of (the argument for) (8.13), and $p^{\text{sos}} = p^{\text{min}}$ by Theorem 6.8, since the quadratic module $\mathbf{M}(g_1, \ldots, g_m, \pm h_1, \ldots, \pm h_{m_0})$ is Archimedean as $\mathcal{J}$ is zero-dimensional. Therefore, $p^{\text{sos}} = p^{\text{mom}} = p^{\min}$.

We now give a direct argument for equality $p^{\text{mom}} = p^{\min}$, relying on Theorem 5.1 (about finite rank moment matrices, instead of Putinar's theorem) and giving an explicit moment SDP formulation for (1.1) using $N \times N$ matrices; see (8.15). Following [96], we use a so-called combinatorial moment matrix which is simply a moment matrix in which some variables are eliminated using the equations $h_j = 0$. For $f \in \mathbb{R}[\mathbf{x}]$, $\text{res}_{\mathcal{B}}(f)$ denotes the unique polynomial in $\text{Span}_{\mathbb{R}}(\mathcal{B})$ such that $f - \text{res}_{\mathcal{B}}(f) \in \mathcal{J}$. Given $y \in \mathbb{R}^N$, define the linear operator L_y on $\text{Span}_{\mathbb{R}}(\mathcal{B})$ ($\simeq \mathbb{R}[\mathbf{x}]/\mathcal{J}$) by $L_y(\sum_{i=1}^N \lambda_i b_i) := \sum_{i=1}^N \lambda_i y_i$ ($\lambda \in \mathbb{R}^N$) and extend L_y to a linear operator on $\mathbb{R}[\mathbf{x}]$ by setting $L_y(f) := L_y(\text{res}_{\mathcal{B}}(f))$ ($f \in \mathbb{R}[\mathbf{x}]$). Then define the $N \times N$ matrix $M_{\mathcal{B}}(y)$ (the *combinatorial moment matrix* of y) whose (i,j)th entry is $L_y(b_i b_j)$. Consider first for simplicity the problem of minimizing $p \in \mathbb{R}[\mathbf{x}]$ over $V_{\mathbb{R}}(\mathcal{J})$, obviously equivalent to minimizing $\text{res}_{\mathcal{B}}(p)$ over $V_{\mathbb{R}}(\mathcal{J})$. With $\text{res}_{\mathcal{B}}(p) := \sum_{i=1}^N c_i b_i$ where $c \in \mathbb{R}^N$, we have $p(v) = [\text{res}_{\mathcal{B}}(p)](v) = c^T \zeta_{\mathcal{B},v}$ $\forall v \in V_{\mathbb{R}}(\mathcal{J})$, after setting $\zeta_{\mathcal{B},v} := (b_i(v))_{i=1}^N$. Hence

$$p^{\min} = \min_{x \in V_{\mathbb{R}}(\mathcal{J})} p(x) = \min c^T y \ \text{ s.t. } y \in \text{conv}(\zeta_{\mathcal{B},v} \mid v \in V_{\mathbb{R}}(\mathcal{J})). \tag{8.14}$$

The next result implies a semidefinite programming formulation for (8.14) and its proof implies $p^{\text{mom}} = p^{\min}$.

PROPOSITION 8.14. *[96, Th. 14] A vector $y \in \mathbb{R}^N$ lies in the polytope conv$(\zeta_{\mathcal{B},v} \mid v \in V_{\mathbb{R}}(\mathcal{J}))$ if and only if $M_{\mathcal{B}}(y) \succeq 0$ and $y_1 = 1$.*

Proof. Let U denote the $N \times |\mathbb{N}^n|$ matrix whose αth column is the vector containing the coordinates of $\text{res}_{\mathcal{B}}(x^\alpha)$ in the basis $\mathcal{B}$. Define $\tilde{y} := U^T y \in \mathbb{R}^{\mathbb{N}^n}$ with $\tilde{y}_\alpha = L_y(\mathbf{x}^\alpha)$ $\forall \alpha \in \mathbb{N}^n$. One can verify that $M(\tilde{y}) = U^T M_{\mathcal{B}}(y) U$, $\mathcal{J} \subseteq \text{Ker}\, M(\tilde{y})$, and $\tilde{y}^T \text{vec}(p) = y^T c$ with $\text{res}_{\mathcal{B}}(p) = \sum_{i=1}^N c_i b_i$. Consider the following assertions (i)-(iv):
(i) $y \in \mathbb{R}_+(\zeta_{\mathcal{B},v} \mid v \in V_{\mathbb{R}}(\mathcal{J}))$; (ii) $M_{\mathcal{B}}(y) \succeq 0$; (iii) $M(\tilde{y}) \succeq 0$; and (iv) $\tilde{y} \in \mathbb{R}_+(\zeta_v \mid v \in V_{\mathbb{R}}(\mathcal{J}))$. Then, (i) $\Longrightarrow$ (ii) [since $M_{\mathcal{B}}(\zeta_{\mathcal{B},v}) = \zeta_{\mathcal{B},v}\zeta_{\mathcal{B},v}^T \succeq 0$]; (ii) $\Longrightarrow$ (iii) [since $M(\tilde{y}) = U^T M_{\mathcal{B}}(y) U$]; (iii) $\Longrightarrow$ (iv) [by Theorem 5.1, since $\text{rank}\, M(\tilde{y}) < \infty$ as $\mathcal{J} \subseteq \text{Ker}\, M(\tilde{y})$]; and (iv) $\Longrightarrow$ (i), because $\tilde{y} = \sum_{v \in V_{\mathbb{R}}(\mathcal{J})} a_v \zeta_v \Longrightarrow y = \sum_{v \in V_{\mathbb{R}}(\mathcal{J})} a_v \zeta_{\mathcal{B},v}$ [since $\sum_v a_v b_i(v) = \sum_v a_v \text{vec}(b_i)^T \zeta_v = \text{vec}(b_i)^T \tilde{y} = \sum_\alpha (b_i)_\alpha L_y(\mathbf{x}^\alpha) = L_y(b_i) = y_i$]. Finally, as $b_1 = 1$, $y_1 = 1$ means $\sum_v a_v = 1$, corresponding to having a convex combination when $a_v \geq 0$. □

Inequalities $g_j \geq 0$ are treated in the usual way; simply add the conditions $M_{\mathcal{B}}(g_j y) \succeq 0$ to the system $M_{\mathcal{B}}(y) \succeq 0$, $y_1 = 1$, after setting $g_j y := M_{\mathcal{B}}(y) c^{(j)}$ where $\text{res}_{\mathcal{B}}(g_j) = \sum_{i=1}^N c_i^{(j)} b_i$, $c^{(j)} = (c_i^{(j)})_{i=1}^N$. Summarizing we have shown

$$p^{\min} = \min\ c^T y \ \text{ s.t. } y_1 = 1, M_{\mathcal{B}}(y) \succeq 0, M_{\mathcal{B}}(g_j y) \succeq 0 \ (\ \forall j \leq m). \tag{8.15}$$

This idea of using equations to reduce the number of variables has been applied e.g. by Jibetean and Laurent [69] in relation with unconstrained minimization. Recall (from Section 7.2, page 116) that for $p \in \mathbb{R}[\mathbf{x}]_{2d}$, $p^{\min} = \inf_{x\in\mathbb{R}^n} p(x)$ can be approximated by computing the minimum of p over the variety $V_{\mathbb{R}}(\mathcal{J})$ with $\mathcal{J} := ((2d{+}2)x_i^{2d+1}{+}\partial p/\partial x_i \ (i = 1,\dots,n))$ for small $\epsilon > 0$. Then $\mathcal{J}$ is zero-dimensional, $\mathcal{B} = \{\mathbf{x}^\alpha \mid 0 \le \alpha_i \le 2d \ \forall i \le n\}$ is a basis of $\mathbb{R}[\mathbf{x}]/\mathcal{J}$, and the equations in $\mathcal{J}$ give a direct algorithm for computing residues modulo $\mathcal{J}$ and thus the combinatorial moment matrix $M_{\mathcal{B}}(y)$. Such computation can however be demanding for large n, d. We now consider the 0/1 case where the residue computation is trivial.

8.2.2. The 0/1 case. A special case, which is particularly relevant to applications in combinatorial optimization, concerns the minimization of a polynomial p over the 0/1 points in a semialgebraic set K. In other words, the equations $\mathbf{x}_i^2 - \mathbf{x}_i = 0$ $(i = 1,\dots,n)$ are present in the description of K; thus $\mathcal{J} = (\mathbf{x}_1^2 - \mathbf{x}_1,\dots,\mathbf{x}_n^2 - \mathbf{x}_n)$ with $V_{\mathbb{C}}(\mathcal{J}) = \{0,1\}^n$. Using the equations $\mathbf{x}_i^2 = \mathbf{x}_i$, we can reformulate all variables y_α $(\alpha \in \mathbb{N}^n)$ in terms of the 2^n variables y_β $(\beta \in \{0,1\}^n)$ via $y_\alpha = y_\beta$ with $\beta_i := \min(\alpha_i, 1)$ $\forall i$.

With $\mathcal{P}(V)$ denoting the collection of all subsets of $V := \{1,\dots,n\}$, the set $\mathcal{B} := \{\mathbf{x}_I := \prod_{i\in I} \mathbf{x}_i \mid I \in \mathcal{P}(V)\}$ is a basis of $\mathbb{R}[\mathbf{x}]/\mathcal{J}$ and $\dim \mathbb{R}[\mathbf{x}]/\mathcal{J} = |\mathcal{P}(V)| = 2^n$. It is convenient to index a combinatorial moment matrix $M_{\mathcal{B}}(y)$ and its argument y by the set $\mathcal{P}(V)$. The matrix $M_{\mathcal{B}}(y)$ has a particularly simple form, since its (I,J)th entry is $y_{I\cup J}$ $\forall I, J \in \mathcal{P}(V)$. Set

$$\Delta_V := \mathrm{conv}(\zeta_{\mathcal{B},v} \mid v \in \{0,1\}^n) \subseteq \mathbb{R}^{\mathcal{P}(V)}. \tag{8.16}$$

We now give a different, elementary, proof[8] for Proposition 8.14.

Lemma 8.15. $\Delta_V = \{y \in \mathbb{R}^{\mathcal{P}(V)} \mid y_\emptyset = 1, M_{\mathcal{B}}(y) \succeq 0\} = \{y \in \mathbb{R}^{\mathcal{P}(V)} \mid y_\emptyset = 1, \sum_{J\subseteq V | I\subseteq J} (-1)^{|J\setminus I|} y_J \ge 0 \ \forall I \subseteq V\}$.

Proof. Let $Z_{\mathcal{B}}$ be the $2^n \times 2^n$ matrix[9] with columns the vectors $\zeta_{\mathcal{B},v} = (\prod_{i\in I} v_i)_{I\in\mathcal{P}(V)}$ $(v \in \{0,1\}^n)$. Given $y \in \mathbb{R}^{\mathcal{P}(V)}$, let D denote the diagonal matrix whose diagonal entries are the coordinates of the vector $Z_{\mathcal{B}}^{-1}y$. As $M_{\mathcal{B}}(y) = Z_{\mathcal{B}} D Z_{\mathcal{B}}^T$ (direct verification, using the fact that $\mathcal{J}$ is radical), $M_{\mathcal{B}}(y) \succeq 0 \iff D \succeq 0 \iff Z_{\mathcal{B}}^{-1}y \ge 0 \iff y = Z_{\mathcal{B}}(Z_{\mathcal{B}}^{-1}y)$ is a conic combination of the vectors $\zeta_{\mathcal{B},v}$ $(v \in \{0,1\}^n)$. Finally use the form of $Z_{\mathcal{B}}^{-1}$ mentioned in the footnote. □

Example 8.16. *Consider the stable set problem. Using the formulation (1.5) for $\alpha(G)$, we derive using Lemma 8.15 that $\alpha(G)$ is given by the*

[8] This proof applies more general to any zero-dimensional radical ideal $\mathcal{J}$ (cf. [96]).

[9] This matrix is also known as the *Zeta matrix* of the lattice $\mathcal{P}(V)$ of subsets of $V = \{1,\dots,n\}$ and its inverse $Z_{\mathcal{B}}^{-1}$ as the Möbius matrix; cf. [102]. This fact motivates the name *Zeta vector* chosen in [96] for the vectors $\zeta_{\mathcal{B},v}$ and by extension for the vectors ζ_v. We may identify each $v \in \{0,1\}^n$ with its support $J := \{i \in \{1,\dots,n\} \mid v_i = 1\}$; the (I,J)th entry of $Z_{\mathcal{B}}$ (resp., of $Z_{\mathcal{B}}^{-1}$) is 1 (resp., is $(-1)^{|J\setminus I|}$) if $I \subseteq J$ and 0 otherwise.

program

$$\max_{y\in\mathbb{R}^{\mathcal{P}(V)}} \sum_{i\in V} y_{\{i\}} \quad s.t. \;\; y_\emptyset = 1,\; M_{\mathcal{B}}(y) \succeq 0,\; y_{\{i,j\}} = 0 \;(ij \in E). \tag{8.17}$$

Thus $\alpha(G)$ can be computed via a semidefinite program with a matrix of size 2^n, or via an LP with 2^n linear inequalities and variables. As this is too large for practical purpose, one can instead consider truncated *combinatorial moment matrices $M_{\mathcal{B}_t}(y)$, indexed by $\mathcal{B}_t := \{\mathbf{x}_I \mid I \in \mathcal{P}(V), |I| \leq t\} \subseteq \mathcal{B}$, leading to the following upper bound on $\alpha(G)$*

$$\max \sum_{i\in V} y_{\{i\}} \quad s.t. \;\; y_\emptyset = 1,\; M_{\mathcal{B}_t}(y) \succeq 0,\; y_{\{i,j\}} = 0 \;(ij \in E). \tag{8.18}$$

For $t = 1$ this upper bound is the well known theta number $\vartheta(G)$ introduced by Lovász [101]. See [93, 99] and references therein for more details.

EXAMPLE 8.17. *Consider the* max-cut *problem, introduced in (1.8). We are now dealing with the ideal $\mathcal{J} = (\mathbf{x}_1^2 - 1, \ldots, \mathbf{x}_n^2 - 1)$ with $V_{\mathbb{C}}(\mathcal{J}) = \{\pm 1\}^n$. The above treatment for the $0/1$ case extends in the obvious way to the ± 1 case after defining $M_{\mathcal{B}}(y) := (y_{I\Delta J})_{I,J\in\mathcal{P}(V)}$ ($I\Delta J$ denotes the symmetric difference of I, J). For any integer t,*

$$\max \sum_{ij\in E} (w_{ij}/2)(1 - y_{\{i,j\}}) \quad s.t. \;\; y_\emptyset = 1,\; M_{\mathcal{B}_t}(y) = (y_{I\Delta J})_{|I|,|J|\leq t} \succeq 0$$

gives an upper bound for $mc(G, w)$, equal to it when $t = n$; moreover, $mc(G, w)$ can reformulated[10] *as*

$$\max \sum_{ij\in E} (w_{ij}/2)(1 - y_{\{i,j\}}) \quad s.t. \;\; y_\emptyset = 1,\; \sum_{J\subseteq V} (-1)^{|I\cap J|} y_J \geq 0 \;\forall I \subseteq V.$$

For $t = 1$, the above moment relaxation is the celebrated SDP relaxation for max-cut used by Goemans and Williamson [47] for deriving the first nontrivial approximation algorithm for max-cut (still with the best performance guarantee as of today). Cf. e.g. [93, 94, 99] and references therein for more details.

Several other combinatorial methods have been proposed in the literature for constructing hierarchies of (LP or SDP) bounds for $p^{\min}$ in the $0/1$ case; in particular, by Sherali and Adams [154] and by Lovász and Schrijver [102]. It turns out that the hierarchy of SOS/moment bounds described here refines these other hierarchies; see [93, 99] for a detailed comparison.

[10] Use here the analogue of Lemma 8.15 for the ±1 case which claims $M_{\mathcal{B}}(y) = (y_{I\Delta J})_{I,J\subseteq V} \succeq 0 \iff \sum_{J\in\mathcal{P}(V)} (-1)^{|I\cap J|} y_J \geq 0$ for all $I \in \mathcal{P}(V)$ (cf. [94]).

8.2.3. Exploiting sparsity in the $0/1$ **case.** Here we revisit exploiting sparsity in the 0/1 case. Namely, consider problem (1.1) where the equations $\mathbf{x}_i^2 = \mathbf{x}_i$ $(i \le n)$ are present in the description of K and there is a sparsity structure, i.e. (8.1), (8.2), (8.3) hold. By Corollary 8.10 there is asymptotic convergence to $p^{\min}$ of the sparse SOS/moment bounds. We now give an elementary argument showing *finite* convergence, as well as a sparse semidefinite programming (and linear programming) formulation for (1.1).

Given $v \in \{0,1\}^n$ with support $J = \{i \in V \mid v_i = 1\}$, it is convenient to rename $\zeta_{\mathcal{B},v}$ as $\zeta_J^V \in \{0,1\}^{\mathcal{P}(V)}$ (thus with Ith entry 1 if $I \subseteq J$ and 0 otherwise, for $I \in \mathcal{P}(V)$). Extend the notation (8.16) to any $U \subseteq V$, setting $\Delta_U := \operatorname{conv}(\zeta_J^U \mid J \subseteq U) \subseteq \mathbb{R}^{\mathcal{P}(U)}$. The next lemma[11] shows that two vectors in Δ_{I_1} and in Δ_{I_2} can be merged to a new vector in $\Delta_{I_1 \cup I_2}$ when certain obvious compatibility conditions hold.

Lemma 8.18. *Assume $V = I_1 \cup \ldots \cup I_k$ where the I_h's satisfy (8.1) and, for $1 \le h \le k$, let $y^{(h)} \in \Delta_{I_h}$ satisfying $y_I^{(h)} = y_I^{(h')}$ for all $I \subseteq I_h \cap I_{h'}, 1 \le h, h' \le k$. Then there exists $y \in \Delta_V$ which is a common extension of the $y^{(h)}$'s, i.e. $y_I = y_I^{(h)}$ for all $I \subseteq I_h$, $1 \le h \le k$.*

Proof. Consider first the case $k = 2$. Set $I_0 := I_1 \cap I_2$ and, for $h = 1, 2$, write $y^{(h)} = \sum_{I \subseteq I_h} \lambda_I^h \zeta_I^{I_h} = \sum_{H \subseteq I_0} \sum_{I \subseteq I_h | I \cap I_0 = H} \lambda_I^h \zeta_I^{I_h}$ for some $\lambda_I^h \ge 0$ with $\sum_{I \subseteq I_h} \lambda_I^h = 1$. Taking the projection on $\mathbb{R}^{\mathcal{P}(I_0)}$, we obtain

$$\sum_{H \subseteq I_0} \Big(\sum_{I \subseteq I_1 | I \cap I_0 = H} \lambda_I^1 \Big) \zeta_H^{I_0} = \sum_{H \subseteq I_0} \Big(\sum_{J \subseteq I_2 | J \cap I_0 = H} \lambda_J^2 \Big) \zeta_H^{I_0},$$

which implies $\sum_{I \subseteq I_1 | I \cap I_0 = H} \lambda_I^1 = \sum_{J \subseteq I_2 | J \cap I_0 = H} \lambda_J^2 =: \lambda_H \ \ \forall H \subseteq I_0$, since the vectors $\zeta_H^{I_0}$ $(H \subseteq I_0)$ are linearly independent. One can verify that

$$y := \sum_{H \subseteq I_0 | \lambda_H > 0} \frac{1}{\lambda_H} \sum_{I \subseteq I_1, J \subseteq I_2 | I \cap I_0 = J \cap I_0 = H} \lambda_I^1 \lambda_J^2 \zeta_{I \cup J}^{I_1 \cup I_2} \in \mathbb{R}^{\mathcal{P}(I_1 \cup I_2)}$$

lies in $\Delta_{I_1 \cup I_2}$ and that y extends each $y^{(h)}$, $h = 1, 2$.

In the general case $k \ge 2$ we show, using induction on j, $1 \le j \le k$, that there exists $z^{(j)} \in \Delta_{I_1 \cup \ldots \cup I_j}$ which is a common extension of $y^{(1)}, \ldots, y^{(j)}$. Assuming $z^{(j)}$ has been found, we derive from the above case $k = 2$ applied to $z^{(j)}$ and $y^{(j+1)}$ the existence of $z^{(j+1)}$. □

Corollary 8.19. *Assume $V = I_1 \cup \ldots \cup I_k$ where (8.1) holds, let $\mathcal{P}_0 := \cup_{h=1}^k \mathcal{P}(I_h)$ and $y \in \mathbb{R}^{\mathcal{P}_0}$ with $y_\emptyset = 1$. Then, y has an extension $\tilde{y} \in \Delta_V \iff M_{\mathcal{B}}(y, I_h) := (y_{I \cup J})_{I, J \in \mathcal{P}(I_h)} \succeq 0$ for all $h = 1, \ldots, k$.*

Proof. Directly from Lemma 8.18 combined with Lemma 8.15. □

[11] Lasserre [85] uses the analogue of this result for non-atomic measures, which is a nontrivial result, while the proof in the 0/1 case is elementary.

As an application one can derive an explicit sparse LP formulation for several graph optimization problems for partial κ-trees; we illustrate this on the stable set and max-cut problems. Let $G = (V, E)$ be a graph satisfying

$$V = I_1 \cup \ldots \cup I_k \text{ and (8.1) holds,} \tag{8.19}$$

$$\forall ij \in E \ \exists h \in \{1, \ldots, k\} \text{ s.t. } i, j \in I_h. \tag{8.20}$$

First consider the formulation (1.5) for the stability number $\alpha(G)$; as (8.20) holds, this formulation satisfies the sparsity assumptions (8.2) and (8.3). Hence, using Lemma 8.15 combined with Corollary 8.19, we deduce that $\alpha(G)$ can be obtained by maximizing the linear objective function $\sum_{i \in V} y_{\{i\}}$ over the set of $y \in \mathbb{R}^{\mathcal{P}_0}$ satisfying $y_\emptyset = 1$, $y_{\{i,j\}} = 0$ for $ij \in E$, and any one of the following equivalent conditions (8.21) or (8.22)

$$M_{\mathcal{B}}(y, I_h) \succeq 0 \text{ for all } 1 \leq h \leq k, \tag{8.21}$$

$$\sum_{J \in \mathcal{P}(I_h) | I \subseteq J} (-1)^{|J \setminus I|} y_J \geq 0 \text{ for all } I \in \mathcal{P}(I_h), 1 \leq h \leq k. \tag{8.22}$$

More generally, given weights c_i ($i \in V$) attached to the nodes of G, one can find $\alpha(G, c)$, the maximum weight $\sum_{i \in S} c_i$ of a stable set S, by maximizing the linear objective function $\sum_{i \in V} c_i y_{\{i\}}$ over the above LP. Analogously, the objective function in the formulation (1.8) of the max-cut problem satisfies (8.2) and thus the max-cut value $\mathrm{mc}(G, w)$ can be obtained by maximizing the linear objective function $\sum_{ij \in E}(w_{ij}/2)(1 - y_{\{i,j\}})$ over the set of $y \in \mathbb{R}^{\mathcal{P}_0}$ satisfying $y_\emptyset = 1$ and

$$\sum_{J \in \mathcal{P}(I_h)} (-1)^{|I \cap J|} y_J \geq 0 \text{ for all } I \in \mathcal{P}(I_h), \ 1 \leq h \leq k. \tag{8.23}$$

With $\max_{h=1}^k |I_h| \leq \kappa$, we find for both the stable set and max-cut problems an LP formulation involving $O(k2^\kappa)$ linear inequalities and variables. This applies in particular when G is a partial κ-tree (i.e. G is a subgraph of a chordal graph with maximum clique size κ). Indeed, then (8.19)-(8.20) hold with $\max_h |I_h| \leq \kappa$ and $k \leq n$, and thus $\alpha(G, c)$, $\mathrm{mc}(G, w)$ can be computed via an LP with $O(n2^\kappa)$ inequalities and variables. As an application, for fixed κ, $\alpha(G, c)$ and $\mathrm{mc}(G, w)$ can be computed in polynomial time[12] for the class of partial κ-trees. This is a well known result; cf. eg. [18, 163, 175].

8.3. Exploiting symmetry. Another useful property that can be exploited to reduce the size of the SOS/moment relaxations is to use the presence of structural symmetries in the polynomials $p, g_1, \ldots, g_m$. This relies on combining ideas from group representation and invariance theory,

[12] in fact, in strongly polynomial time, since all coefficients in (8.22), (8.23) are $0, \pm 1$; see [146, §15.2].

as explained in particular in the work of Gaterman and Parrilo [46] (see also Vallentin [166]). We will only sketch some ideas illlustrated on some examples as a detailed treatment of this topic is out of the scope of this paper.

Group action. Let $\mathcal{G}$ be a finite group acting on $\mathbb{R}^N$ ($N \geq 1$) via an action $\rho_0 : \mathcal{G} \to \mathrm{GL}(\mathbb{R}^N)$. This induces an action $\rho : \mathcal{G} \to \mathrm{Aut}(\mathrm{Sym}_N)$ on Sym_N, the space of $N \times N$ symmetric matrices, defined by $\rho(g)(X) := \rho_0(g)^T X \rho_0(g)$ for $g \in \mathcal{G}$, $X \in \mathrm{Sym}_N$. This also induces an action on PSD_N, the set of $N \times N$ positive semidefinite matrices. We assume here that each $\rho_0(g)$ is an orthogonal matrix. Then, a matrix $X \in \mathbb{R}^{N\times N}$ is invariant under action of $\mathcal{G}$, i.e. $\rho(g)(X) = X \ \forall g \in \mathcal{G}$, if and only if X belongs to the commutant algebra

$$\mathcal{A}^G := \{X \in \mathbb{R}^{N\times N} \mid \rho_0(g)X = X\rho_0(g) \ \forall g \in G\}. \tag{8.24}$$

Note that the commutant algebra also depends on the specific action ρ_0.

Invariant semidefinite program. Consider a semidefinite program

$$\max \langle C, X\rangle \ \text{ s.t. } \langle A_r, X\rangle = b_r \ (r = 1, \ldots, m), X \in \mathrm{PSD}_N, \tag{8.25}$$

in the variable $X \in \mathrm{Sym}_N$, where $C, A_r \in \mathrm{Sym}_N$ and $b_r \in \mathbb{R}$. Assume that this semidefinite program is invariant under action of $\mathcal{G}$; that is, C is invariant, i.e. $C \in \mathcal{A}^G$, and the feasible region is globally invariant, i.e. X feasible for (8.25) $\Longrightarrow \rho(g)(X)$ feasible for (8.25) $\forall g \in \mathcal{G}$. Let X be feasible for (8.25). An important consequence of the convexity of the feasible region is that the new matrix $X_0 := \frac{1}{|G|}\sum_{g\in G} \rho(g)(X)$ is again feasible; moreover X_0 is invariant under action of $\mathcal{G}$ and it has the same objective value as X. Therefore, we can w.l.o.g. require that X is invariant in (8.25), i.e. we can add the constraint $X \in \mathcal{A}^G$ (which is linear in X) to (8.25) and get an equivalent program.

Action induced by permutations. An important special type of action is when $\mathcal{G}$ is a subgroup of $\mathcal{S}_N$, the group of permutations on $\{1, \ldots, N\}$. Then each $g \in \mathcal{S}_N$ acts naturally on $\mathbb{R}^N$ by $\rho_0(g)(x) := (x_{g(i)})_{i=1}^N$ for $x = (x_i)_{i=1}^N \in \mathbb{R}^N$, and on $\mathbb{R}^{N\times N}$ by $\rho(g)(X) := (X_{g(i),g(j)})_{i,j=1}^N$ for $X = (X_{i,j})_{i,j=1}^N$; that is, $\rho(g)(X) = M_g X M_g^T$ after defining M_g as the $N \times N$ matrix with $(M_g)_{i,j} = 1$ if $j = g(i)$ and 0 otherwise.

For $(i, j) \in \{1, \ldots, N\}^2$, its orbit under action of $\mathcal{G}$ is the set $\{(g(i), g(j)) \mid g \in \mathcal{G}\}$. Let ω denote the number of orbits of $\{1, \ldots, N\}^2$ and, for $l = 1, \ldots, \omega$, define the $N \times N$ matrix $\tilde{D}_l$ by $(\tilde{D}_l)_{i,j} := 1$ if the pair (i, j) belongs to the lth orbit, and 0 otherwise. Following de Klerk, Pasechnik and Schrijver [38], define $D_l := \frac{\tilde{D}_l}{\sqrt{\langle \tilde{D}_l, \tilde{D}_l\rangle}}$ for $l = 1, \ldots, \omega$, the multiplication parameters $\gamma_{i,j}^l$ by

$$D_i D_j = \sum_{l=1}^{\omega} \gamma_{i,j}^l D_l \quad \text{for } i,j = 1,\ldots,\omega,$$

and the $\omega \times \omega$ matrices $L_1,\ldots,L_\omega$ by $(L_l)_{i,j} := \gamma_{l,j}^i$ for $i,j,k = 1,\ldots,\omega$. Then the commutant algebra from (8.24) is

$$\mathcal{A}^G = \Big\{ \sum_{l=1}^{\omega} x_l D_l \mid x_l \in \mathbb{R} \Big\}$$

and thus $\dim \mathcal{A}^G = \omega$.

THEOREM 8.20. *[38] The mapping $D_l \mapsto L_l$ is a $*$-isomorphism, known as the regular $*$-representation of $\mathcal{A}^G$. In particular, given $x_1,\ldots,x_\omega \in \mathbb{R}$,*

$$\sum_{l=1}^{\omega} x_l D_l \succeq 0 \Longleftrightarrow \sum_{l=1}^{\omega} x_l L_l \succeq 0. \tag{8.26}$$

An important application of this theorem is that it provides an *explicit* equivalent formulation for an invariant SDP, using only ω variables and a matrix of order ω. Indeed, assume (8.25) is invariant under action of $\mathcal{G}$. Set $c := (\langle C, D_l \rangle)_{l=1}^{\omega}$ so that $C = \sum_{l=1}^{\omega} c_l D_l$, and $a_r := (\langle A_r, D_l \rangle)_{l=1}^{\omega}$. As observed above the matrix variable X can be assumed to lie in $\mathcal{A}^G$ and thus to be of the form $X = \sum_{l=1}^{\omega} x_l D_l$ for some scalars $x_l \in \mathbb{R}$. Therefore, using (8.26), (8.25) can be equivalently reformulated as

$$\max \sum_{l=1}^{\omega} c_l x_l \quad \text{s.t. } a_r^T x = b_r \ (r = 1,\ldots,m), \sum_{l=1}^{\omega} x_l L_l \succeq 0. \tag{8.27}$$

The new program (8.27) involves a $\omega \times \omega$ matrix and ω variables and can thus be much more compact than (8.25). Theorem 8.20 is used in [38] to compute the best known bounds for the crossing number of complete bipartite graphs. It is also applied in [97] to the stable set problem for the class of Hamming graphs as sketched below.

EXAMPLE 8.21. *Given $\mathcal{D} \subseteq \{1,\ldots,n\}$, let $G(n,\mathcal{D})$ be the graph with node set $\mathcal{P}(V)$ (the collection of all subsets of $V = \{1,\ldots,n\}$) and with an edge (I,J) when $|I\Delta J| \in \mathcal{D}$. (Computing the stability number of $G(n,\mathcal{D})$ is related to finding large error correcting codes in coding theory; cf. e.g. [97, 147]). Consider the moment relaxation of order t for $\alpha(G(n,\mathcal{D}))$ as defined in (8.18); note that it involves a matrix of size $O(\binom{|\mathcal{P}(V)|}{t}) = O((2^n)^t)$, which is exponentially large in n. However, as shown in [97], this semidefinite program is invariant under action of the symmetric group $\mathcal{S}_n$, and* *there are $O(n^{2^{2t-1}-1})$ orbits. Hence, by Theorem 8.20, there is an equivalent SDP whose size is $O(n^{2^{2t-1}-1})$, thus polynomial in n for any fixed t, which implies that the moment upper bound on $\alpha(G(n,\mathcal{D}))$ can be computed in polynomial time for any fixed t.*

Block-diagonalization. Theorem 8.20 gives a first, explicit, symmetry reduction for matrices in $\mathcal{A}^G$. Further reduction is possible. Indeed, using Schur's lemma from representation theory (cf. e.g. Serre [152]), it can be shown that all matrices in $\mathcal{A}^G$ can be put in block-diagonal form by a linear change of coordinates. Namely, there exists a unitary complex matrix T and positive integers h, $n_1,\dots,n_h$, $m_1,\dots,m_h$ such that the set $T^*\mathcal{A}^GT := \{T^*XT \mid X \in \mathcal{A}^G\}$ coincides with the set of the block-diagonal matrices

$$\begin{pmatrix} C_1 & 0 & \dots & 0 \\ 0 & C_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C_h \end{pmatrix},$$

where each C_i $(i = 1,\dots,h)$ is a block-diagonal matrix with m_i identical blocks on its diagonal, all equal to some $B_i \in \mathbb{R}^{n_i\times n_i}$. The above parameters have the following interpretation: $N = \sum_{i=1}^h m_i n_i$, $\dim \mathcal{A}^G = \sum_{i=1}^h n_i^2$, there are h nonequivalent irreducible representations $\theta_1,\dots,\theta_h$ for the group $\mathcal{G}$, with respective representation dimensions $n_1,\dots,n_h$ so that $\rho = m_1\theta_1 \oplus \dots \oplus m_h\theta_h$, where $m_1,\dots,m_h$ are the multiplicities. We refer to Gaterman and Parrilo [46], Vallentin [166] for details and further references therein. To be able to apply this for practical computation one needs to know the explicit block-diagonalization. Several examples are treated in detail in [46]. Here is a small (trivial) example as illustration.

Example 8.22. *Consider the semidefinite program*

$$\min\ d+f \ \ s.t.\ \ X := \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix} \succeq 0,\ d+f+2e-b-c = 0 \qquad (8.28)$$

It is invariant under action of the group $\{1,\sigma\} \sim \mathcal{S}_2$, where σ permutes simultaneously the last two rows and columns of X. Thus we may assume in (8.28) that X is invariant under this action, i.e. $d = f$ and $b = c$. This reduces the number of variables from 6 to 4. Next we give the explicit block-diagonalization. Namely, consider the orthogonal matrix $T := \begin{pmatrix} 1 & 0 & 0 \\ 0 & u & u \\ 0 & u & -u \end{pmatrix}$ where $u := 1/\sqrt{2}$, and observe that $T^*XT = \begin{pmatrix} a & \sqrt{2}b & 0 \\ \sqrt{2}b & d+e & 0 \\ 0 & 0 & d-e \end{pmatrix}$.

We now mention the following example due to Schrijver [147], dealing with the block-diagonalization of the Terwilliger algebra.

EXAMPLE 8.23. *Consider the permutation group $\mathcal{S}_n$ acting on $V = \{1,\ldots,n\}$. Then each $g \in \mathcal{S}_n$ acts in the obvious way on $\mathcal{P}(V)$ (by $g(I) := \{g(i) \mid i \in I\}$ for $I \subseteq V$) and thus on matrices indexed by $\mathcal{P}(V)$. The orbit of $(I,J) \in \mathcal{P}(V) \times \mathcal{P}(V)$ depends on the triple $(|I|,|J|,|I \cap J|)$. Therefore, the commutant algebra, consisting of the matrices $X \in \mathbb{R}^{\mathcal{P}(V)\times\mathcal{P}(V)}$ that are invariant under action of $\mathcal{S}_n$, is*

$$\Big\{ \sum_{i,j,t\in\mathbb{N}} \lambda^t_{i,j} M^t_{i,j} \mid \lambda^t_{i,j} \in \mathbb{R} \Big\},$$

known as the Terwilliger algebra. Here $M^t_{i,j}$ denotes the matrix indexed by $\mathcal{P}(V)$ with (I,J)th entry 1 if $|I| = i$, $|J| = j$ and $|I \cap J| = t$, and 0 otherwise. Schrijver [147] has computed the explicit block-diagonalization for the Terwilliger algebra and used it for computing sharp SDP bounds for the stability number $\alpha(G(n,\mathcal{D}))$, also considered in Example 8.21. As explained in [96] this new bound lies between the moment bound of order 1 and the moment bound of order 2. See also [166] for an exposition of symmetry reduction with application to the Terwilliger algebra.

Symmetry in polynomial optimization. When the polynomial optimization problem (1.1) is invariant under action of some finite group $\mathcal{G}$, it is natural to search for relaxation schemes that inherit the symmetry pattern of the polynomials $p, g_1, \ldots, g_m$. For instance, if p is a symmetric polynomial which is a SOS, one may wonder about the existence of a sum of symmetric squares. One has to be careful however. For instance, as noted in [46], the univariate polynomial $p = \mathbf{x}^2 + (\mathbf{x} - \mathbf{x}^3)^2 = \mathbf{x}^6 - 2\mathbf{x}^4 + 2\mathbf{x}^2$ is invariant under the action $x \mapsto -x$, but there is no sum of square decomposition $p = \sum_l u_l^2$ where each u_l is invariant under this action as well (for otherwise, u_l should be a polynomial of degree 3 in $\mathbf{x}^2$, an obvious contradiction). Yet symmetry of p does imply some special symmetry structure for the squares; we refer to Gaterman and Parrilo [46] for a detailed account.

Jansson et al. [67] study how symmetry carries over to the moment relaxations of problem (1.1). Say, the polynomials $p, g_1, \ldots, g_m$ are invariant under action of a group $\mathcal{G} \subseteq \mathcal{S}_n$; i.e. $p(x) = p(\rho_0(g)(x))\ \ \forall g \in \mathcal{G}$, where $\rho_0(g)(x) = (x_{g(i)})_{i=1}^n$, and analogously for the g_j's. For instance the following problem, studied in [67],

$$\min \sum_{i=1}^n x_i^q \ \text{ s.t. } \sum_{i=1}^n x_i^j = b_j \ (j = 1,\ldots,m) \tag{8.29}$$

with $q \in \mathbb{N}$, $b_j \in \mathbb{R}$, falls in this setting with $\mathcal{G} = \mathcal{S}_n$. Then some symmetry carries over to the moment relaxations (6.3). Indeed, if x is a global minimizer of p over K, then each $\rho_0(g)(x)$ (for $g \in \mathcal{G}$) too is a global minimizer.

Thus the sequence y of moments of the measure $\mu := \frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}} \delta_{\rho_0(g)(x)}$ is feasible for any moment relaxation, with optimum value $p^{\min}$. In other words, we can add the invariance condition

$$y_\alpha = y_{\rho_0(g)(\alpha)}, \text{ i.e. } y_{(\alpha_1,\ldots,\alpha_n)} = y_{(\alpha_{g(1)},\ldots,\alpha_{g(n)})} \ \forall g \in \mathcal{G}$$

on the entries of variable y to the formulation of the moment relaxation (6.3) of any order t. For instance, when $\mathcal{G} = \mathcal{S}_n$, one can require that $y_{e_1} = \ldots = y_{e_n}$, i.e. all y_α take a common value for any $|\alpha| = 1$, that all y_α take a common value for any $|\alpha| = 2$, etc. Thus the moment matrix of order 1 is of the form

$$M_1(y) = \begin{pmatrix} a & b & b & b & \ldots & b \\ b & c & d & d & \ldots & d \\ b & d & c & d & \ldots & d \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ b & d & \ldots & d & c & d \\ b & d & \ldots & d & d & c \end{pmatrix}.$$

It is explained in [67] how to find the explicit block-diagonalization for such symmetric $M_t(y)$ ($t = 1, 2$, etc). This is not difficult in the case $t = 1$; using a Schur complement with respect to the upper left corner, one deduces easily that $M_1(y) \succeq 0 \Longleftrightarrow c + (n-1)d - nb^2/a \geq 0$ and $c - d \geq 0$. The details for $t = 2$ are already more complicated and need information about the irreducible representations of the symmetric group $\mathcal{S}_n$.

In conclusion, exploiting symmetry within polynomial optimization and, more generally, semidefinite programming, has spurred recently lots of interesting research activity, with many exciting new developments in various areas. Let us just mention pointers to a few papers dealing with symmetry reduction in various contexts; the list is not exclusive. In particular, Bachoc and Vallentin [3, 4, 5] study the currently best known bounds for spherical codes and the kissing number; Bai et al. [6] deal with truss topology optimization; de Klerk and Sotirov [39] study lower bounds for quadratic assignment; Gvozdenović and Laurent [52, 53] compute approximations for the chromatic number of graphs; Vallentin [165] considers the minimum distortion of embeddings of highly regular graphs in the Euclidean space.

Acknowledgements. I am very grateful to Markus Schweighofer for discussion and helpful remarks about some proofs in the survey, and to Hayato Waki for his careful reading and for his useful questions and suggestions. I also thank Claus Scheiderer, Lars Schewe, Konrad Schmüdgen, Achill Schürmann, Rekha Thomas, Frank Vallentin, for their comments.

9. Bibliography.

REFERENCES

[1] N.I. Akhiezer. *The classical moment problem,* Hafner, New York, 1965.

[2] A.A. Ahmadi and P.A. Parrilo, A convex polynomial that is not sos-convex, arXiv:0903.1287, 2009.

[3] C. Bachoc and F. Vallentin, *New upper bounds for kissing numbers from semidefinite programming*, Journal of the American Mathematical Society **21**:909-924, 2008.

[4] ———, *Semidefinite programming, multivariate orthogonal polynomials, and codes in spherical caps*, European Journal of Combinatorics **30**:625-637, 2009.

[5] ———, *Optimality and uniqueness of the* $(4, 10, 1/6)$*-spherical code*, Journal of Combinatorial Theory Ser. A **116**:195-204, 2009.

[6] Y. Bai, E. de Klerk, D.V. Pasechnik, R. Sotirov, *Exploiting group symmetry in truss topology optimization*, Optimization and Engineering **10(3)**:331-349, 2009.

[7] A. Barvinok. A Course in Convexity. American Mathematical Society, Vol. 54 of Graduate Texts in Mathematics, 2002.

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

[9] C. Bayer and J. Teichmann, *The proof of Tchakaloff's theorem*, Proceedings of the American Mathematical Society **134**:3035–3040, 2006.

[10] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization - Analysis, Algorithms, and Engineering Applications*, MPS-SIAM Series on Optimization, 2001.

[11] C. Berg. *The multidimensional moment problem and semi-groups,* Proc. Symp. Appl. Math. **37**:110–124, 1987.

[12] C. Berg, J.P.R. Christensen, and C.U. Jensen, *A remark on the multidimensional moment problem*, Mathematische Annalen **243**:163–169, 1979.

[13] C. Berg, J.P.R. Christensen, and P. Ressel, *Positive definite functions on Abelian semigroups*, Mathematische Annalen **223**:253–272, 1976.

[14] C. Berg and P.H. Maserick, *Exponentially bounded positive definite functions*, Illinois Journal of Mathematics **28**:162–179, 1984.

[15] J.R.S. Blair and B. Peyton, *An introduction to chordal graphs and clique trees*, In *Graph Theory and Sparse Matrix Completion*, A. George, J.R. Gilbert, and J.W.H. Liu, eds, Springer-Verlag, New York, pp 1–29, 1993.

[16] G. Blekherman, *There are significantly more nonnegative polynomials than sums of squares*, *Isreal Journal of Mathematics*, **153**:355–380, 2006.

[17] J. Bochnak, M. Coste, and M.-F. Roy, *Géométrie Algébrique Réelle*, Springer, Berlin, 1987. (*Real algebraic geometry,* second edition in english, Springer, Berlin, 1998.)

[18] H.L. Bodlaender and K. Jansen, *On the complexity of the maximum cut problem*, Nordic Journal of Computing **7(1)**:14-31, 2000.

[19] H. Bosse, *Symmetric positive definite polynomials which are not sums of squares*, preprint, 2007.

[20] M.-D. Choi and T.-Y. Lam, *Extremal positive semidefinite forms*, Math. Ann. **231**:1–18, 1977.

[21] M.-D. Choi, T.-Y. Lam and B. Reznick, *Real zeros of positive semidefinite forms. I.*, Math. Z. **171(1)**:1–26, 1980.

[22] ———, *Sums of squares of real polynomials*, Proceedings of Symposia in Pure mathematics **58(2)**:103–126, 1995.

[23] A.M. Cohen, H. Cuypers, and H. Sterk (eds.), Some Tapas of Computer Algebra, Springer, Berlin, 1999.

[24] R.M. Corless, P.M. Gianni, B.M. Trager, *A reordered Schur factorization method for zero-dimensional polynomial systems with multiple roots*, Proceedings ACM International Symposium Symbolic and Algebraic Computations, Maui, Hawaii, 133–140, 1997.

[25] D.A. Cox, J.B. Little, and D. O'Shea, *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, Springer, 1997.

[26] ———, *Using Algebraic Geometry*, Graduate Texts in Mathematics **185**, Springer, New York, 1998.

[27] R.E. Curto and L.A. Fialkow, *Recursiveness, positivity, and truncated moment problems*, Houston Journal of Mathematics **17(4)**:603–635, 1991.

[28] ———, *Solution of the truncated complex moment problem for flat data*, Memoirs of the American Mathematical Society **119**(568), 1996.

[29] ———, *Flat extensions of positive moment matrices: recursively generated relations*, Memoirs of the American Mathematical Society, **136**(648), 1998.

[30] ———, *The truncated complex K-moment problem* Transactions of the American Mathematical Society **352**:2825–2855, 2000.

[31] ———, *An analogue of the Riesz-Haviland theorem for the truncated moment problem*, Journal of Functional Analysis **255(10)**:2709-2731, 2008.

[32] R.E. Curto, L.A. Fialkow and H.M. Möller, *The extremal truncated moment problem*, Integral Equations and Operator Theory **60(2)**:177–200, 2008.

[33] E. de Klerk, *Aspects of Semidefinite Programming - Interior Point Algorithms and Selected Applications*, Kluwer, 2002.

[34] E. de Klerk, D. den Hertog and G. Elabwabi, *Optimization of univariate functions on bounded intervals by interpolation and semidefinite programming*, CentER Discussion paper 2006-26, Tilburg University, The Netherlands, April 2006.

[35] E. de Klerk, M. Laurent, and P. Parrilo, *On the equivalence of algebraic approaches to the minimization of forms on the simplex*, In *Positive Polynomials in Control*, D. Henrion and A. Garulli (eds.), Lecture Notes on Control and Information Sciences **312**:121–133, Springer, Berlin, 2005.

[36] ———, *A PTAS for the minimization of polynomials of fixed degree over the simplex*, Theoretical Computer Science **361(2-3)**:210–225, 2006.

[37] E. de Klerk and D.V. Pasechnik, *Approximating the stability number of a graph via copositive programming,* SIAM Journal on Optimization **12**:875–892, 2002.

[38] E. de Klerk, D.V. Pasechnik and A. Schrijver, *Reduction of symmetric semidefinite programs using the regular *-representation*, Mathematical Programming B **109**: 613-624, 2007.

[39] E. de Klerk and R. Sotirov, *Exploiting group symmetry in semidefinite programming relaxations of the quadratic assignment problem*, Mathematical Programming **122(2)**:225-246, 2010.

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

[41] L.A. Fialkow, *Truncated multivariate moment problems with finite variety*, *Journal of Operator Theory* **60**:343–377, 2008.

[42] B. Fuglede, *The multidimensional moment problem*, Expositiones Mathematicae **1**:47–65, 1983.

[43] T. Fujie and M. Kojima. *Semidefinite programming relaxation for nonconvex quadratic programs. Journal of Global Optimization* **10**:367–380, 1997.

[44] K. Fujisawa, M. Fukuda, K. Kobayashi, K. Nakata, and M. Yamashita, *SDPA (SemiDefinite Programming Algorithm) and SDPA-GMP user's manual - Version 7.1.0*, Research Report B-448, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, June 2008.

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

[46] K. Gaterman and P. Parrilo, *Symmetry groups, semidefinite programs and sums of squares*, Journal of Pure and Applied Algebra **192**:95–128, 2004.

[47] M.X. Goemans and D. Williamson, *Improved approximation algorithms for max-*

imum cuts and satisfiability problems using semidefinite programming, Journal of the ACM **42**:1115–1145, 1995.

[48] D. GRIMM, T. NETZER, AND M. SCHWEIGHOFER, *A note on the representation of positive polynomials with structured sparsity*, Archiv der Mathematik **89(5)**:399–403, 2007.

[49] B. GRONE, C.R. JOHNSON, E. MARQUES DE SA, H. WOLKOWICZ, *Positive definite completions of partial Hermitian matrices*, Linear Algebra and its Applications **58**:109–124, 1984.

[50] M. GRÖTSCHEL, L. LOVÁSZ AND A. SCHRIJVER, *Geometric Algorithms and Combinatorial Optimization*, Springer-Verlag, Berlin, 1988.

[51] N. GVOZDENOVIĆ AND M. LAURENT, *Semidefinite bounds for the stability number of a graph via sums of squares of polynomials*, Mathematical Programming **110(1)**:145–173, 2007.

[52] ———, *The operator Ψ for the chromatic number of a graph*, SIAM Journal on Optimization **19(2)**:572-591, 2008.

[53] ———, *Computing semidefinite programming lower bounds for the (fractional) chromatic number via block-diagonalization*, SIAM Journal on Optimization **19(2)**:592-615, 2008.

[54] D. HANDELMAN, *Representing polynomials by positive linear functions on compact convex polyhedra*, Pacific Journal of Mathematics **132(1)**:35–62, 1988.

[55] B. HANZON AND D. JIBETEAN, *Global minimization of a multivariate polynomial using matrix methods*, Journal of Global Optimization **27**:1–23, 2003.

[56] E.K. HAVILAND, *On the momentum problem for distributions in more than one dimension*, American Journal of Mathematics **57**:562–568, 1935.

[57] J.W. HELTON AND J. NIE, *Semidefinite representation of convex sets*, Math. Programming Ser. A **122(1)**:21–64, 2010. [To appear]

[58] J.W. HELTON AND M. PUTINAR, *Positive polynomials in scalar and matrix variables, the spectral theorem and optimization*, arXiv:math/0612103v1, 2006.

[59] D. HENRION. *On semidefinite representations of plane quartics*, LAAS-CNRS Research Report No. 08444, September 2008.

[60] D. HENRION. *Semidefinite representation of convex hulls of rational varieties*, LAAS-CNRS Research Report No. 09001, January 2009.

[61] D. HENRION AND J.-B. LASSERRE, *GloptiPoly: Global optimization over polynomials with Matlab and SeDuMi*, ACM Transactions Math. Soft. **29**:165–194, 2003.

[62] ———, *Detecting global optimality and extracting solutions in GloptiPoly*, In *Positive Polynomials in Control*, D. HENRION AND A. GARULLI (EDS.), Lecture Notes on Control and Information Sciences **312**:293–310, Springer, Berlin, 2005.

[63] D. HENRION, J.-B. LASSERRE, AND J. LÖFBERG, *GloptiPoly 3: moments, optimization and semidefinite programming*, Optimization Methods and Software **24(5 & 5)**:761-779, 2009.

[64] D. HILBERT, *Über die Darstellung definiter Formen als Summe von Formenquadraten*, Mathematische Annalen **32**:342–350, 1888. See Ges. Abh. **2**:154–161, Springer, Berlin, reprinted by Chelsea, New York, 1981.

[65] T. JACOBI AND A. PRESTEL, *Distinguished representations of strictly positive polynomials*, Journal für die Reine und Angewandte Mathematik **532**:223–235, 2001.

[66] J. JAKUBOVIĆ, *Factorization of symmetric matrix polynomials*, Dokl. Akad. Nauk SSSR **194**:532–535, 1970.

[67] L. JANSSON, J.B. LASSERRE, C. RIENER, AND T. THEOBALD, *Exploiting symmetries in SDP-relaxations for polynomial optimization*, Optimization Online, 2006.

[68] D. JIBETEAN, *Algebraic Optimization with Applications to System Theory*, Ph.D thesis, Vrije Universiteit, Amsterdam, The Netherlands, 2003.

[69] D. JIBETEAN AND M. LAURENT, *Semidefinite approximations for global unconstrained polynomial optimization*, SIAM Journal on Optimization **16**:490–514,

2005.

[70] A. Kehrein, M. Kreuzer, and L. Robbiano, *An algebraist's view on border bases*, In *Solving Polynomial Equations - Foundations, Algorithms and Applications*, A. Dickenstein and I.Z. Emiris (eds.), pages 169–202. Springer, 2005.

[71] M. Kojima, S. Kim, and H. Waki, *Sparsity in sums of squares of polynomials*, Mathematical Programming **103**:45–62, 2005.

[72] M. Kojima, M. Muramatsu, *A note on sparse SOS and SDP relaxations for polynomial optimization problems over symmetric cones*, Research Report B-421, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, 2006.

[73] M.G. Krein and A. Nedel'man. *The Markov moment problem and extremal problems*, Transl. Math. Monographs **50**, Americ. Math. Society, Providence, 1977.

[74] J.L. Krivine, *Anneaux préordonnés*, J. Analyse Math. **12**:307–326, 1964.

[75] ———, *Quelques propriétés des préordres dans les anneaux commutatifs unitaires*, C.R. Académie des Sciences de Paris **258**:3417–3418, 1964.

[76] S. Kuhlman and M. Marshall, *Positivity, sums of squares and the multidimensional moment problem*, Transactions of the American Mathematical Society **354**:4285–4301, 2002.

[77] H. Landau (ed.), *Moments in Mathematics*, Proceedings of Symposia in Applied Mathematics **37**, 1–15, AMS, Providence, 1987.

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

[79] ———, *An explicit exact SDP relaxation for nonlinear* $0-1$ *programs* In K. Aardal and A.M.H. Gerards (eds.), Lecture Notes in Computer Science **2081**:293–303, 2001.

[80] ———, *Polynomials nonnegative on a grid and discrete representations*, Transactions of the American Mathematical Society **354**:631–649, 2001.

[81] ———, *Semidefinite programming vs LP relaxations for polynomial programming*, Mathematics of Operations Research **27**:347–360, 2002.

[82] ———, *SOS approximations of polynomials nonnegative on a real algebraic set*, SIAM Journal on Optimization **16**:610–628, 2005.

[83] ———, *Polynomial programming: LP-relaxations also converge*, SIAM Journal on Optimization **15**:383-393, 2005.

[84] ———, *A sum of squares approximation of nonnegative polynomials*, SIAM Journal on Optimization **16**:751–765, 2006.

[85] ———, *Convergent semidefinite relaxations in polynomial optimization with sparsity*, SIAM Journal on Optimization **17**:822–843, 2006.

[86] ———, *Sufficient conditions for a real polynomial to be a sum of squares*, Archiv der Mathematik **89(5)**:390–398, 2007.

[87] ———, *A Positivstellensatz which preserves the coupling pattern of variables*, arXiv:math/0609529, September 2006.

[88] ———, *Convexity in semialgebraic geometry and polynomial optimization*, iSIAM Journal on OPtimization **19**:1995–2014, 2009.

[89] ———, *Representation ofnonnegative convex polynomials*, Archiv der Mathematik **91**:126–130, 2008.

[90] J.B. Lasserre, M. Laurent, and P. Rostalski, *Semidefinite characterization and computation of real radical ideals*, Foundations of Computational Mathematics **8**:607–647, 2008.

[91] ———, *A unified approach for real and complex zeros of zero-dimensional ideals*, In *Emerging Applications of Algebraic Geometry*, Vol. 149 of IMA Volumes in Mathematics and its Applications, M. Putinar and S. Sullivant (eds.), Springer, pages 125-155, 2009.

[92] J.B. Lasserre, and T. Netzer, *SOS approximations of nonnegative polynomials via simple high degree perturbations*, Mathematische Zeitschrift **256**:99–112,

2006.

[93] M. LAURENT, *A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming*, Mathematics of Operations Research **28**(3):470–496, 2003.

[94] ———, *Semidefinite relaxations for Max-Cut*, In The Sharpest Cut: The Impact of Manfred Padberg and His Work. M. Grötschel, ed., pages 257-290, MPS-SIAM Series in Optimization 4, 2004.

[95] ———, *Revisiting two theorems of Curto and Fialkow on moment matrices*, Proceedings of the American Mathematical Society **133**(10):2965–2976, 2005.

[96] ———, Semidefinite representations for finite varieties. *Mathematical Programming*, 109:1–26, 2007.

[97] ———, *Strengthened semidefinite programming bounds for codes*, Mathematical Programming B 109:239–261, 2007.

[98] M. LAURENT AND B. MOURRAIN, *A generalized flat extension theorem for moment matrices*, Archiv der Mathematik **93(1)**:87-98, 2009.

[99] M. LAURENT AND F. RENDL, *Semidefinite Programming and Integer Programming*, In Handbook on Discrete Optimization, K. Aardal, G. Nemhauser, R. Weismantel (eds.), pages 393-514, Elsevier B.V., 2005.

[100] J. LÖFBERG AND P. PARRILO, *From coefficients to samples: a new approach to SOS optimization*, 43rd IEEE Conference on Decision and Control, Vol. 3, pages 3154–3159, 2004.

[101] L. LOVÁSZ, *On the Shannon capacity of a graph.* IEEE Transactions on Information Theory **IT-25**:1–7, 1979.

[102] L. LOVÁSZ AND A. SCHRIJVER, *Cones of matrices and set-functions and* $0-1$ *optimization*, SIAM Journal on Optimization **1**:166–190, 1991.

[103] M. MARSHALL, *Positive polynomials and sums of squares*, Dottorato de Ricerca in Matematica, Dipartimento di Matematica dell'Università di Pisa, 2000.

[104] ———, *Optimization of polynomial functions*, Canadian Math. Bull. **46**:575–587, 2003.

[105] ———, *Representation of non-negative polynomials, degree bounds and applications to optimization,* Canad. J. Math. **61(1)**:205-221, 2009.

[106] ———, *Positive Polynomials and Sums of Squares*, Mathematical Surveys and Monographs, vol. 146, AMS, 2008.

[107] H.M. MÖLLER, *An inverse problem for cubature formulae*, Vychislitel'nye Tekhnologii (Computational Technologies) **9**:13–20, 2004.

[108] H.M. MÖLLER AND H.J. STETTER, *Multivariate polynomial equations with multiple zeros solved by matrix eigenproblems*, Numerische Mathematik **70**:311–329, 1995.

[109] T.S. MOTZKIN AND E.G. STRAUS, *Maxima for graphs and a new proof of a theorem of Túran*, Canadian Journal of Mathematics **17**:533–540, 1965.

[110] B. MOURRAIN, *A new criterion for normal form algorithms*, In *Proceedings of the 13th International Symposium on Applied Algebra, Algebraic Algorithms and Error-Correcting Codes,* H. Imai, S. Lin, and A. Poli (eds.), vol. 1719 of *Lecture Notes In Computer Science*, pages 430–443. Springer Verlag, 1999.

[111] K.G. MURTY AND S.N. KABADI, *Some NP-complete problems in quadratic and nonlinear programming*, Mathematical Programming **39**:117–129, 1987.

[112] Y.E. NESTEROV, *Squared functional systems and optimization problems*, In J.B.G. FRENK, C. ROOS, T. TERLAKY, AND S. ZHANG (EDS.), *High Performance Optimization*, 405–440, Kluwer Academic Publishers, 2000.

[113] Y.E. NESTEROV AND A. NEMIROVSKI, *Interior Point Polynomial Methods in Convex Programming*, Studies in Applied Mathematics, vol. 13, SIAM, Philadelphia, PA, 1994.

[114] J. NIE, *Sum of squares method for sensor network localization*, Optimization Online, 2006,

[115] J. NIE AND J. DEMMEL, *Sparse SOS Relaxations for Minimizing Functions that are Summation of Small Polynomials*, Computational Optimization and Ap-

plications **43(2)**:151–179, 2007.

[116] J. Nie, J. Demmel, and B. Sturmfels, *Minimizing polynomials via sums of squares over the gradient ideal*, Mathematical Programming Series A **106**:587-606, 2006.

[117] J. Nie and M. Schweighofer, *On the complexity of Putinar's Positivstellensatz*, Journal of Complexity **23(1)**:135–150, 2007.

[118] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer Verlag, 2000.

[119] A.E. Nussbaum, *Quasi-analytic vectors,* Archiv. Mat. **6**:179–191, 1966.

[120] P.M. Pardalos and S.A. Vavasis, *Open questions in complexity theory for numerical optimization*, Mathematical Programming **57(2)**:337–339, 1992.

[121] P.A. Parrilo, *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*, PhD thesis, California Institute of Technology, 2000.

[122] ———, *Semidefinite programming relaxations for semialgebraic problems*, Mathematical Programming B **96**:293–320, 2003.

[123] ———, *An explicit construction of distinguished representations of polynomials nonnegative over finite sets*, IfA Technical Report AUT02-02, ETH Zürich, 2002.

[124] ———, *Exploiting algebraic structure in sum of squares programs*, In *Positive Polynomials in Control*, D. Henrion and A. Garulli, eds., LNCIS **312**:181–194, 2005.

[125] P.A. Parrilo and B. Sturmfels, *Minimizing polynomial functions*, In *Algorithmic and Quantitative Real Algebraic geometry*, S. Basu and L. Gonzáles-Vega, eds., DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 60, pp. 83–99, 2003.

[126] G. Pólya, *Über positive Darstellung von Polynomen*, Vierteljahresschrift der Naturforschenden Gesellschaft in Zürich **73**:141–145, 1928. Reprinted in Collected Papers, vol. 2, MIT Press, Cambridge, MA, pp. 309–313, 1974.

[127] V. Powers and B. Reznick, *Polynomials that are positive on an interval*, Transactions of the American Mathematical Society **352(10)**:4677–4692, 2000.

[128] V. Powers and B. Reznick, *A new bound for Pólya's theorem with applications to polynomials positive on polyhedra*, Journal of Pure and Applied Algebra **164**(1-2):221–229, 2001.

[129] V. Powers, B. Reznick, C. Scheiderer, and F. Sottile, *A new approach to Hilbert's theorem on ternary quartics*, C. R. Acad. Sci. Paris, Ser. I **339**:617–620, 2004.

[130] V. Powers and C. Scheiderer, *The moment problem for non-compact semialgebraic sets*, Adv. Geom. **1**:71–88, 2001.

[131] V. Powers and T. Wörmann, *An algorithm for sums of squares of real polynomials*, Journal of Pure and Applied Algebra **127**:99–104, 1998.

[132] S. Prajna, A. Papachristodoulou, P. Seiler and P.A. Parrilo, *SOSTOOLS (Sums of squares optimization toolbox for MATLAB) User's guide*, `http://www.cds.caltech.edu/sostools/`

[133] A. Prestel and C.N. Delzell, *Positive Polynomials – From Hilbert's 17th Problem to Real Algebra*, Springer, Berlin, 2001.

[134] M. Putinar, *Positive polynomials on compact semi-algebraic sets*, Indiana University Mathematics Journal **42**:969–984, 1993.

[135] ———, *A note on Tchakaloff's theorem*, *Proceedings of the American Mathematical Society* **125**:2409–2414, 1997.

[136] J. Renegar, *A Mathematical View of Interior-Point Methods in Convex Optimization*, MPS-SIAM Series in Optimization, 2001.

[137] B. Reznick, *Extremal PSD forms with few terms*, Duke Mathematical Journal **45**(2):363–374, 1978.

[138] ———, *Uniform denominators in Hilbert's Seventeenth Problem,* Mathematische Zeitschrift **220**: 75-98, 1995.

[139] ———, *Some concrete aspects of Hilbert's 17th problem*, In *Real Algebraic Geom-*

etry and Ordered Structures, C.N. Delzell and J.J. Madden (eds.), Contemporary Mathematics **253**:251–272, 2000.

[140] ———, On Hilbert's construction of positive polynomials, 2007, `http://front.math.ucdavis.edu/0707.2156`.

[141] T. Roh and L. Vandenberghe, *Discrete transforms, semidefinite programming and sum-of-squares representations of nonnegative polynomials*, SIAM Journal on Optimization **16**:939–964, 2006.

[142] J. Saxe, *Embeddability of weighted graphs in k-space is strongly NP-hard*, In *Proc. 17th Allerton Conf. in Communications, Control, and Computing*, Monticello, IL, pp. 480–489, 1979.

[143] C. Scheiderer, Personal communication, 2004.

[144] ———*Positivity and sums of squares: A guide to recent results*, In *Emerging Applications of Algebraic Geometry*, Vol. 149 of IMA Volumes in Mathematics and its Applications, M. Putinar and S. Sullivant (eds.), Springer, pages 1–54, 2009.

[145] K. Schmüdgen, *The K-moment problem for compact semi-algebraic sets*, Mathematische Annalen **289**:203–206, 1991.

[146] A. Schrijver, *Theory of Linear and Integer Programming*, Wiley, 1979.

[147] ———, *New code upper bounds from the Terwilliger algebra and semidefinite programming*, IEEE Trans. Inform. Theory **51**:2859–2866, 2005.

[148] M. Schweighofer, *An algorithmic approach to Schmüdgen's Positivstellensatz*, Journal of Pure and Applied Algebra **166**:307–319, 2002.

[149] ———, *On the complexity of Schmüdgen's Positivstellensatz*, Journal of Complexity **20**:529–543, 2004.

[150] ———, *Optimization of polynomials on compact semialgebraic sets*, SIAM Journal on Optimization **15**(3):805–825, 2005.

[151] ———, *Global optimization of polynomials using gradient tentacles and sums of squares*, SIAM Journal on Optimization **17**(3):920–942, 2006.

[152] J.-P. Serre, *Linear Representation of Finite Groups*, Graduate Texts in Mathematics, Vol. 42, Springer Verlag, New York, 1977.

[153] I.R. Shafarevich, *Basic Algebraic Geometry*, Springer, Berlin, 1994.

[154] H.D. Sherali and W.P. Adams, *A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems*, SIAM Journal on Discrete Mathematics, **3**:411–430, 1990.

[155] N.Z. Shor, *An approach to obtaining global extremums in polynomial mathematical programming problems*, Kibernetika, **5**:102–106, 1987.

[156] ———, *Class of global minimum bounds of polynomial functions*, Cybernetics **23(6)**:731–734, 1987. (Russian orig.: Kibernetika **6**:9–11, 1987.)

[157] ———, *Quadratic optimization problems*, Soviet J. Comput. Systems Sci. **25**:1–11, 1987.

[158] ———, *Nondifferentiable Optimization and Polynomial Problems*, Kluwer, Dordrecht, 1998.

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

[160] J. Stochel, *Solving the truncated moment problem solves the moment problem*, Glasgow Journal of Mathematics, **43**:335–341, 2001.

[161] B. Sturmfels, *Solving Systems of Polynomial Equations.* CBMS, Regional Conference Series in Mathematics, Number 97, AMS, Providence, 2002.

[162] V. Tchakaloff, Formules de cubatures mécaniques à coefficients non négatifs. *Bulletin des Sciences Mathématiques*, **81**:123–134, 1957.

[163] J.A. Telle and A. Proskurowski, *Algorithms for vertex partitioning problems on partial k-trees*, SIAM Journal on Discrete Mathematics **10(4)**:529–550, 1997.

[164] M. Todd, *Semidefinite optimization*, Acta Numer. **10**:515–560, 2001.

[165] F. Vallentin, *Optimal Embeddings of Distance Regular Graphs into Euclidean Spaces*, Journal of Combinatorial Theory, Series B **98**:95–104, 2008.

[166] ———, *Symmetry in semidefinite programs*, Linear Algebra and Applications **430**: 360-369, 2009.

[167] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Review **38**:49–95, 1996.

[168] L. Van Den Dries, *Tame topology and o-minimal structures*, Cambridge University Press, Cambridge, 1998.

[169] H.H. Vui and P.T. Son. Minimizing polynomial functions. *Acta Mathematica Vietnamica* **32(1)**:71–82, 2007.

[170] H.H. Vui and P.T. Son. Global optimization of polynomials using the truncated tangency variety and sums of squares. *SIAM Journal on Optimization* 19(2):941–951, 2008.

[171] H. Waki, Personal communication, 2007.

[172] H. Waki, S. Kim, M. Kojima, M. Muramatsu, *Sums of squares and semidefinite programming relaxations for polynomial optimization problems with structured sparsity*, SIAM Journal on Optimization **17**(1):218–242, 2006.

[173] H. Waki, M. Nakata and M. Muramatsu, *Strange behaviors of interior-point methods for solving semidefinite programming problems in polynomial optimization*, Computational Optimization and Applications, to appear.

[174] H. Whitney, *Elementary structure of real algebraic varieties*, Annals of Mathematics **66**(3):545–556, 1957.

[175] T. Wimer, *Linear Time Algorithms on k-Terminal Graphs*, Ph.D. thesis, Clemson University, Clemson, SC, 1988.

[176] H. Wolkowicz, R. Saigal, and L. Vandeberghe (eds.), *Handbook of Semidefinite Programming*, Boston, Kluwer Academic, 2000.

[177] T. Wörmann, *Strikt positive Polynome in der semialgebraischen Geometrie*, Dissertation, Univ. Dortmund, 1998.