Optimization over polynomials: Selected topics

Monique Laurent

Abstract. Minimizing a polynomial function over a region defined by polynomial inequalities models broad classes of hard problems from combinatorics, geometry and optimization. New algorithmic approaches have emerged recently for computing the global minimum, by combining tools from real algebra (sums of squares of polynomials) and functional analysis (moments of measures) with semidefinite optimization. Sums of squares are used to certify positive polynomials, combining an old idea of Hilbert with the recent algorithmic insight that they can be checked efficiently with semidefinite optimization. The dual approach revisits the classical moment problem and leads to algorithmic methods for checking optimality of semidefinite relaxations and extracting global minimizers. We review some selected features of this general methodology, illustrate how it applies to some combinatorial graph problems, and discuss links with other relaxation methods.

Mathematics Subject Classification (2010). Primary 44A60, 90C22, 90C27, 90C30; Secondary 14P10, 13J30, 15A99.

Keywords. Positive polynomial, sum of squares, moment problem, combinatorial optimization, semidefinite optimization, polynomial optimization.

1. Introduction

Polynomial optimization. We consider the following *polynomial optimization problem*: given multivariate polynomials $f, g_1, \ldots, g_m \in \mathbb{R}[\mathbf{x}_1, \ldots, \mathbf{x}_n]$, compute the infimum of the polynomial function f over the basic closed semialgebraic set

$$K = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \ldots, g_m(x) \geq 0\} \tag{1.1}$$

defined by the polynomial inequalities $g_j(x) \geq 0$. That is, compute

$$f_{\text{min}} := \inf_{x \in K} f(x) = \inf\{f(x) : g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}. \tag{P}$$

This is a in general hard, nonlinear and nonconvex optimization problem which models a multitude of problems from combinatorics, geometry, control and many other areas of mathematics and its applications.

Well established methods from nonlinear optimization can be used to tackle problem (P), which however can only guarantee to find *local* minimizers. Exploiting the fact that the functions f, g_j are polynomials, new algorithmic methods have emerged in the past decade that may permit to find *global* minimizers. These methods rely on using algebraic tools

■ Proceedings of International Congress of Mathematicians, 2014, Seoul

(*sums of squares of polynomials*) and analytic tools (*moments of measures*) combined with *semidefinite optimization*.

In a nutshell, sums of squares of polynomials are used to certify positive polynomials, the starting point being that finding $f_{\min}$ amounts to finding the largest scalar λ for which the polynomial $f-\lambda$ is nonnegative on the set K. The key insight is that, while it is hard to test whether a polynomial f is nonnegative, one can test whether f can be written as a sum of squares of polynomials using semidefinite optimization.

Moments of measures are used to model the nonlinearities arising in polynomial functions, the starting point being that finding $f_{\min}$ amounts to finding a positive measure μ on the set K minimizing the integral $\int_K f(x)d\mu = \sum_\alpha f_\alpha \int_K x^\alpha d\mu$. These moments are used to build certain positive semidefinite Hankel type matrices. The key feature of these matrices is that they permit to certify optimality and to find the global minimizers of problem (P) (under certain conditions).

Semidefinite optimization is a wide generalization of the classical tool of linear optimization, where vector variables are replaced by matrix variables constrained to be positive semidefinite. In other words semidefinite optimization is linear optimization over affine sections of the cone of positive semidefinite matrices. The crucial property is that there are efficient algorithms for solving semidefinite programs (to any arbitrary precision).

Sums of squares and moment based methods permit to construct convex relaxations for the original problem (P), whose optimal values can be computed with semidefinite optimization and provide hiearchies of bounds for the global minimum $f_{\min}$. Convergence properties rely on real algebraic results (giving sums of squares certificates for positive polynomials), and optimality conditions and techniques for extracting global minimizers rely on functional analytic results for moment sequences combined with commutative algebra. Hence these methods have their roots in some classical mathematical results, going back to work of Hilbert about positive polynomials and sums of squares and to work on the classical moment problem in the early 1900's. They also use some recent algebraic and functional analytic developments combined with some modern optimization techniques that emerged since a few decades.

Some combinatorial examples. When all polynomials in (P) are linear, problem (P) boils down to linear programming:

$$\min\{c^{\mathsf{T}}x : Ax \geq b\}, \qquad \text{(LP)}$$

well known to be solvable in polynomial time. However, when adding in (LP) the quadratic conditions $x_i^2 = x_i$ on the variables, we get 0/1 integer linear programming (ILP), which is hard. Instances of polynomial optimization problems arise naturally from combinatorial problems.

Consider for instance the *partition problem*, which asks whether a given sequence $a_1, \ldots, a_n$ of integers can be partitioned into two classes with equal sums, well known to be NP-complete [31]. This amounts to deciding whether the minimum over $\mathbb{R}^n$ of the polynomial $f = (\sum_{i=1}^n a_i \mathbf{x}_i)^2 + \sum_{i=1}^n (\mathbf{x}_i^2 - 1)^2$ is equal to 0.

We now mention other NP-hard problems, dealing with cuts, stable sets, graph colorings, and matrix copositivity, to which we will come back later in the paper.

Max-cut. Consider a graph $G=(V,E)$ with edge weights $w=(w_{ij}) \in \mathbb{R}^E$. The *max-cut problem* asks for a partition of the vertices of G into two classes in such a way that the total weight of the edges crossing the partition is maximum. Encoding partitions by vectors in

$\{\pm 1\}^V$, we obtain the following polynomial optimization problem:

$$\mathrm{mc}(G,w) = \max_{x\in\mathbb{R}^V}\{\sum_{\{i,j\}\in E}(w_{ij}/2)(1-x_ix_j) : x_i^2 = 1\ (i\in V)\}, \tag{1.2}$$

which models the max-cut problem. A basic idea to arrive at a semidefinite relaxation of problem (1.2) is to observe that, for any $x \in \{\pm 1\}^V$, the matrix $X = xx^\mathsf{T}$ is positive semidefinite and all its diagonal entries are equal to 1. This leads to the following problem:

$$\mathrm{sdp}(G,w) = \max_{X\in\mathbb{R}^{V\times V}}\{\sum_{\{i,j\}\in E}(w_{ij}/2)(1-X_{ij}) : X_{ii} = 1\ (i\in V),\ X\succeq 0\}, \tag{1.3}$$

where the notation $X \succeq 0$ means that X is symmetric positive semidefinite (i.e., $x^\mathsf{T}Xx \geq 0$ for all $x \in \mathbb{R}^V$). Of course if we would add the condition that X must have rank 1, then this would be a reformulation of the max-cut problem, thus intractable. The program (1.3) is an instance of semidefinite program and it can be solved in polynomial time (to any precision) as will be recalled below. This is the semidefinite program used by Goemans and Williamson [34] in their celebrated 0.878-approximation algorithm for max-cut. They show that for nonnegative edge weights the integrality gap $\mathrm{mc}(G,w)/\mathrm{sdp}(G,w)$ is at least 0.878 and they introduce a novel rounding technique to produce a good cut from an optimal solution to the semidefinite program (1.3). This is a breakthrough application of semidefinite optimization to the design of approximation algorithms, which started much of the research activity in this field (see e.g. [32]).

Stable sets and colorings. A stable set in a graph $G = (V,E)$ is a set of vertices that does not contain any edge. The *stability number* $\alpha(G)$ of G is the maximum cardinality of a stable set in G. It can be computed with any of the following two programs:

$$\alpha(G) = \max_{x\in\mathbb{R}^V}\sum_{i\in V}x_i \ \text{ s.t. } x_ix_j = 0\ (\{i,j\}\in E),\ x_i^2 = x_i\ (i\in V), \tag{1.4}$$

$$\frac{1}{\alpha(G)} = \min_{x\in\mathbb{R}^V}\ x^T(I+A_G)x \ \text{ s.t. } \sum_{i\in V}x_i = 1,\ x_i\geq 0\ (i\in V), \tag{1.5}$$

where A_G is the adjacency matrix of G (see [24] for (1.5)). As computing $\alpha(G)$ is NP-hard, we find again that problem (P) is hard as soon as some nonlinearities occur, either in the constraints (as in (1.4)), or in the objective function (as in (1.5)). Both formulations are useful to construct hierarchies of bounds for $\alpha(G)$.

The *chromatic number* $\chi(G)$ of G is the minimum number of colors needed to color the vertices so that adjacent vertices receive distinct colors. There is a classic reduction to the stability number. Consider the cartesian product $G\Box K_k$ of G and the complete graph on k nodes, whose edges are the pairs $\{(i,h),(j,h')\}$ with $i = j \in V$ and $h \neq h' \in [k]$, or with $\{i,j\} \in E$ and $h = h' \in [k]$. Then a stable set in the cartesian product $G\Box K_k$ corresponds to a subset of vertices of G that can be properly colored with k colors. Hence k colors suffice to properly color all the vertices of G precisely when $\alpha(G\Box K_k) = |V|$. Therefore, $\chi(G)$ is the smallest integer k for which $\alpha(G\Box K_k) = |V|$. This reduction will be useful for deriving hierarchies of bounds for $\chi(G)$ from bounds for $\alpha(G)$.

A well known bound for both $\alpha(G)$ and $\chi(G)$ is provided by the celebrated *theta number*

$\vartheta(G)$ of Lovász [70], defined by the following semidefinite program:

$$\vartheta(G) = \max_{X\in\mathbb{R}^{V\times V}} \{ \sum_{i,j\in V} X_{ij} : \mathrm{Tr}(X) = 1,\ X_{ij} = 0\ (\{i,j\}\in E),\ X \succeq 0\}. \tag{1.6}$$

The following inequalities hold, known as *Lovász' sandwich inequalities*:

$$\alpha(G) \le \vartheta(G) \le \chi(\overline{G}) \ \text{ and }\ \omega(G) \le \vartheta(\overline{G}) \le \chi(G). \tag{1.7}$$

Here, $\overline{G}$ is the complement of G and $\omega(G) = \alpha(\overline{G})$ is the maximum cardinality of a clique (a set of pairwise adjacent vertices) in G. The inequality $\alpha(G) \le \vartheta(G)$ is easy: any stable set S of G gives a feasible solution $X = \chi^S(\chi^S)^{\mathsf{T}}/|S|$ of the program (1.6), where $\chi^S \in \{0,1\}^V$ is the characteristic vector of S.

A graph G is called *perfect* if $\omega(H) = \chi(H)$ for every induced subgraph H of G. Chudnovsky et al. [14] showed that a graph G is perfect if and only if it does not contain an odd cycle of length at least five or its complement as an induced subgraph. In view of (1.7), we have $\alpha(G) = \vartheta(G)$ and $\chi(G) = \vartheta(\overline{G})$ for perfect graphs. Therefore, both parameters $\alpha(G)$ and $\chi(G)$ can be computed in polynomial time for perfect graphs, via the computation of the theta number, using semidefinite optimization. Moreover, maximum stable sets and minimum graph colorings can also be found in polynomial time [36]. This is an early breakthrough application of semidefinite optimization to combinatorial optimization and as of today no other efficient algorithm is known for these problems.

One can strengthen the theta number toward $\alpha(G)$ by adding in program (1.6) the nonnegativity constraint $X \ge 0$ on the entries of X (leading to the parameter $\vartheta'(G)$), and toward $\chi(G)$ by replacing the constraint $X_{ij} = 0$ by $X_{ij} \le 0$ for all edges (leading to the parameter $\vartheta^+(G)$). Thus we have:

$$\alpha(G) \le \vartheta'(G) \le \vartheta(G) \le \vartheta^+(G) \le \chi(\overline{G}). \tag{1.8}$$

We will see how to build hierarchies of bounds toward $\alpha(G)$ and $\chi(G)$ strenghtening the parameters ϑ' and ϑ^+, using the sums of squares and moment approaches.

Copositive matrices. Another interesting instance of unconstrained polynomial optimization is *testing matrix copositivity*, which is a hard problem [27, 74]. Recall that a symmetric $n \times n$ matrix M is called *copositive* if the quadratic form $x^{\mathsf{T}}Mx$ is nonnegative over the nonnegative orthant $\mathbb{R}^n_+$ or, equivalently, the polynomial $f_M = \sum_{i,j=1}^n M_{ij}\mathbf{x}_i^2\mathbf{x}_j^2$ is nonnegative over $\mathbb{R}^n$. Starting with the formulation (1.5) of the stability number $\alpha(G)$, it follows that $\alpha(G)$ can also be computed with the following copositive program:

$$\alpha(G) = \min_{\lambda\in\mathbb{R}} \{\lambda : \lambda(I + A_G) - J \ \text{ is copositive}\}, \tag{1.9}$$

where J is the all-ones matrix. By using sums of squares certificates for certifying matrix copositivity, one can define a hierarchy of cones approximating the copositive cone, which can also be used to define hierarchies of semidefinite bounds for the parameters $\alpha(G)$ and $\chi(G)$.

This paper. The field of polynomial optimization has grown considerably in the recent years. It has roots in early work of Shor [97] and later of Nesterov [75], and the foundations were laid by the groundworks of Lasserre [53, 54] and Parrilo [82, 83]. The monograph of

Lasserre [56], our overview [68] and the handbook [1] can serve as a general source about polynomial optimization. We also refer to the monographs [72, 85] and to the overview [91] for an in-depth treatment of real algebraic aspects, and to the monograph [9] for links to convexity.

In this paper we will discuss only a small selection of results from this field. Inevitably we cannot make full references to the literature and we apologize for all omissions. We will treat some subjects where we have done some (modest) contributions and our choices are biased, in particular, toward properties of the moment relaxations and toward hierarchies of semidefinite bounds for combinatorial problems. Our interest in polynomial optimization was stirred by the work [54] of Lasserre explaining how his method applies to 0/1 linear programming and we are grateful to Jean Lasserre for his inspiring work. We realized that his approach has tight links with lift-and-project methods for combinatorial optimization. This in turn inspired us to show finite convergence for polynomial optimization over finite varieties, to give simple real algebraic proofs for several results about flat extensions of moment matrices, and to investigate hierarchies for combinatorial graph parameters.

The paper is organized as follows. We begin with preliminaries about semidefinite optimization and sums of squares of polynomials. Then we present the sums of squares and moment approaches for polynomial optimization, with a special focus on the properties of moment matrices that allow to certify optimality and extract global optimizers. Then some selected applications are discussed: for computing real roots of polynomial equations, for designing hierarchies of semidefinite approximations for the stability number and the chromatic number, and for approximating matrix copositivity, again with application to approximating graph parameters. We conclude with mentioning some other research directions where hierarchies of semidefinite relaxations are also being increasingly used.

2. Preliminaries

Notation. $\mathbb{N} = \{0, 1, 2, \ldots\}$ is the set of nonnegative integers, $\mathbb{N}^n_t$ consists of the sequences $\alpha \in \mathbb{N}^n$ with $|\alpha| := \sum_{i=1}^n \alpha_i \leq t$ for $t \in \mathbb{N}$ and, for $\alpha \in \mathbb{N}^n$, $\mathbf{x}^\alpha$ denotes the monomial $\mathbf{x}_1^{\alpha_1} \cdots \mathbf{x}_n^{\alpha_n}$ with degree $|\alpha|$. (We use boldface letters $\mathbf{x}, \mathbf{x}_i, ..$ to denote variables.) $\mathbb{R}[\mathbf{x}_1, \ldots, \mathbf{x}_n] = \mathbb{R}[\mathbf{x}]$ is the ring of polynomials in n variables and $\mathbb{R}[\mathbf{x}]_t$ its subspace of polynomials with degree $\leq t$. The vector $[\mathbf{x}]_t = (\mathbf{x}^\alpha)_{\alpha \in \mathbb{N}^n_t}$ lists the monomials of degree at most t (in some given order) and, for a polynomial $f \in \mathbb{R}[\mathbf{x}]_t$, the vector $\mathbf{f} = (f_\alpha)_{\alpha \in \mathbb{N}^n_t}$ lists the coefficients of f (in the same order), so that $f = \sum_\alpha f_\alpha \mathbf{x}^\alpha = \mathbf{f}^\mathsf{T}[\mathbf{x}]_t$.

Given polynomials $g_1, \ldots, g_m$, we let $\mathcal{I} = (g_1, \ldots, g_m)$ denote the ideal that they generate and, for an integer t, $\mathcal{I}_t$ denotes its truncation at degree t, which consists of all polynomials $\sum_{j=1}^m p_j g_j$ with $p_j \in \mathbb{R}[\mathbf{x}]$ and $\deg(p_j g_j) \leq t$.

A polynomial f is a *sum of squares* (*sos*) if $f = g_1^2 + \ldots + g_m^2$ for some polynomials $g_1, \ldots, g_m$. $\Sigma[\mathbf{x}]$ contains all sums of squares of polynomials and we set $\Sigma[\mathbf{x}]_t = \Sigma[\mathbf{x}] \cap \mathbb{R}[\mathbf{x}]_t$. $\mathcal{P}(K)$ contains all polynomials f that are nonnegative over a given set $K \subseteq \mathbb{R}^n$, i.e., $f(x) \geq 0$ for all $x \in K$, also abbreviated as $f \geq 0$ on K.

Ideals and varieties. Consider an ideal $\mathcal{I} \subseteq \mathbb{R}[\mathbf{x}]$. The sets

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

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

are called, respectively, the *radical* and the *real radical* of $\mathcal{I}$. Moreover, the sets

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

are, respectively, the *(complex) variety* and the *real variety* of the ideal $\mathcal{I}$. If $\mathcal{I} = (g_1, \dots, g_m)$ is the ideal generated by a set of polynomials $g_1, \dots, g_m$, then $V(\mathcal{I})$ consists of all their common complex roots while $V_{\mathbb{R}}(\mathcal{I})$ consists of their common real roots. The *vanishing ideal* of a set $V \subseteq \mathbb{C}^n$ is the set of polynomials

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

The sets $\mathcal{I}(V)$, $\sqrt{\mathcal{I}}$ and $\sqrt[\mathbb{R}]{\mathcal{I}}$ are all ideals in $\mathbb{R}[\mathbf{x}]$ and they satisfy the inclusions:

$$\mathcal{I} \subseteq \sqrt{\mathcal{I}} \subseteq \mathcal{I}(V(\mathcal{I})) \ \text{ and } \ \mathcal{I} \subseteq \sqrt[\mathbb{R}]{\mathcal{I}} \subseteq \mathcal{I}(V_{\mathbb{R}}(\mathcal{I})).$$

The ideal $\mathcal{I}$ is called *radical* if $\mathcal{I} = \sqrt{\mathcal{I}}$ and *real radical* if $\mathcal{I} = \sqrt[\mathbb{R}]{\mathcal{I}}$. For instance, the ideal $\mathcal{I} = (\mathbf{x}^2)$ is not radical since $\mathbf{x} \in \sqrt{\mathcal{I}} \setminus \mathcal{I}$, while the ideal $\mathcal{I} = (\mathbf{x}_1^2 + \mathbf{x}_2^2)$ is radical but not real radical since $\mathbf{x}_1, \mathbf{x}_2 \in \sqrt[\mathbb{R}]{\mathcal{I}} \setminus \mathcal{I}$. The following celebrated results relate (real) radical and vanishing ideals.

Theorem 2.1 ([16, 52, 98]). *Let $\mathcal{I}$ be an ideal in $\mathbb{R}[\mathbf{x}]$. Then, $\sqrt{\mathcal{I}} = \mathcal{I}(V(\mathcal{I}))$* **(Hilbert's Nullstellensatz)** *and $\sqrt[\mathbb{R}]{\mathcal{I}} = \mathcal{I}(V_{\mathbb{R}}(\mathcal{I}))$* **(Real Nullstellensatz)**.

As $\mathcal{I} \subseteq \mathcal{I}(V(\mathcal{I})) \subseteq \mathcal{I}(V_{\mathbb{R}}(\mathcal{I}))$, $\mathcal{I}$ real radical implies $\mathcal{I}$ radical and, moreover, $V(\mathcal{I}) = V_{\mathbb{R}}(\mathcal{I}) \subseteq \mathbb{R}^n$ if the real variety $V_{\mathbb{R}}(\mathcal{I})$ is finite. Moreover, an ideal $\mathcal{I}$ is zero-dimensional precisely when $V(\mathcal{I})$ is finite. Then there is a well known relationship between the cardinality of the variety $V(\mathcal{I})$ and the dimension of the quotient space $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ (see e.g. [16]).

Proposition 2.2. *An ideal $\mathcal{I}$ in $\mathbb{R}[\mathbf{x}]$ is zero-dimensional (i.e., the variety $V(\mathcal{I})$ is finite) if and only if the vector space $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ is finite dimensional. Moreover, we have the inequality: $|V(I)| \leq \dim \mathbb{R}[\mathbf{x}]/\mathcal{I}$, with equality if and only if the ideal $\mathcal{I}$ is radical.*

The eigenvalue method for computing the variety $V(\mathcal{I})$. We now recall how to find the variety $V(\mathcal{I})$ of a zero-dimensional ideal $\mathcal{I}$ by computing the eigenvalues of the multiplication operator in the quotient algebra $\mathbb{R}[\mathbf{x}]/\mathcal{I}$, since this technique is used for finding the global minimizers of problem (P) (see [44]). Given a polynomial $h \in \mathbb{R}[\mathbf{x}]$, consider the 'multiplication by h' linear map in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$:

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

and let M_h denote its matrix in a given linear basis $\mathcal{B} = \{b_1, \dots, b_N\}$ of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$.

Theorem 2.3. *Assume $N = \dim \mathbb{R}[\mathbf{x}]/\mathcal{I} < \infty$, let $\mathcal{B} = \{b_1, \dots, b_N\}$ be a linear basis of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$, and let $[v]_{\mathcal{B}} = (b_1(v), \dots, b_N(v))^{\mathsf{T}}$ be the vector of evaluations at $v \in V(\mathcal{I})$ of the polynomials in $\mathcal{B}$. For any $h \in \mathbb{R}[\mathbf{x}]$, the eigenvalues of the multiplication matrix M_h are the evaluations $h(v)$ of h at the points $v \in V(\mathcal{I})$, with corresponding (left) eigenvectors $[v]_{\mathcal{B}}$. That is, $M_h^{\mathsf{T}}[v]_{\mathcal{B}} = h(v)[v]_{\mathcal{B}}$ for all $v \in V(\mathcal{I})$.*

If $\mathcal{I}$ is radical then $|V(\mathcal{I})| = N$ (by Proposition 2.2) and the matrix M_h has a full set of linearly independent eigenvectors ($[v]_{\mathcal{B}}$ for $v \in V(\mathcal{I})$). These vectors can be found by computing the eigenvalues of M_h^T (assuming the values $h(v)$ are pairwise distinct which can be achieved e.g. by selecting a random linear polynomial h) and it is then easy to recover the points $v \in V(\mathcal{I})$ from these vectors $[v]_\mathcal{B}$.

We illustrate this method applied to the univariate case. Say $\mathcal{I} = (p)$, where p is the polynomial: $p = \mathbf{x}^d - p_{d-1}\mathbf{x}^{d-1} - \ldots - p_0$. The set $\mathcal{B} = \{1, \mathbf{x}, \ldots, \mathbf{x}^{d-1}\}$ is a basis of $\mathbb{R}[\mathbf{x}]/(p)$ and with respect to this basis the 'multiplication by $\mathbf{x}$' matrix has the form

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

Its characteristic polynomial is $\det(M_\mathbf{x} - tI) = (-1)^d p(t)$, hence the eigenvalues of the matrix $M_\mathbf{x}$ are the roots of p and indeed $M_\mathbf{x}^\mathsf{T}[v]_\mathcal{B} = v[v]_\mathcal{B}$ holds if $p(v) = 0$.

Semidefinite optimization. $\mathcal{S}^n$ is the set of real symmetric $n \times n$ matrices, equipped with the trace inner product $\langle X, Y \rangle = \mathrm{Tr}(X^\mathsf{T} Y) = \sum_{i,j=1}^n X_{ij} Y_{ij}$. The notation $X \succeq 0$ means that X is positive semidefinite (i.e., $x^\mathsf{T} X x \geq 0$ for all $x \in \mathbb{R}^n$) and $\mathcal{S}^n_+ \subseteq \mathcal{S}^n$ is the cone of positive semidefinite matrices. The cone $\mathcal{S}^n_+$ is self-dual: $X \in \mathcal{S}^n$ is positive semidefinite if and only if $\langle X, Y \rangle \geq 0$ for all $Y \in \mathcal{S}^n_+$.

Given matrices $C, A_1, \ldots, A_m \in \mathcal{S}^n$ and a vector $b \in \mathbb{R}^m$, a *semidefinite program* in standard primal form and its *dual semidefinite program* read:

$$p^* = \sup_{X \in \mathcal{S}^n} \{\langle C, X \rangle : \langle A_j, X \rangle = b_j \ (j \in [m]),\ X \succeq 0\}, \tag{P-SDP}$$

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

Weak duality holds: $p^* \leq d^*$ (since $X, Y = \sum_{j=1}^m y_j A_j - C \succeq 0$ implies $\langle X, Y \rangle \geq 0$). Moreover, if (P-SDP) is bounded and has a positive definite feasible solution X, then strong duality holds: $p^* = d^*$. Semidefinite programs can be solved (approximatively) in polynomial time, using the ellipsoid method (since one can test if a rational matrix is positive semidefinite using Gaussian elimination). However, the ellipsoid method is not efficient in practice, and efficient algorithms used in practical implementations rely on interior-point algorithms. (See e.g. [5, 21, 99, 100].) On the other hand, the exact complexity is not known of the problem of *testing feasibility* of a semidefinite program: given integral matrices $C, A_1, \ldots, A_m \in \mathcal{S}^n$,

$$\text{decide whether there exists } y \in \mathbb{R}^n \text{ such that } C + y_1 A_1 + \ldots + y_m A_m \succeq 0. \tag{F}$$

An obvious difficulty is that there might be only irrational solutions. It is known that (F) belongs to NP if and only if it belongs to co-NP ([88], see also [51]). Moreover, (F) can be solved in polynomial time when fixing either m or n [46] and, when fixing m, one can check in polynomial time if (F) has a rational solution [46].

Recognizing sums of squares of polynomials. It turns out that checking whether a polynomial $f = \sum_{\alpha \in \mathbb{N}^n_{2t}} f_\alpha \mathbf{x}^\alpha$ can be written as a sum of squares of polynomials amounts to checking whether the following semidefinite program:

$$\sum_{\beta, \gamma \in \mathbb{N}^n_t : \beta + \gamma = \alpha} X_{\beta,\gamma} = f_\alpha \quad (\alpha \in \mathbb{N}^n_{2t}),\ \ X \succeq 0, \tag{2.1}$$

(in the matrix variable $X = (X_{\beta,\gamma})_{\beta,\gamma\in\mathbb{N}^n_t}$) admits a feasible solution. To see this, assume $f = \sum_{j=1}^k p_j^2$. Then each p_j has degree at most t and can be written as $p_j = \sum_\alpha (p_j)_\alpha \mathbf{x}^\alpha = \mathbf{p_j}^\mathsf{T}[\mathbf{x}]_t$, where $\mathbf{p_j} = ((p_j)_\alpha)$ is the vector of coefficients of p_j in the monomial basis. Therefore, $f = \sum_{j=1}^k p_j^2 = [\mathbf{x}]_t^\mathsf{T}(\sum_{j=1}^k \mathbf{p_j}\mathbf{p_j}^\mathsf{T})[\mathbf{x}]_t = [\mathbf{x}]_t^\mathsf{T} P[\mathbf{x}]_t$, where the matrix $P = \sum_{j=1}^k \mathbf{p_j}\mathbf{p_j}^\mathsf{T}$ is positive semidefinite. Moreover, by equating the coefficients of both polynomials f and $[\mathbf{x}]_d^\mathsf{T} P[\mathbf{x}]_d$ in the identity $f = [\mathbf{x}]_t^\mathsf{T} P[\mathbf{x}]_t$, it follows that P satisfies the system (2.1). The argument can be easily reversed: any feasible solution of (2.1) gives rise to a sum of squares decomposition of f.

More generally, given polynomials $f, g_1, \ldots, g_m \in \mathbb{R}[\mathbf{x}]$, the problem of finding a decomposition of the form $f = \sigma_0 + \sigma_1 g_1 + \ldots + \sigma_m g_m$, where $\sigma_0, \sigma_1, \ldots, \sigma_m$ are sums of squares with a given degree bound: $\deg(\sigma_0), \deg(\sigma_j g_j) \leq 2t$, can also be cast as a semidefinite program. This program is analogue to (2.1), but it now involves $m+1$ positive semidefinite matrices $X_0, X_1, \ldots, X_m$, where X_0 is indexed by $\mathbb{N}^n_t$ (corresponding to σ_0) and X_j by $\mathbb{N}^n_{t-\lceil \deg(g_j)/2\rceil}$ (corresponding to σ_j). Of course one should adequately define the affine constraints in the semidefinite program.

3. Positive polynomials and sums of squares

3.1. Positivity certificates. Understanding the link between positive polynomials and sums of squares is a classic question which goes back to work of Hilbert around 1890. Hilbert realized that not every nonnegative polynomial is a sum of squares of polynomials and he characterized when this happens.

Theorem 3.1 (Hilbert [45]). *Every nonnegative polynomial of degree $2d$ in n variables is a sum of squares of polynomials if and only if we are in one of the following three cases: $(n = 1, 2d)$, $(n, 2d = 2)$, and $(n = 2, 2d = 4)$.*

In all other cases, Hilbert showed the existence of a nonnegative polynomial which is not sos. The first explicit construction was found only sixty years later by Motzkin: the *Motzkin polynomial* $M = \mathbf{x}_1^2\mathbf{x}_2^2(\mathbf{x}_1^2 + \mathbf{x}_2^2 - 3) + 1$ is nonnegative but not a sum of squares of polynomials. However, the polynomial $(1 + \mathbf{x}_1^2 + \mathbf{x}_2^2)M$ is a sum of squares of polynomials, which certifies the positivity of M. We refer to [89] for an historic account and for more examples. We also refer to [7] for an in-depth study of the two smallest cases $(n = 2, 2d = 6)$ and $(n = 3, 2d = 4)$ when not all nonnegative polynomials are sums of squares.

If we are not in one of the special three cases of Theorem 3.1, then the inclusion $\Sigma[\mathbf{x}]_{2d} \subseteq \mathcal{P}(\mathbb{R}^n) \cap \mathbb{R}[\mathbf{x}]_{2d}$ is strict. Are these two sets far apart or not? That is, are there few or many sums of squares within nonnegative polynomials? The answer depends whether the degree and the number of variables are fixed or not.

On the one hand, sums of squares are dense within nonnegative polynomials if we allow the degree to grow. Lasserre and Netzer [60] show the following explicit sums of squares approximation: if f is nonnegative over the box $[-1,1]^n$ then for any $\epsilon > 0$ there exists $k \in \mathbb{N}$ such that the perturbed polynomial $f + \epsilon(1 + \sum_{i=1}^n \mathbf{x}_i^{2k})$ is a sum of squares of polynomials. (See also Lasserre [55]).

On the other hand, if we fix the degree but let the number of variables grow, then there are significantly more nonnegative polynomials than sums of squares: Blekherman [6] shows that the ratio of volumes of (sections of) the cone of sums of squares and the cone of nonnegative polynomials tends to 0 as n goes to ∞.

At the 1900 International Congress of Mathematicians in Paris, Hilbert asked whether every nonnegative polynomial can be written as a sum of squares of *rational* functions. This question, known as Hilbert's 17th problem, was answered in the affirmative in 1927 by Artin, whose work led the foundations of the field of real algebraic geometry.

Sums of squares certificates (also known as *Positivstellensätze*) are known for characterizing positivity over a general basic closed semialgebraic set K of the form (1.1). They involve weighted combinations of the polynomials $g_1, \ldots, g_m$ describing the set K. The *quadratic module* generated by $g = (g_1, \ldots, g_m)$ is the set

$$\mathcal{Q}(g) = \{\sigma_0 + \sigma_1 g_1 + \ldots + \sigma_m g_m : \sigma_0, \ldots, \sigma_m \in \Sigma[\mathbf{x}]\}, \tag{3.1}$$

the *truncated quadratic module* $\mathcal{Q}_t(g)$ is its subset obtained by restricting the degrees: $\deg(\sigma_j g_j) \leq 2t$ (setting $g_0 = 1$), and the *preordering* $\mathcal{T}(g)$ is the quadratic module generated by the 2^m polynomials $g^e = g_1^{e_1} \cdots g_m^{e_m}$ for $e \in \{0,1\}^m$.

Theorem 3.2 (Krivine [52], Stengle [98]). *Let $f \in \mathbb{R}[\mathbf{x}]$ and K be as in (1.1).*

(i) $f > 0$ *on K if and only if $fq = 1 + p$ for some $p, q \in \mathcal{T}(g)$.*

(ii) $f \geq 0$ *on K if and only if $fq = f^{2k} + p$ for some $p, q \in \mathcal{T}(g)$ and $k \in \mathbb{N}$.*

(iii) $f = 0$ *on K if and only if $-f^{2k} \in \mathcal{T}(g)$ for some $k \in \mathbb{N}$.*

In each case it is clear that the 'if part' gives a certificate that f is positive (nonnegative, or vanishes) on K, the hard part is showing the existence of such a certificate. These certificates use polynomials in $\mathcal{T}(g)$ and thus they can be checked with semidefinite optimization, once a bound on the degrees has been set. However they are not directly useful for our polynomial optimization problem (P). Indeed, in view of Theorem 3.2 (i), one would need to search for the largest scalar λ for which there exist $p, q \in \mathcal{T}(g)$ such that $(f - \lambda)q = 1 + p$, thus involving a quadratic term λq which cannot be dealt with directly using semidefinite optimization.

To go around this difficulty one may instead use the simpler "denominator free" positivity certificates of Schmüdgen and Putinar, which hold in the case when the semialgebraic set K is compact. The following condition:

$$\exists R > 0 \text{ such that } R - \mathbf{x}_1^2 - \ldots - \mathbf{x}_n^2 \in \mathcal{Q}(g), \tag{A}$$

known as the *Archimedean condition*, allows easier positivity certificates using the quadratic module $\mathcal{Q}(g)$. Note that K is compact if (A) holds.

Theorem 3.3 (Schmüdgen [92]). *Assume that the set K in (1.1) is compact. If the polynomial f is positive on K (i.e., $f(x) > 0$ for all $x \in K$), then $f \in \mathcal{T}(g)$.*

Theorem 3.4 (Putinar [86]). *Assume that the Archimedean condition (A) holds. If the polynomial f is positive on K, then $f \in \mathcal{Q}(g)$.*

3.2. Semidefinite relaxations for (P). Motivated by Putinar's result, Lasserre [53] introduced the following relaxations for the polynomial optimization problem (P). For any integer $t \geq d_f = \lceil \deg(f)/2 \rceil$, consider the parameters

$$f_t^{\text{sos}} = \sup_{\lambda \in \mathbb{R}} \{\lambda : f - \lambda \in \mathcal{Q}_t(g)\}, \tag{SOSt}$$

which form a monotone nondecreasing sequence: $f_t^{\text{sos}} \le f_{t+1}^{\text{sos}} \le \ldots \le f_{\min}$.

Each program (SOSt) can be written as a semidefinite program (recall Section 2). Moreover, the dual semidefinite program can be expressed as follows:

$$f_t^{\text{mom}} = \inf_{L \in \mathbb{R}[\mathbf{x}]_{2t}^*} \{L(f) : L(f) = 1,\ L(p) \ge 0\ \forall p \in \mathcal{Q}_t(g)\}, \qquad \text{(MOMt)}$$

where $\mathbb{R}[\mathbf{x}]_{2t}^*$ denotes the set of linear functionals on $\mathbb{R}[\mathbf{x}]_{2t}$. The parameters $f_{\min}$, f_t^{sos} and f_t^{mom} satisfy:

$$f_t^{\text{sos}} \le f_t^{\text{mom}} \le f_{\min}. \qquad (3.2)$$

The inequality $f_t^{\text{sos}} \le f_t^{\text{mom}}$ is easy (by weak duality) and $f_t^{\text{mom}} \le f_{\min}$ is explained below in Section 4.1. There is no duality gap: $f_t^{\text{sos}} = f_t^{\text{mom}}$, for instance if the set K has an interior point. In the compact case the asymptotic convergence of the bounds to the infimum of f is guaranteed by Putinar's theorem.

Theorem 3.5. *(Lasserre [53]) Assume that assumption (A) holds (and thus K is compact). Then,* $\lim_{t\to\infty} f_t^{sos} = \lim_{t\to\infty} f_t^{mom} = f_{min}$.

Proof. For any $\epsilon > 0$, the polynomial $f - f_{\min} + \epsilon$ is positive on K and thus, by Theorem 3.4, it belongs to $\mathcal{Q}_t(g)$ for some t, which implies $f_t^{\text{sos}} \ge f_{\min} - \epsilon$. □

In order to discuss further properties of the dual (moment) programs (MOMt), we need to go in some detail about the moment problem. This is what we do in the next sections and we come back to the hierarchies later in Section 4.4.

4. Moment sequences and moment matrices

4.1. The moment problem. Given a (positive Borel) measure μ on a set $K \subseteq \mathbb{R}^n$, consider the linear functional $L_\mu \in \mathbb{R}[\mathbf{x}]^*$ defined by

$$L_\mu(f) = \int_K f(x)d\mu = \sum_\alpha f_\alpha \left(\int_K x^\alpha d\mu \right) \quad \text{for } f \in \mathbb{R}[\mathbf{x}], \qquad (4.1)$$

which thus depends linearly on the moments $\int_K x^\alpha d\mu$ of the measure μ. The classical moment problem asks to characterize the linear functionals $L \in \mathbb{R}[\mathbf{x}]^*$ admitting such a representing measure μ, i.e., being of the form $L = L_\mu$. The following result (due to Haviland) makes the link to polynomial positivity: $L = L_\mu$ for some measure μ on K if and only if L is nonnegative on $\mathcal{P}(K)$.

Let us go back to problem (P). Following Lasserre [53], we observe that the infimum of f over the set K can be reformulated as

$$f_{\min} = \inf_\mu \{L_\mu(f) : \mu \text{ is a probability measure on } K\}.$$

Indeed, as $f(x) \ge f_{\min}$ for all $x \in K$, by integrating both sides over K for an arbitrary probability measure μ on K, we obtain that $L_\mu(f) \ge f_{\min}$. For the reverse inequality, choose μ to be the Dirac measure at an arbitrary point $x \in K$, so that $L_\mu(f) = f(x)$ and thus $\inf_\mu L_\mu(f) \le f(x)$.

If μ is a probability measure on K, then L_μ is nonnegative on $\mathcal{P}(K)$ and thus on its subset $\mathcal{Q}_t(g)$, which implies the inequality $f_t^{\text{mom}} \leq f_{\min}$ from (3.2). Moreover, the relaxation (MOMt) is exact, i.e., $f_t^{\text{mom}} = f_{\min}$, if it has an optimal solution of the form L_μ where μ is a probability measure on K. This observation motivates searching for sufficient conditions for existence of a representing measure. This is treated in the rest of the section.

If $L \in \mathbb{R}[\mathbf{x}]^*$ has a representing measure then L must be nonnegative on $\mathcal{P}(K)$ and thus on the subcone $\Sigma[\mathbf{x}]$ of all sums of squares. The nonnegativity condition of L over $\Sigma[\mathbf{x}]$ can be conveniently expressed using the following 'Hankel type' matrix $M(L)$:

$$M(L) = (L(\mathbf{x}^\alpha \mathbf{x}^\beta))_{\alpha,\beta\in\mathbb{N}^n},$$

which is indexed by $\mathbb{N}^n$ and called the *moment matrix* of L.

Indeed, note that $L(pq) = \mathbf{p}^\mathsf{T} M(L)\mathbf{q}$ for any $p, q \in \mathbb{R}[\mathbf{x}]$. Therefore, L is nonnegative over $\Sigma[\mathbf{x}]$ if and only if $M(L) \succeq 0$. Moreover, for $g \in \mathbb{R}[\mathbf{x}]$, L is nonnegative on the set $g\Sigma[\mathbf{x}] = \{g\sigma : \sigma \in \Sigma[\mathbf{x}]\}$ if and only if $M(gL) \succeq 0$, where $gL \in \mathbb{R}[\mathbf{x}]^*$ is the new linear functional defined by $(gL)(p) = L(gp)$ for $p \in \mathbb{R}[\mathbf{x}]$.

For example, in the univariate case, L has a representing measure on $\mathbb{R}$ if and only if $M(L) \succeq 0$ (Hamburger's theorem), L has a representing measure on $\mathbb{R}_+$ if and only if $M(L), M(\mathbf{x}L) \succeq 0$ (Stieltjes' theorem), and L has a representing measure on $[0, 1]$ if and only if $M(\mathbf{x}L), M((1 - \mathbf{x})L) \succeq 0$ (Hausdorff's theorem).

Both Theorems 3.3-3.4 have counterparts for the moment problem. If K is compact, then L has a representing measure on K if and only if $L \geq 0$ on $\mathcal{T}(g)$ (Schmüdgen [92]) or, equivalently, $L \geq 0$ on $\mathcal{Q}(g)$ if (A) holds (Putinar [86]).

4.2. Finite rank moment matrices. As we saw above, a necessary condition for $L \in \mathbb{R}[\mathbf{x}]^*$ to have a representing measure is positive semidefiniteness of its moment matrix. Although not sufficient in general, it turns out that this condition is sufficient in the case when $M(L)$ has finite rank ([17], see Theorem 4.1 below). As this result plays a crucial role for studying the finite convergence of the relaxations (MOMt) for (P), we discuss it in detail.

In what follows, $\operatorname{Ker} M(L)$ denotes the kernel of $M(L)$, which consists of the polynomials $p \in \mathbb{R}[\mathbf{x}]$ for which $L(pq) = 0$ for all $q \in \mathbb{R}[\mathbf{x}]$. Hence $\operatorname{Ker} M(L)$ is an ideal in $\mathbb{R}[\mathbf{x}]$. Moreover, $\operatorname{Ker} M(L)$ is real radical if $M(L) \succeq 0$ (since, when $M(L) \succeq 0$, a polynomial p belongs to $\operatorname{Ker} M(L)$ if and only if $L(p^2) = 0$).

Consider a measure μ and the corresponding linear functional L_μ as in (4.1). Its support is contained in the real variety of the polynomials in the kernel of $M(L_\mu)$: $\operatorname{Supp}(\mu) \subseteq V_\mathbb{R}(\operatorname{Ker} M(L_\mu))$. When $\mu = \delta_v$ is the Dirac measure at a point $v \in \mathbb{R}^n$, L_μ is the *evaluation* L_v at v, defined by $L_v(p) = p(v)$ for all $p \in \mathbb{R}[\mathbf{x}]$. If the suppport of μ is finite (i.e., μ is *finite atomic*), say $\operatorname{Supp}(\mu) = \{v_1, \ldots, v_r\}$, then L_μ is a conic combination of evaluations at the v_i's: $L_\mu = \sum_{i=1}^r \lambda_i L_{v_i}$ for some scalars $\lambda_i > 0$. The following theorem shows that this describes all the linear functionals $L \in \mathbb{R}[\mathbf{x}]^*$ with $M(L) \succeq 0$ and $\operatorname{rank} M(L) < \infty$. We present our simple real algebraic proof from [64] (see also [68]).

Theorem 4.1. *(Curto and Fialkow [17]) Let $L \in \mathbb{R}[\mathbf{x}]^*$. Assume that $M(L) \succeq 0$ and that $M(L)$ has finite rank r. Then L has a (unique) representing measure μ. Moreover, μ is finite atomic with r atoms and supported by $V(\operatorname{Ker} M(L))$.*

Proof. As $M(L) \succeq 0$, its kernel $\mathcal{I} := \operatorname{Ker} M(L)$ is a real radical ideal in $\mathbb{R}[\mathbf{x}]$.

Moreover, the quotient space $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ has finite dimension r. This is because we have: $\operatorname{rank} M(L) = r$ and any set of monomials $\mathcal{B}$ indexing a maximal linearly independent set of columns of $M(L)$ is also maximal linearly independent in $\mathbb{R}[\mathbf{x}]/\mathcal{I}$.

Applying Proposition 2.2, we can conclude that the variety of the ideal $\mathcal{I}$ is contained in $\mathbb{R}^n$ and has cardinality r. Set $V(\mathcal{I}) = \{v_1, \dots, v_r\} \subseteq \mathbb{R}^n$.

We consider interpolation polynomials $p_{v_1}, \dots, p_{v_r} \in \mathbb{R}[\mathbf{x}]$ at the points of $V(\mathcal{I})$, i.e., satisfying $p_{v_i}(v_j) = \delta_{i,j}$. As the polynomial $p_{v_i} - p_{v_i}^2$ vanishes on the variety $V(\mathcal{I})$, it belongs to the ideal $\mathcal{I}(V(\mathcal{I}))$, which is equal to $\mathcal{I}$ (since $\mathcal{I}$ is real radical). Hence, $L(p_{v_i}) = L(p_{v_i}^2)$, since $p_{v_i} - p_{v_i}^2 \in \mathcal{I} = \text{Ker } M(L)$. Moreover, $L(p_{v_i}^2) \geq 0$ since $M(L) \succeq 0$. Furthermore, $L(p_{v_i}^2) \neq 0$, since otherwise p_{v_i} would belong to $\text{Ker}\, M(L)$ and thus it would vanish at v_i, a contradiction.

We now claim that $L = \sum_{i=1}^r L(p_{v_i}) L_{v_i}$. Indeed, any $p \in \mathbb{R}[\mathbf{x}]$ can be written as $p = \sum_{i=1}^r p(v_i) p_{v_i} + q$, where $q \in \mathcal{I}$. Hence, $L(q) = 0$ and thus $L(p) = \sum_{i=1}^r p(v_i) L(p_{v_i}) = \sum_{i=1}^r L_{v_i}(p) L(p_{v_i})$. Hence we have shown that L has a finite r-atomic representing measure: $\mu = \sum_{i=1}^r L(p_{v_i}) \delta_{v_i}$, which concludes the proof. □

4.3. Flat extensions of truncated moment matrices. To make the link with the relaxations (MOMt) for problem (P), we introduce the *truncated moment matrix* of $L \in \mathbb{R}[\mathbf{x}]_{2t}^*$, which is the following matrix indexed by $\mathbb{N}_t^n$:

$$M_t(L) = (L(\mathbf{x}^\alpha \mathbf{x}^\beta))_{\alpha,\beta \in \mathbb{N}_t^n}.$$

Following Curto and Fialkow [17] we say that $M_t(L)$ is a *flat extension* of (its principal submatrix) $M_{t-1}(L)$ if

$$\text{rank } M_t(L) = \text{rank } M_{t-1}(L). \tag{4.2}$$

The following result claims that any such moment matrix can be extended to an infinite moment matrix of the same rank.

Theorem 4.2 ([17]). *Let $L \in \mathbb{R}[\mathbf{x}]_{2t}^*$. If $M_t(L)$ is a flat extension of $M_{t-1}(L)$, i.e., (4.2) holds, then there exists $\tilde{L} \in \mathbb{R}[\mathbf{x}]^*$ which extends L (i.e., $L = \tilde{L}$ on $\mathbb{R}[\mathbf{x}]_{2t}$) and has the property that $M(\tilde{L})$ is a flat extension of $M_t(L)$:* $\text{rank } M(\tilde{L}) = \text{rank } M_t(L)$.

The proof is elementary, exploiting the fact that the kernel of $M(\tilde{L})$ is an ideal. Indeed the relations expressing the monomials of degree t in terms of polynomials of degree at most $t-1$ (modulo the kernel of $M_t(L)$) can be used to express recursively any monomial of degree at least $t+1$ in terms of polynomials of degree at most t (modulo the ideal generated by the kernel of $M_t(L)$). Combining Theorems 4.1 and 4.2, we arrive at the following result.

Theorem 4.3. *Let $L \in \mathbb{R}[\mathbf{x}]_{2t}^*$ and assume that $M_t(L) \succeq 0$ and (4.2) holds. Then L has a finite atomic representing measure μ, whose support is given by the variety of the kernel of $M_t(L)$: $V(\text{Ker}\, M_t(L)) = \text{Supp}(\mu) \subseteq \mathbb{R}^n$. Moreover, the ideal generated by the kernel of $M_t(L)$ is equal to the kernel of $M(L_\mu)$: $(\text{Ker}\, M_t(L)) = \text{Ker}\, M(L_\mu)$, and it is a real radical ideal.*

To be able to claim that the representing measure μ is supported within a given semialgebraic set K like (1.1), it suffices to add the *localizing conditions* $M_{t-d_{g_j}}(g_j L) \succeq 0$ (for $j \in [m]$), where g_j are the polynomials defining K and $d_{g_j} = \lceil \deg(g_j)/2 \rceil$, and to assume a stronger flatness condition:

$$\text{rank} M_t(L) = \text{rank} M_{t-d_K}(L), \quad \text{where } d_K = \max\{d_{g_j} : j \in [m]\}. \tag{4.3}$$

Theorem 4.4 ([18]). *Assume $L \in \mathbb{R}[\mathbf{x}]_{2t}^*$ satisfies $M_t(L) \succeq 0$, $M_{t-d_{g_j}}(g_j L) \succeq 0$ for $j \in [m]$, and the flatness condition (4.3). Then L has a representing measure whose support is contained in the set K.*

Proof. We give our simple proof from [64]. We already know that L has a representing measure μ with $\text{Supp}(\mu) =: \{v_1,\dots,v_r\} \subseteq \mathbb{R}^n$, where $r = \text{rank} M_t(L)$ and $L = \sum_{i=1}^r \lambda_i L_{v_i}$ with $\lambda_i = L(p_{v_i}) > 0$. It suffices now to show that each point $v_i \in \text{Supp}(\mu)$ belongs to K, i.e., that $g_j(v_i) \geq 0$ for all $j \in [m]$. For this, the simple but crucial observation is that we can choose the interpolation polynomials p_{v_i} at the v_i's in such a way that they all have degree at most $t - d_K$ (which follows using condition (4.3)). As each polynomial p_{v_i} has degree at most $t - d_K \leq t - d_{g_j}$ and $M_{t-d_{g_j}}(g_j L) \succeq 0$, we can conclude that $0 \leq (g_j L)(p_{v_i}^2) = L(p_{v_i}^2 g_j)$, which implies directly that $g_j(v_i) \geq 0$. □

4.4. The moment relaxations for (P). We now return to the moment relaxation (MOMt) for problem (P) introduced earlier in Section 3.2. First, using truncated moment matrices, it can be reformulated as follows:

$$f_t^{\text{mom}} = \inf_{L \in \mathbb{R}[\mathbf{x}]_{2t}^*} \{L(f) : L(1) = 1,\ M_t(L) \succeq 0,\ M_{t-d_{g_j}}(g_j L) \succeq 0\ (j \in [m])\}, \quad \text{(MOMt)}$$

(explaining the name 'moment' and the notation 'f_t^{mom}'). Recall that $f_t^{\text{mom}} \leq f_{\text{min}}$ from (3.2). Using the preceding results about flat extensions of moment matrices, we can now present the following **optimality certificate** for the relaxation (MOMt), which permits to claim that the infimum of f is reached: $f_t^{\text{mom}} = f_{\text{min}}$.

Theorem 4.5. *Let K_f denote the set of global minimizers of problem (P) and set $d_f = \lceil \deg(f)/2 \rceil$, $d_{g_j} = \lceil \deg(g_j)/2 \rceil$, $d_K = \max\{d_{g_j} : j \in [m]\}$. Let $L \in \mathbb{R}[\mathbf{x}]_{2t}^*$ be an optimal solution of the program (MOMt). Assume that L satisfies the following flatness condition:*

$$\text{rank} M_s(L) = \text{rank} M_{s-d_K}(L) \textit{ for some } s \textit{ satisfying } \max\{d_f, d_K\} \leq s \leq t. \quad (4.4)$$

Then, $f_t^{mom} = f_{min}$ and $V(\text{Ker}\, M_s(L)) \subseteq K_f$. Moreover, if $\text{rank} M_s(L)$ is maximum among all optimal solutions of (MOMt), then equality: $V(\text{Ker}\, M_s(L)) = K_f$ holds and $\mathcal{I}(K_f) = (\text{Ker}\, M_s(L))$.

Proof. Assume $s = t$ (to simplify notation). By Theorem 4.4, L has a representing measure μ with $\text{Supp}(\mu) \subseteq K$. That is, $L = \sum_{i=1}^r \lambda_i L_{v_i}$, where $\lambda_i > 0$, $\sum_i \lambda_i = 1$, and $\{v_1,\dots,v_r\} = V(\text{Ker}\, M_t(L)) \subseteq K$. Then, $f_t^{\text{mom}} = L(f) = \sum_{i=1}^r \lambda_i f(v_i) \geq f_{\text{min}}$. This implies equality $f_t^{\text{mom}} = f_{\text{min}}$ and $f(v_i) = f_{\text{min}}$ for all $i \in [r]$, and thus we can conclude that $V(\text{Ker}\, M_t(L)) = \{v_1,\dots,v_r\} \subseteq K_f$.

Assume now that $M_t(L)$ has maximum rank among the optimal solutions of (MOMt). As the evaluation L_v at any point $v \in K_f$ is also an optimal solution of (MOMt), we deduce that rank $M_t(L_v) \leq$ rank $M_t(L)$, which implies that $\text{Ker}\, M_t(L) \subseteq \text{Ker}\, M_t(L_v) \subseteq \mathcal{I}(v)$ for all $v \in K_f$. Hence, $\text{Ker}\, M_t(L)$ is contained in $\cap_{v \in K_f} \mathcal{I}(v) = \mathcal{I}(K_f)$. By taking the varieties on both sides, we obtain that $K_f \subseteq V(\text{Ker}\, M_t(L))$, which implies $K_f = V(\text{Ker}\, M_t(L))$ and thus $\mathcal{I}(K_f) = (\text{Ker}\, M_s(L))$ (since $(\text{Ker}\, M_t(L))$ is real radical by Theorem 4.3). □

The above result is the theoretical core of the moment approach for problem (P). It has been implemented in the numerical algorithm GloptiPoly. There are several other implementations of the sos/moment approach, including SOSTOOLS, YALMIP, and SparsePOP (tuned to exploit sparsity structure). We conclude with some comments and pointers to a few additional results from the growing literature.

- *The maximality assumption on the rank of the optimal solution is not restrictive.* On the contrary, most interior point algorithms currently used to solve semidefinite programs return an optimal solution lying in the relative interior of the optimal face and thus one with maximum possible rank (see [21]).
- *Under the assumptions of Theorem 4.5, problem (P) has finitely many global minimizers and they can be found using the eigenvalue method from Section 2.* Indeed, we know that the set of global minimizers is $K_f = V(\mathrm{Ker}\, M_s(L))$ and that the quotient space $\mathbb{R}[\mathbf{x}]/(\mathrm{Ker}\, M_s(L))$ has dimension $\mathrm{rank}\ M_s(L) = \mathrm{rank}\ M_{s-d_K}(L)$. Hence any set of monomials indexing a maximal linearly independent set of columns of the matrix $M_{t-d_K}(L)$ is a linear basis of $\mathbb{R}[\mathbf{x}]/(\mathrm{Ker}\, M_s(L))$. So we can construct the multiplication matrices in $\mathbb{R}[\mathbf{x}]/(\mathrm{Ker}\, M_s(L))$ and their eigenvalues/eigenvectors permit to extract the points in $V(\mathrm{Ker}\, M_s(L)) = K_f$.
- The flatness condition (4.4) can be used as a *concrete optimality stopping criterion*: if it is satisfied at a certain order t then the relaxation is exact and the algorithm stops after returning the infimum $f_{\min}$ and the set K_f of global minimizers. Otherwise one may compute the next relaxation of order $t+1$.
- In general, information about the global minimizers can be gained asymptotically from optimal solutions L^t to the relaxations (MOMt). In particular, if (P) has a unique minimizer x^*, then x^* can be found asymptotically as limit point as $t \to \infty$ of the sequences $(L^t(\mathbf{x}_1), \ldots, L^t(\mathbf{x}_n))$ [95]. See [77] for an extension to the case of finitely many global minimizers.

In the compact case, the bounds f_t^{sos}, f_t^{mom} converge asymptotically to $f_{\min}$ (Theorem 3.5). What about **finite convergence**?

- By Theorem 4.5, the flatness condition (4.4) implies the finite convergence of the moment hierarchy (MOMt). Conversely, if the set of global minimizers is nonempty and finite, *the flatness condition (4.4) is also necessary for finite convergence of (MOMt) under some genericity assumptions on the polynomials f, g_j* [77].
- Finite convergence holds in the case when the description of the set K involves some polynomial equations $g_1(x) = 0, \ldots, g_k(x) = 0$ which have finitely many common real roots (since the flatness condition holds) [66, 68, 78].
- Finite convergence also holds in the *convex case*, when $f, -g_1, \ldots, -g_m$ are convex, the set K has a Slater point x_0 (i.e., $g_j(x_0) > 0$ if g_j is not linear), and the Hessian of f is positive definite at the (unique) global minimizer [23].
- Nie [80] shows that, under the Archimedean condition (A), *the Lasserre hierarchy applied to problem (P) has finite convergence generically*. More precisely, finite convergence holds when the classic nonlinear optimality conditions (constraint qualification, strict complementarity, and second order sufficient condition) hold at all global minimizers, and these conditions are satisfied generically.
- Finally we refer to [81] for degree bounds and estimates on the quality of the moment/sos bounds (see [22] for refined results when K is the hypercube).

5. Application to real roots and real radical ideals

The above strategy for computing the global minimizers of (P) was developed and applied by Lasserre, Laurent and Rostalski [57] to the problem of computing the common real roots of a system of polynomial equations: $g_1(x) = 0, \ldots, g_k(x) = 0$.

Computing all *complex* roots is a well studied problem. Several methods exist, including symbolic-numeric methods, which combine symbolic tools (like Gröbner or border bases) with numerical linear algebra (like computing eigenvalues, or univariate root finding), and homotopy continuation methods. As there might be much less real roots than complex ones it is desirable to have methods able to extract directly the real roots without dealing with the complex nonreal ones. This is precisely the feature of the real algebraic method of [57], which can be summarized as follows.

Consider the following instance of (P):

$$\min\{0 : g_1(x) = 0, \ldots, g_k(x) = 0\},$$

which asks to minimize the zero polynomial on the real algebraic variety of the ideal $\mathcal{I} = (g_1, \ldots, g_k)$, so that the set of global minimizers is precisely $V_{\mathbb{R}}(\mathcal{I})$.

Consider the moment relaxations (MOMt) for this problem. [57] shows that the flatness condition (4.4) holds for t large enough, assuming that the set $V_{\mathbb{R}}(\mathcal{I})$ is finite. Hence, by Theorem 4.5, it follows that the real radical ideal of $\mathcal{I}$ is found: $\sqrt[\mathbb{R}]{\mathcal{I}} = (\mathrm{Ker}\, M_s(L))$ and that the variety $V_{\mathbb{R}}(\mathcal{I})$ can be computed using the eigenvalue method applied to the quotient space $\mathbb{R}[\mathbf{x}]/(\mathrm{Ker}\, M_s(L))$ (as explained in the previous section). The fact that the kernel of $M_s(L)$ generates the vanishing ideal of $V_{\mathbb{R}}(\mathcal{I})$ is crucial, since this is the key property which permits to filter out all complex nonreal roots.

We point out that the equality $\sqrt[\mathbb{R}]{\mathcal{I}} = (\mathrm{Ker}\, M_t(L))$ holds for t large enough, even if the variety $V_{\mathbb{R}}(\mathcal{I})$ is infinite. The difficulty, however, is to detect when one has reached such order t, since it is not clear how to detect it algorithmically (as the flatness condition cannot hold when the real variety is not finite).

We refer to [57, 58], [1, Chap.2] for details and extensions. The recent work [59] develops a *sparse* version of the moment method able to work with smaller matrices, indexed by smaller sets of monomials, rather than the full set of monomials of degree at most t. This approach combines the border base method from [73] with the generalized flatness condition from [69].

We conclude with illustrating the method on a small example. Consider the polynomial equation: $\mathbf{x}_1^2 + \mathbf{x}_2^2 = 0$, with a unique real root $(0, 0)$ and infinitely many complex roots. Then the moment relaxation of order $t = 1$ has the constraints

$$M_1(y) = \begin{pmatrix} 1 & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{pmatrix} \succeq 0,\ y_{20} + y_{02} = 0,$$

which imply $y_\alpha = 0$ whenever $\alpha \neq 0$. Therefore the flatness condition holds: $\mathrm{rank} M_1(y) = \mathrm{rank} M_0(y) = 1$. Moreover the kernel of $M_1(y)$ is spanned by the two polynomials $\mathbf{x}_1, \mathbf{x}_2$, which indeed generate the real radical of the ideal $(\mathbf{x}_1^2 + \mathbf{x}_2^2)$.

6. Application to some combinatorial problems

Lift-and-project methods. The polynomial optimization problem (P) contains the general 0/1 linear programming (ILP), asking to optimize a linear function over the $0/1$ solutions to a linear system $Ax \geq b$. Let P denote the integral polytope defined as the convex hull of all $x \in \{0,1\}^n$ satisfying $Ax \geq b$ and let $K = \{x : Ax \geq b\}$ denote its linear relaxation, which can be assumed to lie in the hypercube $[0,1]^n$. A well studied approach in polyhedral combinatorics is to find a (partial) linear inequality description of the polytope P, leading to a new relaxation P' nested between P and K: $P \subseteq P' \subseteq K$, strengthening the initial relaxation K. Several methods have been investigated that construct in a systematic way hierarchies of relaxations nested between P and K, with the property that P is found in finitely many steps. For instance, the classic method in integer programming, which consists of iteratively adding Gomory-Chvátal cuts, finds the integral polytope P in $O(n^2 \log n)$ steps [30], but linear optimization over the first Gomory-Chvátal closure is a hard problem [29]. On the other hand, the lift-and-project methods of Sherali and Adams [96] and of Lovász and Schrijver [71] produce hierarchies of LP and SDP relaxations P_t that find the integral polytope in n steps and with the property that linear optimization over the t-th relaxation P_t is polynomial time for any *fixed* t. They are all based on the following basic strategy:

(a) Generate new polynomial constraints by multiplying the polynomial inequalities $a_j^{\mathsf{T}} x - b_j \geq 0$ of the system $Ax \geq b$ by x_i or $1 - x_i$ (and their products) and eliminate all squared variables replacing each x_i^2 by x_i.

(b) Linearize all monomials $\prod_{i \in I} x_i$ by introducing new variables y_I, so that the constraints generated in (a) form a linear system in the variables (x, y).

(c) Project back on the x-variables space, which gives a polyhedron P' nested between P and K.

The construction may allow the addition of positive semidefiniteness constraints, leading to stronger semidefinite relaxations. This is the case for the construction of Lovász and Schrijver [71], which we now briefly describe.

Suppose the vector $x \in \{0,1\}^n$ satisfies the system $Ax \geq b$. Consider the new vector $\hat{x} = (1, x) \in \mathbb{R}^{n+1}$ (where the additionnal entry is indexed by '0') and the matrix $Y = \hat{x}\hat{x}^{\mathsf{T}} \in \mathcal{S}^{n+1}$. Then the matrix Y satisfies the following conditions: (i) $Y \succeq 0$, (ii) $Y_{00} = 1$, (iii) $Y_{0i} = Y_{ii}$ for all $i \in [n]$, and (iv) the vectors $Y^{(i)}$, $Y^{(0)} - Y^{(i)}$ (for $i \in [n]$) satisfy the linear system: $Ax - bx_0 \geq 0$ (where $Y^{(i)} \in \mathbb{R}^{n+1}$ denotes the i-th column of Y). Let $M^+(K)$ denote the set of matrices $Y \in \mathcal{S}^{n+1}$ satisfying the above conditions (i)-(iv), define its projection

$$N^+(K) = \{x \in \mathbb{R}^n : \exists Y \in M^+(K) \text{ such that } x_i = Y_{0i} \ (i \in [n])\},$$

and define analogously $N(K)$ by omitting the positive semidefiniteness condition (i) in the definition of $M^+(K)$. Then, $P \subseteq N^+(K) \subseteq N(K) \subseteq K$. For an integer $t \geq 2$, one can iteratively define $N_t(K) = N(N_{t-1}(K))$, $N_t^+(K) = N^+(N_{t-1}^+(K))$ (setting $N_1(K) = N(K)$ and $N_1^+(K) = N^+(K)$). This leads to hierarchies of linear and semidefinite relaxations, that find P in n steps: $P \subseteq N_t^+(K) \subseteq N_t(K)$, with equality for $t = n$. From the optimization point of view, these hierarchies behave well: if linear optimization over K can be done in polynomial time then the same holds for linear optimization over $N_t(K)$ and $N_t^+(K)$ for any fixed $t \geq 1$ [71].

The paper [71] also investigates in detail how the construction applies to the stable set problem. Given a graph $G = (V = [n], E)$, let $K \subseteq \mathbb{R}^n$ be defined by nonnegativity $x \geq 0$ and the edge inequalities $x_i + x_j \leq 1$ ($\{i,j\} \in E$), so that the corresponding polytope $P = \text{conv}(K \cap \{0,1\}^n)$ is the stable set polytope of G. The first linear relaxation $N(K)$ is completely understood: $N(K)$ is the polyhedron defined by nonnegativity $x \geq 0$ and the odd cycle inequalities $\sum_{i \in O} x_i \leq (|O| - 1)/2$ for all $O \subseteq V$ inducing an odd cycle in G. The relaxation $N^+(K)$ is much stronger. Indeed, for any clique C of G, the corresponding clique inequality $\sum_{i \in C} x_i \leq 1$ is valid for $N^+(K)$, while the first order t for which it is valid for the linear relaxation $N_t(K)$ is $t = |C| - 2$. Moreover the stable set polytope P is found after $\alpha(G)$ steps of the semidefinite hierarchy, compared to $n - \alpha(G) - 1$ steps of the linear hierarchy. These results have motivated much of the interest in these lift-and-project semidefinite relaxations for combinatorial optimization.

The Lasserre approach. The general moment approach applied to (ILP) also produces a hierarchy of semidefinite relaxations $L_t(K)$ converging to P [54]. As explained in [61], the relaxation $L_t(K)$ can easily be described in a direct way following the above lift-and-project strategy. We just indicate here how to apply the previously described general moment method. We start with the set K defined by the polynomial inequalities $g_j = a_j^{\top}\mathbf{x} - b_j \geq 0$ ($j \in [m]$) and the polynomial equations $\mathbf{x}_i^2 - \mathbf{x}_i = 0$ ($i \in [n]$). Then $L_t(K)$ is defined as the set of all vectors $x \in \mathbb{R}^n$ of the form $x = (L(\mathbf{x}_1), \ldots, L(\mathbf{x}_n))$ for some linear functional $L \in \mathbb{R}[\mathbf{x}]_{2t}^*$ satisfying the moment relaxation (MOMt), i.e., the conditions (i) $L(1) = 1$, (ii) $M_t(L) \succeq 0$, (ii) $M_{t-1}(g_j L) \succeq 0$ ($j \in [m]$), and (iii) $L(f) = 0$ for all polynomials f in the truncated ideal $(\mathbf{x}_1^2 - \mathbf{x}_1, \ldots, \mathbf{x}_n^2 - \mathbf{x}_n)_{2t}$.

What the above condition (iii) says is that one can simplify the Lasserre relaxation by eliminating variables and working with smaller moment matrices. Indeed, instead of considering the moment matrix $M_t(L)$ indexed by *all* monomials of degree at most t, it suffices to consider its principal submatrix indexed by all *square-free* monomials of degree at most t (of the form $\prod_{i \in I} x_i$ for $I \in \binom{V}{\leq t}$), and to consider only variables $y_J := L(\prod_{i \in J} x_i)$ for sets $J \in \binom{V}{\leq 2t}$. Here $\binom{V}{\leq t}$ denotes the collection of subsets of $V = [n]$ with cardinality at most t.

As a direct consequence, the flatness condition (4.3) holds at order $t = n+1$: $\text{rank}\, M_{n+1}(L) = \text{rank}\, M_n(L)$. Hence the Lasserre relaxation of order $n + 1$ is exact: $L_{n+1}(K) = P$ (which follows by applying Theorem 4.5). There is also a simple direct proof for this claim or, alternatively, this claim follows from the fact that the Lasserre hierarchy refines the Lovász-Schrijver hierarchy. Namely, for any $t \geq 2$, we have: $L_t(K) \subseteq N(L_{t-1}(K))$, which thus implies the inclusion $L_t(K) \subseteq N_{t-1}(K)$. Moreover, the Lasserre hierarchy also refines the Sherali-Adams hierarchy. We refer to [61] for the above results, and we refer e.g. to the recent work [2] for a comprehensive treatment and further references, also about other lift-and-project hierarchies. We now indicate how the Lasserre hierarchy applies to maximum stable sets, minimum graph colorings and max-cut.

Lasserre hierarchies for $\alpha(G)$ and $\chi(G)$. As an illustration, the moment relaxation (MOMt) for the stable set problem (1.4) reads:

$$\text{las}_t(G) = \max_{y \in \binom{V}{\leq 2t}} \{\sum_{i \in V} y_i : (y_{I \cup J})_{I,J \in \binom{V}{\leq t}} \succeq 0,\ y_{ij} = 0\ (\{i,j\} \in E),\ y_{\emptyset} = 1\}. \quad (6.1)$$

For $t=1$, we find Lovász' theta number from (1.6): $\mathrm{las}_1(G)=\vartheta(G)$. Moreover, the Lasserre bound is exact: $\mathrm{las}_t(G)=\alpha(G)$ for $t\ge\alpha(G)$. On the dual side, the sos relaxation (SOSt) asks for the smallest scalar λ for which the polynomial $\lambda-\sum_{i\in V}\mathbf{x}_i$ can be written as a sum of squares of degree at most $2t$ modulo the ideal generated by the polynomials $\mathbf{x}_i\mathbf{x}_j$ (for $\{i,j\}\in E$) and $\mathbf{x}_i^2-\mathbf{x}_i$ (for $i\in V$). We refer to Gouveia et al. [35] for a detailed study of the hierarchies from this point of view of sums of squares, also in the setting of general polynomial ideals.

In [39] we investigate Lasserre type bounds for the chromatic number $\chi(G)$. A first possibility is to consider the following analogue of the bounds in (6.1):

$$\psi_t(G)=\min_{y\in\binom{V}{\le 2t}}\{y_\emptyset:(y_{I\cup J})_{I,J\in\binom{V}{\le t}}\succeq 0,\ y_{ij}=0\ (\{i,j\}\in E),\ y_i=1\ (i\in V)\}. \quad (6.2)$$

Then, $\psi_1(G)=\vartheta(\overline{G})\le\psi_t(G)\le\chi(G)$. However, these bounds cannot in general reach the chromatic number since they all remain below the *fractional chromatic number* $\chi_f(G)$: $\psi_t(G)\le\chi_f(G)$, with equality if $t\ge\alpha(G)$.

To define a hierarchy of semidefinite bounds able to reach the chromatic number $\chi(G)$, one can use the reduction of $\chi(G)$ to the stability number of the cartesian product $G\square K_k$ described in the Introduction. Namely, $\chi(G)$ is equal to the smallest integer k for which $\alpha(G\square K_k)=|V(G)|$. This motivates defining the parameter $\mathrm{Las}_t(G)$ as the smallest integer k for which $\mathrm{las}_t(G\square K_k)=|V(G)|$. Then, we have the inequality: $\mathrm{Las}_t(G)\le\chi(G)$, with equality for $t=n$. Note that, for $t=1$, we find again the (rounded) theta number: $\mathrm{Las}_1(G)=\lceil\vartheta(\overline{G})\rceil$.

An easy way to strengthen the various bounds is by adding the nonnegativity constraint $y\ge 0$ to the program (6.1), call $\mathrm{las}'_t(G)$ the resulting parameter. Analogously, define $\mathrm{Las}'_t(G)$ as the smallest integer k for which $\mathrm{las}'_t(G\square K_k)=|V|$. Then, we have: $\alpha(G)\le\mathrm{las}'_t(G)\le\mathrm{las}_t(G)$ and $\mathrm{Las}_t(G)\le\mathrm{Las}'_t(G)\le\chi(G)$. It turns out that the parameters $\mathrm{las}'_1(G)$ and $\mathrm{Las}'_1(G)$ coincide, respectively, with the parameters $\vartheta'(G)$ and $\vartheta^+(\overline{G})$ (recall (1.8)).

The bounds $\mathrm{las}_t(G)$ (and $\mathrm{las}'_t(G)$) have been used in particular to upper bound the cardinality of error correcting codes. When dealing with binary codes of length N, one needs to find the stability number of a Hamming graph G, with vertex set $V=\{0,1\}^N$ and where two vertices $u,v\in V$ are adjacent if their Hamming distance does not belong to some prescribed set. Thus this graph G has 2^N vertices. Fortunately it has a large automorphism group which can be used to compute the parameter $\mathrm{las}_t(G)$ with a semidefinite program involving smaller matrices of size $O(N^{2^t-1})$ (polynomial in N for fixed t), while the original formulation (6.1) involves matrices of size $O(|V|^t=2^{tN})$ (exponential in N). This is shown in [67] using symmetry reduction techniques from [25]. Moreover, Schrijver [93] shows that the semidefinite bound $\mathrm{las}'_1(G)=\vartheta'(G)$ of order $t=1$ coincides with the well known linear programming bound of Delsarte, which is expressed by a linear program of size N. Furthermore, Schrijver [94] shows that the semidefinite bound of the next order 2 (more precisely, some variation in-between the bounds of order 1 and 2) can be computed with a semidefinite program involving (roughly) $N/2$ matrices of size at most N, which he shows using block-diagonalization techniques for matrix algebras. Numerical computations using these parameters and some strengthenings give the currently best known bounds for codes (see [33, 67, 94] and references therein). Computations for the chromatic number using the bounds $\mathrm{Las}_t(G)$ (and variations) can be found in [39, 41].

The Lasserre hierarchy for max-cut. As another illustration let us apply the Lasserre hierarchy to the max-cut problem (1.2). The equations $\mathbf{x}_i^2 = 1$ permit to express the relaxation (MOMt) as

$$\max_{y\in\mathbb{R}^{\binom{V}{\le 2t}}} \{ \sum_{\{i,j\}\in E} (w_{ij}/2)(1-y_{ij}) : (y_{I\Delta J})_{I,J\in\binom{V}{\le t}} \succeq 0,\ y_\emptyset = 1\}.$$

For $t = 1$ this is the relaxation (1.3) used by Goemans and Williamson [34] for their 0.878-approximation algorithm for max-cut. More details about geometric properties of the Lasserre hierarchy for max-cut can be found in [63]. A natural question is how many steps are needed to solve max-cut using the hierarchy. In [62] we show that, for the all-ones weight function, the relaxation is exact if and only if $t \ge t_n := \lceil n/2 \rceil$ and we conjecture that t_n iterations suffice for arbitrary weights w. Equivalently, we conjecture that the polynomial $f_w = \mathrm{mc}(G,w) - \sum_{\{i,j\}\in E}(w_{ij}/2)(1-\mathbf{x}_i\mathbf{x}_j)$ can be written as a sum of squares of degree at most $2t_n$ modulo the ideal $(\mathbf{x}_i^2 - 1 : i \in [n])$. Recently, Blekherman et al. [8] show that this is indeed true when allowing “denominators", i.e., they show that there exists a polynomial p such that $p^2 f_w$ has such a decomposition.

Copositive based hierarchies. Let $\mathcal{C}^n$ denote the copositive cone, consisting of all matrices $M \in \mathcal{S}^n$ for which the polynomial $f_M = \sum_{i,j=1}^n M_{ij}\mathbf{x}_i^2\mathbf{x}_j^2$ is nonnegative over $\mathbb{R}^n$. As mentioned in the Introduction, the stability number $\alpha(G)$ of a graph G can be obtained from the program (1.9), which is linear optimization over the copositive cone $\mathcal{C}^n$. As we indicate below this formulation leads to another type of hierarchies.

Motivated by the fact that testing matrix copositivity is a hard problem, Parrilo [82] introduced a hierarchy of sufficient conditions, which can be tested using semidefinite optimization and leads to the hierarchy of cones $\mathcal{K}_t$ considered by de Klerk and Pasechnik [24]. Namely, $\mathcal{K}_t$ consists of the matrices $M \in \mathcal{S}^n$ for which the polynomial $f_M(\sum_{i=1}^n \mathbf{x}_i^2)^t$ is a sum of squares. The cone K_0 consists precisely of the matrices M that can be written as the sum of a positive semidefinite matrix and an entrywise nonnegative matrix. Clearly, the cones $\mathcal{K}_t$ form a hierarchy of subcones of $\mathcal{C}^n$: $\mathcal{K}_t \subseteq \mathcal{K}_{t+1} \subseteq \mathcal{C}^n$. Parrilo [82] shows that they cover the interior of $\mathcal{C}^n$: if $f_M(x) > 0$ for all nonzero $x \in \mathbb{R}^n$ then M belongs to some $\mathcal{K}_t$. His proof uses the following result of Pólya: if $g \in \mathbb{R}[\mathbf{x}]$ is a homogeneous polynomial satisfying $g(x) > 0$ for all nonzero $x \in \mathbb{R}^n_+$, then there exists an integer $t \in \mathbb{N}$ for which all the coefficients of the polynomial $(\sum_{i=1}^n \mathbf{x}_i)^t g$ are nonnegative.

The cones $\mathcal{K}_t$ lead to another hierarchy of bounds for the stability number $\alpha(G)$. Starting from relation (1.9), De Klerk and Pasechnik [24] define the parameter

$$\vartheta_t(G) = \min\{\lambda : \lambda(I + A_G) - J \in \mathcal{K}_t\}. \tag{6.3}$$

They show that the first bound is the theta number: $\vartheta_0(G) = \vartheta'(G)$, and they show convergence *after rounding*: $\lfloor \vartheta_t(G) \rfloor = \alpha(G)$ for $t \ge \alpha(G)^2$. Moreover, they conjecture that finite convergence: $\alpha(G) = \vartheta_t(G)$ holds for $t \ge \alpha(G) - 1$, which would mirror the known finite convergence in $\alpha(G)$ steps for the Lasserre bounds $\mathrm{las}_t(G)$. In [38] we give a partial proof and prove this conjecture for all graphs with $\alpha(G) \le 8$.

This approach also gives lower bounds $\Theta_t(G)$ for the chromatic number $\chi(G)$. Namely, define $\Theta_t(G)$ as the smallest integer k for which $\vartheta_t(G \square K_k) = |V(G)|$. In [38] we compare both types of hierarchies and we show that the Lasserre hierarchies refine these ‘copositive based’ hierarchies. Namely, we show that $\mathrm{las}'_t(G) \le \vartheta_{t-1}(G)$ and thus $\Theta_{t-1}(G) \le$

$\mathrm{Las}'_t(G)$ for any $t \geq 1$. Hence, the Lasserre hierarchy may give better bounds and moreover it seems much easier to handle. For instance its finite convergence is easy, while the finite convergence of the copositive hierarchy is still open. A reason might be that the Lasserre construction uses explicitly the presence of binary variables, while the copositive based construction does not. Nevertheless copositive based approximations have gained popularity in the recent years and they open the way to other types of approaches for approximating hard problems. We refer e.g. to [11, 28] and references therein.

7. Conclusions

We have presented the general approach permitting to construct semidefinite relaxations for polynomial optimization problems by using sums of squares representations for positive polynomials and moment matrices. We reviewed some basic properties regarding in particular their convergence properties. We also discussed how the general methodology applies for building hierarchies of semidefinite relaxations for combinatorial problems in graphs. We have only discussed a small piece of this rapidly expanding research area. We now mention a few other research areas, where this type of methods are also being increasingly used.

Semidefinite optimization and in particular the Lasserre hierarchy are playing a growing role in theoretical computer science for the design of efficient approximation algorithms. Understanding the power and limitations of the Lasserre hierarchy is a fundamental question, which has tight links with complexity theory. For instance, assuming the unique game conjecture [48], Khot et al. [49] show that one cannot beat the Goemans-Williamson 0.878-approximation guarantee for max-cut, which is based on the Lasserre relaxation of smallest order. Yet recent results of Guruswami and Sinop [37] exploit higher order relaxations to give improved approximation algorithms for graph partition problems, depending on spectral properties of the graph. We refer e.g. to [32, 65], the recent overview by Chlemtac and Tulsiani [1, Chap. 6] and references therein.

Semidefinite bounds are also used to attack geometric problems, like the kissing number problem and the problem of coloring the Euclidean space [3, 4]. These problems lead to maximum stable set and minimum coloring problems in infinite graphs. For instance, the kissing number problem is finding a maximum stable set, where the vertex set is the unit sphere with two points being adjacent depending on their spherical distance. Bachoc and Vallentin [3] use low order bounds in the Lasserre hierarchy to give the best known bounds for the kissing number problem, a crucial ingredient in their approach is exploiting symmetry in order to get computable semidefinite programs.

Hierarchies of semidefinite relaxations have also been used recently to attack polynomial optimization problems in noncommutative variables. Such problems arise when, instead of instantiating variables to scalars, one allows variables to be matrices (or bounded operators on some Hilbert space) and they have applications in many areas of quantum phsyics. Given a symmetric polynomial f in n noncommutative variables, one can consider the following two kinds of positivity: f is said to be matrix-positive if $f(X_1, \ldots, X_n) \succeq 0$ when evaluating f at arbitrary matrices $X_1, \ldots, X_n \in \mathcal{S}^d$ ($d \geq 1$), and f is said to be trace-positive if $\mathrm{Tr}(f(X_1, \ldots, X_n)) \geq 0$ for all $X_1, \ldots, X_n \in \mathcal{S}^d$ ($d \geq 1$). These two notions lead to different noncommutative polynomial optimization problems. For both problems analogues of the moment and sums of squares approaches have been investigated, we refer to [12, 20, 84] and references therein.

By Hilbert's theorem, not all nonnegative polynomials are sums of squares. However, Helton [42] shows the following remarkable result: a symmetric polynomial is matrix-positive if and only if it is a sum of Hermitian squares. Moreover, Helton and McCullough [43] show a result characterizing matrix-positivity on a compact set which can be seen as an analogue of Putinar's result (Theorem 3.4). On the other hand, the analogue result for trace-positive polynomials is still open, and it is in fact related to a deep conjecture of Connes [15] in operator algebra. Indeed, Klep and Schweighofer [50] show that Connes' embedding conjecture is equivalent to a real algebraic conjecture characterizing the trace-positive polynomials on all contraction matrices.

Problems in quantum information have led in the recent years to some quantum analogues of the classical graph parameters $\alpha(G)$ and $\chi(G)$. These quantum parameters require to find positive semidefinite matrices satisfying certain polynomial conditions and, as in the classical case, the theta number serves also as bound for them (see [10, 13] and further references therein). Investigating how to construct hierarchies of stronger semidefinite bounds for these quantum graph parameters is a natural direction that we are currently investigating.

References

[1] Anjos, M.A., Lasserre, J.B. (eds), *Handbook on Semidefinite, Conic and Polynomial Optimization*, International Ser. Oper. Res. & Management Sci. **166** (2012), Springer.

[2] Hu Hin Au, Tuncel, L., *A comprehensive analysis of polyhedral lift-and-project methods*, arXiv:1312.5972.

[3] Bachoc, C., Vallentin, F., *New upper bounds for kissing numbers from semidefinite programming*, J. Amer. Math. Soc. **21** (2008), 909–924.

[4] Bachoc, C., Nebe, G., de Oliveira Filho, F.M., Vallentin, F., *Lower bounds for measurable chromatic numbers*, Geom. Funct. Anal. **19** (2009), 645–661.

[5] Ben-Tal, A., Nemirovski, A., *Lectures on Modern Convex Optimization - Analysis, Algorithms, and Engineering Applications*, MOS-SIAM Series Optim. **2** (2001).

[6] Blekherman, G., *There are significantly more nonnegative polynomials than sums of squares*, Isreal Journal of Mathematics **153** (2006), 355–380.

[7] Blekherman, G., *Nonnegative polynomials and sums of squares*, Journal of the Amer. Math. Soc. **25** (2012), 617–635.

[8] Blekherman, G., Gouveia, J., Pfeiffer, J., *Sums of squares on the hypercube*, arXiv:1402.4199 (2014).

[9] Blekherman, G., Thomas, R.R., Parrilo, P.A. (eds.), *Semidefinite Optimization and Convex Algebraic Geometry*, MOS-SIAM Series Optim. **13** (2012).

[10] Briët, J., Buhrman, H., Laurent, M., Piovesan, T., Scarpa, G., *Zero-error source-channel coding with entanglement*, arXiv:1308.4283 (2013).

[11] Burer, S., *On the copositive representation of binary and continuous nonconvex quadratic programs*, Math. Program. **120** (2009), 479–495.

[12] Burgdorf, S., *Trace-Positive Polynomials, Sums of Hermitian Squares and The Tracial Moment Problem*, PhD thesis, Universität Konstanz & Université de Rennes 1 (2011).

[13] Cameron, P.J., Montanaro, A., Newman, M.W., Severini, S., Winter, A., *On the quantum chromatic number of a graph*, Electr. J. Comb. **14**-1(R81) (2007).

[14] Chudnovsky, M., Robertson, N., Seymour, P., Thomas, R., *The strong perfect graph theorem*, Annals Math. **164**(1) (2006), 51–229.

[15] Connes, A., *Classification of injective factors. Cases* $\Pi_1, \Pi_\infty, \Pi_\lambda$, $\lambda \neq 1$, Ann. of Math. **104**(2) (1976), 73–115.

[16] Cox, D.A., Little, J.B., O'Shea, D., *Ideals, Varieties and Algorithms*, Springer, 1997.

[17] Curto, R., Fialkow, L., *Solution of the truncated complex moment problem for flat data*, Memoirs Amer. Math. Soc. **119**(568) (1996).

[18] Curto, R., Fialkow, L., *Flat extensions of positive moment matrices: recursively generated relations*, Memoirs Amer. Math. Soc. **136**(648) (1998).

[19] Curto, R., Fialkow, L., *The truncated complex K-moment problem*, Trans. Amer. Math. Soc. **352** (2000), 2825–2855.

[20] Doherty, A.C., Liang, Y.-C., Toner, B., Wehner, S., *The quantum moment problem and bounds on entangled multi-prover games*, Proc. CCC'08 (2008), 199–210.

[21] De Klerk, E., *Aspects of Semidefinite Programming - Interior Point Algorithms and Selected Applications*, Kluwer, 2002.

[22] De Klerk, E., Laurent, M., *Error bounds for some semidefinite programming approaches to polynomial minimization on the hypercube*, SIAM J. Optim. **20**(6) (2010), 3104–3120.

[23] De Klerk, E., Laurent, M., *On the Lasserre hierarchy of semidefinite programming relaxations of convex polynomial optimization problems*, SIAM J. Optim. **21** (2011), 824–832.

[24] De Klerk, E., Pasechnik, D.V., *Approximating the stability number of a graph via copositive programming*, SIAM J. Optim. **12** (2002), 875–892.

[25] De Klerk, E., Pasechnik, D.V., Schrijver, A., *Reduction of symmetric semidefinite programs using the regular *-representation*, Math. Program. Ser. B **109** (2007), 613–624.

[26] De Oliveira Filho, F., *New Bounds for Geometric Packing and Coloring via Harmonic Analysis and Optimization*, PhD thesis, University of Amsterdam, 2009.

[27] Dickinson, P., Gijben, L., *On the computational complexity of membership problems for the completely positive cone and its dual*, Comp. Opt. and Appl. **57**(2) (2014), 403–415.

[28] Dür, M., *Copositive programming - a survey*, In Recent Advances in Optimization and its Applications in Engineering, M. Diehl et al. (eds.), Springer, 2010, pp. 3–20.

[29] Eisenbrand, F., *On the membership problem for the elementary closure of a polyhedron*, Combinatorica **19** (2000), 299–300.

[30] Eisenbrand, F., Schulz, A.S., *Bounds on the Chvátal rank of polytopes in the 0/1 cube*, In G. Cornuéjols et al. IPCO 1999, LNCS **1610** (1999), 137–150.

[31] Garey, M.R., Johnson, D.S., *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman & Company Publishers (1979).

[32] Gärtner, B., Matousek, J., *Approximation Algorithms and Semidefinite Programming,* Springer (2012).

[33] Gijswijt, D.C., Mittelmann, H.D., Schrijver, A., *Semidefinite code bounds based on quadruple distances*, IEEE Trans. Inform. Theory **58** (2012), 2697–2705.

[34] Goemans, M.X., Williamson, D. *Improved approximation algorithms for maximum cuts and satisfiability problems using semidefinite programming*, J. of the ACM **42** (1995), 1115–1145.

[35] Gouveia, J., Thomas, R.R., Parrilo, P.A., *Theta bodies for polynomial ideals*, SIAM J. Optim. **20(4)** (2010), 2097–2118.

[36] Grötschel, M., Lovász, L., Schrijver, A., *The ellipsoid method and its consequences in combinatorial optimization*, Combinatorica **1(2)** (1981), 169–197.

[37] Guruswami, V., Sinop, A., *Lasserre hierarchy, higher eigenvalues, and approximation schemes for quadratic integer programming with PSD objectives*, FOCS 2011, 482–491.

[38] Gvozdenović, N., Laurent, M., *Semidefinite bounds for the stability number of a graph via sums of squares of polynomials*, Math. Program. Ser. B **110(1)** (2007), 145–173.

[39] Gvozdenović, N., Laurent, M., *The operator Ψ for the chromatic number of a graph*, SIAM J. Optim. **19(2)** (2008), 572-591.

[40] Gvozdenović, N., Laurent, M., *Computing semidefinite programming lower bounds for the (fractional) chromatic number via block-diagonalization*, SIAM J. Optim. **19(2)** (2008), 592-615.

[41] Gvozdenović, N., Laurent, M., Vallentin, F., *Block-diagonal semidefinite programming hierarchies for 0/1 programming*, Oper. Res. Letters **37** (2009), 27-31.

[42] Helton, J.W., *"Positive" noncommutative polynomials are sums of squares*, Ann. of Math. **156** (2002), 675–694.

[43] Helton, J.W., McCullough, S., *A Positivstellensatz for non-commutative polynopmials*, Trans. Amer. Math. Soc. **356** (2004), 3721–3737.

[44] Henrion, D., Lasserre, J.B., *Detecting global optimality and extracting solutions in GloptiPoly*, In Positive Polynomials in Control, *LNCIS* **312** (2005), 293–310.

[45] Hilbert, D., *Über die Darstellung definiter Formen als Summe von Formenquadraten*, Math. Annalen **32** (1888), 342–350.

[46] Khachiyan, L., Porkolab, L., *Computing integral points in convex semi-algebraic sets*, In FOCS (1997), 162–171.

[47] Porkolab, L., Khachiyan, L., *On the complexity of semidefinite programs*, J. Global Opt. **10** (1997), 351–365.

[48] Khot, S., *On the power of unique 2-prover 1-round games*, In Proc. 34th Ann. ACM Symp. on the Theory of Computing, (2002), 767–775.

[49] Khot, S., Kindler, G., Mossel, E., O'Donnell, R., *Optimal inapproximability results for MAX-CUT and other 2-variable CSPs?*, In FOCS 2004, 146–154.

[50] Klep, I., Schweighofer, M., *Connes' embedding conjecture and sums of Hermitian squares*, Adv. Math. **217(4)** (2008), 1816–1837.

[51] Klep, I., Schweighofer, M., *An exact duality theory for semidefinite programming based on sums of squares*, Math. Oper. Res. **38** (2013), 569-590.

[52] Krivine, J.L., *Anneaux préordonnés*, J. Analyse Math. **12** (1964), 307–326.

[53] Lasserre, J.B., *Global optimization with polynomials and the problem of moments*, SIAM J. Optim. **11** (2001), 796–817.

[54] Lasserre, J.B., *An explicit exact SDP relaxation for nonlinear* $0-1$ *programs*, In K. Aardal and A.M.H. Gerards (eds.), LNCS **2081** (2001), 293–303.

[55] Lasserre, J.B., *A sum of squares approximation of nonnegative polynomials*, SIAM J. Optim. **16** (2006), 751–765.

[56] Lasserre, J.B., *Moments, Positive Polynomials and Their Applications*, Imperial College Press (2009).

[57] Lasserre, J.B., Laurent, M., Rostalski, P., *Semidefinite characterization and computation of real radical ideals*, Foundations Comput. Math. **8** (2008), 607–647.

[58] Lasserre, J.B., Laurent, M., Rostalski, P., *A unified approach for real and complex zeros of zero-dimensional ideals*, pages 125–155 in [87].

[59] Lasserrre, J.B., Laurent, M., Mourrain, B., Rostalski, P., Trebuchet, P., *Moment matrices, border bases and radical computation*, J. Symb. Comput. **51** (2013), 63-85.

[60] Lasserre, J.B., Netzer, T., *SOS approximations of nonnegative polynomials via simple high degree perturbations*, Math. Zeitschrift **256** (2006), 99–112.

[61] Laurent, M., *A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming*, Math. Oper. Res. **28**(3) (2003), 470–496.

[62] Laurent, M., *Lower bounds for the number of iterations in semidefinite hierarchies for the cut polytope*, Math. Oper. Res. **28(4)** (2003), 871–883.

[63] Laurent., M., *Semidefinite relaxations for Max-Cut*, In The Sharpest Cut, M. Grötschel (ed.), MOS-SIAM Series Optim. **4** (2004), 257–290.

[64] Laurent, M., *Revisiting two theorems of Curto and Fialkow on moment matrices*, Proc. Amer. Math. Soc. **133(10)** (2005), 2965–2976.

[65] Laurent, M., Rendl, F., *Semidefinite Programming and Integer Programming*, In Handbook on Discrete Optimization, K. Aardal et al. (eds.), Elsevier (2005), 393-514.

[66] Laurent, M., *Semidefinite representations for finite varieties*, Math. Program. Ser. A **109** (2007), 1–26.

[67] Laurent, M., *Strengthened semidefinite programming bounds for codes*, Math. Program. Ser. B **109** (2007), 239–261.

[68] Laurent, M., *Sums of squares, moment matrices and optimization over polynomials*, pages 157–270 in [87].

[69] Laurent, M., Mourrain, B., *A generalized flat extension theorem for moment matrices*, Archiv Math. **93(1)** (2009), 87-98.

[70] Lovász, L., *On the Shannon capacity of a graph*, IEEE Trans. Inform. Theory **IT-25** (1979), 1–7.

[71] Lovász, L., Schrijver, A., *Cones of matrices and set-functions and* $0-1$ *optimization*, SIAM J. Optim. **1** (1991), 166–190.

[72] Marshall, M., *Positive Polynomials and Sums of Squares*, Mathematical Surveys and Monographs, Amer. Math. Soc., **146** (2008).

[73] Mourrain, B., *A new criterion for normal form algorithms*, In H. Imai et al. (eds.), LNCS **1719** (1999), 430–443.

[74] Murty, K.G., Kabadi, S.N., *Some NP-complete problems in quadratic and nonlinear programming*, Math. Program. Ser. A, **39** (1987), 117–129.

[75] Nesterov, Y.E., *Squared functional systems and optimization problems*, In High Performance Optimization, J.B.G. Frenk et al. (eds.), Kluwer (2000), 405–440.

[76] Nesterov, Y.E., Nemirovski, A., *Interior Point Polynomial Methods in Convex Programming*, SIAM, Studies in Applied Mathematics **13** (1994).

[77] Nie, J., *Certifying convergence of Lasserre's hierarchy via flat truncation*, Math. Program. Ser. A **142** (2013), 485–510.

[78] Nie, J., *Polynomial optimization with real varieties*, SIAM J. Optim. **23(3)** (2013), 1634-1646.

[79] Nie, J., *An exact Jacobian SDP relaxation for polynomial optimization*, Math. Program. Ser. A **137** (2013), 225–255.

[80] Nie, J., *Optimality conditions and finite convergence of Lasserre's hierarchy*, arXiv: 1206.0319v2 (2013).

[81] Nie, J., Schweighofer, M., *On the complexity of Putinar's Positivstellensatz*, J. Complexity **23(1)** (2007), 135–150.

[82] Parrilo, P.A., *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, PhD thesis, Caltech (2000).

[83] Parrilo, P.A., *Semidefinite programming relaxations for semialgebraic problems*, Math. Program. Ser. B **96** (2003), 293–320.

[84] Pironio, S., Navascués, M., Acín, A., *Convergent relaxations of polynomial optimization problems in non-commutative variables*, SIAM J. Optim. **20(5)** (2010), 2157-2180.

[85] Prestel, A., Delzell, C.N., *Positive Polynomials - From Hilbert's 17th Problem to Real Algebra*, Springer (2001).

[86] Putinar, M., *Positive polynomials on compact semi-algebraic sets*, Indiana University Math. J. **42** (1993), 969–984.

[87] Putinar, M., Sullivant, S. (eds.), *Emerging Applications of Algebraic Geometry*, IMA Volumes in Mathematics and its Applications **149** (2009).

[88] Ramana, M.W., *An exact duality theory for semidefinite programming and its complexity implications*, Math. Program. **77** (1997), 129–162.

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

[90] Scheiderer, C., *Sums of squares of regular functions on real algebraic varieties*, Trans. Amer. Math. Soc. **352** (1999), 1039–1069.

[91] Scheiderer, C., *Positivity and sums of squares: A guide to recent results*, pages 1–54 in [87].

[92] Schmüdgen, K., *The K-moment problem for compact semi-algebraic sets*, Math. Annalen **289** (1991), 203–206.

[93] Schrijver, A., *A comparison of the Delsarte and Lovász bounds*, IEEE Trans. Inform. Theory **25** (1979), 425–429.

[94] Schrijver, A., *New code upper bounds from the Terwilliger algebra and semidefinite programming*, IEEE Trans. Inform. Theory **51** (2005), 2859–2866.

[95] Schweighofer, M., *Optimization of polynomials on compact semialgebraic sets*, SIAM J. Optim. **15(3)** (2005), 805–825.

[96] Sherali, H.D., Adams, W.P., *A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems*, SIAM J. Disc. Math. **3** (1990), 411–430.

[97] Shor, N.Z., *An approach to obtaining global extremums in polynomial mathematical programming problems*, Kibernetika **5** (1987), 102–106.

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

[99] Vandenberghe, L., Boyd, S., *Semidefinite programming*, SIAM Rev. **38**(1996),49–95.

[100] Wolkowicz, H., Saigal, R., Vandeberghe, L. (eds.), *Handbook of Semidefinite Programming*, Kluwer (2000).

Centrum Wiskunde & Informatica (CWI), Science Park 123, 1098 XG Amsterdam, The Netherlands; Tilburg University, Department of Econometrics and Operations Research, PO Box 90153, 5000 LE Tilburg, The Netherlands.

E-mail: M.Laurent@cwi.nl