Math. Program., Ser. A 109, 1–26 (2007)

Digital Object Identifier (DOI) 10.1007/s10107-004-0561-4

Monique Laurent[⋆]

Semidefinite representations for finite varieties

Received: January 9, 2003 / Accepted: September 24, 2004
Published online: June 2, 2006 – © Springer-Verlag 2006

Abstract. We consider the problem of minimizing a polynomial over a set defined by polynomial equations and inequalities. When the polynomial equations have a finite set of complex solutions, we can reformulate this problem as a semidefinite programming problem. Our semidefinite representation involves combinatorial moment matrices, which are matrices indexed by a basis of the quotient vector space $\mathbb{R}[x_1, \ldots, x_n]/I$, where I is the ideal generated by the polynomial equations in the problem. Moreover, we prove the finite convergence of a hierarchy of semidefinite relaxations introduced by Lasserre. Semidefinite approximations can be constructed by considering truncated combinatorial moment matrices; rank conditions are given (in a grid case) that ensure that the approximation solves the original problem to optimality.

1. Introduction

A central problem in combinatorial optimization and other areas of mathematics concerns the optimization of a linear or, more generally, polynomial function over a basic closed semi-algebraic set. It can be formulated as

$$p^* := \inf\ f(x) \text{ s.t. } h_1(x) = 0,\ \ldots, h_m(x) = 0,\ h_{m+1}(x) \geq 0,\ \ldots, h_{m+k}(x) \geq 0 \tag{1}$$

where $f, h_1, \ldots, h_{m+k} \in \mathbb{R}[x_1, \ldots, x_n]$. Set

$$V := \{x \in \mathbb{C}^n \mid h_1(x) = 0, \ldots, h_m(x) = 0\}, \tag{2}$$

$$S := V \cap \mathbb{R}^n,\quad F := S \cap \{x \mid h_{m+1}(x) \geq 0,\ \ldots, h_{m+k}(x) \geq 0\}. \tag{3}$$

We assume throughout the paper that the set V is *finite* and we let I denote the ideal generated by the polynomials $h_1, \ldots, h_m$; thus V is the complex variety associated to I. In typical combinatorial applications, $V = \{0, 1\}^n$, which corresponds to the case when the polynomials $h_1, \ldots, h_m$ are the quadratic polynomials $x_i^2 - x_i$ $(i = 1, \ldots, n)$, and F is determined by imposing additional polynomial constraints. For example, in the maximum stable set problem, F is the set of $0/1$ points satisfying $x_i x_j = 0$ for all pairs $ij \in E$, where E is the set of edges of a graph; alternatively, F is is the set of $0/1$ points satisfying $x_i + x_j \leq 1$ for all edges $ij \in E$.

Write the polynomial $f(x)$ as $f(x) = \sum_{\alpha \in S_d} f_\alpha x^\alpha$, where d is its degree, and let $f = (f_\alpha)_{\alpha \in S_d}$ denote the vector consisting of the coefficients of the polynomial $f(x)$. Here and below, S_d denotes the set of sequences $\alpha \in \mathbb{Z}_+^n$ with $|\alpha| := \sum_{i=1}^n \alpha_i \leq d$ and

M. Laurent: CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands

[⋆] Supported by the Netherlands Organisation for Scientific Research grant NWO 639.032.203.

$\mathbb{Z}_+$ is the set of nonnegative integers. Given an integer $t \geq d$, one can also view f as a vector in $\mathbb{R}^{S_t}$ by setting $f_\alpha := 0$ for all $\alpha \in S_t \setminus S_d$.

A classical approach for solving problem (1) is to linearize the objective function by introducing variables $y_\alpha = x^\alpha$ for the monomials x^α present in the problem. This allows us to write $f(x) = f^T y$. Given an integer $t \geq \deg(f)$, the set:

$$P := \text{conv}((v^\alpha)_{\alpha \in S_t} \mid v \in F). \tag{4}$$

is a polytope as F is a finite set. Therefore, problem (1) can be reformulated as the problem: $\min\limits_{y \in P} f^T y$, of minimizing the linear objective function $f^T y$ over P. Various methods have been proposed in the literature for constructing relaxations of the polytope P; that is, for generating new valid constraints from the given constraints $h_j(x) = 0$, $h_j(x) \geq 0$.

Combinatorial methods. Consider the 0/1 case, when the equations: $x_i^2 - x_i = 0$ ($i = 1, \dots, n$) are present in the description of F. A possibility for generating new constraints is to multiply the inequalities $h_j(x) \geq 0$ by certain products of the variables x_i and $1 - x_i$; then linearization is applied, after having replaced each occurrence of a square x_i^2 by x_i. In this way, hierarchies of linear relaxations: $P \subseteq \dots \subseteq P^n \subseteq \dots \subseteq P^t \subseteq \dots$ for the original problem are obtained. The lift-and-project method of Balas, Ceria and Cornuéjols [1], the matrix-cut method of Lovász and Schrijver [19], and the reformulation-linearization technique of Sherali and Adams [29] can all be cast in this framework. Lovász and Schrijver [19] propose moreover a stronger hierarchy of semidefinite relaxations. We do not give here the exact details of applicability of these methods; see [14] for a comparative presentation of these methods. A common feature is that the original problem (1) is solved exactly at step $t = n$ as a linear programming problem over a 2^n-dimensional simplex, and that each intermediary relaxation gives an efficiently computable bound for any fixed t (under some assumptions).

Algebraic methods via moments and sums of squares of polynomials. Several other authors have proposed methods for constructing semidefinite relaxations of the problem (1), based on results about moment sequences and (the dual theory of) representations of nonnegative polynomials as sums of squares. See Nesterov [21], Lasserre [11], Parrilo [22, 23], Parrilo and Sturmfels [25], Shor [30] and the recent papers of Marshall [20] and Schweighofer [28]. It turns out that, in the 0/1 case, these constructions yield relaxations that are at least as strong as the relaxations obtained via the above mentioned combinatorial methods. Moreover, they apply to the case when F is an arbitrary compact semi-algebraic set. Details about the links between the various methods can be found in [14].

Contents of the paper. The paper considers the problem of minimizing a polynomial function $f(x)$ over the semi-algebraic set F from (3), assuming that the equations in the description of F have a finite set V of complex solutions. One possibility would be to solve the problem using the theory of quantifier elimination (with a polynomial running time for a fixed number n of variables; see [2]). One can also compute all the points of V using the eigenvalue method sketched in Section 1.2 (together with exact symbolic computations using univariate representations for algebraic real numbers, see [2]), and then evaluate $f(v)$ for all $v \in F$. In this paper, we propose an alternative - more elementary - method, which does not need the enumeration of the points of V. Instead, we will reformulate problem (1) as a semidefinite program. As this program involves matrices of size $|\mathcal{B}| \geq |V|$, it can be solved in practice only for small $|\mathcal{B}|$. However, one can define (in the grid case) a hierarchy of semidefinite relaxations of the original problem whose low order members yield efficiently computable bounds to the original problem; some results in this direction are given in Section 4.

The paper is organized as follows. Section 1.1 contains some algebraic preliminaries about polynomial ideals and Section 1.2 recalls some well known results for the related problem of solving a system of polynomial equations. We present in Section 1.3 the method of Lasserre [11]. In particular, we introduce the hierarchy (15) of relaxations for (1), involving trucated moment matrices, and the dual hierarchy (16), in terms of sums of squares of polynomials with bounded degrees. We observe that, under the assumption that V is finite, problem (1) can be reformulated as (10) (involving infinite moment matrices) or (12) (involving sums of squares). We also introduce some results by Curto and Fialkow (Theorem 9), Putinar (Theorem 10) and Parrilo (Theorem 11), which play a central role in the paper.

In Section 2 we present the new semidefinite representation (22) for problem (1). It involves combinatorial moment matrices $M_{\mathcal{B}}(y)$, which are matrices indexed by a basis $\mathcal{B}$ of the quotient space $\mathbb{R}[x_1, \dots, x_n]/I$. Roughly speaking, this representation can be seen as a finite analogue of the program (10), obtained by 'factoring through the ideal I' generated by the polynomial equations entering the description of the semi-algebraic set F. The new formulation does not contain any semidefinite constraint for the equations $h_j(x) = 0$ ($j \leq m$), as they are used for the construction of the combinatorial moment matrix $M_{\mathcal{B}}(y)$. The ideas underlying this construction were already mentioned in the grid case ([13–15]). The equivalence of (1) and (22) follows using the result of Curto and Fialkow from Theorem 9. In the radical case, an alternative proof is based on a simple combinatorial identity involving the Zeta matrix of the ideal I (see Lemma 17); this is a direct extension of the corresponding result given in [19] and in [14] for the 0/1 and ±1 cases (which also underlies the convergence results for combinatorial lift-and-project methods). Combinatorial moment matrices turn out to be closely related to some algebraic notions (Hermite's form, multiplication operator) used for solving zero-dimensional systems of polynomial equations; see Sections 1.2 and 2.2 for details.

In Section 3, we prove the finite convergence of the bounds μ_t^* provided by the semidefinite hierarchy (15) (see Theorem 22). If I is radical, or if the polynomials $h_1, \dots, h_m$ form a Groebner basis of I, then there is also finite convergence of the bounds σ_t^* provided by the dual hierarchy (16). Under the second assumption, we can prove estimates on the order t for which $\sigma_t^* = \mu_t^* = p^*$; in the grid case these estimates are sharper than those given in [13]. (See Theorem 23 and Example 24.)

In Section 4, we consider approximations for problem (1) in the case when S is the set $\{x \mid (x_i - a_i)(x_i - b_i) = 0 \; \forall i = 1, \dots, n\}$ (where $a_i \neq b_i$ are given real numbers) (thus including the $0/1$ and ± 1 cases). These approximations are obtained by considering truncated combinatorial moment matrices $M_{\mathcal{B}_t}(y)$, where $M_{\mathcal{B}_t}(y)$ is indexed by all square free monomials of degree $\leq t$. Our main result is that, if $\mathcal{M}_{\mathcal{B}_t}(y) \succeq 0$ and rank $M_{\mathcal{B}_s}(y) \leq \sum_{i=1}^{s} \binom{t}{i}$ for some $1 \leq s \leq t$, then y is a convex combination of the vectors $(v^\alpha)_{\alpha\in\mathcal{B}_{2t}}$ $(v \in S)$. Moreover, the number of vectors entering the convex combination is at most 2^{t-1} if rank $M_{\mathcal{B}_1}(y) \leq t$. (See Theorem 25.) Therefore, if the optimum solution $\mathcal{M}_{\mathcal{B}_t}(y)$ of the relaxed problem satisfies the above rank condition, then it in fact solves the original problem (1) at optimality.

We also mention a result of the same flavour for the maximum stable set problem. The theta number:

$$\vartheta(G) := \max \sum_{i=1}^{n} y_i \text{ s.t. } M_{\mathcal{B}_1}(y) \succeq 0,\ y_{ij} = 0\ (ij \in E),\ y_0 = 1 \tag{5}$$

is a well known upper bound on the stability number of a graph $G = (\{1, \dots, n\}, E)$, introduced by Lovász [18]. If we impose that $M_{\mathcal{B}_1}(y)$ has rank 1 in (5), then the program solves the maximum stable set problem exactly. We prove that the same holds if we require only that rank $M_{\mathcal{B}_1}(y) \leq 2$ (see Proposition 30). This is an analogue of a result of Burer, Monteiro and Zhang [4] given for another formulation of the theta number.

1.1. Polynomial ideals and varieties

We group here some preliminaries on polynomial ideals and varieties. For more information on the material in the present and the next subsection, see, e.g., [2], [5], [6], [31]. Given an ideal I in $\mathbb{R}[x_1, \dots, x_n]$, the set V from (2) is the *complex variety* associated to I. If I is generated by $h_1, \dots, h_m$, then V consists of the common complex zeros of $h_1, \dots, h_m$. When V is finite, the ideal I is said to be *zero-dimensional.*

As the polynomials $h_1, \dots, h_m$ are real valued, V is closed under complex conjugation; that is, V can be partitioned as $S \cup T \cup \overline{T}$, where $S = V \cap \mathbb{R}^n$ and $\overline{T} = \{\overline{v} \mid v \in T\}$. When V is finite, one can find complex polynomials p_v (for $v \in V$) satisfying $p_v(v) = 1$ and $p_v(v') = 0$ for $v' \in V \setminus \{v\}$ as well as $p_{\overline{v}} = \overline{p_v}$ $(v \in V)$; the p_v's are known as *interpolation polynomials* at the points of V. Therefore, given complex numbers a_v $(v \in V)$ such that $a_{\overline{v}} = \overline{a_v}$ $(v \in V)$, one can find a real polynomial taking the prescribed values a_v at the points $v \in V$. This fact will be used in the proofs of Proposition 8 and Theorem 11.

The set

$$I(V) := \{f \in \mathbb{R}[x_1, \dots, x_n] \mid f(x) = 0 \text{ for all } x \in V\}$$

is an ideal in $\mathbb{R}[x_1, \dots, x_n]$ that contains the ideal I. When equality $I = I(V)$ holds, the ideal I is said to be *radical*; this corresponds, roughly speaking, to the case when all solutions $x \in V$ have single multiplicities. For instance, the ideal I in $\mathbb{R}[x]$ generated by $h(x) = x^2$ is not radical since $V = \{0\}$ and $f(x) = x$ belongs to $I(V) \setminus I$.

The monomials in $\mathbb{R}[x_1, \dots, x_n]$ are denoted as $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ for $\alpha \in \mathbb{Z}_+^n$. The *degree* of x^α is $|\alpha| = \sum_{i=1}^n \alpha_i$. Given a nonzero polynomial $f(x) = \sum_\alpha f_\alpha x^\alpha$, its *terms* are the quantities $f_\alpha x^\alpha$ with $f_\alpha \neq 0$ and its *degree* $\deg(f)$ is the maximum degree of a term of f. A *monomial ordering* '$<$' is a total ordering of the set of monomials which is a well-ordering and satisfies the condition: $x^\alpha < x^\beta \Longrightarrow x^{\alpha+\gamma} < x^{\beta+\gamma}$. Examples of monomial ordering are the *lexicographic order* '$<_{lex}$', where $x^\alpha <_{lex} x^\beta$ if $\alpha < \beta$

for a lexicographic order on $\mathbb{Z}_+^n$, or the *graded lexicographic order* '$<_{grlex}$', where $x^\alpha <_{grlex} x^\beta$ if $|\alpha| < |\beta|$, or $|\alpha| = |\beta|$ and $x^\alpha <_{lex} x^\beta$.

Fix a monomial ordering on $\mathbb{R}[x_1, \ldots, x_n]$. Given a nonzero polynomial $f(x) = \sum_\alpha f_\alpha x^\alpha$, its *leading term* $\mathrm{LT}(f)$ is defined as $f_\alpha x^\alpha$, where x^α is the maximum (with respect to the given ordering) monomial for which $f_\alpha \neq 0$. Let I be an ideal in $\mathbb{R}[x_1, \ldots, x_n]$. A finite subset $G \subseteq I$ is called a *Groebner basis* of I if the leading term of every nonzero polynomial in I is divisible by the leading term of some polynomial in G. It is known that a Groebner basis always exists. A monomial x^α is called a *standard monomial* if it not divisible by the leading term of any polynomial in I or, equivalently, if x^α is not divisible by the leading term of any poynomial in a Groebner basis.

Once a monomial ordering is fixed, one can apply the division algorithm. Let $h_1, \ldots, h_m$ and f be nonzero polynomials. If one divides f by the polynomials $h_1, \ldots, h_m$, one obtains polynomials $u_1, \ldots, u_m$ and r satisfying $f = \sum_{j=1}^m u_j h_j + r$, no term of r is divisible by $\mathrm{LT}(h_j)$ $(j = 1, \ldots, m)$ if $r \neq 0$, and $\mathrm{LT}(f) \geq \mathrm{LT}(u_j h_j)$ if $u_j \neq 0$. When the polynomials $h_1, \ldots, h_m$ form a Groebner basis of I, the remainder r is uniquely determined and it is a linear combination of the set $\mathcal{B}$ of standard monomials; that is, $r(x) = \sum_{x^\beta \in \mathcal{B}} r_\beta x^\beta$ where $r_\beta \in \mathbb{R}$. Moreover, $f \in I$ if and only if $r = 0$. Therefore, $\mathbb{R}[x_1, \ldots, x_n]/I$ and $\mathbb{R}^{\mathcal{B}}$ are isomorphic vector spaces.

Definition 1. *A set $\mathcal{B} := \{f_1, \ldots, f_N\}$ of polynomials forms a basis of $\mathbb{R}[x_1, \ldots, x_n]/I$ if, for every polynomial f, there exists a unique set of real numbers $\lambda_1^{(f)}, \ldots, \lambda_N^{(f)}$ such that $f - \sum_{i=1}^N \lambda_i^{(f)} f_i \in I$. When $\mathcal{B}$ contains only monomials, we call $\mathcal{B}$ a* monomial basis. *The polynomial $\sum_{i=1}^N \lambda_i^{(f)} f_i$ is called the* residue of f modulo I w.r.t. the basis $\mathcal{B}$ *and we set $\lambda^{(f)} := (\lambda_i^{(f)})_{i=1}^N \in \mathbb{R}^{|\mathcal{B}|}$. Moreover, for $v \in V$, define the vector $\zeta_v^{\mathcal{B}} := (f_i(v))_{i=1}^N \in \mathbb{R}^{|\mathcal{B}|}$; thus $f(v) = (\lambda^{(f)})^T \zeta_v^{\mathcal{B}}$.*

For instance, the set of standard monomials (w.r.t. some monomial ordering) is a (monomial) basis of $\mathbb{R}[x_1, \ldots, x_n]/I$. A fundamental property that we will use in the paper is the following, which can be found, e.g., in ([6], Theorem 2.10).

Theorem 2. *Let I be an ideal in $\mathbb{R}[x_1, \ldots, x_n]$ with complex variety V (as in (2)). Then, the variety V is finite if and only if the vector space $\mathbb{R}[x_1, \ldots, x_n]/I$ has finite dimension N. Moreover, $|V| \leq N$, with equality if and only if the ideal I is radical.*

Example 3. ([5], p. 227) Let I be the ideal in $\mathbb{R}[x, y]$ generated by $xy^3 - x^2$ and $x^3y^2 - y$. W.r.t. the lexicographic order with $y > x$, the polynomials $y - x^7$ and $x^{12} - x^2$ form a Groebner basis with corresponding set of standard monomials $\mathcal{B}_1 = \{1, x, x^2, \ldots, x^{11}\}$. W.r.t. the graded lexicographic order with $y > x$, the polynomials $x^3y^2 - y$, $x^4 - y^2$, $xy^3 - x^2$, $y^4 - xy$ form a Groebner basis with corresponding set of standard monomials $\mathcal{B}_2 = \{1, x, x^2, x^3, y, xy, x^2y, x^3y, y^2, xy^2, x^2y^2, y^3\}$. Hence, $|\mathcal{B}_1| = |\mathcal{B}_2| = 12$, and the maximum degree of a standard monomial is 11 in $\mathcal{B}_1$ and 4 in $\mathcal{B}_2$. The complex variety of I is $V = \{(0, 0)\} \cup \{(x, x^7) \mid x^{10} = 1\}$, with cardinality 11. As $|V| < |\mathcal{B}_1|$, the ideal I is not radical.

The quotient space $\mathbb{R}[x_1, \ldots, x_n]/I$ has simultaneously the structure of a vector space and of a ring. In order to specify the multiplication operation, it suffices to give the multiplication table w.r.t. a basis $\mathcal{B} = \{f_1, \dots, f_N\}$; this is the $\mathcal{B} \times \mathcal{B}$ matrix $T_{\mathcal{B}}$ whose (f_i, f_j)th entry is the residue w.r.t. $\mathcal{B}$ of the product $f_i f_j$; that is, for $f_i, f_j \in \mathcal{B}$,

$$T_{\mathcal{B}}(f_i, f_j) = \sum_{k=1}^{N} \lambda_k^{(f_i f_j)} f_k. \tag{6}$$

The multiplication table $T_{\mathcal{B}}$ can be computed using the division algorithm.

Example 4. Let I_1 be the ideal in $\mathbb{R}[x, y]$ generated by $x^2 - x$ and $y^2 - y$ and I_2 be the ideal generated by $x^2 - 1$ and $y^2 - 1$. Then, for $i = 1, 2$, $\mathcal{B} := \{1, x, y, xy\}$ is a monomial basis of $\mathbb{R}[x, y]/I_i$ whose multiplication table T_i is as follows:

$$T_1 = \begin{array}{c} \\ 1 \\ x \\ y \\ xy \end{array} \begin{array}{c} \begin{array}{cccc} 1 & x & y & xy \end{array} \\ \begin{pmatrix} 1 & x & y & xy \\ x & x & xy & xy \\ y & xy & y & xy \\ xy & xy & xy & xy \end{pmatrix} \end{array}, \quad T_2 = \begin{array}{c} \\ 1 \\ x \\ y \\ xy \end{array} \begin{array}{c} \begin{array}{cccc} 1 & x & y & xy \end{array} \\ \begin{pmatrix} 1 & x & y & xy \\ x & 1 & xy & y \\ y & xy & 1 & x \\ xy & y & x & 1 \end{pmatrix} \end{array}.$$

1.2. Solving systems of polynomial equations

Our paper is dealing with the problem of optimizing a polynomial function over a finite set of points arising as the solution set of a system of polynomial equations and inequalities. A related problem, which has received considerable attention in the literature, is the problem of solving a system of polynomial equations: $h_1(x) = 0, \dots, h_m(x) = 0$. Let I be the ideal generated by $h_1, \dots, h_m$ with complex variety V, assumed to be finite. Then the so-called eigenvalue method (also known as the Stetter-Möller method) can be applied for finding V, which relies on finding the eigenvalues of the multiplication operator. Given a polynomial $h \in \mathbb{R}[x_1, \dots, x_n]$, let

$$\begin{aligned} m_h : \mathbb{R}[x_1, \dots, x_n]/I &\longrightarrow \mathbb{R}[x_1, \dots, x_n]/I \\ f &\longmapsto fh \end{aligned}$$

be the 'multiplication by h' operator. Given a basis $\mathcal{B}$ of $\mathbb{R}[x_1, \dots, x_n]/I$, let M_h denote the matrix of m_h w.r.t. $\mathcal{B}$; thus M_h is the $\mathcal{B} \times \mathcal{B}$ matrix whose fth column is the vector $\lambda^{(hf)} \in \mathbb{R}^{\mathcal{B}}$, giving the residue w.r.t. $\mathcal{B}$ of the product hf, for $f \in \mathcal{B}$. When $h(x) = x_i$, the multiplication matrices $M_{x_i} =: M_i$ are known as the *companion matrices* of the ideal I. As the matrices $M_1, \dots, M_n$ pairwise commute, they generate a commutative algebra $\mathbb{R}[M_1, \dots, M_n]$, which is isomorphic to the ring $\mathbb{R}[x_1, \dots, x_n]/I$ via the correspondance $x_i \longrightarrow M_i$. It is easy to verify that, for $h(x) = \sum_\alpha h_\alpha x^\alpha$, $M_h = \sum_\alpha h_\alpha M_1^{\alpha_1} \cdots M_n^{\alpha_n}$. The next result shows that the coordinates of the points $v \in V$ can be evaluated by computing the eigenvalues of the companion matrices.

Theorem 5 (Stickelberger's theorem). *Assume that V is finite.*

(i) *The complex zeros of I are the vectors of joint eigenvalues of the companion matrices $M_1, \dots, M_n$; that is, $V = \{\lambda \in \mathbb{C}^n \mid \exists u \in \mathbb{C}^n \setminus \{0\}$ s.t. $M_i u = \lambda_i u \ \forall i = 1, \dots, n\}$.*

(ii) *Given $h \in \mathbb{R}[x_1, \dots, x_n]$, $Tr(M_h) = \sum_{v \in V} m(v)h(v)$, where $m(v)$ $(v \in V)$ are the multiplicities of the zeros of I. In particular, $Tr(M_h) = \sum_{v \in V} h(v)$ when I is radical.*

A related relevant result is the following result (see, e.g., [2], section 4.5) permitting to count the number of real zeros of a system of polynomial equations having finitely many complex zeros. It involves the bilinear map:

$$\begin{aligned} her_h : \mathbb{R}[x_1, \dots, x_n]/I \times \mathbb{R}[x_1, \dots, x_n]/I &\longrightarrow \mathbb{R} \\ (f, g) &\longmapsto Tr(M_{fgh}) \end{aligned}$$

whose associated quadratic form: $Her_h(f) := her_h(f, f)$ for $f \in \mathbb{R}[x_1, \dots, x_n]/I$, is known as the multivariate *Hermite's form*. Let $H_h = (Tr(M_{fgh}))_{f,g \in \mathcal{B}}$ denote the matrix of the Hermite's form w.r.t. a basis $\mathcal{B}$ of $\mathbb{R}[x_1, \dots, x_n]/I$; thus H_h is real symmetric.

Theorem 6 (Counting real zeros via the Hermite's form). *Let σ_+ (resp., σ_-) denote the number of positive (resp., negative) eigenvalues of H_h, and set $s_+ := |\{v \in V \cap \mathbb{R}^n \mid h(v) > 0\}|$, $s_- := |\{v \in V \cap \mathbb{R}^n \mid h(v) < 0\}|$, and $2t := |\{v \in V \setminus \mathbb{R}^n \mid h(v) \neq 0\}|$. Then, $\sigma_+ - \sigma_- = s_+ - s_-$, and* rank *$H_h = |\{v \in V \mid h(v) \neq 0\}| = s_+ + s_- + 2t$. Therefore, $\sigma_+ = s_+ + t$ and $\sigma_- = s_- + t$.*

As we will see in Section 2.2, there is a close relationship between the Hermite's matrix H_h and the notion of combinatorial moment matrix considered in this paper. The following observations will be useful for establishing this link. In view of Theorem 5 (ii), $H_h = (\sum_{v \in V} m(v)f(v)g(v)h(v))_{f,g \in \mathcal{B}}$; that is, $H_h = \sum_{v \in V} m(v)h(v)\zeta_v^{\mathcal{B}}(\zeta_v^{\mathcal{B}})^T$, where $\zeta_v = (f(v))_{f \in \mathcal{B}}$ (as in Definition 1). In fact, any matrix H of the form $H = \sum_{v \in V} a_v \zeta_v^{\mathcal{B}}(\zeta_v^{\mathcal{B}})^T$, where the scalars $a_v \in \mathbb{C}$ satisfy $a_{\overline{v}} = \overline{a_v}$ for $v \in V$, is the Hermite's matrix H_h of some polynomial $h \in \mathbb{R}[x_1, \dots, x_n]$, as one can find such polynomial h satisfying $h(v)m(v) = a_v$ for all $v \in V$. Thus Theorem 6 applies to any such matrix H. In particular,

$$H \succeq 0 \Longleftrightarrow s_- = t = 0 \Longleftrightarrow a_v \geq 0 \ (v \in V \cap \mathbb{R}^n), \ \ a_v = 0 \ (v \in V \setminus \mathbb{R}^n). \tag{7}$$

[One can easily prove this fact directly, along the same lines as for the proof of Lemma 2.5 in [17].]

1.3. Moments and sums of squares of polynomials

We present here the method of Lasserre [11] for solving problem (1). A basic observation underlying Lasserre's construction is the fact that

$$p^* := \inf_{x \in F} f(x) = \inf_{\mu} \int f(x)\mu(dx), \tag{8}$$

where the second infimum is taken over all probability measures μ on $\mathbb{R}^n$ supported by F (i.e., μ is a nonnegative measure with mass 0 outside of F and with total mass

$\int_F \mu(dx) = 1$). Note that $\displaystyle\int f(x)\mu(dx) = \sum_\alpha f_\alpha \int x^\alpha \mu(dx)$. For $\alpha \in \mathbb{Z}_+^n$, the quantity $y_\alpha := \displaystyle\int x^\alpha \mu(dx)$ is called the *moment of order* α of the measure μ, and $y = (y_\alpha)_{\alpha \in \mathbb{Z}_+^n}$ is called the sequence of moments of μ; one also says that μ is a *representing measure* for y. The measure μ is a probability measure when $y_0 = 1$. Therefore, (8) can be reformulated as

$$p^* = \inf\ f^T y \ \text{ s.t. } y \text{ has a representing measure supported by } F. \tag{9}$$

When F is a finite set, every probability measure μ supported by F is atomic; that is, μ can be written as $\mu = \sum_{v \in F} \lambda_v \delta_v$, where $\lambda_v \geq 0$, $\sum_{v \in F} \lambda_v = 1$, and δ_v is the Dirac measure at v (with mass 1 at v and zero elsewhere). Then the moment of order α of μ is equal to $\sum_{v \in F} \lambda_v v^\alpha$. Therefore, the points of the polytope P from (4) are precisely the sequences having a probability representing measure supported by F and relaxations of P can be obtained by considering necessary conditions for the existence of such measures. Characterizing sequences of moments of measures is the object of the classical theory of moments; see, e.g., [7], [8], [10]. Moment matrices $M(y)$ and the so-called shift operator $h * y$ are two classical notions used in characterizations of moment sequences, as the next lemma shows. In what follows, $\mathbb{R}^{\mathbb{Z}_+^n}$ denotes the vector space consisting of the sequences $y = (y_\alpha)_{\alpha \in \mathbb{Z}_+^n}$.

Given $y \in \mathbb{R}^{\mathbb{Z}_+^n}$, its *moment matrix* is the (infinite) matrix $M(y) := (y_{\alpha+\beta})_{\alpha,\beta \in \mathbb{Z}_+^n}$ and, given a polynomial $h \in \mathbb{R}[x_1, \dots, x_n]$, $h * y := M(y)h \in \mathbb{R}^{\mathbb{Z}_+^n}$, with entries $(h * y)_\alpha = \sum_\beta h_\beta y_{\alpha+\beta}$ ($\alpha \in \mathbb{Z}_+^n$). For a real symmetric matrix X, the notation: $X \succeq 0$ means that X is positive semidefinite. If A is a principal submatrix of X, one says that X is a *flat extension* of A if rank $X =$ rank A; then, $X \succeq 0 \Longleftrightarrow A \succeq 0$.

Lemma 7. (i) *If $y \in \mathbb{R}^{\mathbb{Z}_+^n}$ has a representing measure μ, then $M(y) \succeq 0$. Moreover, if $p(x)$ is a polynomial such that $M(y)p = 0$, then the support of μ is contained in the set of zeros of $p(x)$.*

(ii) *Let $h(x)$ be a polynomial and $F := \{x \in \mathbb{R}^n \mid h(x) \geq 0\}$. If $y \in \mathbb{R}^{\mathbb{Z}_+^n}$ has a representing measure supported by F, then $M(h * y) \succeq 0$.*

(iii) *If $M(y) \succeq 0$, then* Ker $M(y)$ *is an ideal in* $\mathbb{R}[x_1, \dots, x_n]$.

Proof. (i) follows from the fact that, for any polynomial p, $p^T M(y) p = \sum_{\alpha,\beta} p_\alpha p_\beta y_{\alpha+\beta} = \sum_{\alpha,\beta} p_\alpha p_\beta \int x^{\alpha+\beta} \mu(dx) = \int p(x)^2 \mu(dx)$ and (ii) follows from the fact that $p^T M(h * y) p = \int_F h(x) p(x)^2 \mu(dx)$.

(iii) Let p, q be polynomials such that $M(y)p = 0$ and set $f := pq$, $g := pq^2$; we show that $M(y) f = 0$. One can easily verify that $f^T M(y) f = g^T M(y) p$; thus, $f^T M(y) f = 0$, which implies $M(y) f = 0$ since $M(y) \succeq 0$. □

Based on this, one can define the following bound for p^*:

$$\begin{aligned} \mu^* := \inf\ f^T y \ \text{ s.t. } \ & M(y) \succeq 0,\ M(h_j * y) = 0 \ (j = 1, \dots, m) \\ & M(h_j * y) \succeq 0 \ (j = m+1, \dots, m+k),\ y_0 = 1. \end{aligned} \tag{10}$$

By Lemma 7 (i),(ii), $\mu^* \leq p^*$. In fact, equality: $\mu^* = p^*$ holds, as the next result shows.

Proposition 8. *Let F be as in (3) and assume that V is finite. The following assertions are equivalent for $y \in \mathbb{R}^{\mathbb{Z}_+^n}$:*

(i) *y has a representing measure supported by F.*
(ii) *$M(y) \succeq 0, M(h_j{*}y) = 0 (j = 1, \ldots, m), M(h_j{*}y) \succeq 0 (j = m{+}1, \ldots, m{+}k)$.*

The proof relies on the following result of Curto and Fialkow.

Theorem 9. [7] *Given $y \in \mathbb{R}^{\mathbb{Z}_+^n}$, if $M(y)$ is positive semidefinite and has finite rank, then y has a (unique) representing measure.*

Proof of Proposition 8. (i) $\Longrightarrow$ (ii) follows from Lemma 7.
(ii) $\Longrightarrow$ (i) Note first that $I \subseteq$ Ker $M(y)$, since Ker $M(y)$ is an ideal containing the polynomials $h_1, \ldots, h_m$, as $0 = h_j * y = M(y)h_j$ $(j = 1, \ldots, m)$. Let $\mathcal{B}$ be a monomial basis of $\mathbb{R}[x_1, \ldots, x_n]/I$. As any monomial x^α is congruent modulo I to a polynomial which is a linear combination of monomials in $\mathcal{B}$, the αth column of $M(y)$ can be expressed as a linear combination of the columns indexed by $\mathcal{B}$. Therefore, the matrix $M(y)$ has finite rank. By Theorem 9, y has a representing measure μ. As $I \subseteq$ Ker $M(y)$, it follows from Lemma 7 that the support of μ is contained in $V \cap \mathbb{R}^n = S$. Say, $\mu = \sum_{v \in S} a_v \delta_v$ with $a_v \geq 0$. Remains to verify that $a_v = 0$ if $v \in S \setminus F$. Let $v \in S \setminus F$ and let $j \geq m + 1$ such that $h_j(v) < 0$. Let p be a real polynomial such that $p(v) = 1$ and $p(v') = 0$ for $v' \in S \setminus \{v\}$. Then, $0 \leq p^T M(h_j * y)p = a_v h_j(v)$, which implies that $a_v = 0$. Thus the support of μ is contained in F. □

Call a polynomial u a *s.o.s.* if u can be written as a sum of squares of polynomials. Define

$$\mathcal{M}(F) := \{u_0 + \sum_{j=1}^{m+k} u_j h_j \mid u_j \in \mathbb{R}[x_1, \ldots, x_n], u_0, u_{m+1}, \ldots, u_{m+k} \text{ are s.o.s.}\} \tag{11}$$

As $p^* = \sup\ \rho\ \text{ s.t. } f(x) - \rho \geq 0\ \forall x \in F$, one can derive a lower bound on p^* by replacing the nonnegativity condition: $f(x) - \rho \geq 0\ \forall x \in F$ by the stronger condition: $f(x) - \rho \in \mathcal{M}(F)$; namely,

$$\sigma^* := \sup\ \rho\ \text{ s.t. } f(x) - \rho \in \mathcal{M}(F). \tag{12}$$

Then, $\sigma^* \leq \mu^*$. Indeed, $f^T y \geq \rho$ if y is feasible for (10) and if ρ is feasible for (12), which can be seen using the following facts: Any polynomial u can be written as the difference of two s.o.s. (e.g., $u = \frac{1}{4}((u + 1)^2 - (u - 1)^2)$); given two polynomials p and h and $g := hp^2$, then $y^T g = p^T M(h * y)p$ (easy to check). In fact, equality:

$$\sigma^* = \mu^* = p^* \tag{13}$$

holds, which can be proved using the following result of Putinar [27].

Theorem 10. [27] *Let $F := \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_L(x) \geq 0\}$ for some polynomials $g_1, \ldots, g_L$, and $\mathcal{M}(F) := \{u_0 + \sum_{\ell=1}^{L} u_\ell g_\ell \mid u_0, u_\ell$ are s.o.s.$\}$. If F is compact and if there exists a polynomial $u \in \mathcal{M}(F)$ for which the set $U := \{x \in \mathbb{R}^n \mid u(x) \geq 0\}$ is compact, then every* positive *polynomial on F belongs to $\mathcal{M}(F)$.*

In the present case, the assumptions of Theorem 10 hold. Indeed, the polynomial $u(x) := -\sum_{j=1}^{m} h_j(x)^2$ belongs to $\mathcal{M}(F)$ and $U = V \cap \mathbb{R}^n$ is finite. Therefore, given $\epsilon > 0$, as the polynomial $f(x) - p^* + \epsilon$ is positive on F, it belongs to $\mathcal{M}(F)$ and thus $p^* - \epsilon$ is feasible for (12), implying $\sigma^* \geq p^* - \epsilon$. Letting ϵ go to 0, we find that $\sigma^* \geq p^*$ and thus (13) holds.

Comment. Equality: $\mu^* = p^*$ can be proved using Curto and Fialkow's result (Theorem 9) and it is implied by (13), which follows using Putinar's result (Theorem 10). The original proofs for these two results are based on functional analytic tools (including Riesz representation theorem and the spectral theorem). An alternative simple proof for Theorem 9 has been given recently in [17] which, beside Hilbert's Nullstellensatz, uses only elementary linear algebra; a key observation is that the kernel of a positive semidefinite moment matrix is a *radical* ideal. Although a more elementary proof for Theorem 10 has been given in [28], it remains however more involved than the proof for Theorem 9.

Parrilo [24] shows the following extension of Putinar's result to *nonnegative* polynomials on F, in the case when the polynomials $h_j(x)$ $(j = 1, \ldots, m)$ generate a radical ideal. The proof is elementary and is included for completeness. Thus, in the radical case, (13) follows directly from Theorem 11.

Theorem 11. [24] *Let F be as in (3). Assume that V is finite and that the ideal generated by $h_1, \ldots, h_m$ is radical. Then every nonnegative polynomial on F belongs to $\mathcal{M}(F)$.*

Proof. Write $V = S \cup T \cup \bar{T}$, where $S = V \cap \mathbb{R}^n$, $\bar{T} = \{\bar{v} \mid v \in T\}$ and $T \cup \bar{T} = V \setminus \mathbb{R}^n$. Suppose first that f is a real polynomial nonnegative on the set S. For $v \in S \cup T$, let $\gamma_v = \sqrt{f(v)}$ (thus, $\gamma_v \in \mathbb{R}_+$ if $v \in S$) and define the real polynomials $q_v := \gamma_v p_v$ $(v \in S)$ and $q_v := \gamma_v p_v + \overline{\gamma_v p_v}$ $(v \in T)$. The polynomial $f - \sum_{v \in S \cup T} (q_v)^2$ vanishes at all points of V; hence it belongs to $I(V)$, which is equal to I since I is radical. This shows that $f = \sigma + q$, where σ is a s.o.s. and $q \in I$.

Suppose now that f is nonnegative on the set F. We define real polynomials s_0, $s_{m+1}, \ldots, s_{m+k}$ 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 $s_j(v) = 0$ $(j = m+1, \ldots, m+k)$. Otherwise, $v \notin F$ and thus $h_{j_v}(v) < 0$ for some $j_v \in \{m+1, \ldots, m+k\}$; then $s_{j_v}(v) = \frac{f(v)}{h_{j_v}(v)}$ and $s_0(v) = s_j(v) = 0$ for all remaining j. Then, each of $s_0, s_{m+1}, \ldots, s_{m+k}$ is nonnegative on S; by the above, $s_j = \sigma_j + q_j$, where σ_j is a s.o.s. and $q_j \in I$. By construction, the polynomial $f - s_0 - \sum_{j=m+1}^{m+k} s_j h_j$ vanishes at all points of V and thus belongs to I. Therefore, $f - \sigma_0 - \sum_{j=m+1}^{m+k} \sigma_j h_j \in I$. □

Practically, one can use the programs (10) and (12) for computing p^* in the following way. Define

$$d_j := \lceil \deg(h_j)/2 \rceil \ (j = 1, \ldots, m+k), \quad d := \max(d_1, \ldots, d_{m+k}). \qquad (14)$$

Following Lasserre [11], one can formulate the following semidefinite relaxations for (1), obtained from (10) by replacing the (infinite) moment matrix $M(y)$ by its leading principal submatrix (truncation) $M_t(y) := (y_{\alpha+\beta})_{\alpha,\beta\in S_t}$, indexed by the set S_t corresponding to the set of monomials with degree at most t.

$$\begin{aligned} \mu_t^* := \inf \quad & f^T y \\ \text{s.t.} \quad & M_t(y) \succeq 0, \quad M_{t-d_j}(h_j * y) = 0 \; (j = 1, \dots, m) \\ & M_{t-d_j}(h_j * y) \succeq 0 \; (j = m+1, \dots, m+k), \quad y_0 = 1 \end{aligned} \tag{15}$$

for any $t \geq \max(d, \lceil \deg(f)/2 \rceil)$. Obviously, $\mu_t^* \leq \mu_{t+1}^* \leq \mu^* \leq p^*$. By bounding the degrees of the unknown polynomials $u_j(x)$ in (12), one obtains the hierarchy:

$$\begin{aligned} \sigma_t^* := \; & \sup \rho \\ & \text{s.t. } f(x) - \rho = u_0(x) + \sum_{j=1}^{m+k} u_j(x) h_j(x) \; \text{ with } u_0, u_{m+1}, \dots, u_{m+k} \text{ s.o.s.} \\ & \text{and } \deg(u_0), \; \deg(u_j h_j) \leq 2t \; (j = 1, \dots, m+k) \end{aligned} \tag{16}$$

for $t \geq \max(d, \lceil \deg(f)/2 \rceil)$. Obviously, $\sigma_t^* \leq \sigma_{t+1}^* \leq \sigma^* \leq p^*$. As sums of squares of polynomials can be tested using semidefinite programming (see [26]), (16) can be reformulated as a semidefinite program. Any polynomial u can be decomposed as $u' - u''$, where u' and u'' are s.o.s. with degree at most $\deg(u)$; hence the program (16) is equivalent to the usual formulation, where the polynomials u_j $(j = 1, \dots, m)$ are required to be of the form $u'_j - u''_j$, with u'_j, u''_j s.o.s. of degree at most $2(t - d_j)$.

Lasserre [11] shows that the program (16) is the semidefinite dual of (15); hence,

$$\sigma_t^* \leq \mu_t^* \leq p^* \tag{17}$$

by weak duality. By Theorem 10, the bounds σ_t^* (and thus μ_t^*) converge to p^* as $t \to \infty$. When the ideal I generated by $h_1, \dots, h_m$ is radical, Theorem 11 implies the *finite* convergence of the bounds σ_t^* (and thus μ_t^*) to p^*. Lasserre [12, 13] has shown the *finite* convergence of the bounds σ_t^*, μ_t^* in the case when F is contained in $\{0, 1\}^n$ or is equal to the set of points in a grid; the result follows alternatively from Parrilo's result since the ideal is radical in the grid case. In the non-radical case, we show the finite convergence of μ_t^* (see Theorem 22) and, in the case when the polynomials $h_1, \dots, h_m$ form a Groebner basis of I (w.r.t. some monomial ordering), the finite convergence of σ_t^* (see Theorem 23). Furthermore, we give a semidefinite representation for problem (1) (see Corollary 16), which is more concise than (15), as it involves matrices of size smaller than $|S_t|$ for any t ensuring the finite convergence of (15).

2. A Semidefinite Representation for a Finite Variety

2.1. Approach via combinatorial moment matrices

Consider the problem (1) of minimizing a polynomial f over the semi-algebraic set F from (3). As before, I is the ideal generated by the polynomials $h_1, \dots, h_m$ and V is its complex variety, assumed to be finite. Then, $\mathbb{R}[x_1, \dots, x_n]/I$ has finite dimension $N \geq |V|$ (by Theorem 2). Let $\mathcal{B} := \{f_1, \dots, f_N\}$ be a basis of $\mathbb{R}[x_1, \dots, x_n]/I$.

Problem (1) remains unchanged if we replace the polynomial f by its residue $r(x) := \sum_{i=1}^{N} \lambda_i^{(f)} f_i(x)$ modulo I w.r.t. $\mathcal{B}$ (recall Definition 1). That is,

$$p^* = \min_{v \in F} (\lambda^{(f)})^T \zeta_v^{\mathcal{B}} = \min \; y^T \lambda^{(f)} \;\text{ s.t. } y \in P_{\mathcal{B}}(F) := \mathrm{conv}(\zeta_v^{\mathcal{B}} \mid v \in F). \tag{18}$$

We give here a semidefinite representation for the polytope $P_{\mathcal{B}}(F)$, which can be seen as a finite analogue of the program (10); it does not need the explicit knowledge of the variety V, but only the knowledge of a basis $\mathcal{B}$ of $\mathbb{R}[x_1, \dots, x_n]/I$ and of its associated multiplication table.

The key ingredient in the proof of Proposition 8 was the fact that the kernel of a matrix $M(y)$ feasible for (10) contains the ideal I; this implies that $M(y)$ is a flat extension of its principal submatrix $M_{\mathcal{B}}(y)$ indexed by a monomial basis $\mathcal{B}$ of $\mathbb{R}[x_1, \dots, x_n]/I$. In particular, all entries of $M(y)$ can be expressed in terms of the entries of y indexed by $\mathcal{B}$, using the equations $h(x) = 0$ provided by $h \in I$. We now formalize this idea and introduce the 'combinatorial' analogues $M_{\mathcal{B}}(y)$ and $h * y \in \mathbb{R}^{|\mathcal{B}|}$ (for $y \in \mathbb{R}^{|\mathcal{B}|}$) of the corresponding notions in the classical case.

Given $y = (y_1, \dots, y_N)^T \in \mathbb{R}^{|\mathcal{B}|}$, its *combinatorial moment matrix* $M_{\mathcal{B}}(y)$ is the $|\mathcal{B}| \times |\mathcal{B}|$ matrix indexed by $\mathcal{B}$, whose (f_i, f_j)th entry is $y^T \lambda^{(f_i f_j)} = \sum_{k=1}^{N} \lambda_k^{(f_i f_j)} y_k$ for $f_i, f_j \in \mathcal{B}$. Thus $M_{\mathcal{B}}(y)$ is obtained from the multiplication table of $\mathcal{B}$ (recall (6)) by 'linearizing', where 'linearizing' means replacing each occurrence of f_k by y_k.

Given a polynomial $h \in \mathbb{R}[x_1, \dots, x_n]$, define $h * y := M_{\mathcal{B}}(y)\lambda^{(h)} \in \mathbb{R}^{|\mathcal{B}|}$. One can verify that $h * y = M_h^T y$, where M_h is the 'multiplication by h' operator introduced in Section 1.2.

Let U denote the $N \times |\mathbb{Z}_+^n|$ matrix whose rows (resp., columns) are indexed by $\mathcal{B}$ (resp., by $\mathbb{Z}_+^n$) and whose (i, α)-entry is equal to $\lambda_i^{(x^\alpha)}$. In words, the αth column of U contains the coordinates of the residue of x^α in the basis $\mathcal{B}$. For a polynomial h, its residue modulo I w.r.t. $\mathcal{B}$ is $\sum_{i=1}^{N} \lambda_i^{(h)} f_i$, where

$$\lambda^{(h)} = Uh. \tag{19}$$

Indeed, $\sum_{i=1}^{N} \lambda_i^{(h)} f_i \equiv h$ mod. I, while $h = \sum_\alpha h_\alpha x^\alpha \equiv \sum_\alpha h_\alpha (\sum_{i=1}^{N} U_{i,\alpha} f_i)$ which is equal to $\sum_{i=1}^{N} (\sum_\alpha h_\alpha U_{i,\alpha}) f_i = \sum_{i=1}^{N} (Uh)_i f_i$, thus showing that $\lambda_i^{(h)} = (Uh)_i$ for $i = 1, \dots, N$.

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

$$\tilde{y} := U^T y \in \mathbb{R}^{\mathbb{Z}_+^n}. \tag{20}$$

Hence, for $\alpha \in \mathbb{Z}_+^n$, $\tilde{y}_\alpha$ is obtained by 'linearizing' the residue of x^α modulo I w.r.t. $\mathcal{B}$.

Example 12. Consider the 0/1 case, when I is generated by the polynomials $x_j^2 - x_j$ $(j = 1, \dots, n)$ and $V = \{0, 1\}^n$. W.r.t. the graded lexicographic order, the set of standard monomials is

$$\mathcal{B} = \{x^A := \prod_{i \in A} x_i \mid A \subseteq \{1, \dots, n\}\}. \tag{21}$$

For $x^A, x^B \in \mathcal{B}$, the (x^A, x^B)th entry of the multiplication table is $x^{A \cup B}$, since $x^{A \cup B}$ is the residue of the product $x^A x^B$ modulo I. Hence the combinatorial moment matrix

$M_{\mathcal{B}}(y)$ of a vector $y = (y_A)_{A\subseteq\{1,\dots,n\}}$ is the matrix indexed by all subsets of $\{1,\dots,n\}$ with (A,B)th entry $y_{A\cup B}$. The extension $\tilde{y}$ of y satisfies: $\tilde{y}_\alpha = y_A$, where $A = \{i \in \{1,\dots,n\} \mid \alpha_i \geq 1\}$.

In the ±1 case, I is generated by the polynomials $x_j^2 - 1$ $(j = 1,\dots,n)$, $V = \{\pm 1\}^n$, and (21) is again the set of standard monomials. The (x^A, x^B)th entry of the multiplication table is now equal to $x^{A\Delta B}$, and the combinatorial moment matrix $M_{\mathcal{B}}(y)$ of a vector $y = (y_A)_{A\subseteq\{1,\dots,n\}}$ has its (A,B)th entry equal to $y_{A\Delta B}$. The extension $\tilde{y}$ of y satisfies: $\tilde{y}_\alpha = y_A$, where $A = \{i \in \{1,\dots,n\} \mid \alpha_i \text{ odd }\}$.

Lemma 13. *Let $y \in \mathbb{R}^{|\mathcal{B}|}$ and let $\tilde{y} \in \mathbb{R}^{\mathbb{Z}_+^n}$ be its extension as in (20).*

(i) *$M(\tilde{y}) = U^T M_{\mathcal{B}}(y) U$. Hence, when $\mathcal{B}$ is a monomial basis, $M_{\mathcal{B}}(y)$ is the principal submatrix of $M(\tilde{y})$ indexed by $\mathcal{B}$ and $M(\tilde{y})$ is a flat extension of $M_{\mathcal{B}}(y)$.*
(ii) *The extension of $h * y$ is equal to $h * \tilde{y}$.*
(iii) $I \subseteq \operatorname{Ker} M(\tilde{y})$.
(iv) *For a polynomial h, $h^T\tilde{y} = (\lambda^{(h)})^T y$.*

Proof. (i) Given $\alpha, \beta \in \mathbb{Z}_+^n$, the (α,β)th entry of the matrix $U^T M_{\mathcal{B}}(y) U$ is equal to $\sum_{i,j=1}^N U_{i,\alpha} U_{j,\beta} M_{\mathcal{B}}(y)_{ij} = \sum_{i,j=1}^N U_{i,\alpha} U_{j,\beta} \left(\sum_{k=1}^N \lambda_k^{(f_i f_j)} y_k\right) = \sum_{k=1}^N \left(\sum_{i,j=1}^N U_{i,\alpha} U_{j,\beta} \lambda_k^{(f_i f_j)}\right) y_k$. On the other hand, $x^\alpha \equiv \sum_{i=1}^N U_{i,\alpha} f_i$, $x^\beta \equiv \sum_{j=1}^N U_{j,\beta} f_j$ modulo I, which implies that $x^\alpha x^\beta \equiv \sum_{i,j=1}^N U_{i,\alpha} U_{j,\beta} f_i f_j \equiv \sum_{k=1}^N \left(\sum_{i,j=1}^N U_{i,\alpha} U_{j,\beta} \lambda_k^{(f_i f_j)}\right) f_k$ modulo I. Therefore, the (α,β)th entry of $M_{\mathcal{B}}(y)$ is equal to $\sum_{k=1}^N \left(\sum_{i,j=1}^N U_{i,\alpha} U_{j,\beta} \lambda_k^{(f_i f_j)}\right) y_k$.
(ii) By definition, the extension of $h * y$ is $U^T(h * y) = U^T(M_{\mathcal{B}}(y)\lambda^{(h)})$, while $h * \tilde{y} = M(\tilde{y})h = U^T M_{\mathcal{B}}(y) U h$ (using (i) above), which is equal to $U^T M_{\mathcal{B}}(y)\lambda^{(h)}$ (using (19)).
(iii) If $h \in I$, then $0 = \lambda^{(h)} = Uh$ (by (19)), implying $M(\tilde{y})h = U^T M_{\mathcal{B}}(y) U h = 0$.
(iv) We have: $h^T\tilde{y} = h^T(U^T y) = (Uh)^T y = (\lambda^{(h)})^T y$ (using (19)). □

Theorem 14. *The following assertions are equivalent for $y \in \mathbb{R}^{|\mathcal{B}|}$ and its extension $\tilde{y} \in \mathbb{R}^{\mathbb{Z}_+^n}$.*

(i) *The vector y belongs to the cone generated by the vectors $\zeta_v^{\mathcal{B}} = (f_i(v))_{i=1}^N$ $(v \in F)$.*
(ii) *$M_{\mathcal{B}}(y) \succeq 0$, $M_{\mathcal{B}}(h_j * y) \succeq 0$ $(j = m+1,\dots,m+k)$.*
(ii) *$M(\tilde{y}) \succeq 0$, $M(h_j * \tilde{y}) = 0$ $(j = 1,\dots,m)$, $M(h_j * \tilde{y}) \succeq 0$ $(j = m+1,\dots,m+k)$.*
(iv) *The vector $\tilde{y}$ belongs to the cone generated by the vectors $\zeta_v = (v^\alpha)_{\alpha\in\mathbb{Z}_+^n}$ $(v \in F)$.*

Proof. (i) $\Longrightarrow$ (ii) To see it, let $y := \zeta_v^{\mathcal{B}}$ for $v \in F$. Then, $M_{\mathcal{B}}(y) = yy^T \succeq 0$. Indeed, the (f_i, f_j)th entry of yy^T is $f_i(v) f_j(v)$, while the (f_i, f_j)th entry of $M_{\mathcal{B}}(y)$ is equal to $\sum_{k=1}^N \lambda_k^{(f_i f_j)} f_k(v)$ and thus to $f_i(v) f_j(v)$, since $f_i f_j \equiv \sum_{k=1}^N \lambda_k^{(f_i f_j)} f_k$ modulo I. Moreover, $M_{\mathcal{B}}(h_j * y) = h_j(v) yy^T \succeq 0$, for $j \geq m+1$.
(ii) $\Longrightarrow$ (iii) As $M(\tilde{y}) = U^T M_{\mathcal{B}}(y) U$, it follows that $M(\tilde{y}) \succeq 0$. For $j = 1,\dots,m$, $h_j * \tilde{y} = M(\tilde{y}) h_j = 0$, since $I \subseteq \operatorname{Ker} M(\tilde{y})$ (by Lemma 13 (iii)). For $j = m+1,\dots .m+k$, $M(h_j * \tilde{y}) = M(\widetilde{h_j * y}) = U^T M_{\mathcal{B}}(h_j * y) U \succeq 0$ (use Lemma 13 (i),(ii)).

(iii) $\Longrightarrow$ (iv) follows from Proposition 8.
(iv) $\Longrightarrow$ (i) Say, $\tilde{y} = \sum_{v \in F} a_v \zeta_v$ with $a_v \geq 0$. Let $i = 1, \ldots, N$. As $\tilde{y} = U^T y$, $f_i^T \tilde{y} = f_i^T(U^T y) = (Uf_i)^T y = y_i$. On the other hand, as $f_i^T \zeta_v = f_i(v) = (\zeta_v^{\mathcal{B}})_i$, $f_i^T \tilde{y} = \sum_{v \in F} a_v f_i(v)$ is the ith coordinate of $\sum_{v \in F} a_v \zeta_v^{\mathcal{B}}$. Hence, $y = \sum_{v \in F} a_v \zeta_v^{\mathcal{B}}$. □

Corollary 15. *Assume that V is finite and let $\mathcal{B}$ be a monomial basis of $\mathbb{R}[x_1, \ldots, x_n]/I$ containing the constant monomial* 1. *Then, problem (1) is equivalent to*

$$\min\ r^T y \ \ s.t.\ M_{\mathcal{B}}(y) \succeq 0,\ M_{\mathcal{B}}(h_j * y) \succeq 0\ (j = m+1, \ldots, m+k),\ y_0 = 1, \quad (22)$$

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

Proof. Directly from Theorem 14 since, for $y = \sum_v a_v \zeta_v^{\mathcal{B}}$, $\sum_v a_v = 1 \Longleftrightarrow y_0 = 1$. □

We have formulated for simplicity the above corollary for a monomial basis although it holds for any basis containing the constant polynomial 1. Therefore, (22) provides a concrete finite semidefinite program permitting to solve the infinite program (10). We now consider the dual semidefinite program of the semidefinite program (22). For convenience, the same symbol $\mathcal{B}$ denotes the set of exponents β for which $x^\beta \in \mathcal{B}$ and, for a polynomial $q(x)$, the notation: $q \in \mathbb{R}^{\mathcal{B}}$ means that $q(x)$ is a linear combination of monomials in $\mathcal{B}$, i.e., $q(x) = \sum_{\beta \in \mathcal{B}} q_\beta x^\beta$. Setting $h_0(x) := 1$ and $M_{\mathcal{B}}(h_j * y) = \sum_{\beta \in \mathcal{B}} C_\beta^j y_\beta$ for $j = 0, 1, \ldots, m+k$, the dual semidefinite program of (22) reads:

$$\begin{aligned} \rho^* := \ & \sup r_0 - \langle C_0^0, Z_0 \rangle - \sum_{j=m+1}^{m+k} \langle C_0^j, Z_j \rangle \\ & \text{s.t. } \langle C_\beta^0, Z_0 \rangle + \sum_{j=m+1}^{m+k} \langle C_\beta^j, Z_j \rangle = r_\beta \ (\beta \in \mathcal{B} \setminus \{0\}) \\ & Z_0, Z_{m+1}, \ldots, Z_{m+k} \succeq 0. \end{aligned} \quad (23)$$

One can verify that (23) is equivalent to

$$\begin{aligned} \rho^* = \ & \sup\ \rho \\ & \text{s.t. } r(x) - \rho = (\sum_{i_0} q_{0,i_0}^2) + \sum_{j=m+1}^{m+k} h_j(\sum_{i_j} q_{j,i_j}^2)\ + q \\ & \text{where all } q_{j,i_j} \text{ belong to } \mathbb{R}^{\mathcal{B}} \text{ and } q \in I \end{aligned}$$

and thus to the program: $\rho^* = \sup\ \rho$ s.t. $f(x) - \rho \in \mathcal{M}(F)$. (The latter equivalence follows using the fact that any sum of squares: $s = \sum_\ell p_\ell^2$ can be written as $s = \sum_\ell r_\ell^2 + q$, where $r_\ell \in \mathbb{R}^{\mathcal{B}}$ and $q \in I$, by replacing p_ℓ by its residue r_ℓ modulo I w.r.t. $\mathcal{B}$.) This program being identical to (23), we find that $\sigma^* = \rho^*$. By (13), $\rho^* = p^*$ and thus there is no duality gap between (22) and its dual (23).

Corollary 16. *Assume that V is finite and let $\mathcal{B}$ be a monomial basis of $\mathbb{R}[x_1, \ldots, x_n]/I$ containing* 1. *Then the programs (1), (22) and (23) are equivalent.*

2.2. *The radical case and the link with the Hermite's form*

When the ideal I is radical, one can prove the equivalence of (i) and (ii) in Theorem 14 without using the result of Curto and Fialkow. A first alternative is to use Parrilo's result (Theorem 11). A second alternative is to extend some arguments used in [14], [15] in the 0/1 and ±1 cases; it goes as follows. Given a basis $\mathcal{B} = \{f_1, \dots, f_N\}$ of $\mathbb{R}[x_1, \dots, x_n]/I$, let $Z_{\mathcal{B}}$ denote the complex $|\mathcal{B}| \times |V|$ matrix with columns $\zeta_v^{\mathcal{B}} = (f_i(v))_{i=1}^N$ $(v \in V)$.

In the $0/1$ case, when $\mathcal{B}$ is the basis from (21), the matrix $Z_{\mathcal{B}}$ is known as the Zeta matrix of the lattice of subsets of the set $\{1, \dots, n\}$ and the inverse matrix $Z_{\mathcal{B}}^{-1}$ as its Möbius matrix. By analogy, in the general case, we may call $Z_{\mathcal{B}}$ the *Zeta matrix of the ideal I* w.r.t. $\mathcal{B}$.

When I is radical, $Z_{\mathcal{B}}$ is nonsingular and the residue of a polynomial $f \in \mathbb{R}[x_1, \dots, x_n]$ w.r.t. $\mathcal{B}$ is the polynomial $\sum_{i=1}^N \lambda_i^{(f)} f_i$, where $\lambda^{(f)} = ((Z_{\mathcal{B}})^T)^{-1}(f(v))_{v\in V}$. The polytope $P_{\mathcal{B}}(F)$ from (18) is the convex hull of the columns of $Z_{\mathcal{B}}$. Thus, in the radical case,

$$P_{\mathcal{B}}(F) = \{y \in \mathbb{R}^{|\mathcal{B}|} \mid (Z_{\mathcal{B}}^{-1} y)_v \geq 0\,(v \in F),\ (Z_{\mathcal{B}}^{-1} y)_v = 0\,(v \in V\backslash F),\ e^T Z_{\mathcal{B}}^{-1} y = 1\}.$$

The next result shows how to express the combinatorial moment matrix $M_{\mathcal{B}}(y)$ in terms of the vector $Z_{\mathcal{B}}^{-1} y$.

Lemma 17. *Assume that I is radical. Then, for $y \in \mathbb{R}^{|\mathcal{B}|}$, $M_{\mathcal{B}}(y) = Z_{\mathcal{B}} \mathrm{diag}(Z_{\mathcal{B}}^{-1} y)\,(Z_{\mathcal{B}})^T$.*

Proof. Given $f_i, f_j \in \mathcal{B}$, the (f_i, f_j)th entry of the matrix $Z_{\mathcal{B}}\mathrm{diag}(Z_{\mathcal{B}}^{-1} y)(Z_{\mathcal{B}})^T$ is equal to $\sum_{v\in V} f_i(v) f_j(v) (Z_{\mathcal{B}}^{-1} y)_v = (Z_{\mathcal{B}}^{-1} y)^T (f_i(v) f_j(v))_{v\in V} = y^T (Z_{\mathcal{B}}^{-1})^T (f_i(v) f_j(v))_{v\in V}$, which in turn is equal to $y^T \lambda^{(f_i f_j)}$ and thus to the (f_i, f_j)th entry of $M_{\mathcal{B}}(y)$. □

When $V \subseteq \mathbb{R}^n$, $Z_{\mathcal{B}}$ and $Z_{\mathcal{B}}^{-1} y$ are real valued and thus $M_{\mathcal{B}}(y) \succeq 0 \Longleftrightarrow Z_{\mathcal{B}}^{-1} y \geq 0$. When V contains complex points, we can apply (7) and conclude that $M_{\mathcal{B}}(y) \succeq 0 \Longleftrightarrow (Z_{\mathcal{B}}^{-1} y)_v \geq 0$ $(v \in V \cap \mathbb{R}^n)$ and $(Z_{\mathcal{B}}^{-1} y)_v = 0$ $(v \in V \setminus \mathbb{R}^n)$. Therefore, Theorem 14 holds.

Comment. An analogous technique is used in [17] for proving Theorem 9 (and thus Proposition 8), which can be sketched as follows. As $M(y) \succeq 0$, its kernel $I :=$ Ker $M(y)$ is a *radical* ideal in $\mathbb{R}[x_1, \dots, x_n]$ and, as rank $M(y) < \infty$, I is zero-dimensional and thus its variety V is finite. The proof relies then on the identity: $M(y) = Z^T \mathrm{diag}(\tilde{Z} y) Z$, analogue to Lemma 17. Here, Z is the $|\mathbb{Z}_+^n| \times |V|$ matrix with columns $\zeta_v = (v^\alpha)_{\alpha\in\mathbb{Z}_+^n}$ $(v \in V)$, and $\tilde{Z}$ is the $|V| \times |\mathbb{Z}_+^n|$ matrix with rows p_v $(v \in V)$, where p_v are interpolation polynomials at the points of V (see Section 1.1).

Link between combinatorial moment matrices and Hermite's forms. As observed at the end of Section 1.2, the class of Hermite's matrices coincides with the class of matrices $H = Z_{\mathcal{B}} \mathrm{diag}(a)(Z_{\mathcal{B}})^T = \sum_{v\in V} a_v \zeta_v^{\mathcal{B}} (\zeta_v^{\mathcal{B}})^T$, where $a \in \mathbb{C}^V$ satisfies: $a_{\overline{v}} = \overline{a_v}$ for $v \in V$. Hence, any Hermite's matrix H is a combinatorial moment matrix; namely, $H = M_{\mathcal{B}}(y)$, after setting $y := \sum_{v\in V} a_v \zeta_v^{\mathcal{B}}$. Conversely, a combinatorial moment matrix $M_{\mathcal{B}}(y)$ is a Hermite's matrix in any of the following two cases: if $M_{\mathcal{B}}(y) \succeq 0$ (by Theorem 14), or if I is radical (by Lemma 17). On the other hand, if I is not radical, we give in the next example an instance $y \in \mathbb{R}^{|\mathcal{B}|}$ for which the matrix $M_{\mathcal{B}}(y)$ is not a Hermite's matrix.

Example 18. Let I be the ideal in $\mathbb{R}[x]$ generated by $h(x) := x^2$. Then, $\mathcal{B} = \{1, x\}$ is a basis of $\mathbb{R}[x]/I$ and $V(I) = \{v := 0\}$. For $y = (y_0, y_1)^T \in \mathbb{R}^{|\mathcal{B}|}$, we have: $M_{\mathcal{B}}(y) = \begin{array}{c} \\ 1 \\ x \end{array}\!\!\begin{array}{c} \begin{array}{cc} 1 & x \end{array} \\ \begin{pmatrix} y_0 & y_1 \\ y_1 & 0 \end{pmatrix} \end{array}$. Hence, $M_{\mathcal{B}}(y) \succeq 0$ implies that $y_1 = 0$ and thus $M_{\mathcal{B}}(y) = y_0 \zeta_v^{\mathcal{B}} (\zeta_v^{\mathcal{B}})^T$. When $y_1 \neq 0$, $M_{\mathcal{B}}(y)$ is not a multiple of $\zeta_v^{\mathcal{B}} (\zeta_v^{\mathcal{B}})^T$ and thus it is not a Hermite's matrix.

Consider now the problem of computing $p^* = \min\ x$ s.t. $x^2 = 0$. Obviously, $p^* = 0$. By Corollary 16, $0 = \min y_1$ s.t. $M_{\mathcal{B}}(y) \succeq 0,\ y_0 = 1$ and $0 = \rho^*$. Indeed, for any $\epsilon > 0$, $x + \epsilon = \left(\sqrt{\epsilon} + \frac{1}{2\sqrt{\epsilon}}x\right)^2 - \frac{1}{4\epsilon}x^2$, which shows that $\rho^* \geq -\epsilon$. This also shows that the relaxation (16) is exact at order $t = 1$; that is, $\sigma_1^* = p^* = 0$. Observe that the program (23) does not attain its optimum, since the polynomial x cannot be written as $u_0 + ux^2$, where $u_0, u \in \mathbb{R}[x]$ and u_0 is a s.o.s. This also shows that Parrilo's result in Theorem 11 does not hold when I is not radical.

We conclude with an example illustrating the notions introduced in this section.

Example 19 ([5], p. 229*).* Consider an optimization problem over the set

$$\begin{aligned} F := \{(x, y) \mid h_1(x) = x^2 + y - 1 = 0,\ h_2(x) \\ = xy - 2y^2 + 2y = 0,\ h(x) = x^2 - y + 1/2 \geq 0\}. \end{aligned}$$

Let I be the ideal in $\mathbb{R}[x, y]$ generated by h_1, h_2. W.r.t. the lexicographic order with $x > y$, the polynomials h_1, h_2 and $h_3(x) = y^3 - \frac{7}{4}y^2 + \frac{3}{4}y$ form a Groebner basis of I. Hence, the set of standard monomials is $\mathcal{B} = \{1, x, y, y^2\}$, $V = S = \{v_1 = (1, 0), v_2 = (-1, 0), v_3 = (0, 1), v_4 = (-\frac{1}{2}, \frac{3}{4})\}$, and $F = V \setminus \{v_3\}$. The Zeta matrix w.r.t. $\mathcal{B}$ and its inverse read:

$$Z_{\mathcal{B}} = \begin{array}{c|cccc} & v_1 & v_2 & v_3 & v_4 \\ \hline 1 & 1 & 1 & 1 & 1 \\ x & 1 & -1 & 0 & -\frac{1}{2} \\ y & 0 & 0 & 1 & \frac{3}{4} \\ y^2 & 0 & 0 & 1 & \frac{9}{16} \end{array}, \quad Z_{\mathcal{B}}^{-1} = \begin{array}{c|cccc} & 1 & x & y & y^2 \\ \hline v_1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{6} & -\frac{2}{3} \\ v_2 & \frac{1}{2} & -\frac{1}{2} & -\frac{5}{2} & 2 \\ v_3 & 0 & 0 & -3 & 4 \\ v_4 & 0 & 0 & \frac{16}{3} & -\frac{16}{3} \end{array}$$

and the multiplication table in $\mathbb{R}[x, y]/I$ is

$$\begin{array}{c|cccc} & 1 & x & y & y^2 \\ \hline 1 & 1 & x & y & y^2 \\ x & x & 1 - y & 2y^2 - 2y & \frac{3}{2}y^2 - \frac{3}{2}y \\ y & y & 2y^2 - 2y & y^2 & \frac{7}{4}y^2 - \frac{3}{4}y \\ y^2 & y^2 & \frac{3}{2}y^2 - \frac{3}{2}y & \frac{7}{4}y^2 - \frac{3}{4}y & \frac{37}{16}y^2 - \frac{21}{16}y \end{array}.$$

The ‘multiplication by h’ matrix reads:

$$M_h = \begin{array}{c} \\ 1 \\ x \\ y \\ y^2 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & x & y & y^2 \end{array} \\ \begin{pmatrix} 3/2 & 0 & 0 & 0 \\ 0 & 3/2 & 0 & 0 \\ -2 & 4 & 3/2 & 3/2 \\ 0 & -4 & -2 & -2 \end{pmatrix} \end{array}.$$

Therefore, a vector $u = (1, a, b, c) \in \mathbb{R}^{\mathcal{B}}$ belongs to the polytope $P_{\mathcal{B}}(F)$, the convex hull of the columns of $Z_{\mathcal{B}}$ indexed by F, if and only if $M_{\mathcal{B}}(u) \succeq 0$ and $M_{\mathcal{B}}(h * u) \succeq 0$, where $h * u = M_h^T u = (3/2 - 2b, 3a/2 + 4b - 4c, 3b/2 - 2c, 3b/2 - 2c)^T$. We find again that $h * u = M_h^T u = M_{\mathcal{B}}(u)\lambda^{(h)}$, where $\lambda^{(h)} = -2y + 3/2$ is the residue of h. Recall that

$$M_{\mathcal{B}}(u) = \begin{array}{c|cccc} & 1 & x & y & y^2 \\ \hline 1 & 1 & a & b & c \\ x & a & 1-b & 2c-2b & \frac{3}{2}c - \frac{3}{2}b \\ y & b & 2c-2b & c & \frac{7}{4}c - \frac{3}{4}b \\ y^2 & c & \frac{3}{2}c - \frac{3}{2}b & \frac{7}{4}c - \frac{3}{4}b & \frac{37}{16}c - \frac{21}{16}b \end{array} \succeq 0 \iff \begin{cases} \frac{1}{2} + \frac{1}{2}a + \frac{1}{6}b - \frac{2}{3}c \geq 0 \\ \frac{1}{2} - \frac{1}{2}a - \frac{5}{2}b + 2c \geq 0 \\ -3b + 4c \geq 0 \\ \frac{16}{3}b - \frac{16}{3}c \geq 0. \end{cases}$$

3. Finite Convergence of Lasserre’s Hierarchy

As mentioned in Section 1.3, there is finite convergence of the Lasserre hierarchies (15) and (16) when the ideal I is radical. We extend here the finite convergence result for (15) to the non-radical case and the finite convergence result for (16) to the case when $h_1, \dots, h_m$ form a Groebner basis of I (w.r.t. some monomial ordering). In that case we can also prove estimates on the order t for which $\sigma_t^* = \mu_t^* = p^*$; in the grid case, these estimates are sharper than those given in [12, 13]. We use the following result of Curto and Fialkow [7].

Theorem 20. [7] *Given $y \in \mathbb{R}^{S_{2t}}$, assume that $M_t(y) \succeq 0$ and that $M_t(y)$ is a flat extension of $M_{t-1}(y)$. Then, y is the sequence of moments of a nonnegative measure.*

Hence a strategy for proving equivalence of the programs (1) and (15) is to show that $M_t(y)$ is a flat extension of $M_{d_{\mathcal{B}}}(y)$, where $d_{\mathcal{B}}$ is the maximum degree of a monomial in a monomial basis $\mathcal{B}$ of $\mathbb{R}[x_1, \dots, x_n]/I$, which allows us to apply Theorem 20. A way to achieve this is to show that, for all $\alpha \in S_t$, the polynomials $f^{(\alpha)}(x) := x^\alpha - \lambda^{(\alpha)}(x)$, where $\lambda^{(\alpha)}(x)$ is the residue modulo I of x^α w.r.t. $\mathcal{B}$, belong to the kernel of $M_t(y)$. For this, the following property of the kernel of moment matrices will be useful; it comes from [7] and its proof is included for completeness.

Lemma 21. *Assume that $M_t(y) \succeq 0$ and let f, g be two polynomials whose product $h = fg$ has degree* $\deg(h) \leq t - 1$. *Then, $M_t(y)f = 0$ implies $M_t(y)h = 0$.*

Proof. It suffices to show the result for $g(x) = x_i$ since the general result follows from repeated applications of this special case. Assume $\deg(f) \leq t - 2$. We have: $h(x) = \sum_\alpha f_\alpha x^{\alpha+e_i} = \sum_{\alpha|\alpha \geq e_i} f_{\alpha-e_i} x^\alpha$. As $\deg(h) \leq t - 1$ and $M_t(y) \succeq 0$, $M_t(y)h = 0$ holds if $(M_t(y)h)_\alpha = 0$ for all $\alpha \in S_{t-1}$. Let $\alpha \in S_{t-1}$. The αth component of $M_t(y)h$ is equal to $\sum_\gamma h_\gamma y_{\alpha+\gamma} = \sum_{\gamma|\gamma \geq e_i} f_{\gamma-e_i} y_{\alpha+\gamma} = \sum_\gamma f_\gamma y_{\alpha+\gamma+e_i}$, which is equal to the $(\alpha + e_i)$th component of $M_t(y)f$ and thus to zero. □

Recall the definition of d_j, d from (14) and define $d_0 := \max(d_1, \ldots, d_m)$, $d_+ := \max(d_{m+1}, \ldots, d_{m+k})$, $d_+ = 0$ if $k = 0$; thus, $d = \max(d_0, d_+)$.

Theorem 22. *Assume that V is finite. There is finite convergence of the Lasserre hierarchy (15); that is, $\mu_t^* = p^*$ for t large enough.*

Proof. Let $\mathcal{B}$ be a monomial basis of $\mathbb{R}[x_1, \ldots, x_n]/I$ and let $d_{\mathcal{B}}$ be the maximum degree of a monomial in $\mathcal{B}$. Set $t_0 := \max(\lceil \deg(f)/2 \rceil, d_{\mathcal{B}} + 1, d_{\mathcal{B}} + d_+, 2d_0)$. Given $\alpha \in S_{t_0}$, let $\lambda^{(\alpha)}(x)$ be the residue of x^α modulo I w.r.t. $\mathcal{B}$. Then, $f^{(\alpha)}(x) = x^\alpha - \lambda^{(\alpha)}(x)$ belongs to I. Say, $f^{(\alpha)}(x) = \sum_{j=1}^m u_j^{(\alpha)} h_j$, where $u_j^{(\alpha)} \in \mathbb{R}[x_1, \ldots, x_n]$. Next, define $T_0 := \max(t_0, 1 + \deg(u_j^{(\alpha)} h_j)$ for $j = 1, \ldots, m$, $\alpha \in S_{t_0})$. We now show that $\mu_t^* = p^*$ for $t \geq T_0$. Fix $t \geq T_0$ and let $y \in S_{2t}$ be a feasible solution to (15). As $t \geq 2d_0$, $M_{t-d_j}(h_j * y) = 0$ implies that $M_t(y)h_j = 0$ for all $j = 1, \ldots, m$. Therefore, $M_t(y) f^{(\alpha)} = 0$ for all $\alpha \in S_{t_0}$ (we can apply Lemma 21, since $\deg(u_j^{(\alpha)} h_j) \leq T_0 - 1 \leq t - 1$ for all j). This implies that $M_{t_0}(y)$ is a flat extension of $M_{d_{\mathcal{B}}}(y)$ (in fact, of its submatrix indexed by $\mathcal{B}$). As $t_0 \geq d_{\mathcal{B}} + 1$, we deduce from Theorem 20 that $(y_\alpha)_{|\alpha| \leq 2t_0}$ has a representing measure μ. Now, for $j = 1, \ldots, m$, $M_t(y)h_j = 0$ implies that $M_{t_0}(y)h_j = 0$ since $t_0 \geq 2d_0$. Hence, the support of μ is contained in $S = V \cap \mathbb{R}^n$. Say, $\mu = \sum_{v \in S} a_v \delta_v$ where $a_v \geq 0$. For $v \in S \setminus F$, let $p \in \mathbb{R}^{\mathcal{B}}$ be a polynomial such that $p(v) = 1$ and $p(v') = 0$ for $v' \in S \setminus \{v\}$, and let $j \geq m + 1$ for which $h_j(v) < 0$. Then, $p^T M_{t_0 - d_j}(h_j * y) p = \int p(x)^2 h_j(x) \mu(dx) = a_v h_j(v) \geq 0$, which implies that $a_v = 0$ (we have used here the fact that $t_0 - d_j \geq d_{\mathcal{B}}$ as $t_0 \geq d_{\mathcal{B}} + d_+$). Hence, the support of μ is contained in F. Therefore, $f^T y = \sum_{v \in F} a_v f(v) \geq p^*$, which implies that $\mu_t^* \geq p^*$ and thus $\mu_t^* = p^*$. □

Theorem 23. *Assume that V is finite and that $h_1, \ldots, h_m$ form a Groebner basis of I (w.r.t. some monomial ordering) and let $d_{\mathcal{B}}$ be the maximum degree of a polynomial in a basis $\mathcal{B}$ of $\mathbb{R}[x_1, \ldots, x_n]/I$. For $t \geq \max(d_{\mathcal{B}} + d_+, d_0, \lceil \deg(f)/2 \rceil)$, $\sigma_t^* = \mu_t^* = p^*$.*

Proof. Fix $t \geq \max(d_{\mathcal{B}} + d_+, d_0, \lceil \deg(f)/2 \rceil)$. We show that, for any $\epsilon > 0$, $\sigma_t^* \geq p^* - \epsilon$, which implies $\sigma_t^* \geq p^*$ and thus $\sigma_t^* = p^*$. As the polynomial $f(x) - p^* + \epsilon$ is positive on F, Theorem 10 implies that $f(x) - p^* + \epsilon \in \mathcal{M}(F)$; that is, $f(x) - p^* + \epsilon = s_0 + \sum_{j=m+1}^{m+k} s_j h_j + q$, where s_j are s.o.s. and $q \in I$. Such a decomposition exists where each s_j is a sum of squares of polynomials that are linear combinations of members of $\mathcal{B}$. Hence, $\deg(s_0)$, $\deg(s_j h_j) \leq 2(d_{\mathcal{B}} + d_+) \leq 2t$, implying $\deg(q) \leq 2t$. As $h_1, \ldots, h_m$ is a Groebner basis of I, one can find (using the division algorithm) a decomposition $q = \sum_{j=1}^m u_j h_j$, where $\deg(u_j h_j) \leq \deg(q) \leq 2t$. Therefore, $p^* - \epsilon$ is feasible for the program (16), which implies that $\sigma_t^* \geq p^* - \epsilon$. □

Therefore, one obtains a tighter bound on the order t at which finite convergence takes place if one can find a basis $\mathcal{B}$ of $\mathbb{R}[x_1, \ldots, x_n]/I$ with small maximum degree $d_{\mathcal{B}}$. For instance, we have seen in Example 3 two instances of bases with respective maximum degrees $d_{\mathcal{B}} = 4$ and 11.

Example 24. **The grid case.** In the 0/1 case, Theorem 23 gives the finite convergence result in $d_{\mathcal{B}} + d_+ = n + d_+$ steps from [12]. Consider now the general grid case as in [13]. For $j = 1, \ldots, n$, let $a_1^{(j)}, \ldots, a_{m_j}^{(j)}$ be distinct real numbers and define the polynomial $h_j(x) = \prod_{i=1}^{m_j} (x_j - a_i^{(j)})$ of degree m_j. W.r.t. the graded lexicographic order,

$h_1, \dots, h_n$ form a Groebner basis of I and the set $\mathcal{B}$ of standard monomials consists of the monomials x^β with $0 \le \beta_j \le m_j - 1$ for all j; thus, $d_\mathcal{B} = \sum_{j=1}^n (m_j - 1)$. Here, $d = d_0 = \max_{j=1,\dots,n} \lceil m_j/2 \rceil$. Note that the ideal I is radical, since $|V| = |\mathcal{B}| = \prod_{j=1}^n m_j$. Lasserre [13] shows the finite convergence of σ_t^* to p^* in $d + d_\mathcal{B}$ steps; Theorem 23 shows the finite convergence in $\max(d_\mathcal{B}, d)$ steps (assuming the polynomial f in the objective has degree at most $\max(2d_\mathcal{B}, 2d)$).

4. Semidefinite approximations of low order

From a practical point of view, the semidefinite formulation (22) for problem (1) is not very useful, since it involves matrices of size $|\mathcal{B}| \times |\mathcal{B}|$, with $|\mathcal{B}| \ge |S|$ being at least as large as the size of the set of points to be searched. One can instead consider semidefinite approximations of the problem (1), obtained by restricting our attention to some principal submatrix $M_\mathcal{A}(y)$ of the combinatorial moment matrix $M_\mathcal{B}(y)$ indexed by a small subset $\mathcal{A}$ of $\mathcal{B}$. For instance, when the polynomial f in the objective function is linear or quadratic and $V = \{0, 1\}^n$ or $\{\pm 1\}^n$ (the most common case in combinatorial applications), one can choose $\mathcal{A} = \mathcal{B}_t$ consisting of the residues modulo I of all the monomials x^α for $\alpha \in S_t$, for any given $t \ge 1$. Increasing values of t yield tighter approximations of (1) at an increasing computational cost, however. The approximation solves problem (1) exactly at $t = n$, but this may sometimes already happen for smaller values of t. We present below some conditions on the rank of the optimum matrix $M_{\mathcal{B}_t}(y)$ ensuring that the corresponding semidefinite relaxation solves, in fact, problem (1) exactly. We give our result in the case when

$$S = \{x \in \mathbb{R}^n \mid (x_j - a_j)(x_j - b_j) = 0 \ \text{ for all } j = 1, \dots, n\}, \tag{24}$$

where $a_1 \ne b_1, \dots, a_n \ne b_n$ are given real numbers. Therefore, this contains both the $0/1$ and ± 1 cases, which are very important in combinatorial optimization. Define the polynomials

$$h_j(x) = (x_j - a_j)(x_j - b_j) = x_j^2 - s_j x_j - t_j, \ \text{ setting } s_j := a_j + b_j,\ t_j := -a_j b_j \tag{25}$$

for $j = 1, \dots, n$. Let I be the ideal generated by $h_1, \dots, h_n$ and $S = V$ the associated variety, given by (24). W.r.t. the graded lexicographic order, the set of standard monomials is the set $\mathcal{B}$ (also denoted as $\mathcal{B}(n)$) from (21). Given an integer $t = 1, \dots, n$, $\mathcal{B}_t$ (also denoted as $\mathcal{B}_t(n)$) consists of the standard monomials $x^A = \prod_{i \in A} x_i$ having degree $|A| \le t$ and can be identified with the collection of subsets of the set $\mathcal{N} := \{1, \dots, n\}$ having size $\le t$. Given $y \in \mathbb{R}^{\mathcal{B}_{2t}}$, we can define its truncated combinatorial moment matrix $M_{\mathcal{B}_t}(y)$ as the matrix indexed by $\mathcal{B}_t$, whose (A, B)th entry is equal to $r^T y$, where r is the residue modulo I of $x^A x^B$, and thus to

$$\sum_{C \subseteq A \cap B} \left(\prod_{i \in C} s_i \prod_{i \in (A \cap B) \setminus C} t_i \right) y_{(A \Delta B) \cup C}. \tag{26}$$

Indeed, $x^A x^B = x^{A\Delta B}(x^{A\cap B})^2 \equiv x^{A\Delta B} \cdot \prod_{i\in A\cap B}(s_i x_i + t_i)$ (modulo I), which is equal to $x^{A\Delta B}\left(\sum_{C\subseteq A\cap B}\prod_{i\in C}(s_i x_i)\prod_{i\in(A\cap B)\setminus C} t_i\right) = \sum_{C\subseteq A\cap B}\left(\prod_{i\in C} s_i \prod_{i\in(A\cap B)\setminus C} t_i\right)$ $x^{(A\Delta B)\cup C}$. In the 0/1 case, $a_i = 0$, $b_i = 1$, $s_i = 1$, $t_i = 0$ for all i and thus (26) reads $y_{A\cup B}$; in the ± 1 case, $a_i = -1$, $b_i = 1$, $s_i = 0$, $t_i = 1$ for all i and thus (26) reads $y_{A\Delta B}$.

Theorem 25. *Let S and $h_1, \dots, h_n$ be as in (24) and (25). Let $y \in \mathbb{R}^{\mathcal{B}_{2t}(n)}$ with $y_0 = 1$ and $n \geq t \geq 1$. If $M_{\mathcal{B}_t}(y) \succeq 0$ and* rank $M_{\mathcal{B}_s}(y) \leq \sum_{i=1}^{s}\binom{t}{s}$ *for some $1 \leq s \leq t$, then y is the sequence of moments of a probability measure μ supported by S. If, moreover,* rank $M_{\mathcal{B}_1}(y) \leq t$, *then μ is supported by a subset of S of size at most 2^{t-1}.*

We first establish a stronger version of Lemma 21, valid in the combinatorial setting.

Lemma 26. *Let $n \geq t$, $y \in \mathbb{R}^{\mathcal{B}_{2t}(n)}$, assume that $M_{\mathcal{B}_t}(y) \succeq 0$, and let f, g be two polynomials in $\mathbb{R}^{\mathcal{B}_t}$ such that the residue h modulo I of their product fg satisfies* $\deg(h) \leq t$. *Then, $M_{\mathcal{B}_t}(y)f = 0$ implies $M_{\mathcal{B}_t}(y)h = 0$.*

Proof. It suffices to show the result for $g(x) = x_i$. Write $f(x) = \sum_{m\in M} f_m x^{A_m} + \sum_{\ell\in L} f_\ell x^{A_\ell}$, where the A_m (resp., A_ℓ) are distinct subsets of $\mathcal{N}$ containing (resp., not containing) the element i, and $f_m, f_\ell \neq 0$. Then, $x_i f(x) = \sum_{m\in M} f_m x^{A_m\setminus i} x_i^2 + \sum_{\ell\in L} f_\ell x^{A_\ell+i}$ is congruent modulo I to the polynomial $\sum_{m\in M} f_m x^{A_m\setminus i}(s_i x_i + t_i) + \sum_{\ell\in L} f_\ell x^{A_\ell+i}$. Therefore, $h(x) = s_i \sum_{m\in M} f_m x^{A_m} + t_i \sum_{m\in M} f_m x^{A_m\setminus i} + \sum_{\ell\in L} f_\ell x^{A_\ell+i}$. As $M_{\mathcal{B}_t}(y) \succeq 0$, $M_{\mathcal{B}_t}(y)h = 0$ holds if we can show that its components indexed by A_m, $A_m \setminus i$ $(m \in M)$ and $A_\ell + i$ $(\ell \in L)$ are equal to 0. Let X denote the principal submatrix of $M_{\mathcal{B}_t}(y)$ indexed by the sets A_m $(m \in M)$, $A_m \setminus i$ $(m \in M)$, and $A_\ell + i$ $(\ell \in L)$, and set $u := (f_m)_{m\in M}$, $v := (f_\ell)_{\ell\in L}$. We have to prove that $X\begin{pmatrix} s_i u \\ t_i u \\ v\end{pmatrix} = 0$. By assumption, $M_{\mathcal{B}_t}(y)f = 0$. Therefore, $Y\begin{pmatrix} u \\ v\end{pmatrix} = 0$, where Y is the submatrix of $M_{\mathcal{B}_t}(y)$ with columns indexed by A_m $(m \in M)$ and A_ℓ $(\ell \in L)$, and with rows indexed by A_m $(m \in M)$, $A_m \setminus i$ $(m \in M)$, A_ℓ $(\ell \in L)$, and $A_\ell + i$ $(\ell \in L)$. The matrices have block decompositions:

$$Y = \begin{array}{c|cc} & A_m & A_\ell \\ \hline A_m & C & D \\ A_m\setminus i & E & F \\ A_\ell & D^T & G \\ A_\ell + i & H^T & K \end{array}, \quad X = \begin{array}{c|ccc} & A_m & A_m\setminus i & A_\ell + i \\ \hline A_m & C & E & H \\ A_m\setminus i & E & R & S \\ A_\ell + i & H^T & S^T & T \end{array}.$$

Here are some equations relating the above matrices: (a) $S = D$; (b) $H = s_i D + t_i F$; (c) $C = s_i E + t_i R$; (d) $T = s_i K + t_i G$. To see this, we use the fact that the (A, B)th entry of $M_{\mathcal{B}_t}(y)$ is obtained by computing the residue modulo I of $x^A x^B$ and linearizing, which means replacing any occurrence of x^C by y_C. In this way, equation: $x^{A_m\setminus i} x^{A_\ell+i} = x^{A_m} x^{A_\ell}$ yields (a). Moreover, $x^{A_m} x^{A_\ell+i} = x^{A_m\setminus i} x^{A_\ell} x_i^2$ is congruent modulo I to $x^{A_m\setminus i} x^{A_\ell}(s_i x_i + t_i) = s_i x^{A_m} x^{A_\ell} + t_i x^{A_m\setminus i} x^{A_\ell}$, which yields (b). Finally, for $m, m' \in M$, $x^{A_m} x^{A_{m'}} \equiv x^{A_m\setminus i} x^{A_{m'}\setminus i}(s_i x_i + t_i)$ and, for $\ell, \ell' \in L$, $x^{A_\ell+i} x^{A_{\ell'}+i} \equiv$

$x^{A_\ell}x^{A_{\ell'}}(s_ix_i + t_i)$, yielding (c) and (d). We can now show that $X\begin{pmatrix} s_iu \\ t_iu \\ v\end{pmatrix} = 0$. Indeed, $s_iCu + t_iEu + Hv = s_iCu + t_iEu + (s_iD + t_iF)v$ (using (b)) which is equal to $s_i(Cu + Dv) + t_i(Eu + Fv)$ and thus to 0, using the fact that $Y\begin{pmatrix} u \\ v\end{pmatrix} = 0$. The identities: $s_iEu + t_iRu + Sv = 0$ and $s_iH^Tu + t_iS^Tu + Tv = 0$ follow analogously. □

Lemma 27. *Let $y \in \mathbb{R}^{\mathcal{B}(n)}$ and $\mathcal{B}(n) = \mathcal{B}$ as in (21). Assume that $M_{\mathcal{B}}(y) \succeq 0$ and* rank $M_{\mathcal{B}_s}(y) \le \sum_{i=1}^{s}\binom{n}{i}$ *for some $1 \le s \le n$. Then, the column of $M_{\mathcal{B}}(y)$ indexed by the set $\mathcal{N} = \{1, \ldots, n\}$ is a linear combination of the remaining columns.*

Proof. By assumption, rank $M_{\mathcal{B}_s}(y) < |\mathcal{B}_s|$. Therefore, there exists a nonzero vector u belonging to the kernel of $M_{\mathcal{B}_s}(y)$; let $u(x) = \sum_{I\in\mathcal{B}_s} u_Ix^I$ be the corresponding polynomial with degree $k \le s$ and let K with $u_K \ne 0$ and $|K| = k$. Multiplying $u(x)$ by $x^{\mathcal{N}\setminus K}$ yields the polynomial $u_Kx^{\mathcal{N}} + \sum_{I\in\mathcal{B}_s, I\ne K} u_Ix^Ix^{\mathcal{N}\setminus K}$, whose residue modulo I belongs to the kernel of $M_{\mathcal{B}}(y)$ (by Lemma 26, applied with $n = t$). The result now follows since, for each $I \ne K$ with $u_I \ne 0$, the residue modulo I of $x^Ix^{\mathcal{N}\setminus K}$ is a linear combination of monomials x^H with $H \subseteq I \cup (\mathcal{N} \setminus K)$ which is strictly contained in $\mathcal{N}$.□

Proof of Theorem 25. Suppose first that $t = 1$. Then, rank $M_{\mathcal{B}_1}(y) = 1$ by assumption. The submatrix of $M_{\mathcal{B}_1}(y)$ indexed by $\emptyset$ and i has the form $\begin{pmatrix} 1 & y_i \\ y_i & s_iy_i + t_i\end{pmatrix}$; as its rank is 1, this implies that $y_i^2 = s_iy_i + t_i$ and thus $y_i = a_i$ or b_i. The submatrix of $M_{\mathcal{B}_1}(y)$ indexed by $\emptyset, i, j$ has the form $\begin{pmatrix} 1 & y_i & y_j \\ y_i & y_i^2 & y_{ij} \\ y_j & y_{ij} & y_j^2\end{pmatrix}$; as it has rank 1, this implies that $y_{ij} = y_iy_j$. Therefore, $y = (v^A)_{A\in\mathcal{B}_2}$ for some $v \in S$ and the conclusion of Theorem 25 holds.

We can now assume that $t \ge 2$. Given a subset $A \subseteq \mathcal{N}$ with $|A| = t$, we deduce from Lemma 27 (applied (with $n = t$) to the principal submatrix of $M_{\mathcal{B}_t}(y)$ indexed by all subsets of A) that the column of $M_{\mathcal{B}_t}(y)$ indexed by A is a linear combination of columns indexed by sets of size $\le t - 1$. Therefore, $M_{\mathcal{B}_t}(y)$ is a flat extension of $M_{\mathcal{B}_{t-1}}(y)$. We can extend y to a vector of $\mathbb{R}^{S_{2t}}$, again denoted by y, in such a way that $M_t(y)$ is a flat extension of $M_{\mathcal{B}_t}(y)$. [Indeed, for $\alpha \in S_{2t}$, let $r(x) = \sum_{\beta\in\mathcal{B}} r_\beta x^\beta$ be the residue of x^α modulo I w.r.t. $\mathcal{B}$. As $r(x)$ uses only monomials of degree $\le 2t$, one can define $y_\alpha := \sum_{\beta\in\mathcal{B}_{2t}} r_\beta y_\beta$.] As $M_t(y)$ is a flat extension of $M_{t-1}(y)$, Theorem 20 implies that $(y_\alpha)_{\alpha\in S_{2t}}$ is the sequence of moments of a nonnegative measure μ. By construction, each polynomial $h_i(x) = x_i^2 - s_ix_i - t_i$ belongs to the kernel of $M_t(y)$ and, therefore, by Lemma 7 (i), the support of μ is contained in the grid S. Say, $\mu = \lambda_1\delta_{v^1} + \ldots + \lambda_L\delta_{v^L}$, where $\lambda_1, \ldots, \lambda_L > 0$ and $v^1, \ldots, v^L$ are distinct points of S.

Remains to show that $L \le 2^{t-1}$ in the case when rank $M_{\mathcal{B}_1}(y) \le t$. We have that $M_{\mathcal{B}_1}(y) = \sum_{\ell=1}^{L}\lambda_\ell z_\ell z_\ell^T$, setting $z_\ell := (1, v_1^\ell, \ldots, v_n^\ell)^T$. Let X denote the $L \times (n + 1)$ matrix with $z_1^T, \ldots, z_L^T$ as rows. Each component v_i^ℓ is equal to a_i or b_i for all ℓ. Therefore, the 0th column of X is the all ones vector χ^W (setting $W := \{1, \ldots, L\}$), and the ith column of X is of the form $a_i\chi^{W_i} + b_i\chi^{W\setminus W_i}$, for some set $W_i \subseteq W$. The two matrices $M_{\mathcal{B}_1}(y)$ and X have the same rank, which is also equal to the rank of the set of vectors $\{\chi^W, \chi^{W_1}, \dots, \chi^{W_n}\}$; this rank is at most t by assumption. The next claim permits to conclude that $L = |W| \leq 2^{t-1}$, which finishes the proof of Theorem 25. □

Claim 28. Let $A_1, \dots, A_L$ be subsets of $\{1, \dots, n\}$ and let Y be the $L \times (n+1)$ matrix whose rows are $(1, (\chi^{A_1})^T), \dots, (1, (\chi^{A_L})^T)$. If rank $Y \leq t$, then $L \leq 2^{t-1}$.

Proof of the Claim. The proof is by induction on $n \geq t$. Suppose first that $n = t$. As rank $Y \leq n$, there exists a nonzero vector $u \in \ker Y$. This vector can be chosen to be integral valued with at least one of $u_1, \dots, u_n$ odd. Let O denote the set of $i = 1, \dots, n$ for which u_i is odd. Then $|A_\ell \cap O| \equiv u_0$ (modulo 2) for all $\ell = 1, \dots, L$. This implies that $L \leq 2^{n-1}$. Suppose now that $n \geq t+1$ and that the claim holds for $n-1$. Set $L_0 := \{\ell \mid n \notin A_\ell\}$ and $L_1 := \{\ell \mid n \in A_\ell\}$; for $i = 0, 1$, let Y_i be the submatrix of Y with rows those indexed by L_i and Y_i' the submatrix of Y_i with columns those indexed by $0, 1, \dots, n-1$. If $L_0 = \emptyset$, then rank $Y_1' =$ rank $Y_1 \leq t$; by the induction assumption, this implies that $L = |L_1| \leq 2^{t-1}$; analogously, if $L_1 = \emptyset$. If $L_0, L_1 \neq \emptyset$, then rank Y_0, rank $Y_1 \leq t-1$, since no row of Y_0 (resp., of Y_1) is a linear combination of rows of Y_1 (resp., of Y_0). Using the induction assumption, we find that $|L_1|, |L_2| \leq 2^{t-2}$ and thus $L = |L_1| + |L_2| \leq 2^{t-1}$. □

In the case when $S = \{\pm 1\}^n$ and $s = 1$, the result from Theorem 25 was proven[1] in [15]; the proof given there is elementary (by induction on the number n of variables) and, in particular, it does not use Curto and Fialkow's result from Theorem 20. In our proof of Theorem 25, we have used the following fact:

Theorem 29. *Let S and $h_1, \dots, h_n$ be as in (24) and (25). Given $y \in \mathbb{R}^{\mathcal{B}_{2t}}$, if $M_{\mathcal{B}_t}(y)$ is a flat extension of $M_{\mathcal{B}_{t-1}}(y)$, then y is the sequence of moments (of order $\beta \in \mathcal{B}_{2t}$) of a nonnegative measure supported by S.*

This result can be proved as an application of Theorem 20. Alternatively, one can proceed as follows: First, show that y has an extension to $\mathbb{R}^{\mathcal{B}_{2t+2}}$, again denoted as y, for which the matrix $M_{\mathcal{B}_{t+1}}(y)$ is a flat extension of $M_{\mathcal{B}_t}(y)$, iterate and apply Theorem 14 to conclude.

The condition about the rank of $M_{\mathcal{B}_s}(y)$ in Theorem 25 is best possible. For instance, in the 0/1 case with $n = 2$ and $s = t = 1$, the matrix
$$M_{\mathcal{B}_1}(y) := \begin{array}{c|ccc} & \emptyset & 1 & 2 \\ \hline \emptyset & 1 & 3/4 & 3/4 \\ 1 & 3/4 & 3/4 & 3/8 \\ 2 & 3/4 & 3/8 & 3/4 \end{array}$$
is positive semidefinite with rank 2, but $y = (1, 3/4, 3/4, 3/8) \in \mathbb{R}^{\mathcal{B}_2}$ is not the sequence of moments of a nonnegative measure supported by $S = \{0, 1\}^2$. Indeed, the matrix
$$M_{\mathcal{B}_2}(y) = \begin{array}{c|cccc} & \emptyset & 1 & 2 & 12 \\ \hline \emptyset & 1 & 3/4 & 3/4 & 3/8 \\ 1 & 3/4 & 3/4 & 3/8 & 3/8 \\ 2 & 3/4 & 3/8 & 3/4 & 3/8 \\ 12 & 3/8 & 3/8 & 3/8 & 3/8 \end{array}$$
is not positive semidefinite, since the vector

[1] The paper [15] proves the result in the special case when the entries of y indexed by odd sets are all equal to 0 (because the paper is devoted to an application to the max-cut problem, in which the odd indexed entries do not play a role). The proof given there extends, however, to the general case when y is arbitrary.

$x := (-3, 2, 2)$ belongs to the kernel of $M_{\mathcal{B}_1}(y)$ while $(x, 0)$ does not belong to the kernel of $M_{\mathcal{B}_2}(y)$.

In combinatorial applications, the polynomial in the objective function of problem (1) is often linear. This is the case, e.g., for the maximum stable set problem, in which case we are, in fact, interested in finding some conditions on $M_{\mathcal{B}_t}(y)$ ensuring that the vector $(y_1, \ldots, y_n)^T$ (i.e., the projection of y on the space indexed by the singletons) can be written as a convex combination of stable sets. [The (A, B)th entry of $M_{\mathcal{B}_t}(y)$ being defined as $y_{A\cup B}$, as we are in the $0/1$ setting.] The following result has this flavor.

Proposition 30. *Let $G = (\mathcal{N}, E)$ be a graph with node set $\mathcal{N} = \{1, \ldots, n\}$. Let $y \in \mathbb{R}^{\mathcal{B}_2(n)}$ with $y_0 = 1$ and $y_{ij} = 0$ for all $ij \in E$. If $M_{\mathcal{B}_1}(y) \succeq 0$ and* rank *$M_{\mathcal{B}_1}(y) \le 2$, then $(y_1, \ldots, y_n)^T$ is a convex combination of incidence vectors of stable sets of G.*

Proof. We can assume w.l.o.g. that $y_i > 0$ for all $i \in V$ (if needed, delete the nodes with $y_i = 0$). We can also assume that $E \neq \emptyset$ (for, otherwise, the result is trivial). Let $C_0, C_1, \ldots, C_n$ denote the columns of $M_{\mathcal{B}_1}(y)$. Then,

(a) $\quad y_i + y_j = 1$ and $C_0 = C_i + C_j$ for any edge $ij \in E$.

Indeed, the submatrix of $M_{\mathcal{B}_1}(y)$ indexed by $\emptyset$, i and j has the form $\begin{pmatrix} 1 & y_i & y_j \\ y_i & y_i & 0 \\ y_j & 0 & y_j \end{pmatrix}$. As it has rank ≤ 2, its determinant is equal to 0, which implies that $y_i + y_j = 1$. Now the vector $(1, -1, -1)^T$ belongs to the kernel of this submatrix and thus to the kernel of $M_{\mathcal{B}_1}(y)$, whch shows (a). Denote by $V_1, \ldots, V_q$ the connected components of G. By (a), there exists $\alpha_p \in [0, 1]$ such that $y_i \in \{\alpha_p, 1 - \alpha_p\}$ for all $i \in V_p$, for each $p = 1, \ldots, q$. We claim that

(b) $\quad$ the subgraph of G induced by V_p is bipartite.

If (b) holds, then V_p can partitioned into two stable sets S_p and T_p such that $y_i = \alpha_p$ for $i \in S_p$, $y_i = 1 - \alpha_p$ for $i \in T_p$. Say, $\alpha_1 \le \ldots \le \alpha_p$. Then, the vector $(y_1, \ldots, y_n)^T$ is equal to $\sum_{p=1}^{q} \alpha_p \chi^{S_p} + (1 - \alpha_p)\chi^{T_p}$ and thus to $\alpha_1 \chi^{S_1\cup\ldots\cup S_p} + \sum_{q=1}^{p-1}(\alpha_{q+1} - \alpha_q)\chi^{T_1\cup\ldots T_q\cup S_{q+1}\cup\ldots\cup S_p} + (1 - \alpha_p)\chi^{T_1\cup\ldots\cup T_p}$. This concludes the proof since the latter is a convex combination of incidence vectors of stable sets.

Remains to verify (b). Suppose, for contradiction, that C is an odd circuit contained in V_p. Choose such circuit of minimum length; then C is chordless. Moreover, $y_i = \frac{1}{2}$ for all $i \in V_p$ (using (a) applied to the edges in C). Say, $C = (1, 2, \ldots, 2k+1)$. We have from (a) that $C_0 = C_i + C_{i+1}$ for $i = 1, \ldots, 2k + 1$ (taking indices modulo $2k + 1$). It is not difficult to see that this implies that y_{ij} is equal to $\frac{1}{2}$ (resp., to 0) if the distance between nodes i and j along the circuit C is even (resp., odd). Therefore, $y_{1,k+2} = y_{2,k+2}$ since the two pairs $(1, k + 2)$ and $(2, k + 2)$ are at the same distance along C. We now reach a contradiction, since $C_0 = C_1 + C_2$ implies that $\frac{1}{2} = y_{k+2} = y_{1,k+2} + y_{2,k+2}$. □

Therefore, if we add the condition that rank $M_{\mathcal{B}_1}(y) \le 2$ in the formulation (5) of the theta number $\vartheta(G)$, we obtain a program which is, in fact, equivalent to the maximum stable set problem. Another formulation for $\vartheta(G)$ is given by

$$\vartheta(G) = \max \sum_{i,j=1}^{n} X_{ij} \text{ s.t. } X \succeq 0,\ \sum_{i=1}^{n} X_{ii} = 1,\ X_{ij} = 0\ (ij \in E).$$

Burer, Monteiro and Zhang [4] proved that adding the constraint $\text{rank}(X) \leq 2$ in the above program yields again a formulation of the stable set problem. Thus, Proposition 30 is an analogue of their result.

Let us conclude with a question related to the max-cut problem. Given weights $w = (w_{ij})_{1\leq i<j\leq n}$, this is the problem: $\max_{x_1^2=1,\dots,x_n^2=1} \sum_{1\leq i<j\leq n} w_{ij}(1-x_ix_j)$. Therefore, it can be reformulated as

$$\max \sum_{1\leq i<j\leq n} w_{ij}(1-y_{ij}) \text{ s.t. } M_{\mathcal{B}}(y) \succeq 0,\ y_0 = 1,$$

where $M_{\mathcal{B}}(y)$ is the combinatorial moment matrix indexed by all subsets of $\{1,\dots,n\}$, with (A,B)th entry $y_{A\Delta B}$. (See [15] for a detailed treatment.) Heuristics for max-cut based on low rank formulations are explored in [3]. It would be of interest to determine the smallest integer t for which the semidefinite relaxation:

$$\max \sum_{1\leq i<j\leq n} w_{ij}(1-y_{ij}) \text{ s.t. } M_{\mathcal{B}_t}(y) \succeq 0,\ y_0 = 1$$

solves the max-cut problem exactly, for any weight function w. ($M_{\mathcal{B}_t}(y)$ being the truncation of $M_{\mathcal{B}}(y)$ indexed by all subsets of size at most t.) It is shown in [16] that $t \geq \lceil \frac{n}{2} \rceil$ and equality is conjectured (it holds for $n \leq 7$). In other words, it is conjectured that, for $t = \lceil \frac{n}{2} \rceil$,

$$M_{\mathcal{B}_t}(y) \succeq 0,\ y_0 = 1 \Longrightarrow (y_{ij})_{1\leq i<j\leq n} \text{ is a convex combination of the 'cut' vectors } (v_iv_j)_{1\leq i<j\leq n}\ (v \in \{\pm 1\}^n).$$

5. Concluding Remarks

In this paper, we have given a semidefinite representation for the problem (1) of computing the minimum value p^* taken by a polynomial over the set of real solutions to a system of polynomial equations and inequalities, assuming the equations have a finite set V of complex solutions. Our semidefinite representation involves combinatorial moment matrices $M_{\mathcal{B}}(y)$, which are indexed by a basis $\mathcal{B}$ of the quotient space $\mathbb{R}[x_1,\dots,x_n]/I$. We also show the finite convergence of the hierarchy (15) of semidefinite relaxations introduced by Lasserre [11] and, in the case when the polynomial equations form a Groebner basis of I, the finite convergence of the dual hierarchy (16). The matrices $M_{\mathcal{B}}(y)$ involved in the new semidefinite representation have size $|\mathcal{B}|$ which is usually much smaller than the size $|S_t|$ of the classical moment matrices $M_t(y)$ appearing in the program (15) for any t ensuring the finite convergence. In the non-radical case, $|\mathcal{B}| > |V|$. In order to give a semidefinite representation involving matrices of smaller size $|V|$, it suffices to replace the ideal I by the larger ideal $I(V)$, consisting of the polynomials vanishing at all points of V. One can find $I(V)$ by some Groebner bases computations (see, e.g., Proposition 2.7 in [6]). A more efficient alternative might be to use the fact that $I(V)$ is the radical $\{f \in \mathbb{R}[x_1,\dots,x_n] \mid her_h(f,g) = 0\ \forall g \in \mathbb{R}[x_1,\dots,x_n]\}$ of the Hermite's form Her_h, where h is the constant polynomial 1 (see [2], Theorem 4.71).

We also consider in this paper semidefinite approximations for problem (1) obtained by taking truncated combinatorial moment matrices and we have given rank conditions ensuring that the relaxation solves the original problem to optimality. A concrete application of the technique developed in the present paper to unconstrained global optimization can be found in the recent paper [9].

The new semidefinite representation for problem (1) in terms of combinatorial moment matrices can be proved using a result of Curto and Fialkow (Theorem 9) about finite rank (infinite) moment matrices. In the radical case, it can also be proved using a combinatorial identity expressing a combinatorial moment matrix in terms of the Zeta matrix of the ideal (see Lemma 17); this is a direct extension of an idea used in the 0/1 or ±1 cases. As mentioned in the Comment in Section 2.2, an analogous combinatorial identity underlies the new proof given in [17] for Theorem 9. Hence, in some sense, the non-radical case reduces to the radical case. Let us say a few words about the 'history' of the present paper. An earlier version of the paper (posted on the home webpage of the author in December 2002) was dealing exclusively with the radical case, as the proofs were based on the combinatorial facts about the Zeta matrix which need the radicality assumption. We realized later that this idea could also be used for proving Theorem 9, which led to the paper [17], and that Theorem 9 could be used in the non-radical case, which led to the present version of this paper.

Acknowledgements. We are grateful to a referee, Dima Pasechnik, for several suggestions that helped improve the presentation of the paper. In particular, we thank Dima for suggesting us to investigate how our results about combinatorial moment matrices relate with known results about the Hermite's form for solving systems of polynomial equations in the zero-dimensional case. We also thank Dima for questioning the necessity of the radicality assumption in an earlier version of the paper.

References

1. Balas, E., Ceria, S., Cornuéjols, G.: A lift-and-project cutting plane algorithm for mixed 0-1 programs. Mathematical Programming **58**, 295–324 (1993)
2. Basu, S., Pollack, R., Roy, M.-F.: Algorithms in Real Algebraic Geometry. Springer, 2003
3. Burer, S., Monteiro, R.D.C., Zhang, Y.: Rank-two heuristics for max-cut and other binary quadratic programs. SIAM Journal on Optimization **12**, 503–521 (2002)
4. Burer, S., Monteiro, R.D.C., Zhang, Y.: Maximum stable set formulations and heuristics based on continuous optimization. Mathematical Programming **94**, 137–166 (2002)
5. Cox, D.A., Little, J.B., O'Shea, D.: Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra. Springer, 1997
6. Cox, D.A., Little, J.B., O'Shea, D.: Using Algebraic Geometry. Graduate Texts in Mathematics, Number 185, Springer, New York, 1998
7. Curto, R.E., Fialkow, L.A.: Solution of the truncated complex moment problem for flat data. Memoirs of the American Mathematical Society vol. **119**, n. 568 (1996)
8. Fuglede, B.: The multidimensional moment problem. Expositiones Mathematicae **1**, 47–65 (1983)
9. Jibetean, D., Laurent, M.: Semidefinite approximations for global unconstrained polynomial optimization. SIAM Journal on Optimization **16**, 490–514 (2005)
10. Landau, H.: Classic background of the moment problem. In Moments in Mathematics, Proceedings of Symposia in Applied Mathematics, vol. **37**, AMS, Providence, 1987, pp. 1–15
11. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM Journal on Optimization **11**, 796–817 (2001)
12. Lasserre, J.B.: An explicit exact SDP relaxation for nonlinear $0-1$ programs. In: K. Aardal, A.M.H. Gerards, (eds.), Lecture Notes in Computer Science **2081**, 293–303 (2001)
13. Lasserre, J.B.: Polynomials nonnegative on a grid and discrete representations. Transactions of the American Mathematical Society **354**, 631–649 (2001)

14. Laurent, M.: A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming. Mathematics of Operations Research **28**, 470–496 (2003)
15. Laurent, M.: Semidefinite relaxations for Max-Cut. In: The Sharpest Cut: The Impact of Manfred Padberg and His Work. M. Grötschel, ed. MPS-SIAM Series in Optimization **4**, 257–290 (2004)
16. Laurent, M.: Lower bound for the number of iterations in semidefinite relaxations for the cut polytope. Mathematics of Operations Research **28**, 871–883 (2003)
17. Laurent, M.: Revisiting two theorems of Curto and Fialkow on moment matrices. Preprint, 2004. To appear in Proceedings of the American Mathematical Society **133**, 2965–2976 (2005)
18. Lovász, L.: On the Shannon capacity of a graph. IEEE Transactions on Information Theory **IT-25**, 1–7 (1979)
19. Lovász, L., Schrijver, A.: Cones of matrices and set-functions and $0-1$ optimization. SIAM Journal on Optimization **1**, 166–190 (1991)
20. Marshall, M.: Optimization of polynomial functions. Canad. Math. Bull. **46**, 575–587 (2003)
21. Nesterov, Y.: Squared functional systems and optimization problems. In: J.B.G. Frenk, C. Roos, T. Terlaky, S. Zhang, (eds.), High Performance Optimization, Kluwer Academic Publishers, 2000, pp. 405–440
22. Parrilo, P.A.: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. PhD thesis, California Institute of Technology, May, 2000
23. Parrilo, P.A.: Semidefinite programming relaxations for semialgebraic problems. Mathematical Programming B **96**, 293–320 (2003)
24. Parrilo, P.A.: An explicit construction of distinguished representations of polynomials nonnegative over finite sets. Preprint, ETH, Zürich, 2002
25. Parrilo, P., Sturmfels, B.: Minimizing polynomial functions. In: Algorithmic and quantitative real algebraic geometry, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. **60**, AMS, 2003, pp. 83–99
26. Powers, V., Wörmann, T.: An algorithm for sums of squares of real polynomials. Journal of Pure and Applied Algebra **127**, 99–104 (1998)
27. Putinar, M.: Positive polynomials on compact semi-algebraic sets. Indiana University Mathematics Journal **42**, 969–984 (1993)
28. Schweighofer, M.: Optimization of polynomials on compact semialgebraic sets. SIAM Journal on Optimization **15**, 805–825 (2005)
29. Sherali, H.D., Adams, W.P.: 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)
30. Shor, N.Z.: An approach to obtaining global extremums in polynomial mathematical programming problems. Kibernetika **5**, 102–106 (1987)
31. Sturmfels, B.: Solving Systems of Polynomial Equations. CBMS, Regional Conference Series in Mathematics, Number 97, AMS, Providence, 2002