Semidefinite Representations for Finite Varieties

Monique Laurent[*†]

October 24, 2003

Abstract

We consider the problem of minimizing a polynomial over a semi-algebraic set defined by polynomial equalities and inequalities. When the polynomial equalities have a finite number of complex solutions and define a radical ideal $\mathcal{I}$, we can reformulate this problem as a semidefinite programming problem. This semidefinite program involves combinatorial moment matrices, which are matrices indexed by a basis of the quotient vector space $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$. Our arguments are elementary and extend known facts for the grid case including 0/1 and ± 1 polynomial programming. They also relate to known algebraic tools for solving polynomial systems of equations with finitely many complex solutions. 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 at 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 semi-algebraic set. It can be formulated as

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

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

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

and

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

In typical applications, $S = \{0,1\}^n$, 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

[*]CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands
[†]Supported by the Netherlands Organisation for Scientific Research grant NWO 639.032.203.

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 $\mathbb{Z}_+$ 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 classic 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 permits to write $f(x) = f^T y$. Given an integer $t \geq d = \deg(f)$, define the convex set:

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

Thus, P is a polytope when F is a finite set. Therefore, problem (1) can be reformulated as the problem:

$$\min_{y \in P} \ f^T y$$

of minimizing the linear objective function $f^T y$ over P. Various methods have been proposed in the literature for constructing relaxations of 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 polynomial equations: $x_i^2 - x_i = 0$ $(i = 1, \ldots, n)$ are present in the description of F. Then, 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: $P \subseteq \ldots \subseteq P^n \subseteq \ldots \subseteq P^t \subseteq \ldots$ 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 [20], and the reformulation-linearization technique of Sherali and Adams [29] can all be cast in this framework. Lovász and Schrijver [20] propose moreover a stronger hierarchy of semidefinite relaxations. We do not give here the exact details of applicability of these methods; see [16] 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. 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 [13], Parrilo [22, 23], Shor [30]). 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. Details about the links between the various methods can be found in [16].

We focus in this paper on the algebraic methods, as they apply to the case when F is an arbitrary compact semialgebraic set. We review the main ideas in the next subsection, following the exposition by Lasserre [13].

1.1 Preliminaries: Moments and sums of squares of polynomials

Assume F is a compact set. A basic observation underlying Lasserre's construction is the fact that

$$p^* := \min_{x \in F} f(x) = \min_{\mu} \int f(x)\mu(dx), \tag{5}$$

where the second minimum 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 $\int_F \mu(dx) = 1$). Indeed, if μ is such a measure, then $\int f(x)\mu(dx) = \int_F f(x)\mu(dx) \geq p^*$, which shows the inequality: $\min\limits_{\mu} \int f(x)\mu(dx) \geq p^*$. To show the reverse inequality, let $v \in F$ be a minimizer of $f(x)$ over F and let $\mu := \delta_v$ be the Dirac measure at v (with mass 1 at v and 0 elsewhere). Then, $\int f(x)\mu(dx) = f(v) = p^*$.

Note that $\int f(x)\mu(dx) = \sum\limits_{\alpha} f_\alpha \int x^\alpha \mu(dx)$. For $\alpha \in \mathbb{Z}_+^n$, the quantity $y_\alpha := \int x^\alpha \mu(dx)$ is called the *moment of order* α of the measure μ. Therefore, it follows from (5) that

$$p^* = \min f^T y \ \text{ s.t. } \quad \begin{array}{l} y \text{ is the sequence of moments of a probability} \\ \text{measure supported by } F. \end{array} \tag{6}$$

When F is a finite set, every probability measure μ supported by F is atomic; that is, μ can be written as a convex combination of Dirac measures at the points of F. Say, $\mu = \sum_{v \in F} \lambda_v \delta_v$ where $\lambda_v \geq 0$ and $\sum_{v \in V} \lambda_v = 1$. Then, the moment of order α of μ is equal to $\sum_{v \in V} \lambda_v v^\alpha$. Therefore, the members of the polytope P defined earlier in (4) are precisely the sequences of moments (of order $\alpha \in S_t$) of probability measures supported by F. Hence, relaxations of P can be constructed by considering necessary conditions that are satisfied by moment sequences of probability measures supported by F.

Characterizing sequences of moments of measures is the object of the classic theory of moments. (See, e.g., [10], [12].) Moment matrices $M_t(y)$ and the so-called shift operator $h * y$ are two classic notions used in characterizations of moment sequences, as shows Lemma 2 below. For a real symmetric matrix M, the notation: $M \succeq 0$ means that M is positive semidefinite.

DEFINITION 1 **(i)** *Given an integer* $t \geq 1$ *and* $y \in \mathbb{R}^{S_{2t}}$, *the* moment matrix of order t *of* y *is the matrix* $M_t(y)$ *indexed by* S_t, *with* (α, β)*th entry* $y_{\alpha+\beta}$, *for* $\alpha, \beta \in S_t$.

(ii) *Given $h, y \in \mathbb{R}^{\mathbb{Z}^n_+}$ with h having a finite support (i.e., h is the vector of coefficients of a polynomial), the* shifted vector *$h * y$ is the vector of $\mathbb{R}^{\mathbb{Z}^n_+}$ whose αth entry is $(h*y)_\alpha := \sum_\beta h_\beta y_{\alpha+\beta}$, for $\alpha \in \mathbb{Z}^n_+$. The notation $h*y$ will also be used for denoting the truncated vector $((h*y)_\alpha)_{\alpha \in S_t}$ of $\mathbb{R}^{S_t}$ for an integer $t \geq 1$.*

LEMMA 2 *Let $t \geq 1$ be an integer and let $y \in \mathbb{R}^{S_t}$.*

(i) *If y is the sequence of moments (up to order $2t$) of a nonnegative measure μ, then $M_t(y) \succeq 0$. Moreover, if $p(x)$ is a polynomial of degree $\leq t$ such that $M_t(y)p = 0$, then the support of μ is contained in the set of zeros of $p(x)$.*

(ii) *Let $h(x)$ be a polynomial of degree $2v$ or $2v-1$ ($v \geq 1$) and $F := \{x \mid h(x) \geq 0\}$. If $y \in \mathbb{R}^{S_{2t}}$ is the sequence of moments of a nonnegative measure μ supported by F, then $M_{t-v}(h*v) \succeq 0$.*

PROOF. (i) Let $p(x)$ be a polynomial of degree $\leq t$ and $p \in \mathbb{R}^{S_t}$ its vector of coefficients. Then,

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

showing that $M_t(y) \succeq 0$. If $M_t(y)p = 0$, then $0 = p^T M_t(y) p = \int p(x)^2 \mu(dx)$. As μ is nonnegative, this implies that the support of μ is contained in the set of zeros of $p(x)$.
(ii) Let $p(x)$ be a polynomial of degree $\leq t-v$ and $p \in \mathbb{R}^{S_{t-v}}$ its vector of coefficients. Then,

$$\begin{aligned} p^T M_{t-v}(h*y) p &= \sum_{\alpha,\beta \in S_{t-v}} p_\alpha p_\beta (h*y)_{\alpha+\beta} = \sum_{\alpha,\beta \in S_{t-v}} p_\alpha p_\beta \sum_{\gamma \in S_{2v}} h_\gamma y_{\alpha+\beta+\gamma} \\ &= \sum_{\alpha,\beta \in S_{t-v},\ \gamma \in S_{2t-2v}} p_\alpha p_\beta h_\gamma \int x^{\alpha+\beta+\gamma} \mu(dx) = \int_F h(x) p(x)^2 \mu(dx) \geq 0, \end{aligned}$$

which shows that $M_{t-v}(h*y) \succeq 0$. ∎

COROLLARY 3 *Let F be the semialgebraic set from (3) and, for $j = 1, \ldots, m+k$, denote the degree of the polynomial $h_j(x)$ as $2v_j$ or $2v_j - 1$ ($v_j \geq 1$). If $y \in \mathbb{R}^{S_t}$ is the sequence of moments (up to order $2t$) of a probability measure supported by F, then*

$$y_0 = 1,\ M_{t-v_j}(h*y) = 0\ (j = 1, \ldots, m),\ M_{t-v_j}(h*y) \succeq 0\ (j = m+1, \ldots, m+k).$$

For $j = 1, \ldots, m+k$, let $v_j = \lceil \deg(h_j)/2 \rceil$ (as in Corollary 3) and $v := \max(v_1, \ldots, v_{m+k})$. Following Lasserre [13], one can formulate the following semidefinite relaxation for (1):

$$\begin{array}{rll} p_t^* := & \min & \sum_\alpha f_\alpha y_\alpha \\ & \text{s.t.} & M_t(y) \succeq 0 \\ & & M_{t-v_j}(h_j * y) = 0\ (j = 1, \ldots, m) \\ & & M_{t-v_j}(h_j * y) \succeq 0\ (j = m+1, \ldots, m+k) \\ & & y_0 = 1 \end{array} \tag{7}$$

for any $t \geq \max(v, \lceil \frac{\deg(f)}{2} \rceil)$. Obviously,

$$p_t^* \leq p_{t+1}^* \leq p^*.$$

Another way to get a lower bound for p^* is express it as the maximum scalar ρ for which the polynomial $f(x) - p^*$ is nonnegative on the set F; that is,

$$p^* = \max \ \rho \ \text{ s.t. } f(x) - \rho \geq 0 \ \forall x \in F.$$

One can now replace the nonnegativity condition by the following stronger condition:

$$\begin{aligned}
\rho^* := \max \ \rho \ \text{ s.t.} \quad & f(x) - \rho = u_0(x) + \sum_{j=1}^{m+k} u_j(x) h_j(x) \ \text{ where} \\
& u_0(x), u_{m+1}(x), \ldots, u_{m+k}(x) \text{ are sums of squares of polynomials.}
\end{aligned}$$

Obviously, $\rho^* \leq p^*$. By bounding the degrees of the unknown polynomials $u_j(x)$, we get the following hierarchy of bounds for p^*:

$$\begin{aligned}
\rho_t^* := \quad & \max \quad \rho \\
& \text{s.t.} \quad f(x) - \rho = u_0(x) + \sum_{j=1}^{m+k} u_j(x) h_j(x) \\
& \qquad \text{where } u_0(x), u_{m+1}(x), \ldots, u_{m+k}(x) \text{ are sums of squares of} \\
& \qquad \text{polynomials and } \deg(u_0),\ \deg(u_j h_j) \leq 2t \text{ for } j = 1, \ldots, m+k.
\end{aligned} \tag{8}$$

Obviously,

$$\rho_t^* \leq \rho_{t+1}^* \leq p^*.$$

A crucial property is that sums of squares of polynomials can be tested using semidefinite programming (see [26] and Remark 4 below). Therefore, program (8) can be reformulated as a semidefinite program.

REMARK 4 *Note that program (8) slightly differs from the usual formulation, where the polynomial u_j ($j = 1, \ldots, m$) is required to be of the form $u_j' - u_j''$, where u_j', u_j'' are sums of squares with degree $\leq 2(t - v_j)$. The two formulations are, however, equivalent. One can indeed verify that every polynomial u with $\deg(u) \leq 2t$ can be decomposed as $u = u' - u''$, where the polynomials u' and u'' are sums of squares of polynomials with $\deg(u'), \deg(u'') \leq 2t$.*

To see it, write $u(x) = \sum_{\gamma \in S_{2t}} u_\gamma x^\gamma$ and consider the following linear system in the variables $X_{\alpha,\beta}$ ($\alpha, \beta \in S_t$):

$$\sum_{\alpha,\beta | \alpha+\beta=\gamma} X_{\alpha,\beta} = u_\gamma \quad (\gamma \in S_{2t}) \tag{9}$$

$$X_{\alpha,\beta} = X_{\beta,\alpha} \quad (\alpha \neq \beta \in S_t). \tag{10}$$

This linear system is feasible since distinct equations in (9) involve disjoint sets of variables. (For instance, for each $\gamma \in S_{2t}$, choose a pair (α_0, β_0) with $\alpha_0 + \beta_0 = \gamma$; set $X_{\alpha_0,\beta_0} = X_{\beta_0,\alpha_0} = \frac{1}{2}u_\gamma$ if $\alpha_0 \neq \beta_0$ and $X_{\alpha_0,\beta_0} = u_\gamma$ if $\alpha_0 = \beta_0$, and set $X_{\alpha,\beta} = 0$ for all remaining pairs (α, β).) As the matrix $X = (X_{\alpha,\beta})_{\alpha,\beta\in S_t}$ is real symmetric, it can be decomposed as $X = \sum_{i=1}^{p} u'_i(u'_i)^T - \sum_{j=1}^{q} u''_j(u''_j)^T$ for some vectors $u'_i, u''_j \in \mathbb{R}^{S_t}$. This implies that the polynomial $u(x)$ can be decomposed as $\sum_{i=1}^{p}(u'_i(x))^2 - \sum_{j=1}^{q}(u''_j(x))^2$, where $u'_i(x)$ is defined as $\sum_\alpha (u'_i)_\alpha x^\alpha$ and analogously for $u''_j(x)$.

The above argument also shows that the polynomial $u(x)$ is a sum of squares of polynomials if and only if the system (9)-(10) has a solution X which is positive semidefinite.

It is proved in [13] that program (8) is, in fact, the semidefinite dual of (7); thus, weak semidefinite duality implies that

$$\rho_t^* \leq p_t^* \leq p^*. \tag{11}$$

A result in real algebraic geometry (Schmüdgen [28], Putinar [27], Jacobi and Prestel [11]) asserts that, when F is compact (and satisfies some additional technical condition), then every *positive* polynomial on F has a decomposition:

$$u_0 + \sum_{j=1}^{m+k} u_j h_j \quad \text{where } u_0, u_{m+1}, \ldots, u_{m+k} \text{ are sums of squares of polynomials.} \tag{12}$$

This result implies the asymptotic convergence of the lower bounds ρ_t^*, p_t^* to p^* as t goes to infinity [13].

In the special case when F is contained in $\{0,1\}^n$ or is equal to the set of points in a grid, Lasserre [14, 15] shows the *finite* convergence of the bounds ρ_t^*, p_t^* to p^*. His proof is based on a result by Curto and Fialkow [8, 9] about flat extensions of moment matrices (which uses the Riesz representation theorem) (see Corollary 30). As a consequence, every polynomial *nonnegative* on F has a decomposition (12) with a priori bounds on the degrees of the polynomials u_j.

When the ideal $\mathcal{I}$ generated by the polynomials $h_1, \ldots, h_m$ defining S is radical and the associated complex variety $V(\mathcal{I})$ is finite, Parrilo [24] proves (in an elementary way) that every polynomial *nonnegative* on F has a decomposition (12) with a priori bounds on the degrees of $u_0, u_{m+1}, \ldots, u_{m+k}$ (see Theorem 26). In fact, a priori bounds can also be given on the degrees of $u_1, \ldots, u_m$ when $h_1, \ldots, h_m$ form a Groebner basis of $\mathcal{I}$.

In this paper, we observe that the finite convergence result of Lasserre for the grid case extends to the same setting as Parrilo's result. That is, we prove that when $\mathcal{I}$ is a radical ideal and $V(\mathcal{I})$ is finite, there is finite convergence of the bounds p_t^* and ρ_t^* to the optimum p^* of problem (1). Our argument does not use Curto and Fialkow's result, but Parrilo's result which is more elementary. Moreover, it establishes finite convergence in a smaller number of steps. Furthermore, we give a semidefinite representation for problem (1) (see (34)), which is more concise than (7). This representation involves 'combinatorial' moment matrices $M_{\mathcal{B}}(y)$, which are matrices indexed by a basis $\mathcal{B}$ of the quotient space $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$; note that $|\mathcal{B}| < |S_t|$ for any t ensuring finite convergence of the Lasserre relaxations (7). (See Sections 2, 3 for details.) The next subsection introduces some necessary background on polynomial ideals and the quotient ring $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$.

1.2 Preliminaries: 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 subsections, see, e.g., [2], [5], [6], [7], [31]. Given an ideal $\mathcal{I}$ in $\mathbb{R}[x_1, \ldots, x_n]$, the set

$$V = V(\mathcal{I}) := \{x \in \mathbb{C}^n \mid h(x) = 0 \text{ for all } h \in \mathcal{I}\}, \tag{13}$$

is the complex variety associated to $\mathcal{I}$. If $\mathcal{I}$ is generated by $h_1, \ldots, h_m$, then V consists of the common complex zeros of $h_1, \ldots, h_m$. When V is finite, the ideal $\mathcal{I}$ is said to be zero-dimensional. The set

$$I(V) := \{f \in \mathbb{R}[x_1, \ldots, x_n] \mid f(x) = 0 \;\text{ for all } x \in V\} \tag{14}$$

is an ideal in $\mathbb{R}[x_1, \ldots, x_n]$ that contains the ideal $\mathcal{I}$. When equality $\mathcal{I} = I(V)$ holds, the ideal $\mathcal{I}$ 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 $\mathcal{I}$ in $\mathbb{R}[x]$ generated by $h(x) = x^2$ is not radical since $V = \{0\}$ and $f(x) = x$ belongs to $I(V) \setminus \mathcal{I}$.

Given an ideal $\mathcal{I}$ in $\mathbb{R}[x_1, \ldots, x_n]$, the relation $f \equiv g$ if $f - g \in \mathcal{I}$ defines an equivalence relation on $\mathbb{R}[x_1, \ldots, x_n]$ whose equivalence classes form the quotient space $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$. It will be convenient to have a systematic way of chosing a representative polynomial in each equivalence class of $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$. This can be achieved if we choose some monomial ordering in $\mathbb{R}[x_1, \ldots, x_n]$.

The monomials in $\mathbb{R}[x_1, \ldots, x_n]$ 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: $x^\alpha < x^\beta \Longrightarrow x^{\alpha+\gamma} < x^{\beta+\gamma}$. An example of monomial ordering is the *lexicographic order* '$<_{lex}$', corresponding to the lexicographic order for n-tuples; that is, $x^\alpha <_{lex} x^\beta$ if the left-most nonzero entry of $\beta - \alpha$ is positive. Another example is the *graded lexicographic order* '$<_{grlex}$', which takes into account the total degree of monomials; that is, $x^\alpha <_{grlex} x^\beta$ if $|\alpha| < |\beta|$, or $|\alpha| = |\beta|$ and $x^\alpha <_{lex} x^\beta$.

In what follows we 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 $f_\alpha \neq 0$. Once a monomial ordering is fixed, we can apply the division algorithm. Given nonzero polynomials $h_1, \ldots, h_m$ and f, the division algorithm applied for dividing f by the h_j's produces polynomials $u_1, \ldots, u_m$ and r such that

$$f = \sum_{j=1}^m u_j h_j + r, \tag{15}$$

$$\text{no term of } r \text{ is divisible by } \mathrm{LT}(h_j)(j = 1, \ldots, m) \text{ if } r \neq 0, \tag{16}$$

$$\deg(f) \geq \deg(u_j h_j) \text{ if } u_j \neq 0. \tag{17}$$

The remainder r of f produced by the division algorithm is, in general, not uniquely determined. In particular, it can happen that f belongs to the ideal $\mathcal{I}$ generated by $h_1, \ldots, h_m$ and that, yet, the remainder r is nonzero. Unicity of the remainder holds precisely when the polynomials $h_1, \ldots, h_m$ form a Groebner basis of $\mathcal{I}$. A finite subset $G \subseteq \mathcal{I}$ being called a *Groebner basis* of $\mathcal{I}$ if the leading term of every nonzero polynomial in $\mathcal{I}$ is divisible by the leading term of some polynomial in G.

Given an ideal $\mathcal{I}$ in $\mathbb{R}[x_1, \ldots, x_n]$, its *leading term ideal* $\langle \mathrm{LT}(\mathcal{I}) \rangle$ is the ideal generated by the leading terms of polynomials in $\mathcal{I}$. Set

$$\mathcal{B} := \{x^\alpha \mid x^\alpha \notin \langle \mathrm{LT}(\mathcal{I}) \rangle\}; \tag{18}$$

monomials in $\mathcal{B}$ are called *standard monomials.* In other words, $x^\alpha \in \mathcal{B}$ if and only if x^α is not divisible by the leading term of any polynomial in $\mathcal{I}$ or, equivalently, if x^α is not divisible by the leading term of any polynomial in a Groebner basis of $\mathcal{I}$.

Therefore, if we divide a polynomial f by a set of polynomials $h_1, \ldots, h_m$ forming a Groebner basis, the division algorithm produces an expression (15), where the remainder r is a linear combination of the standard monomials in $\mathcal{B}$, i.e.,

$$r(x) = \sum_{\beta \in \mathcal{B}} r_\beta x^\beta.$$

Moreover, $f \in \mathcal{I}$ if and only if $r = 0$, and the remainder r is uniquely determined for all f. As $f - r \in \mathcal{I}$, the reminder r is also called the *residue of f modulo $\mathcal{I}$ with respect to $\mathcal{B}$.* It can be seen as a standard representative of the equivalence class (the coset) of f in the quotient space $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$. Therefore, $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$ and $\mathbb{R}^\mathcal{B}$ are isomorphic vector spaces.

Definition 5 *A set $\mathcal{B} := \{f_1, \ldots, f_N\}$ of polynomials forms a basis of $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$ if, for every polynomial f, there exists a unique set of real numbers $\lambda_1^{(f)}, \ldots, \lambda_N^{(f)}$ such that*

$$f - \sum_{i=1}^{N} \lambda_i^{(f)} f_i \in \mathcal{I}. \tag{19}$$

The polynomial $\sum_{i=1}^N \lambda_i^{(f)} f_i$ is called the residue of f modulo $\mathcal{I}$ w.r.t. the basis $\mathcal{B}$.

For instance, the set of standard monomials (w.r.t. some monomial ordering) is a basis of $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$, also called a *monomial basis.* A fundamental property that we will use here is the following, which can be found, e.g., in ([5], chap. 2) or ([6], chap. 5).

Theorem 6 *Let $\mathcal{I}$ be an ideal in $\mathbb{R}[x_1, \ldots, x_n]$ with complex variety V (as in (13)). Then, the variety V is finite if and only if the vector space $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$ has finite dimension N. Moreover, $|V| \leq N$, with equality if and only if the ideal $\mathcal{I}$ is radical, i.e., if*

$$\{f \in \mathbb{R}[x_1, \ldots, x_n] \mid f(v) = 0 \ \forall v \in V\} = \mathcal{I}. \tag{20}$$

Example 7 *([6], p. 227) Let $\mathcal{I}$ 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 $\mathcal{I}$ is $V = \{(0,0)\} \cup \{(x, x^7) \mid x^{10} = 1\}$, with cardinality 11. As $|V| < |\mathcal{B}_1|$, the ideal $\mathcal{I}$ is not radical. Indeed, we see that $(0,0)$ is a zero with double multiplicity.

The quotient space $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$ 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 $\mathcal{B}$; 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 $f_i, f_j \in \mathcal{B}$,

$$T_{\mathcal{B}}(f_i, f_j) = \sum_{k=1}^{N} \lambda_k^{(f_i f_j)} f_k. \tag{21}$$

When $\mathcal{B}$ is a monomial basis, $T_{\mathcal{B}}$ can be computed using the division algorithm.

Example 8 *Let $\mathcal{I}_1$ be the ideal in $\mathbb{R}[x,y]$ generated by $x^2 - x$ and $y^2 - y$ and $\mathcal{I}_2$ 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 $\mathbb{R}[x,y]/\mathcal{I}_i$ whose multiplication table T_i is as follows:*

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

1.3 Preliminaries: 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 equalities and inequalities. A related problem which has received considerable attention in the literature, is the problem of solving a system of polynomial equations:

$$h_1(x) = 0, \ldots, h_m(x) = 0.$$

As before, $\mathcal{I}$ denotes the ideal generated by $h_1, \ldots, h_m$, V is the associated complex variety, and $S = V \cap \mathbb{R}^n$ is the set of real roots. We assume that V is finite. Then, the so-called eigenvalue method (also known as the Stetter-Möller method) can be applied for finding V. We first need to define the multiplication operator.

DEFINITION 9 *Given a polynomial $h \in \mathbb{R}[x_1, \ldots, x_n]$, let*

$$\begin{array}{rccc} m_h: & \mathbb{R}[x_1, \ldots, x_n]/\mathcal{I} & \longrightarrow & \mathbb{R}[x_1, \ldots, x_n]/\mathcal{I} \\ & f & \longmapsto & fh \end{array}$$

be the 'multiplication by h' operator. It is easy to verify that this linear operator is well-defined. Given a monomial basis $\mathcal{B}$ of $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$, 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 x^βth column is the vector $\lambda^{(hx^\beta)} \in \mathbb{R}^{\mathcal{B}}$, giving the residue w.r.t. $\mathcal{B}$ of the product hx^β.

When $h(x) = x_i$, the multiplication matrices $M_{x_i} =: M_i$ are also known as the *companion matrices* of the ideal $\mathcal{I}$. It follows from the definition that the matrices $M_1, \ldots, M_n$ pairwise commute. Hence, they generate a commutative algebra $\mathbb{R}[M_1, \ldots, M_n]$, which is isomorphic to the ring $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$ via the correspondance $x_i \longrightarrow M_i$. It is easy to verify that, for $h(x) = \sum_\alpha h_\alpha x^\alpha$, $M_h = \sum_\alpha h_\alpha M_1^{\alpha_1} \cdots M_n^{\alpha_n}$. It turns out that the coordinates of the points $v \in V$ can be evaluated by computing the eigenvalues of the companion matrices.

THEOREM 10 *(Stickelberger's theorem) Assume that V is finite.*

(i) *The complex zeros of $\mathcal{I}$ are the vectors of joint eigenvalues of the companion matrices $M_1, \ldots, M_n$; that is,*

$$V = \{\lambda \in \mathbb{C}^n \mid \exists x \in \mathbb{C}^n \setminus \{0\} \ s.t. \ M_i x = \lambda_i x \ \forall i = 1, \ldots, n\}.$$

(ii) *When $\mathcal{I}$ is radical, the left eigenvectors of M_h are the vectors $\zeta_v := (v^\beta)_{\beta \in \mathcal{B}}$ with corresponding eigenvalue $h(v)$, for $v \in V$; that is, $M_h^T \zeta_v = h(v)\zeta_v$ for $v \in V$.*

(iii) *The trace of the multiplication matrix M_h is equal to $\sum_{v \in V} m(v)h(v)$, where $m(v)$ are the multiplicities of the zeros of I. In particular, $Tr(M_h) = \sum_{v \in V} h(v)$ when $\mathcal{I}$ is radical.*

Another relevant result is the following result from [25] permitting to count the number of real zeros, which involves the following bilinear form s_h (known as the trace form or the Hermite form):

$$\begin{array}{rccc} s_h: & \mathbb{R}[x_1, \ldots, x_n]/\mathcal{I} \times \mathbb{R}[x_1, \ldots, x_n]/\mathcal{I} & \longrightarrow & \mathbb{R} \\ & (f, g) & \longmapsto & Tr(M_{fgh}) \end{array}$$

Let S_h denote the matrix of s_h w.r.t. a monomial basis $\mathcal{B}$. Thus, S_h is a real symmetric matrix.

THEOREM 11 *[25] Let σ_+ (resp., σ_-) denote the number of positive (resp., negative) eigenvalues of S_h, and set $s_+ := |\{v \in V \cap \mathbb{R}^n \mid h(v) > 0\}|$, $s_- := |\{v \in V \cap \mathbb{R}^n \mid h(v) < 0\}|$. Then, $\sigma_+ - \sigma_- = s_+ - s_-$, and the rank of S_h is equal to the number of $v \in V$ with $h(v) \neq 0$.*

1.4 Contents of the paper

We are dealing in this paper with the problem of minimizing a polynomial function $f(x)$ over a semialgebraic set F, under the assumption that the equations entering the description of F have a finite set V of complex solutions and generate a radical ideal. Hence, one obvious possible solution would be to find all points of V using the eigenvalue method sketched in the previous subsection, and to compute $f(v)$ for all $v \in F$. A disavantage of this method might be the propagation of numerical errors. In this paper, we want to propose an alternative 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 $|V|$, it can be solved only for small $|V|$. However, a feature of our method is that it leads naturally to a hierarchy of semidefinite relaxations of the original problem whose low order members yield efficiently computable bounds to the original problem. (See Section 4 for results in this direction.)

A second feature of our method is that is makes a bridge between classic algebraic results (about the companion matrices and the trace form) and elementary combinatorial results previously known in the 0/1 case and used in the combinatorial methods (about the Zeta matrix of the lattice of subsets of a finite set). Not surprisingly, our semidefinite representation for problem (1) uses concepts closely related to the multiplication operator M_h and to the trace operator S_h. In particular, the matrix M_h will be used for defining the shift vector $h * y$ (see Definition 19) and we will prove two results (Lemma 20 and Proposition 17) which will give, respectively, Theorem 10 in the radical case and Theorem 11. It is an objective of this paper to highlight the links existing between these two groups of techniques.

Finally, our semidefinite representation for problem (1) in terms of combinatorial moment matrices provides the rationale and an elementary proof for the finite convergence result of the Lasserre hierarchy in the zero-dimensional radical case.

Our results need the assumption that the ideal $\mathcal{I}$ generated by the polynomial equations is radical. However, when $\mathcal{I}$ is not radical, one can work instead with the radical ideal $I(V)$ (from (14)), which can be found via some Groebner bases computations (see, e.g., ([7], prop. 2.7)).

The paper is organized as follows. In Section 2 we present a new semidefinite representation for problem (1) (see (34)). It involves 'combinatorial' moment matrices $M_{\mathcal{B}}(y)$, which are matrices indexed by a basis $\mathcal{B}$ of the quotient space $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$. Roughly speaking, this representation can be seen as an analogue of the Lasserre relaxation (7), obtained by 'factoring through the ideal $\mathcal{I}$' generated by the polynomial equations entering the description of the semialgebraic set F. The 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 ([15, 16, 17]). The equivalence of (1) and (34) is an easy result, which follows from a simple combinatorial identity involving the Zeta matrix of the ideal $\mathcal{I}$ (see Lemma 21). This is a direct extension of the corresponding result given in [20] and in [16] for the 0/1 and ± 1 cases (which also underlies the convergence results for combinatorial lift-and-project methods). Alternatively, the equivalence of (1) and (34) follows, using semidefinite programming duality, from a result of Parrilo about representations of nonnegative polynomials over a semialgebraic set. The two approaches are described, respectively, in Sections 2.1 and 2.2.

In Section 3, we elaborate on the links existing between the new semidefinite representation (involving combinatorial moment matrices) and Lasserre's hierarchy (involving classic moment matrices). A central notion there is that of 'flat extension' (see Definition 28). Roughly speaking, for t large enough, the matrices occurring in Lasserre's relaxation of order t are flat extensions of the combinatorial moment matrices. As a consequence, one can prove better convergence results for the Lasserre hierarchy.

In Section 4, we consider approximations for problem (1) in the grid case, when S is the set $\{x \mid x_i = a_i, b_i \ \forall i = 1, \dots, n\}$ (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 35.) We can prove this result in two ways; either by using Curto and Fialkow's results from Theorem 29 and Corollary 30, or by using their (elementary) combinatorial analogues provided by Theorem 39 and Corollary 40. As Curto and Fialkow's result is quite involved, we give a self-contained proof for its combinatorial analogue from Theorem 39 in the Appendix.

We also mention a result of the same flavour for the maximum stable set problem. The theta number $\vartheta(G)$ is a well known upper bound on the stability number of a graph $G = (\{1, \dots, n\}, E)$, introduced by Lovász [19]; it can be obtained as optimum value of the semidefinite program:

$$\max \sum_{i=1}^{n} y_i \text{ s.t. } M_{\mathcal{B}_1}(y) \succeq 0, \ y_{ij} = 0 \ (ij \in E), \ y_0 = 1. \tag{22}$$

If we impose that $M_{\mathcal{B}_1}(y)$ has rank 1 in (22), 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 41). This is an analogue of a result of Burer, Monteiro and Zhang [4] given for another formulation of the theta number.

2 A Semidefinite Representation for a Finite Variety

2.1 Approach via combinatorial moment matrices

In this section, we consider the problem (1) of minimizing a polynomial f over the semialgebraic set F from (3). Let $\mathcal{I}$ be the ideal generated by the polynomials $h_1, \dots, h_m$, and let V, $S = V \cap \mathbb{R}^n$ as in (13) and (2). We assume throughout that V is a finite set and that the ideal $\mathcal{I}$ is radical. Then, (20) holds and the dimension N of the vector space $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$ satisfies $N = |V|$. Let $\mathcal{B} := \{f_1, \ldots, f_N\}$ be a basis of $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$ (e.g., $\mathcal{B}$ is the set of standard monomials w.r.t. some monomial ordering). We begin with introducing the following matrix Z, which will play a fundamental role in our treatment.

Definition 12 *Let Z denote the $\mathcal{B} \times V$ matrix with entries*

$$Z(f, v) = f(v) \text{ [the evaluation at } v \text{ of the polynomial } f(x)] \quad \text{for } f \in \mathcal{B},\ v \in V \tag{23}$$

and let ζ_v denote the column of Z indexed by $v \in V$. By analogy to the 0/1 case (see Example 15 below), we may call the matrix Z the Zeta matrix *of the ideal $\mathcal{I}$ (w.r.t. the basis $\mathcal{B}$).*

Lemma 13 **(i)** *The matrix Z is nonsingular.*

(ii) *For a polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$, its residue w.r.t. the basis $\mathcal{B}$ (recall Definition 5) is the polynomial $\sum_{i=1}^N \lambda_i^{(f)} f_i$, where $\lambda^{(f)} = (Z^T)^{-1}(f(v))_{v \in V}$.*

Proof. By definition of the residue, $f - \sum_{i=1}^N \lambda_i^{(f)} f_i \in \mathcal{I}$. As $\mathcal{I}$ is radical, this is equivalent to $f(v) = \sum_{i=1}^N \lambda_i^{(f)} f_i(v)$ for all $v \in V$, i.e., $(\lambda^{(f)})^T Z = (f(v))_{v \in V}$. Therefore, $(\lambda^{(f)})^T Z = 0$ if and only if $f \in \mathcal{I}$ which, in turn, is equivalent to $\lambda^{(f)} = 0$. This shows that (i) Z is nonsingular as well as assertion (ii). ∎

The problem (1) remains obviously unchanged if we replace the polynomial f by its residue $r(x) = \sum_{i=1}^N \lambda_i^{(f)} f_i(x)$ modulo $\mathcal{I}$ (w.r.t. the basis $\mathcal{B}$). That is, $p^* = \min_{x \in F} f(x) = \min_{x \in F} r(x)$. For $v \in V$, $r(v) = \sum_{i=1}^N \lambda_i^{(f)} f_i(v) = \zeta_v^T \lambda^{(f)}$. Therefore,

$$p^* = \min_{v \in F} \ \zeta_v^T \lambda^{(f)} = \min \ y^T \lambda^{(f)} \ \text{ s.t. } y \in P_{\mathcal{B}} := \operatorname{conv}(\zeta_v \mid v \in F). \tag{24}$$

Hence, we are left with the task of characterizing the polytope $P_{\mathcal{B}}$, which lives in the N-dimensional space $\mathbb{R}^{\mathcal{B}}$ indexed by the given basis $\mathcal{B}$. Obviously,

$$\begin{aligned} y \in P_{\mathcal{B}} &\Longleftrightarrow y = Z\mu \text{ for some } \mu \in \mathbb{R}_+^V \text{ with } \mu_v = 0 \ \forall v \in V \setminus F \text{ and } \textstyle\sum_{v \in V} \mu_v = 1 \\ &\Longleftrightarrow Z^{-1}y \geq 0,\ (Z^{-1}y)_v = 0 \ \forall v \in V \setminus F,\ e^T Z^{-1} y = 1. \end{aligned} \tag{25}$$

This gives, therefore, a linear inequality description for the polytope $P_{\mathcal{B}}$ and a formulation of problem (1) as a linear programming problem over an N-dimensional simplex. This description, however, relies on the Zeta matrix Z whose description needs the explicit knowledge of the variety V. We present below an alternative semidefinite description for the polytope $P_{\mathcal{B}}$ that does not need the explicit knowledge of V. It needs only the knowledge of a basis $\mathcal{B}$ of $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$ and of its associated multiplication table. Another advantage of the semidefinite representation is that it will naturally lead to a hierarchy of semidefinite relaxations for the polytope $P_{\mathcal{B}}$ (see Section 4). In order to define it, we need to introduce the following notion of of 'combinatorial' moment matrix.

Definition 14 *Given* $y = (y_1, \ldots, y_N)^T \in \mathbb{R}^{\mathcal{B}}$, *the combinatorial moment matrix of* y *is the* $\mathcal{B} \times \mathcal{B}$ *symmetric matrix* $M_{\mathcal{B}}(y)$ *whose* (f_i, f_j)*th entry is equal to* $y^T \lambda^{(f_i f_j)} = \sum_{k=1}^N \lambda_k^{(f_i f_j)} y_k$ *for* $f_i, f_j \in \mathcal{B}$. *In other words,* $M_{\mathcal{B}}(y)$ *is obtained from the multiplication table of* $\mathcal{B}$ *(recall (21)) by 'linearizing', which means replacing each occurrence of* f_k *by the variable* y_k.

Example 15 *Consider the* $0/1$ *case, when* $\mathcal{I}$ *is generated by the polynomials* $x_j^2 - x_j$ $(j = 1, \ldots, n)$*; thus,* $V = \{0,1\}^n$. *W.r.t. the graded lexicographic order, the set of standard monomials is*

$$\mathcal{B} = \{x^A := \prod_{i \in A} x_i \mid A \subseteq \{1, \ldots, n\}\}. \tag{26}$$

W.r.t. this monomial basis $\mathcal{B}$, *the matrix* Z *is known as the Zeta matrix of the lattice of subsets of* $\{1, \ldots, n\}$ *and* Z^{-1} *as its Möbius matrix. For* $A, B \subseteq \{1, \ldots, n\}$, *the* (A, B)*th entry of* Z *(resp., of* Z^{-1}*) is equal to 1 (resp., to* $(-1)^{|B \setminus A|}$*) if* $A \subseteq B$, *and to 0 (resp., 0) otherwise (see [20]). 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* $\mathcal{I}$. *Hence, the combinatorial moment matrix* $M_{\mathcal{B}}(y)$ *of a vector* $y = (y_A)_{A \subseteq \{1, \ldots, n\}}$ *is the matrix indexed by all subsets of* $\{1, \ldots, n\}$ *with* (A, B)*th entry* $y_{A \cup B}$.

In the ± 1 *case,* $\mathcal{I}$ *is generated by the polynomials* $x_j^2 - 1$ $(j = 1, \ldots, n)$, $V = \{\pm 1\}^n$, *and (26) is again a monomial basis. One can verify that the Zeta matrix* Z *satisfies:* $Z^{-1} = 2^{-n} Z$ *(see [17]). 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* $y = (y_A)_{A \subseteq \{1, \ldots, n\}}$ *has its* (A, B)*th entry equal to* $y_{A \Delta B}$.

The next lemma shows the link existing between $M_{\mathcal{B}}(y)$ and $Z^{-1}y$.

Lemma 16 $M_{\mathcal{B}}(y) = Z \mathrm{diag}(Z^{-1}y) Z^T$.

Proof. Given $f_i, f_j \in \mathcal{B}$, the (f_i, f_j)th entry of the matrix $Z \mathrm{diag}(Z^{-1}y) Z^T$ is equal to

$$\sum_{v \in V} f_i(v) f_j(v) (Z^{-1}y)_v = (Z^{-1}y)^T (f_i(v) f_j(v))_{v \in V} = y^T (Z^{-1})^T (f_i(v) f_j(v))_{v \in V},$$

which is equal to $y^T \lambda^{(f_i f_j)}$ (by Lemma 13 (ii)) and thus to the (f_i, f_j)th entry of $M_{\mathcal{B}}(y)$. ∎

Proposition 17 below[1] will tell us how positive semidefiniteness of the matrix $M_{\mathcal{B}}(y)$ can be expressed in terms of the vector $Z^{-1}y$. It also implies the counting result from Theorem 11 for the real roots of systems of polynomial equations.

[1] Strictly speaking, we only need here the last part of Proposition 17 characterizing when M is positive semidefinite, for which we could have given a slightly shorter proof. However, we decided to prove a more general result as it implies Theorem 11. This result might already be known in the literature although we did not find it. The proof given here has anyway the advantage of using only elementary linear algebra and thus of being easily accessible.

As the ideal $\mathcal{I}$ is generated by polynomials with real coefficients, the complex conjugate $\overline{v}$ of $v \in V$ also belongs to V. Thus, V can be partitioned into $V = S \cup T \cup \overline{T}$, where $S = V \cap \mathbb{R}^n$, $T \cup \overline{T} = V \setminus \mathbb{R}^n$, and $\overline{T} = \{\overline{v} \mid v \in T\}$.

Proposition 17 *Let $M := Z\mathrm{diag}(w)Z^T$, where $w \in \mathbb{C}^V$ satisfies $w_{\overline{v}} = \overline{w_v}$ for all $v \in V$. Then, M is a real symmetric matrix. Let σ_+ (resp., σ_-) denote the number of positive (resp., negative) eigenvalues of M and set $s_+ := |\{v \in S \mid w_v > 0\}|$, $s_- := |\{v \in S \mid w_v < 0\}|$, and $t := |\{v \in T \mid w_v \neq 0\}| = |\{v \in V \setminus \mathbb{R}^n \mid w_v \neq 0\}|/2$. Then,*

$$\sigma_+ = s_+ + t,\ \sigma_- = s_- + t.$$

In particular, the rank of M is equal to the number of $v \in V$ with $w_v \neq 0$, and M is positive semidefinite if and only if $w_v \geq 0$ for $v \in S$ and $w_v = 0$ for $v \in V \setminus S$.

Proof. W.r.t. the partition $V = S \cup T \cup \overline{T}$, w and Z have the block decompositions:

$$w = \begin{array}{c} S \\ T \\ \overline{T} \end{array}\begin{pmatrix} a \\ b \\ \overline{b} \end{pmatrix},\quad Z = \begin{array}{ccc} S & T & \overline{T} \\ \multicolumn{3}{c}{\begin{pmatrix} A & B & \overline{B} \end{pmatrix}} \end{array},$$ where a, A have real entries. Thus,

$$M = A\mathrm{diag}(a)A^T + B\mathrm{diag}(b)B^T + \overline{B\mathrm{diag}(b)B^T}. \tag{27}$$

Write $a = a_+ - a_-$, where $a_+, a_- \geq 0$ and a_+ (resp., a_-) has s_+ (resp., s_-) positive entries. We have:

$$A\mathrm{diag}(a)A^T = A\mathrm{diag}(a_+)A^T - A\mathrm{diag}(a_-)A^T. \tag{28}$$

Consider now the matrix $B\mathrm{diag}(b)B^T$. Set $T_1 := \{v \in T \mid w_v \neq 0\}$ with cardinality t, let B_1 be the submatrix of B consisting of its columns indexed by T_1, set $c_1 := (\sqrt{b_v})_{v \in T_1}$ and $C := B_1\mathrm{diag}(c_1)$. Then, C has t columns and $B\mathrm{diag}(b)B^T = CC^T$. Decomposing C as $C = E + iF$ where E, F are real matrices, we find:

$$B\mathrm{diag}(b)B^T + \overline{B\mathrm{diag}(b)B^T} = CC^T + \overline{CC^T} = 2EE^T - 2FF^T. \tag{29}$$

Combining (27), (28) and (29), we obtain:

$$M = (A\mathrm{diag}(a_+)A^T + 2EE^T) - (A\mathrm{diag}(a_-)A^T + 2FF^T) = P - Q \tag{30}$$

where $P := A\mathrm{diag}(a_+)A^T + 2EE^T \succeq 0$ and $Q := A\mathrm{diag}(a_-)A^T + 2FF^T \succeq 0$.

On the other hand, as M has σ_+ positive eigenvalues and σ_- negative eigenvalues, we can write

$$M = \sum_{i=1}^{\sigma_+} u_i u_i^T - \sum_{j=1}^{\sigma_-} v_j v_j^T, \tag{31}$$

where $u_1, \ldots, u_{\sigma_+}, v_1, \ldots, v_{\sigma_-}$ are pairwise orthogonal vectors. We claim that

$$\sigma_+ \leq s_+ + t. \tag{32}$$

For this, observe first that if a vector x lies in the linear span of the vectors $u_1, \ldots, u_{\sigma_+}$ and belongs to the kernel of P, then $x = 0$. Indeed, using (31), we find that $x^T M x = x^T(\sum_i u_i u_i^T)x \geq 0$ and, using (30), we find that $x^T M x = -x^T Q x \leq 0$. Therefore, $x^T M x = 0$ which implies that $x = 0$ as the u_i's are pairwise orthogonal. Therefore, $\sigma_+ \leq \text{rank} P \leq s_+ + t$. Similarly, one can verify that

$$\sigma_- \leq s_- + t. \tag{33}$$

Summing relations (32) and (33), we obtain that $\sigma_+ + \sigma_- \leq s_+ + s_- + 2t$. Now, equality holds since both sides of the inequality are equal to the rank of M. Therefore, equality also holds in (32) and in (33). This concludes the proof, as the remaining statements in Proposition 17 follow as a direct consequence. ∎

Remark 18 *Theorem 11 can be derived from Proposition 17. This follows from the fact that the matrix S_h can be expressed as $S_h = Z\text{diag}(w)Z^T$, after setting $w_v := h(v)$ for $v \in V$. This identity follows directly from the definition of S_h and using Theorem 10 (iii). If the ideal $\mathcal{I}$ is not radical, i.e., if the roots $v \in V$ have a multiplicity $m(v) \geq 1$, one should define Z to be the matrix with column ζ_v repeated $m(v)$ times, so that Z is a square matrix of size $\mathcal{B}$, and one should assume that the entries of w indexed by various copies of $v \in V$ are identical. It is not difficult to see that the proof of Proposition 17 extends to this setting.*

Lemma 16 together with Proposition 17 permit to conclude that, for $y \in \mathbb{R}^{\mathcal{B}}$,

$$M_{\mathcal{B}}(y) \succeq 0 \Longleftrightarrow (Z^{-1}y)_v \geq 0 \ \forall v \in S, \ (Z^{-1}y)_v = 0 \ \forall v \in V \setminus \mathbb{R}^n.$$

In order to get a complete characterization of the polytope $P_{\mathcal{B}}$ from (34), we need to ensure that $(Z^{-1}y)_v = 0$ for $v \in S \setminus F$. This can be done by involving the inequality constraints $h_j(x) \leq 0$ ($j = m+1, \ldots, m+k$) present in the description of F and requiring positive semidefiniteness of the combinatorial moment matrices $M_{\mathcal{B}}(h_j * y)$ ($j = m+1, \ldots, m+k$). For this, we need an appropriate definition for the shifted operator $h * y$. If we recall the definition: $(h*y)_\alpha = \sum_\beta h_\beta y_{\alpha+\beta}$ given in Definition 1 (ii) in the classic case of a sequence y indexed by all monomials, we note that this can be seen as applying the 'multiplication by h' operator to the vector y. As we are now working in the quotient space $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$, we adopt the following definition.

Definition 19 *Given $y \in \mathbb{R}^{\mathcal{B}}$, define $h * y \in \mathbb{R}^{\mathcal{B}}$ by $h * y = M_h^T y$.*

Using the definition of the residue from Lemma 13 (ii), one can easily verify the following expression for M_h.

Lemma 20 $M_h = (Z^T)^{-1}\text{diag}(h(v))_{v \in V} Z^T$.

From this follows the following extension of Lemma 16 to the matrix $M_{\mathcal{B}}(h * y)$, which generalizes results from [20] and [16] for the 0/1 and ± 1 cases.

LEMMA 21 *Given* $y \in \mathbb{R}^{\mathcal{B}}$ *and* $h \in \mathbb{R}[x_1, \ldots, x_n]$*, define* $w \in \mathbb{C}^V$ *by* $w_v := (Z^{-1}y)_v h(v)$ *for* $v \in V$*. Then,* $M_{\mathcal{B}}(h * y) = Z\mathrm{diag}(w)Z^T$*.*

We can now state the main result of this section.

THEOREM 22 *Assume that* $\mathcal{I}$ *is radical and* V *is finite, and let* $\mathcal{B}$ *be a basis of* $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$*. Given* $y \in \mathbb{R}^{\mathcal{B}}$*, the following assertions are equivalent.*

(i) *The vector* y *is a conic combination of the columns of the matrix* Z *indexed by points in* F*; that is,* $Z^{-1}y \geq 0$, $(Z^{-1}y)_v = 0$ *for all* $v \in V \setminus F$*.*

(ii) $M_{\mathcal{B}}(y) \succeq 0$ *and* $M_{\mathcal{B}}(h_j * y) \succeq 0$ *for all* $j = m + 1, \ldots, m + k$*.*

PROOF. Directly from Lemmas 16, 21 and Proposition 17. ∎

COROLLARY 23 *Assume that* $\mathcal{I}$ *is radical and* V *is finite, and let* $\mathcal{B}$ *be a monomial basis of* $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$ *(w.r.t. some monomial ordering). Then, problem (1) (or (24)) is equivalent to*

$$\min r^T y \quad s.t. \; M_{\mathcal{B}}(y) \succeq 0, \; M_{\mathcal{B}}(h_j * y) \succeq 0 \; (j = m + 1, \ldots, m + k), \; y_0 = 1, \tag{34}$$

where $r(x) = \sum_{\beta \in \mathcal{B}} r_\beta x^\beta$ *denotes the residue of the polynomial* $f(x)$ *w.r.t.* $\mathcal{B}$*.*

PROOF. Directly from Theorem 22 after noting that the condition $e^T Z^{-1} y = 1$ is equivalent to $y_0 = 1$, since the row of Z indexed by the degree 0 monomial (which does belong to $\mathcal{B}$) is the all-ones vector. ∎

EXAMPLE 24 *Given* $w \in \mathbb{R}^{\binom{n}{2}}$*, the* max-cut problem *is the problem:*

$$\max \sum_{1 \leq i < j \leq n} w_{ij}(1 - x_i x_j) \quad s.t. \; x_i^2 = 1 \; (i = 1, \ldots, n),$$

which can therefore be reformulated as

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

where $M_{\mathcal{B}}$ *is the combinatorial moment matrix indexed by all subsets of* $\{1, \ldots, n\}$*, with* (A, B)*th entry* $y_{A \Delta B}$*. (See [17] for a detailed treatment.)*

Example 25 *([6], p. 229) Consider an optimization problem over the set*

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

Let $\mathcal{I}$ 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 $\mathcal{I}$. 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:

$$Z = \begin{array}{c} \\ 1 \\ x \\ y \\ y^2 \end{array}\!\!\begin{array}{c} \begin{array}{cccc} v_1 & v_2 & v_3 & v_4 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 0 & -\frac{1}{2} \\ 0 & 0 & 1 & \frac{3}{4} \\ 0 & 0 & 1 & \frac{9}{16} \end{pmatrix} \end{array}, \quad Z^{-1} = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array}\!\!\begin{array}{c} \begin{array}{cccc} 1 & x & y & y^2 \end{array} \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{6} & -\frac{2}{3} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{5}{2} & 2 \\ 0 & 0 & -3 & 4 \\ 0 & 0 & \frac{16}{3} & -\frac{16}{3} \end{pmatrix} \end{array}$$

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

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

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 $\underline{u} = (1, a, b, c) \in \mathbb{R}^{\mathcal{B}}$ belongs to the polytope $P_{\mathcal{B}}$ (i.e., the convex hull of the columns of Z indexed by F) if and only if $M_{\mathcal{B}}(\underline{u}) \succeq 0$ and $M_{\mathcal{B}}(h * \underline{u}) \succeq 0$, where $h * \underline{u} = M_h^T \underline{u} = (3/2 - 2b, 3a/2 + 4b - 4c, 3b/2 - 2c, 3b/2 - 2c)^T$. Recall that

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

2.2 Approach via sums of squares modulo the ideal $\mathcal{I}$

Let us now consider the dual semidefinite program of the semidefinite program (34). The basis $\mathcal{B}$ is now assumed to be the set of standard monomials w.r.t. some monomial ordering.

For convenience, $\mathcal{B}$ also denotes the set of exponents β for which x^β is a standard monomial. Moreover, for a polynomial $q(x)$, the notation: $q \in \mathbb{R}^{\mathcal{B}}$ means that $q(x)$ uses only standard monomials, i.e., $q(x) = \sum_{\beta\in\mathcal{B}} q_\beta x^\beta$. In other words, we identify as usual a polynomial and its sequence of coefficients.

Write $M_{\mathcal{B}}(h_j * y) = \sum_{\beta\in\mathcal{B}} C^j_\beta y_\beta$; setting $h_0(x) := 1$, $M_{\mathcal{B}}(y) = M_{\mathcal{B}}(h_0 * y)$ holds. The dual semidefinite program of (34) reads:

$$\begin{array}{lll} \rho^* := & \max & r_0 - \langle C^0_0, Z_0\rangle - \sum_{j=m+1}^{m+k}\langle C^j_0, Z_j\rangle \\ & \text{s.t.} & \langle C^0_\beta, Z_0\rangle + \sum_{j=m+1}^{m+k}\langle C^j_\beta, Z_j\rangle = r_\beta \ (\beta \in \mathcal{B}\setminus\{0\}) \\ & & Z_0, Z_{m+1}, \ldots, Z_{m+k} \succeq 0. \end{array}$$

Analogously to the fact that the dual of (7) is equivalent to (8), the dual program of (34) is equivalent to

$$\begin{array}{ll} \rho^* = & \max \rho \\ & \text{s.t. } r(x) - \rho = (\sum_{i_0} q^2_{0,i_0}) + \sum_{j=m+1}^{m+k} h_j(\sum_{i_j} q^2_{j,i_j}) \ + q \\ & \text{where all } q_{j,i_j} \text{ belong to } \mathbb{R}^{\mathcal{B}} \text{ and } q \in \mathcal{I}. \end{array} \tag{35}$$

Indeed, let Z_j $(j = 0, m+1, \ldots, m+k)$ be a dual feasible solution with objective value ρ. Then, for all j, $Z_j = \sum_{i_j} q_{j,i_j} q^T_{j,i_j}$ for some $q_{j,i_j} \in \mathbb{R}^{\mathcal{B}}$ and

$$\begin{array}{l} r(x) - \rho = \sum_{\beta\neq 0} r_\beta x^\beta + r_0 - \rho = \sum_{\beta\neq 0} x^\beta(\sum_j \langle C^j_\beta, Z_j\rangle) + \sum_{j\geq m+1}\langle C^j_0, Z_j\rangle \\ = \sum_j \langle Z_j, \sum_\beta C^j_\beta x^\beta\rangle = \sum_j \sum_{i_j}\langle q_{j,i_j} q^T_{j,i_j}, \sum_\beta C^j_\beta x^\beta\rangle, \end{array}$$

which is congruent modulo $\mathcal{I}$ to $\sum_j h_j(x)(\sum_{i_j} q_{j,i_j}(x)^2)$. Indeed, $\sum_\beta C^j_\beta x^\beta = M_{\mathcal{B}}(h_j * y)$, after setting $y_\beta := x^\beta$ for $\beta \in \mathcal{B}$. Moreover, for $\alpha \in \mathbb{Z}^n_+$, the polynomial $h_j * y(\alpha)$ is congruent modulo $\mathcal{I}$ to $h_j(x)x^\alpha$. Therefore, for $q \in \mathbb{R}^{\mathcal{B}}$, $\langle qq^T, \sum_\beta C^j_\beta x^\beta\rangle = q^T M_{\mathcal{B}}(h_j * y)q$ is congruent modulo $\mathcal{I}$ to $\sum_{\beta,\beta'\in\mathcal{B}} q_\beta q_{\beta'} h_j(x) x^{\beta+\beta'} = h_j(x)q(x)^2$. The above argument can be reversed, thus showing equality in (35).

Weak semidefinite duality implies $\rho^* \leq p^*$. The following result of Parrilo [24] implies equality: $\rho^* = p^*$. Hence, there is no duality gap between (34) and its dual (35) (although (34) has no strictly feasible solution if $S \neq V$).

Theorem 26 *([24]) Assume that $\mathcal{I}$ is radical and V is finite. Then, every nonnegative polynomial f on F can be written as $f = (\sum_{i_0} q^2_{0,i_0}) + \sum_{j=m+1}^{m+k} h_j(\sum_{i_j} q^2_{j,i_j}) + q$, where $q \in \mathcal{I}$ and $q_{j,i_j} \in \mathbb{R}^{\mathcal{B}}$ for all j, i_j.*

Proof. In a first step, we show that every polynomial f nonnegative on S has a decomposition $q + \sum_i q_i^2$, where $q \in \mathcal{I}$ and $q_i \in \mathbb{R}^{\mathcal{B}}$. Let p_v be interpolation polynomials for $v \in V$; that is, $p_v(v) = 1$ and $p_v(v') = 0$ for $v' \in V \setminus \{v\}$. Such polynomials are provided by the rows of the matrix Z^{-1} and, thus, they can be chosen in $\mathbb{C}^{\mathcal{B}}$. We have $p_{\overline{v}} = \overline{p_v}$ for all $v \in V$. Partition V into $S \cup T \cup \overline{T}$, where $\overline{T} = \{\overline{v} \mid v \in T\}$, set $\gamma_v := \sqrt{f(v)}$ for $v \in V$, and $q_v := \gamma_v p_v$ for $v \in S$, $q_v := \gamma_v p_v + \overline{\gamma_v} p_{\overline{v}}$ for $v \in T$. Then, $q_v \in \mathbb{R}^{\mathcal{B}}$ and $f - \sum_{v \in S \cup T} (q_v)^2$ belongs to $\mathcal{I}$ since it vanishes at all $v \in V$.

Suppose now that f is nonnegative on F. We define polynomials $s_0, s_{m+1}, \ldots, s_{m+k} \in \mathbb{R}^{\mathcal{B}}$ satisfying the following prescribed values at 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 $j_v \geq m+1$; then $s_{j_v}(v) = \frac{f(v)}{h_{j_v}(v)}$ and $s_0(v) = s_j(v) = 0$ for all remaining j. (Note that the unique polynomial $s \in \mathbb{C}^{\mathcal{B}}$ taking prescribed values λ_v at $v \in V$ is given by $s = (Z^{-1})^T \lambda$.) Then, each of $s_0, s_{m+1}, \ldots, s_{m+k}$ is a real polynomial which is nonnegative on S; by the above, it can be decomposed modulo $\mathcal{I}$ as a sum of squares of polynomials in $\mathbb{R}^{\mathcal{B}}$. The result now follows since $f - s_0 - \sum_{j=m+1}^{m+k} s_j h_j$ belongs to $\mathcal{I}$. ∎

Corollary 27 *Assume that $\mathcal{I}$ is radical, V is finite, and $\mathcal{B}$ is a monomial basis of $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$. Then the programs (1), (34) and (35) are equivalent.*

Proof. As the polynomial $r(x) - p^*$ is nonnegative on F, we deduce using Theorem 26 that p^* is feasible for (35) and, thus, $p^* \leq \rho^*$. ∎

3 Relation with Other Semidefinite Representations

In this section, we relate our result about the semidefinite representation (34) for problem (1) and the result of Lasserre [14, 15] showing the finite convergence of the programs (7) and (8) in the grid case. The following notion of flat extension (introduced in [8]) will be useful.

Definition 28 *Let X be a symmetric matrix with block decomposition $X = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$. One says that X is a* flat extension *of A if* rank X = rank A *or, equivalently, if $B = AW$ and $C = B^T W = W^T A W$ for some matrix W. Observe that, if X is a flat extension of A, then $X \succeq 0 \iff A \succeq 0$.*

Curto and Fialkow [8] show the following results about flat extensions of moment matrices.

Theorem 29 *[8] 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, one can extend y to a vector in $\mathbb{R}^{S_{2t+2}}$, again denoted by y, in such a way that $M_{t+1}(y)$ is a flat extension of $M_t(y)$.*

Corollary 30 *[8] 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, one can extend y to a vector in $\mathbb{R}^{\mathbb{Z}^n_+}$ which is the sequence of moments of a nonnegative measure.*

PROOF. By repeatedly applying Theorem 29, we find an extension $y \in \mathbb{R}^{\mathbb{Z}^n_+}$ such that the matrix $M(y) = (y_{\alpha+\beta})_{\alpha,\beta \in \mathbb{Z}^n_+}$ is positive semidefinite with a finite rank. Then, using the Riesz representation theorem, Curto and Fialkow ([8], chapter 4) prove the existence of a nonnegative measure whose sequence of moments is y. ∎

The above mentioned result of Curto and Fialkow plays a central role in the finite convergence result in the 0/1 and grid cases proved by Lasserre [14, 15]; see Proposition 32 and Example 33 below. Given an integer $t \geq 1$, a strategy for proving equivalence of the programs (1) and (7) is indeed to show that $M_t(y)$ is a flat extension of $M_d(y)$, where d is the maximum degree of a monomial in a monomial basis $\mathcal{B}$ (w.r.t. a given monomial ordering). A way to achieve this is to show that, for all $\alpha \in S_t$, the polynomials $f^{(\alpha)}(x) := x^\alpha - r^{(\alpha)}(x)$, where $r^{(\alpha)}(x)$ is the residue modulo $\mathcal{I}$ 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 [8] and its proof is included for completeness.

LEMMA 31 *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 α-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 $(\alpha+e_i)$-component of $M_t(y)f$ and, thus, to zero. ∎

Recall that $v_j = \lceil \frac{\deg(h_j)}{2} \rceil$ for all j. Set

$$v_0 := \max(v_1, \ldots, v_m) \quad \text{and} \quad v_+ := \max(v_{m+1}, \ldots, v_{m+k}); \quad \text{thus, } v = \max(v_0, v_+)$$

($v_+ = 0$ if $k = 0$). Following an analogous reasoning as in [15], we can prove:

PROPOSITION 32 *Assume that $\mathcal{I}$ is radical, $V = S$ is finite, and $h_1, \ldots, h_m$ form a Groebner basis of $\mathcal{I}$ with respect to some monomial ordering. Let d be the maximum degree of a monomial in the corresponding basis of standard monomials, and let $t \geq \max(d+2, 2v_0+1, d+v_+)$. The problems (1) and (7) are equivalent when the polynomial f in the objective function has degree $\leq 2t-2$.*

PROOF. Let $y \in S_{2t}$ be a feasible solution to (7). As $t \geq 2v_0$, $M_{t-v_j}(h_j * y) = 0$ implies $M_t(y)h_j = 0$ for all $j \leq m$. This implies $M_t(y)h = 0$ for any $h \in \mathcal{I}$ with $\deg(h) \leq t-1$. Indeed, such h can be decomposed as $\sum_{j=1}^m h_j u_j$ where $\deg(u_j h_j) \leq \deg(h) \leq t-1$ if $u_j \neq 0$ (since $h_1, \ldots, h_m$ form a Groebner basis of $\mathcal{I}$); Lemma 31 implies $M_t(y)h = 0$. In particular, for any $\alpha \in S_{t-1}$, the polynomial $f^{(\alpha)}(x) = x^\alpha - r^{(\alpha)}(x)$ belongs to the kernel of $M_t(y)$. This implies that $M_{t-1}(y)$ is a flat extension of $M_d(y)$. Now Corollary 30 implies that $(y_\alpha)_{\alpha\in S_{2t-2}}$ is the sequence of moments of a nonnegative measure μ. By Lemma 2 (i), the support of μ is contained in the set of zeros of each h_j $(j = 1,\dots,m)$ (since $M_{t-1}(y)h_j = 0$ as $t \geq 2v_0 + 1$) and, thus, in S. Say, $\mu = \sum_{v\in S} \lambda_v\delta_v$ where $\lambda_v \geq 0$. For $v \in S\setminus F$, let $p \in \mathbb{R}^{\mathcal{B}}$ be an interpolation polynomial for v [i.e., $p(v) = 1$ and $p(v') = 0$ for $v' \in V \setminus \{v\}$] and let $j \geq m + 1$ for which $h_j(v) < 0$. Then, $p^T M_{t-v_j}(h_j * y)p = \int p(x)^2 h_j(x)\mu(dx) = \lambda_v h_j(v) \geq 0$, which implies that $\lambda_v = 0$ (we have used here the fact that $t \geq d + v_+$). Hence, the support of μ is contained in F. From this follows that $p^*_{t-1} \geq p^*$ and, thus, $p^* = p^*_{t-1} = p^*_t$. ∎

Example 33 The grid case. *For $j = 1,\dots,n$, let $a^{(j)}_1,\dots,a^{(j)}_{m_j}$ be distinct real numbers and define the polynomial $h_j(x) = \prod_{i=1}^{m_j}(x_j - a^{(j)}_i)$ of degree m_j. All common zeros of $h_1,\dots,h_n$ are real and they form an n-dimensional grid S. W.r.t. the graded lexicographic order, the set $\mathcal{B}$ of standard monomials consists of the monomials x^α with $\alpha_j \leq m_j - 1$ for all j. Thus, $d = \sum_{j=1}^n(m_j - 1)$. The ideal $\mathcal{I}$ generated by $h_1,\dots,h_n$ is radical. (To see it, it suffices to show that the Zeta matrix Z from (23) is nonsingular. This can be done using induction on $n \geq 1$; see the Appendix in [15].) Moreover, $h_1,\dots,h_n$ form a Groebner basis of $\mathcal{I}$. Thus, the assumptions in Proposition 32 hold and we can conclude that $p^* = p^*_t$ for $t \geq \max(d+2, 2v+1)$. Finite convergence in $v+d$ steps is shown in [15] and in $\max(d+1, v)$ steps in Proposition 34 below; note that $v+d \geq \max(d+2, 2v+1) \geq \max(d+1, v)$, assuming $v \geq 2$ for the first inequality.*

We now give a sharpening of Proposition 32, based on Parrilo's result from Theorem 26.

Proposition 34 *Assume that $\mathcal{I}$ is radical, V is finite, and $h_1,\dots,h_m$ form a Groebner basis of $\mathcal{I}$ with respect to some monomial ordering. Let d be the maximum degree of a monomial in the corresponding basis $\mathcal{B}$ of standard monomials, and let $t \geq \max(d + v_+ + 1, v_0)$ (resp., $t \geq \max(d + v_+, v_0)$ if $h_1,\dots,h_m$ have an even degree).*

(i) *Problems (1) and (7) are equivalent when the polynomial f in the objective function satisfies $\deg(f) \leq 2(d + v_+)$.*

(ii) *For $y \in \mathbb{R}^{S_{2t}}$, if* **(a)** *$M_t(y) \succeq 0$, $M_{t-v_j}(h_j * y) = 0$ $(j = 1,\dots,m)$, $M_{t-v_j}(h_j * y) \succeq 0$ $(j = m+1,\dots,m+k)$, $y_0 = 1$, then* **(b)** *$(y_\alpha)_{\alpha\in S_{2d+2v_+}}$ is the sequence of moments of a probability measure supported by F (and, thus, its projection $y := (y_\beta)_{\beta\in\mathcal{B}}$ is feasible for (34)).*

Proof. (i) The polynomial $f(x) - p^*$ is nonnegative on F. Therefore, by Theorem 26, it can be decomposed as $q + s_0 + \sum_{j=m+1}^{m+k} s_j h_j$, where $q \in \mathcal{I}$, s_0, s_{m+1}, s_{m+k} are sums of squares of polynomials belonging to $\mathbb{R}^{\mathcal{B}}$. Thus, $\deg(q) \leq 2d + 2v_+$. As $h_1,\dots,h_m$ form a Groebner basis of $\mathcal{I}$, the division of q by $h_1,\dots,h_m$ yields a decomposition: $q = \sum_{j=1}^m u_j h_j$ satisfying $\deg(u_j h_j) \leq \deg(q) \leq 2d+2v_+$ if $u_j \neq 0$. Hence, $\deg(u_j) \leq 2d+2v_+ - \deg(h_j) <$

$2(d+v_+ +1-v_j)$. As $\deg(s_0) \leq 2d \leq 2d+2v_+$ and $\deg(s_j) \leq 2d \leq 2(d+v_+ -v_j)$ for $j \geq m+1$, the scalar p^* is feasible for (8) when $t \geq \max(d+v_+ +1, v_0)$. Hence, $p^* \leq \rho_t^*$ and, thus, equality holds in (11). When $\deg(h_j) = 2v_j$ for all $j = 1, \ldots, m$, p^* is feasible for the program (8) with $t \geq \max(d+v_+, v_0)$, yielding equality in (11).
(ii) Let A denote the set of vectors $y \in \mathbb{R}^{S_{2t}}$ satisfying (a), let B be the set of vectors $y \in \mathbb{R}^{S_{2d+2v_+}}$ satisfying (b), and let π denote the projection from $\mathbb{R}^{S_{2t}}$ onto the subspace $\mathbb{R}^{S_{2d+2v_+}}$. Obviously, $B \subseteq \pi(A)$. We show that $\pi(A) = B$. Equivalently, we show that for any $f \in \mathbb{R}^{S_{2d+2v_+}}$, minimizing the objective function $f^T y$ over $\pi(A)$ and over B yields the same optimum value. Indeed, the optimum over B is p^* while the optimum over $\pi(A)$ is p_t^* and both values coincide by (i). ∎

For instance, in the 0/1 case, we find the finite convergence result in $d+v_+ = n+v_+$ steps from [14].

4 Semidefinite Approximations of Low Order

From a practical point of view, the semidefinite formulation (34) for problem (1) is not very useful, since it involves matrices of size $|V| \times |V|$, with $|V| \geq |S|$ 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 (the most common case in combinatorial applications), one can choose $\mathcal{A} = \mathcal{B}_t$ consisting of the residues modulo $\mathcal{I}$ of all the monomials x^α for $\alpha \in S_t$, for any given $t \geq 1$. Increasing values of t yield tighter approximations of (1) at an increasing computational cost, however. By Theorem 22, the approximation solves problem (1) exactly at $t = d$. 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

$$S = \{x \in \mathbb{R}^n \mid x_j = a_j, b_j \ \text{ for all } j = 1, \ldots, n\}, \tag{36}$$

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

$$h_j(x) = (x_j - a_j)(x_j - b_j) = x_j^2 - s_j x_j - t_j, \ \text{ setting } s_j := a_j + b_j, \ t_j = -a_j b_j \tag{37}$$

for $j = 1, \ldots, n$. Let $\mathcal{I}$ be the ideal generated by $h_1, \ldots, h_n$ and $S = V$ the associated variety, given by (36). W.r.t. the graded lexicographic order, the set of standard monomials is the set $\mathcal{B}$ (also denoted as $\mathcal{B}(n)$) from (26). Given an integer $t = 1, \ldots, n$, $\mathcal{B}_t$ (also denoted as $\mathcal{B}_t(n)$) consists of the standard monomials $x^A = \prod_{i \in A} x_i$ having degree $|A| \leq t$ and can be identified with the collection of subsets of the set $N := \{1, \ldots, n\}$ having size

$\leq t$. 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 $\mathcal{I}$ of $x^A x^B$, and thus to

$$\sum_{C \subseteq A \cap B} \left(\prod_{i \in C} s_i \prod_{i \in (A \cap B) \setminus C} t_i \right) y_{(A \Delta B) \cup C}. \tag{38}$$

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 $\mathcal{I}$), 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, (38) 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, (38) reads $y_{A \Delta B}$.

THEOREM 35 *Let S and $h_1, \ldots, h_n$ be as in (36) and (37). 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 $\operatorname{rank} M_{\mathcal{B}_s}(y) \leq \sum_{i=1}^{s} \binom{t}{s}$ for some $1 \leq s \leq t$, then y is the sequence of moments of a probability measure supported by S. If, moreover, $\operatorname{rank} M_{\mathcal{B}_1}(y) \leq t$, then y has a representing measure supported by a subset of S of size $\leq 2^{t-1}$.*

We first establish a stronger version of Lemma 31, valid in the combinatorial setting.

LEMMA 36 *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 $\mathcal{I}$ 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 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 $\mathcal{I}$ 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:

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

Here are some equations relating the above matrices: (a) $S = D$; (b) $H = s_iD + t_iF$; (c) $C = s_iE + t_iR$; (d) $T = s_iK + t_iG$. 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 $\mathcal{I}$ of x^Ax^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 $\mathcal{I}$ to $x^{A_m\setminus i}x^{A_\ell}(s_ix_i + t_i) = s_ix^{A_m}x^{A_\ell} + t_ix^{A_m\setminus i}x^{A_\ell}$, which yields (b). Finally, for $m, m' \in M$, $x^{A_m}x^{A_{m'}} \equiv x^{A_m\setminus i}x^{A_{m'}\setminus i}(s_ix_i + t_i)$ 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. ∎

The next result will be used in the proof of Theorem 35.

Lemma 37 *Let $y \in \mathbb{R}^{\mathcal{B}(n)}$ and $\mathcal{B}(n) = \mathcal{B}$ as in (26). 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 $N = \{1, \ldots, n\}$ 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 $u_K \neq 0$ and $|K| = k$. Multiplying $u(x)$ by $x^{N\setminus K}$ yields the polynomial $u_Kx^N + \sum_{I\in\mathcal{B}_s, I\neq K} u_Ix^Ix^{N\setminus K}$, whose residue modulo $\mathcal{I}$ belongs to the kernel of $M_{\mathcal{B}}(y)$ (by Lemma 36, applied with $n = t$). The result now follows since, for each $I \neq K$ with $u_I \neq 0$, the residue modulo $\mathcal{I}$ of $x^Ix^{N\setminus K}$ is a linear combination of monomials x^H with $H \subseteq I \cup (N \setminus K)$ which is strictly contained in N. ∎

Proof of Theorem 35. 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 35 holds.

We can now assume that $t \ge 2$. Given a subset $A \subseteq N$ with $|A| = t$, we deduce from Lemma 37 (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 $\mathcal{I}$ w.r.t. $\mathcal{B}$. As $r(x)$ uses only monomials of degree $\leq 2t$, one can define $y_\alpha := \sum_{\beta\in\mathcal{B}_{2t}} r_\beta y_\beta$.] As $M_t(y)$ is a flat extension of $M_{t-1}(y)$, Corollary 30 implies that $(y_\alpha)_{\alpha\in S_{2t}}$ is the sequence of moments of a nonnegative measure μ. By construction, each polynomial $h_i(x) = x_i^2 - s_i x_i - t_i$ belongs to the kernel of $M_t(y)$ and, therefore, by Lemma 2 (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 $L \leq 2^{t-1}$ in the case when rank $M_{\mathcal{B}_1}(y) \leq t$. 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 0-th column of X is the all ones vector χ^W (setting $W := \{1, \ldots, L\}$), and the i-th 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 $\{\chi^W, \chi^{W_1}, \ldots, \chi^{W_n}\}$; this rank is $\leq t$ by assumption. The next claim permits to conclude that $L = |W| \leq 2^{t-1}$, which finishes the proof of Theorem 35.

Claim 38 *Let $A_1, \ldots, A_L$ be subsets of $\{1, \ldots, n\}$ and let Y be the $L \times (n+1)$ matrix whose rows are $(1, (\chi^{A_1})^T), \ldots, (1, (\chi^{A_L})^T)$. 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 $u_1, \ldots, u_n$ odd. Let O denote the set of $i = 1, \ldots, n$ for which u_i is odd. Then $|A_\ell \cap O| \equiv u_0$ (modulo 2) for all $\ell = 1, \ldots, L$. 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 $0, 1, \ldots, n-1$. 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, $L = |L_1| + |L_2| \leq 2^{t-1}$. ∎

The result from Theorem 35 in the case when $S = \{\pm 1\}^n$ and $s = 1$ was proven[2] in [17]. The proof given there is elementary (by induction on the number n of variables); in particular, it does not use Curto and Fialkow's result from Corollary 30. We can also give a proof for the more general result from Theorem 35 not using Curto and Fialkow's result. Indeed, the following combinatorial analogue of Theorem 29 holds, whose proof is delayed till the Appendix.

[2]The paper [17] 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.

Theorem 39 *Let S and $h_1, \ldots, h_n$ be as in (36) and (37). 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 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 $M_{\mathcal{B}_t}(y)$.*

Corollary 40 *Let S and $h_1, \ldots, h_n$ be as in (36) and (37). 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.*

Proof. Directly from Theorems 22 and 39. ∎

The condition about the rank of $M_{\mathcal{B}_s}(y)$ in Theorem 35 is best possible. For instance, in the 0/1 case with $n = 2$ and $s = t = 1$, the matrix $M_{\mathcal{B}_1}(y) := \begin{array}{c} \\ \emptyset \\ 1 \\ 2 \end{array}\!\!\begin{array}{c} \begin{array}{ccc} \emptyset & 1 & 2 \end{array} \\ \begin{pmatrix} 1 & 3/4 & 3/4 \\ 3/4 & 3/4 & 3/8 \\ 3/4 & 3/8 & 3/4 \end{pmatrix} \end{array}$ 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 $M_{\mathcal{B}_2}(y) = \begin{array}{c} \\ \emptyset \\ 1 \\ 2 \\ 12 \end{array}\!\!\begin{array}{c} \begin{array}{cccc} \emptyset & 1 & 2 & 12 \end{array} \\ \begin{pmatrix} 1 & 3/4 & 3/4 & 3/8 \\ 3/4 & 3/4 & 3/8 & 3/8 \\ 3/4 & 3/8 & 3/4 & 3/8 \\ 3/8 & 3/8 & 3/8 & 3/8 \end{pmatrix} \end{array}$ is not positive semidefinite, as the vector $x := (-3, 2, 2)$ 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 41 *Let $G = (N, E)$ be a graph with node set $N = \{1, \ldots, n\}$. 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 *$M_{\mathcal{B}_1}(y) \leq 2$, 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) $$y_i + y_j = 1 \text{ and } C_0 = C_i + C_j \text{ for any edge } ij \in E.$$

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) 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, $\alpha_1 \leq \ldots \leq \alpha_p$. 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 a 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 $M_{\mathcal{B}_1}(y) \leq 2$ in the formulation (22) 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

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

Burer, Monteiro and Zhang [4] proved that adding the constraint rank$(X) \leq 2$ in the above program yields again a formulation of the stable set problem. Thus, Proposition 41 is an analogue of their result.

Heuristics for max-cut based on low rank formulations are explored in [3]. Let us conclude with a question about the smallest integer t for which the semidefinite relaxation:

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

solves the max-cut problem exactly, for any weight function w. [The (A,B)th entry of $M_{\mathcal{B}_t}(y)$ being now defined as $y_{A\Delta B}$, as we are in the ± 1 setting.] It is shown in [18] that the smallest such integer satisfies $t \geq \lceil \frac{n}{2} \rceil$ and equality is conjectured (it holds for $n \leq 7$). In other words, it is conjectured that, for $t = \lceil \frac{n}{2} \rceil$,

$$M_{\mathcal{B}_t}(y) \succeq 0,\; y_0 = 1,\; y_I = 0 \;(|I| \text{ odd}) \Longrightarrow \begin{array}{l} (y_{ij})_{1\leq i<j\leq n} \text{ is a convex combination of the} \\ \text{`cut' vectors } (v_i v_j)_{1\leq i<j\leq n} \;(v \in \{\pm 1\}^n). \end{array}$$

Acknowledgments. We are grateful to a referee, Dimitri Pasechnik, for several suggestions that helped improve the presentation of the paper. In particular, he suggested to investigate how our results about combinatorial moment matrices relate with known results about the (Hermite) trace form for solving systems of polynomial equations in the 0-dimensional case.

References

[1] E. Balas, S. Ceria, and G. Cornuéjols. A lift-and-project cutting plane algorithm for mixed 0-1 programs. *Mathematical Programming*, 58:295–324, 1993.

[2] S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in Real Algebraic Geometry.* Springer, 2003.

[3] S. Burer, R.D.C. Monteiro, and Y. Zhang. Rank-two heuristics for max-cut and other binary quadratic programs. *SIAM Journal on Optimization*, 12:503–521, 2002.

[4] S. Burer, R.D.C. Monteiro, and Y. Zhang. Maximum stable set formulations and heuristics based on continuous optimization. Technical Report TR00-34, Computational and Applied Mathematics, Rice University, Houston, USA, 2000. [*Mathematical Programming*, to appear.]

[5] A.M. Cohen, H. Cuypers, and H. Sterk (eds.). *Some Tapas of Computer Algebra.* Springer, Berlin, 1999.

[6] D. A. Cox, J. B. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra.* Springer, 1997.

[7] D. A. Cox, J. B. Little, and D. O'Shea. *Using Algebraic Geometry.* Graduate Texts in Mathematics, Number 185, Springer, New York, 1997.

[8] R.E. Curto and L.A. Fialkow. Solution of the truncated complex moment problem for flat data. *Memoirs of the American Mathematical Society,* vol. 119, n. 568, 1996.

[9] R.E. Curto and L.A. Fialkow. The truncated complex K-moment problem. *Transactions of the American Mathematical Society*, 352:2825–2855, 2000.

[10] B. Fuglede. The multidimensional moment problem. *Expositiones Mathematicae*, 1:47–65, 1983.

[11] T. Jacobi and A. Prestel. Distinguished representations of strictly positive polynomials. *Journal für die Reine und Angewandte Mathematik*, 532:223–235, 2001.

[12] H. Landau. Classic background of the moment problem. In *Moments in Mathematics*, Proceedings of Symposia in Applied Mathematics, vol. 37, pages 1–15, AMS, Providence, 1987.

[13] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11:796–817, 2001.

[14] J.B. Lasserre. An explicit exact SDP relaxation for nonlinear $0-1$ programs. In K. Aardal and A.M.H. Gerards, eds., *Lecture Notes in Computer Science*, 2081:293–303, 2001.

[15] J.B. Lasserre. Polynomials nonnegative on a grid and discrete representations. *Transactions of the American Mathematical Society*, 354:631–649, 2001.

[16] M. Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming. *Mathematics of Operations Research*, 28:470–496, 2003.

[17] M. Laurent. Semidefinite relaxations for Max-Cut. In *The Sharpest Cut, Festschrift in Honor of M. Padberg's 60th Birthday,* M. Grötschel, ed., pages 291–327, MPS-SIAM. [To appear.]

[18] M. Laurent. Lower bound for the number of iterations in semidefinite relaxations for the cut polytope. *Mathematics of Operations Research.* [To appear.]

[19] L. Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, IT-25:1–7, 1979.

[20] L. Lovász and A. Schrijver. Cones of matrices and set-functions and $0-1$ optimization. *SIAM Journal on Optimization*, 1:166–190, 1991.

[21] Y. Nesterov. Squared functional systems and optimization problems. In J.B.G. Frenk, C. Roos, T. Terlaky, and S. Zhang, eds., *High Performance Optimization*, pages 405–440. Kluwer Academic Publishers, 2000.

[22] P.A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization.* PhD thesis, California Institute of Technology, May 2000.

[23] P.A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming B*, 96:293–320, 2003.

[24] P.A. Parrilo. An explicit construction of distinguished representations of polynomials nonnegative over finite sets. Preprint, ETH, Zürich, 2002.

[25] P. Pedersen, M.-F. Roy, and A. Szpirglas. Counting real zeros in the multivariate case. In *Computational Algebraic Geometry*, F. Eyssette and A. Galligo, eds., Birkhäuser, Boston, pages 203–224, 1993.

[26] V. Powers and T. Wörmann. An algorithm for sums of squares of real polynomials. *Journal of Pure and Applied Algebra*, 127:99–104, 1998.

[27] M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42:969–984, 1993.

[28] K. Schmüdgen. The K-moment problem for compact semi-algebraic sets. *Mathematische Annalen*, 289:203–206, 1991.

[29] H.D. Sherali and W.P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3:411–430, 1990.

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

[31] B. Sturmfels. *Solving Systems of Polynomial Equations.* CBMS, Regional Conference Series in Mathematics, Number 97, AMS, Providence, 2002.

Appendix: Proof of Theorem 39

Let $y \in \mathbb{R}^{\mathcal{B}_{2t}}$, set $M := M_{\mathcal{B}_t}(y)$ and let M_I denote the Ith column of M for $I \in \mathcal{B}_t$. By assumption, $M \succeq 0$ and M is a flat extension of $M_{\mathcal{B}_{t-1}}(y)$. This implies:

$$\text{For } \lambda \in \mathbb{R}^{\mathcal{B}_t}, \lambda \in \ker M \Longleftrightarrow \left(\sum_{I \in \mathcal{B}_t} \lambda_I M_I \right)_H = 0 \ \text{ for all } H \in \mathcal{B}_{t-1}. \tag{39}$$

Let $\mathcal{P}_{2t}$ denote the set of polynomials $f \in \mathbb{R}[x_1, \ldots, x_n]$ whose residue r modulo $\mathcal{I}$ belongs to $\mathbb{R}^{\mathcal{B}_{2t}}$, i.e., is of the form $\sum_{I \in \mathcal{B}_{2t}} r_I x^I$. We let φ denote the linear mapping from $\mathcal{P}_{2t}$ to $\mathbb{R}$ obtained by setting $\varphi(f) = \sum_{I \in \mathcal{B}_{2t}} r_I y_I$; in words, $\varphi(f)$ is obtained by linearization of the residue of f, where linearizing means replacing x^I by y_I. By assumption, M is the combinatorial moment matrix of y, which means that, for $A, B \in \mathcal{B}_t$, $M(A,B) = \varphi(x^A x^B)$, whose explicit expression is given in (38). Our objective is to construct an extension of y to $\mathbb{R}^{\mathcal{B}_{2t+2}}$ in such a way that $\tilde{M} = M_{\mathcal{B}_{t+1}}(y)$ is a flat extension of M. We proceed in two steps: First we construct a flat extension $\tilde{M}$ of M and then we show that $\tilde{M}$ is a combinatorial moment matrix.

Constructing a flat extension $\tilde{M}$ of M. We first construct a flat extension $\tilde{M} = \begin{pmatrix} M & B \\ B^T & C \end{pmatrix}$ of M, indexed by $\mathcal{B}_{t+1}$. For this, it suffices to set $B = MW$ and $C = B^T W$ for some matrix W. Let $K \subseteq \{1, \ldots, n\}$ with $|K| = t+1$ and let $k \in K$. As M is a flat extension of $M_{\mathcal{B}_{t-1}}(y)$,

$$M_{K \setminus k} = M\lambda \ \text{ for some } \lambda \in \mathbb{R}^{\mathcal{B}_{t-1}}; \tag{40}$$

that is, the polynomial $x^{K\setminus k} - \lambda(x)$ belongs to $\ker M$. Here and below, we identify again a polynomial and its sequence of coefficients; i.e., $\lambda(x) = \sum_{I \in \mathcal{B}_{t-1}} \lambda_I x^I$. Lemma 36 leads us to defining $\tilde{M}$ in such a way that the residue modulo $\mathcal{I}$ of the polynomial $x_k(x^{K\setminus k} - \lambda(x))$ belongs to its kernel. That is, the K-th column B_K of B is defined by

$$B_K = Mr, \ \text{ where } r \text{ is the residue modulo } \mathcal{I} \text{ of } x_k \lambda(x). \tag{41}$$

Then,

$$\tilde{M}(H,K) = \varphi(x^H r(x)) \text{ for } H \in \mathcal{B}_t. \tag{42}$$

Indeed, $\tilde{M}(K,H) = (Mr)_H = \sum_{I\in\mathcal{B}_t} M(H,I)r_I = \sum_{I\in\mathcal{B}_t} \varphi(x^H x^I)r_I = \varphi(x^H r(x))$. Moreover,

$$\varphi(x^H r(x)) = \varphi(x^H x^K) \text{ for } H \in \mathcal{B}_{t-1}. \tag{43}$$

Indeed, $\varphi(x^H r(x)) = \varphi(x^H x_k \lambda(x))$. Suppose first that $k \notin H$. Then, $\varphi(x^H x_k \lambda(x)) = \varphi(x^{H+k}\lambda(x)) = \sum_I \lambda_I \varphi(x^{H+k}x^I)$, which is equal to $\sum_I \lambda_I M(H+k,I) = (M\lambda)_{H+k} = M_{K\setminus k}(H+k) = M(H+k, K\setminus k) = \varphi(x^{K\setminus k}x^{H+k}) = \varphi(x^K x^H)$. Suppose now that $k \in H$. [Note that the following argument also holds for $|H| = t$.] Then, $\varphi(x^H x_k \lambda(x)) = \varphi(x^{H\setminus k}x_k^2\lambda(x)) = s_k\varphi(x^H\lambda(x)) + t_k\varphi(x^{H\setminus k}\lambda(x))$, which can be verified to be equal to $s_k M(H, K\setminus k) + t_k M(H\setminus k, K\setminus k) = s_k\varphi(x^{K\setminus k}x^H) + t_k\varphi(x^{K\setminus k}x^{H\setminus k}) = \varphi(x^{K\setminus k}x^{H\setminus k}(s_k x_k + t_k)) = \varphi(x^{K\setminus k}x^{H\setminus k}(x_k)^2) = \varphi(x^K x^H)$.

We can now verify that B_K is well defined, i.e., that the definition in (41) does not depend on the choice of λ or of $k \in K$. Indeed, if $\lambda' \in \mathbb{R}^{\mathcal{B}_{t-1}}$ also satisfies (40), then $\lambda - \lambda' \in \ker M$ which implies that the residue of $x_k(\lambda(x) - \lambda'(x))$ belongs to $\ker M$ (by Lemma 36, since $\lambda - \lambda'$ has degree $\leq t-1$). We now verify that B_K does not depend on the choice of $k \in K$. For this, let $k' \in K\setminus k$, $M_{K\setminus k'} = M\lambda'$ with $\lambda' \in \mathbb{R}^{\mathcal{B}_{t-1}}$, and r' the residue modulo $\mathcal{I}$ of $x_{k'}\lambda'(x)$. We show that $Mr = Mr'$. By (39), it suffices to show that $(Mr)_H = (Mr')_H$ for $H \in \mathcal{B}_{t-1}$. This holds, since $(Mr)_H = \sum_I M(H,I)r_I = \sum_I \varphi(x^H x^I)r_I = \varphi(x^H r(x))$, which is equal to $\varphi(x^H x^K)$ by (43) and, similarly, $(Mr')_H = \varphi(x^H x^K)$.

Next, we claim:

$$\tilde{M}(H,K) = \varphi(x^K x^H) \text{ for all } H \in \mathcal{B}_t \text{ with } |H \cup K| \leq 2t. \tag{44}$$

Indeed, by (42), $\tilde{M}(K,H) = \varphi(x^H r(x))$ is equal to $\varphi(x^H x^K)$ when $|H| \leq t-1$ by (43). When $|H| = t$, then $H \cap K \neq \emptyset$ and we can choose $k \in H \cap K$. Then, $\tilde{M}(H,K) = \varphi(x^H x^K)$ holds again by the argument used in the proof of (43).

$\tilde{M}$ is a combinatorial moment matrix. As above, let $|K| = t+1$, $k \in K$, and $\lambda \in \mathbb{R}^{\mathcal{B}_{t-1}}$ such that

$$\tilde{M}_{K\setminus k} = \tilde{M}\lambda,\ \tilde{M}_K = \tilde{M}r \text{ where } r \text{ is the residue modulo } \mathcal{I} \text{ of } x_k\lambda(x). \tag{45}$$

In order to show that $\tilde{M}$ is a combinatorial moment matrix, we have to prove that $\tilde{M}(H,K)$ depends only on the residue modulo $\mathcal{I}$ of $x^H x^K$, for all $|H| \leq t+1$. By the above, this holds for $|H| \leq t-1$. In view of (44), the case when $|H| = t$ will be settled if the following holds:

$$\tilde{M}(H,K) = \tilde{M}(H',K') \text{ for } |H|,|H'| = t, |K|,|K'| = t+1,\ H\cup K = H'\cup K',\ H\cap K = \emptyset. \tag{46}$$

It is sufficient to show that $\tilde{M}(H,K) = \tilde{M}(H+k, K\setminus k)$ for any $k \in K\setminus H$, since any pair (H',K') as in (46) can be reached from (H,K) by a sequence of such local exchanges. By

(45), $\tilde{M}(K \setminus k, H+k) = (\tilde{M}\lambda)_{H+k} = \sum_I \lambda_I \tilde{M}(H+k, I) = \sum_I \lambda_I \varphi(x^{H+k}x^I)$ (by (44)), which is equal to $\varphi(x^{H+k}\lambda(x)) = \varphi(x^H r(x)) = \tilde{M}(K, H)$ (by (42)).

Therefore, we can extend y to $\mathbb{R}^{\mathcal{B}_{2t+1}}$ in such a way that $\tilde{M}(H,K) = y_{H \cup K}$ if $|H \cup K| = 2t+1$, which also permits to extend the mapping φ to $\mathbb{R}^{\mathcal{B}_{2t+1}}$ by setting $\varphi(x^L) = y_L$ for $L \in \mathcal{B}_{2t+1}$. In other words, relation (44) remains valid for $|H \cup K| = 2t+1$; that is,

$$\tilde{M}(H,K) = \varphi(x^K x^H) \text{ for all } H \in \mathcal{B}_t. \tag{47}$$

Next, we claim:

$$\tilde{M}(H,K) = \varphi(x^H x^K) \text{ for } |H| = |K| = t+1 \text{ with } H \cap K \neq \emptyset. \tag{48}$$

Indeed, choose $k \in H \cap K$. Then, $\tilde{M}(H,K) = (\tilde{M}_K)_H = (\tilde{M}r)_H = \sum_{I \in \mathcal{B}_t} \tilde{M}(H,I) r_I$ is equal to $\sum_I \varphi(x^H x^I) r_I$ (by (47)) and thus to $\varphi(x^H r(x))$. Now, as in the proof of (43), $\varphi(x^H r(x)) = \varphi(x^H x_k \lambda(x)) = s_k \varphi(x^H \lambda(x)) + t_k \varphi(x^{H \setminus k}\lambda(x)) = s_k \tilde{M}(H, K \setminus k) + t_k \tilde{M}(H \setminus k, K \setminus k) = s_k (\tilde{M}\lambda)_H + t_k (\tilde{M}\lambda)_{H \setminus k} = s_k \tilde{M}(H, K \setminus k) + t_k \tilde{M}(H \setminus k, K \setminus k) = s_k \varphi(x^{K \setminus k} x^H) + t_k \varphi(x^{K \setminus k} x^{H \setminus k}) = \varphi(x^{K \setminus k} x^{H \setminus k}(s_k x_k + t_k)) = \varphi(x^K x^H)$.

The proof will be complete if we can show that

$$\tilde{M}(H,K) = \tilde{M}(H',K') \text{ for } |H|, |H'|, |K|, |K'| = t+1,\ H \cup K = H' \cup K',\ H \cap K = \emptyset. \tag{49}$$

For this, it suffices to show that $\tilde{M}(H,K) = \tilde{M}(H \setminus h + k, K \setminus k + h)$ for all $h \in H \setminus K$ and $k \in K \setminus H$, since any pair (H', K') as in (49) can be obtained from (H,K) by a sequence of such local exchanges. Let $\tilde{M}_K = \tilde{M}r$ as in (45) and $\tilde{M}_{K \setminus k + h} = \tilde{M}s$, where $s(x)$ is the residue modulo $\mathcal{I}$ of $x_h \lambda(x)$. Then, $\tilde{M}(K \setminus k + h, H \setminus h + k) = (\tilde{M}s)_{H \setminus h + k} = \sum_I \tilde{M}(H \setminus h + k, I) s_I = \varphi(x^{H \setminus h + k} s(x)) = \varphi(x^{H \setminus h + k} x_h \lambda(x)) = \varphi(x^H x_k \lambda(x)) = \varphi(x^H r(x))$ which is, therefore, equal to $\tilde{M}(H,K)$. This concludes the proof of Theorem 39.