Semidefinite Representations for Finite Varieties

Monique Laurent
CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands
monique@cwi.nl
February 13, 2003

Abstract

We present a concise semidefinite formulation for the problem of minimizing a polynomial over a semi-algebraic set defined by polynomial equalities and inequalities. When the polynomial equalities define a radical ideal $\mathcal{I}$ with a finite variety, this representation involves combinatorial moment matrices, indexed by a basis of $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$. The arguments are elementary and extend known facts for the grid case including 0/1 and ± 1 programming. Semidefinite approximations are obtained by considering truncated combinatorial moment matrices; rank conditions are given that ensure that the approximation solves the original problem at optimality.

1 Introduction

1.1 Preliminaries: Moments and sums of squares of polynomials

A central problem in combinatorial optimization 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

$$S = \{x \in \mathbb{R}^n \mid h_j(x) = 0 \ \text{ for } j = 1, \ldots, m\} \tag{2}$$

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 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; an alternative way to describe F is to require $x_i + x_j \leq 1$ for all $ij \in E$.

A classic approach is to linearize the problem (1) by introducing variables $y_\alpha = x^\alpha$ for the monomials x^α present in the problem. Thus, $f(x) = f^T y$, where $f = (f_\alpha)$ denotes the vector of coefficients of the polynomial $f(x)$. Therefore, problem (1) is the problem of minimizing $f^T y$ over the polytope P whose vertices are the vectors $(v^\alpha)_{\alpha \in S_t}$ for $v \in F$, where t is any integer $\geq \deg(f)$. Here and below, S_t denotes the set of sequences $\alpha \in \mathbb{Z}_+^n$ with $|\alpha| := \sum_{i=1}^n \alpha_i \leq t$. Various methods have been proposed in the literature for constructing relaxations of this polytope; that is, for generating new valid constraints from the given constraints $h_j(x) = 0$, $h_j(x) \geq 0$.

Combinatorial methods. In the 0/1 case, 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 [15], and the reformulation-linearization technique of Sherali and Adams [22] can all be cast in this framework. (We do not give here the exact details of applicability of these methods; see [8] 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).

Lovász and Schrijver [15] propose stronger semidefinite relaxations, obtained by requiring moreover semidefiniteness of the matrix $(y_{\alpha+\beta})_{\alpha,\beta \in S_1}$, after setting $y_{2e_i} = y_{e_i}$ for $i = 1, \ldots, n$.

Algebraic methods. Several other authors have proposed methods for constructing semidefinite relaxations of the problem (1), based on results about moment sequences and representations of nonnegative polynomials as sums of squares (see Nesterov [16], Lasserre [11], Parrilo [17, 18], Shor [23]); the basic ideas are reviewed below. It turns out that, in the 0/1 case, they yield relaxations that are at least as strong as the relaxations obtained via the above mentioned methods of Lovász and Schrijver, and Sherali and Adams. Details about the links between the various methods can be found in [8].

Problem (1) can be formulated as the problem of minimizing $\int f(x)\mu(dx)$ subject to μ being a probability measure supported by F. Therefore, (1) is equivalent to the problem:

$$\begin{array}{ll} \min & f^T y \\ \text{s.t.} & y \text{ is the sequence of moments of a probability measure supported by } F. \end{array} \tag{4}$$

Recall that, for a measure μ, its *moment of order* α is $\int x^\alpha \mu(dx)$.

If δ_v is the *Dirac measure* at $v \in \mathbb{R}^n$, with mass 1 at v and zero elsewhere, then its sequence of moments is the vector $(v^\alpha)_{\alpha \in \mathbb{Z}_+^n}$. When F is a finite set, every probability measure supported by F is atomic and can be written as a convex combination of Dirac measures at points of F. Therefore, we find again that (4) (and thus (1)) is the problem of minimizing $f^T y$ over the polytope whose vertices are the vectors $(v^\alpha)_{\alpha \in S_t}$ $(v \in F)$, for $t \geq \deg(f)$.

Given an integer $t \geq 1$ and a vector $y \in \mathbb{R}^{S_{2t}}$, its *moment matrix* of order t is the matrix $M_t(y)$ indexed by S_t with (α, β)-entry $y_{\alpha+\beta}$. Given two vectors $h, y \in \mathbb{R}^{\mathbb{Z}_+^n}$ with h having a finite support, define $h * y \in \mathbb{R}^{\mathbb{Z}_+^n}$ by

$$h * y(\alpha) = \sum_\beta h_\beta y_{\alpha+\beta} \quad \text{for all } \alpha \in \mathbb{Z}_+^n.$$

If y is the sequence of moments of a nonnegative measure μ supported by F, then $M_t(y) \succeq 0$, $M_t(h_j * y) = 0$ for $j = 1, \ldots, m$, $M_t(h_j * y) \succeq 0$ for $j = m+1, \ldots, m+k$, and $t \geq 1$. Indeed, for $p \in \mathbb{R}^{S_t}$, $p^T M_t(y) p = \int p(x)^2 \mu(dx) \geq 0$ and $p^T M_t(h_j * y) p = \int_F h_j(x) p(x)^2 \mu(dx)$ is 0 if $j \leq m$ and ≥ 0 if $j \geq m+1$. From this follows the next result which we collect in a lemma for further reference.

Lemma 1 *Let y be the sequence of moments of order $\leq 2t$ of some nonnegative measure μ. If $M_t(y)p = 0$ for some $p \in \mathbb{R}^{S_t}$, then the support of the measure μ is contained in the set of zeros of the polynomial $p(x)$.*

For $j = 1, \ldots, m+k$, set $v_j := \lceil \frac{\deg(h_j)}{2} \rceil$, and $v := \max(v_1, \ldots, v_{m+k})$. Following Lasserre [11], one can formulate the following semidefinite relaxation for (1):

$$\begin{array}{rll} p_t^* := & \min & \sum_\alpha f_\alpha y_\alpha \\ & \text{s.t.} & M_t(y) \succeq 0 \\ & & M_{t-v_j}(h_j * y) = 0 \ (j = 1, \ldots, m) \\ & & M_{t-v_j}(h_j * y) \succeq 0 \ (j = m+1, \ldots, m+k) \\ & & y_0 = 1 \end{array} \tag{5}$$

for any $t \geq \max(v, \lceil \frac{\deg(f)}{2} \rceil)$. Thus, $p_t^* \leq p^*$.

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. One can now relax this condition by requiring that $f(x) - p^*$ has a decomposition

$$u_0 + \sum_{j=1}^{m+k} u_j h_j \text{ where } u_0, u_{m+1}, \ldots, u_{m+k} \text{ are sums of squares of polynomials.} \tag{6}$$

A crucial observation is that sums of squares of polynomials can be tested using semidefinite programming (see, e.g., [17]), which leads to the following semidefinite relaxation of (1):

$$\begin{array}{rll} \rho_t^* := & \max & \rho \\ & \text{s.t.} & f(x) - \rho = u_0 + \sum_{j=1}^{m+k} h_j u_j \\ & & \text{where } u_0, u_{m+1}, \ldots, u_{m+k} \text{ are sums of squares of po lynomials} \\ & & \text{and } \deg(u_0) \leq 2t, \ \deg(u_j) \leq 2(t - v_j) \text{ for } j = 1, \ldots, m+k. \end{array} \tag{7}$$

Remark 2 *Note that program (7) 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* $X_{\alpha,\beta}$ $(\alpha, \beta \in S_t)$*:*

$$\sum_{\alpha,\beta | \alpha+\beta=\gamma} X_{\alpha,\beta} = u_\gamma \quad (\gamma \in S_{2t}) \tag{8}$$

$$X_{\alpha,\beta} = X_{\beta,\alpha} \quad (\alpha \neq \beta \in S_t). \tag{9}$$

This linear system is feasible since distinct equations in (8) 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* $u''_j(x)$*.*

It is proved in [11] that program (7) is, in fact, the semidefinite dual of (5); thus, weak semidefinite duality implies that

$$\rho_t^* \leq p_t^* \leq p^*. \tag{10}$$

A result in real algebraic geometry (Schmüdgen [21], Putinar [20], Jacobi and Prestel [7]) asserts that, when F is compact (and satisfies some additional technical condition), then every *positive* polynomial on F has a decomposition (6). This result implies the asymptotic convergence of the lower bounds ρ_t^*, p_t^* to p^* as t goes to infinity [11].

In the special case when F is contained in $\{0,1\}^n$ or is equal to the set of points in a grid, Lasserre [12, 13] shows the *finite* convergence of the bounds ρ_t^*, p_t^* to p^*. His proof is based on a result by Curto and Fialkow [5, 6] about flat extensions of moment matrices (which uses the Riesz representation theorem) (see Corollary 13). As a consequence, every polynomial *nonnegative* on F has a decomposition (6) 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 [19] proves (in an elementary way) that every polynomial *nonnegative* on F has a decomposition (6) with a priori bounds on the degrees of $u_0, u_{m+1}, \ldots, u_{m+k}$ (see Theorem 9). 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 $\mathcal{I}$.

This paper. In this paper, we observe that the result of Lasserre for the grid case extends to the same setting as Parrilo's result. That is, 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. See Section 3 for details.

Moreover, we give a semidefinite representation (21) for problem (1), more concise than (5). This representation involves 'combinatorial' moment matrices $M_\mathcal{B}(y)$, which are indexed by a set of monomials $\mathcal{B}$ forming a basis of the quotient space $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$; such set is smaller than S_t, for any t ensuring finite convergence of the Lasserre relaxations (5). The formulation does not contain any semidefinite constraint for the equations $h_j(x) = 0$ ($j \leq m$), as they have already been used for the construction of the moment matrix $M_{\mathcal{B}}(y)$. Such concise representation was mentioned in the grid case ([8, 9, 13]). Equivalence of (1) and (21) is an easy result, which follows from a simple combinatorial identity involving the Zeta matrix of the ideal $\mathcal{I}$ (see Lemma 6); this is a direct extension of the corresponding result given in [15] and [8] for the 0/1 and ± 1 cases (which also underlies the convergence result for combinatorial lift-and-project methods). Alternatively, the equivalence of (1) and (21) follows from Parrilo's result using semidefinite programming duality.

We present this 'combinatorial' argument in Section 2 and elaborate on the links with Lasserre's results in Section 3.

In Section 4, we consider approximations for problem (1) in the grid case, when S is the set $\{x \mid x_i = a_i, b_i \ \forall i = 1, \ldots, n\}$ (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 18.) We can prove this result in two ways; either by using Curto and Fialkow's results from Theorem 12 and Corollary 13, or by using their (elementary) combinatorial analogues provided by Theorems 8 and 22.

We 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 $G = (\{1, \ldots, n\}, E)$, introduced by Lovász [14]; it can be obtained as optimum value of the semidefinite program:

$$\max \sum_{i=1}^{n} y_i \text{ s.t. } M_{\mathcal{B}_1}(y) \succeq 0, \ y_{ij} = 0 \ (ij \in E), \ y_0 = 1. \tag{11}$$

If we impose that $M_{\mathcal{B}_1}(y)$ has rank 1 in (11), then the program solves the maximum stable set problem exactly. The same holds if we require only that rank $M_{\mathcal{B}_1}(y) \leq 2$; see Proposition 23, which is an analogue of a result of Burer, Monteiro and Zhang [3] given for another formulation of the theta number.

1.2 Preliminaries: Polynomial ideals and varieties

We group here some preliminaries on polynomial ideals. The interested reader may consult [4] for more information. Given an ideal $\mathcal{I}$ in $\mathbb{R}[x_1, \ldots, x_n]$, the set

$$V = V(\mathcal{I}) := \{x \in \mathbb{C}^n \mid h(x) = 0 \text{ for all } h \in \mathcal{I}\}, \tag{12}$$

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 $h_1, \ldots, h_m$. The set

$$I(V) := \{f \in \mathbb{R}[x_1, \ldots, x_n] \mid f(x) = 0 \ \text{ for all } x \in V\} \tag{13}$$

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

$$f = \sum_{j=1}^m u_j h_j + r, \tag{14}$$

$$\text{no term of } r \text{ is divisible by } \mathrm{LT}(h_j)(j = 1, \ldots, m) \text{ if } r \neq 0, \tag{15}$$

$$\deg(f) \geq \deg(u_j h_j) \text{ if } u_j \neq 0. \tag{16}$$

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

$$\mathcal{B} := \{x^\alpha \mid x^\alpha \notin \langle \mathrm{LT}(\mathcal{I}) \rangle\}; \tag{17}$$

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 $\mathcal{I}$.

Example 3 *([4], 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 $x^{12} - x$ 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 $\mathcal{B}_2 = \{1, x, x^2, x^3, y, xy, x^2y, x^3y, y^2, xy^2, x^2y^2, y^3\}$. Of course, $|\mathcal{B}_1| = |\mathcal{B}_2|$, but the maximum degree of a standard monomial is 11 in $\mathcal{B}_1$ and 4 in $\mathcal{B}_2$. The associated variety is $V = \{(0,0)\} \cup \{(x, x^7) \mid x^{11} = 1\}$.*

In what follows, we will often identify a polynomial $f(x) = \sum_\alpha f_\alpha x^\alpha$ with its vector of coefficients $f = (f_\alpha)_\alpha$. Moreover, we will use the same symbol $\mathcal{B}$ for denoting the set of sequences $\alpha \in \mathbb{Z}_+^n$ for which $x^\alpha \in \mathcal{B}$. Thus, $\mathbb{R}^{\mathcal{B}}$ denotes the space of polynomials of the form $r(x) = \sum_{\beta \in \mathcal{B}} r_\beta x^\beta$ as well as the space of their coefficient vectors $r = (r_\beta)_{\beta \in \mathcal{B}}$. Note that the only polynomial in $\mathbb{R}^{\mathcal{B}} \cap \mathcal{I}$ is the zero polynomial.

Every polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ is equivalent modulo $\mathcal{I}$ to a unique polynomial $r \in \mathbb{R}^{\mathcal{B}}$, called its *residue modulo $\mathcal{I}$*; this polynomial r will be chosen as representative of the class of f in the quotient space $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$. Therefore, the spaces $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$ and $\mathbb{R}^{\mathcal{B}}$ are isomorphic as $\mathbb{R}$-vector spaces. Moreover, they are finite-dimensional if and only if the variety V is a finite set, in which case $|V| \le |\mathcal{B}|$. Equality $|V| = |\mathcal{B}|$ holds if $\mathcal{I}$ is a radical ideal.

2 A Semidefinite Representation for a Finite Variety

In this section, we consider the problem (1) of minimizing a polynomial f over the semi-algebraic set F. Let $\mathcal{I}$ be the ideal generated by the polynomials $h_1, \ldots, h_m$ and let V be the associated complex variety as in (12). Thus,

$$S = V \cap \mathbb{R}^n.$$

We fix a monomial ordering and consider the associated set $\mathcal{B}$ of standard monomials as basis of $\mathbb{R}[x_1, \ldots, x_n]/\mathcal{I}$. Let d denote the highest degree of a monomial in $\mathcal{B}$. We assume from now on that the variety V is a finite set and that the ideal $\mathcal{I}$ is radical. (The latter assumption can be made without loss of generality since, if $\mathcal{I}$ is not radical, we can simply replace $\mathcal{I}$ by the radical ideal $I(V)$ defined in (13).) Then, $|V| = |\mathcal{B}|$.

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

$$Z(\beta, v) = v^\beta \quad \text{for } \beta \in \mathcal{B},\ v \in V \tag{18}$$

and let ζ_v denote its column indexed by $v \in V$. Thus, ζ_v is the sequence of moments of order $\beta \in \mathcal{B}$ of the Dirac measure δ_v. We call Z the *Zeta matrix* of the ideal $\mathcal{I}$. For $r \in \mathbb{R}^{\mathcal{B}}$, $r^T\zeta_v$ is equal to $r(v)$, the evaluation at $v \in V$ of the polynomial $r(x)$. The matrix Z is nonsingular. Indeed, for $r \in \mathbb{R}^{\mathcal{B}}$, $Z^Tr = 0$ means that $r(v) = 0$ for all $v \in V$ and, thus, $r \in \mathcal{I}$ (since $\mathcal{I}$ is radical), which implies that $r = 0$.

For instance, in the 0/1 case, $\mathcal{I}$ is generated by the polynomials $x_j^2 - x_j$ $(j = 1, \ldots, n)$; w.r.t. the graded lexicographic order, the set of standard monomials is

$$\mathcal{B} = \{x^A := \prod_{i \in A} x_i \mid A \subseteq \{1, \ldots, n\}\}; \tag{19}$$

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 [15]). In the ± 1 case, $\mathcal{I}$ is generated by the polynomials $x_j^2 - 1$ $(j = 1, \ldots, n)$ with (19) as set of standard monomials and $Z^{-1} = 2^{-n} Z$ (see [9]).

The problem (1) remains obviously unchanged if we replace the polynomial f by its residue $r \in \mathbb{R}^{\mathcal{B}}$ modulo $\mathcal{I}$. Therefore, it is equivalent to

$$\min\ r(v) \text{ s.t. } v \in F \quad = \quad \min\ r^T \underline{y} \ \text{ s.t. } \underline{y} \in \operatorname{conv}(\zeta_v \mid v \in F). \tag{20}$$

We now give a semidefinite representation for (1):

$$\min r^T \underline{y} \ \text{ subject to } M_{\mathcal{B}}(\underline{y}) \succeq 0,\ M_{\mathcal{B}}(h_j * \underline{y}) \succeq 0 \ (j = m+1, \ldots, m+k),\ \underline{y}_0 = 1 \tag{21}$$

where the variable $\underline{y}$ belongs to $\mathbb{R}^{\mathcal{B}}$ and $M_{\mathcal{B}}(\underline{y})$ is its 'combinatorial' moment matrix, indexed by $\mathcal{B}$. For $\beta, \beta' \in \mathcal{B}$, the (β, β')-th entry should be $\underline{y}_{\beta+\beta'}$ which is, thus, undefined if $\beta + \beta' \notin \mathcal{B}$. This value can, however, be defined by 'factoring through $\mathcal{I}$', i.e., by computing the residue modulo $\mathcal{I}$ of the monomial $x^{\beta+\beta'}$ and 'linearizing' by replacing each occurrence of x^δ $(\delta \in \mathcal{B})$ by $\underline{y}_\delta$. Namely,

Definition 4 *Given $\alpha \in \mathbb{Z}_+^n$, let $r^{(\alpha)} \in \mathbb{R}^{\mathcal{B}}$ denote the residue of the monomial x^α modulo $\mathcal{I}$, defined by*

$$r^{(\alpha)} = (Z^T)^{-1} \nu^{(\alpha)}, \quad \text{where } \nu^{(\alpha)} = (v^\alpha)_{v \in V} \in \mathbb{R}^V. \tag{22}$$

Given a vector $\underline{y} \in \mathbb{R}^{\mathcal{B}}$, extend $\underline{y}$ to a vector $y = (y_\alpha)_{\alpha \in \mathbb{Z}_+^n}$ by setting

$$y_\alpha = (r^{(\alpha)})^T \underline{y} \ \text{ for } \alpha \in \mathbb{Z}_+^n \tag{23}$$

and define the $\mathcal{B} \times \mathcal{B}$ matrix $M_{\mathcal{B}}(\underline{y})$, called the combinatorial moment matrix *of $\underline{y}$, with (β, β')-entry $y_{\beta+\beta'} = (r^{(\beta+\beta')})^T \underline{y}$ for $\beta, \beta' \in \mathcal{B}$.*

Example 5 *Given $w \in \mathbb{R}^{\binom{n}{2}}$, the* max-cut problem *is the problem:*

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

The set of standard monomials is the set $\mathcal{B}$ from (19), $d = n$, and the combinatorial moment matrix $M_{\mathcal{B}}(\underline{y})$ has its (A, B)-th entry equal to $\underline{y}_{A \Delta B}$ for all $A, B \subseteq N$. See [9] for a detailed treatment.

Given a graph $G = (N := \{1, \ldots, n\}, E)$ and $w \in \mathbb{R}^n$, the stable set problem *is the problem:*

$$\max \sum_{i=1}^{n} w_i x_i \quad s.t.\ x_i^2 - x_i = 0\ (i = 1, \ldots, n),\ x_i x_j = 0\ (ij \in E).$$

Then, the standard monomials are the monomials $\prod_{i \in A} x_i$ for which the set $A \subseteq N$ is stable in G. Thus, $d = \alpha(G)$, the maximum cardinality of a stable set in G. The combinatorial moment matrix $M_{\mathcal{B}}(\underline{y})$ is indexed by the stable sets and its (A, B)-th entry is $\underline{y}_{A \cup B}$ if $A \cup B$ is stable and 0 otherwise.

Hence, the equalities $h_j(x) = 0$ $(j \leq m)$ in the description of F are being used for the construction of the combinatorial moment matrix $M_{\mathcal{B}}(y)$. Each constraint $h(x) \geq 0$ (for $h = h_{m+1}, \ldots, h_{m+k}$) can be modeled by using the combinatorial moment matrix $M_{\mathcal{B}}(h * \underline{y})$. Indeed, every α-entry of $h * \underline{y}$ is $\sum_\gamma h_\gamma y_{\alpha+\gamma}$ which, using (23), can be expressed in terms of $\underline{y}$. We have: $h * \underline{y}(\alpha) = \sum_\gamma h_\gamma (Z^{-1}\underline{y})^T \nu^{(\alpha+\gamma)} = \sum_\gamma h_\gamma (\sum_{v \in V} v^{\alpha+\gamma} (Z^{-1}\underline{y})_v)$; that is,

$$h * \underline{y}(\alpha) = \sum_{v \in V} v^\alpha (Z^{-1}\underline{y})_v h(v). \tag{24}$$

The next lemmas give explicit information about the signature of a combinatorial moment matrix; they extend results from [15] and [8] for the 0/1 and ± 1 cases.

Lemma 6 *Given $\underline{y} \in \mathbb{R}^{\mathcal{B}}$ and $h \in \mathbb{R}[x_1, \ldots, x_n]$, define $w \in \mathbb{R}^V$ by $w_v := (Z^{-1}\underline{y})_v h(v)$ for $v \in V$. Then,*

$$M_{\mathcal{B}}(h * \underline{y}) = Z \operatorname{diag}(w) Z^T.$$

In particular, $M_{\mathcal{B}}(h * \zeta_v) = h(v) \zeta_v \zeta_v^T$ for any column ζ_v of the Zeta matrix Z.

Proof. The (β, β')-entry of the matrix $Z\operatorname{diag}(w)Z^T$ is equal to $\sum_{v \in V} Z_{\beta,v} Z_{\beta',v} w_v = \sum_{v \in V} v^\beta v^{\beta'} (Z^{-1}\underline{y})_v h(v)$, which is equal to the (β, β')-entry of $M_{\mathcal{B}}(h * \underline{y})$ by (24). ∎

Lemma 7 *Given $w \in \mathbb{C}^V$, if $Z\operatorname{diag}(w)Z^T$ is a real positive semidefinite matrix, then $w_v \geq 0$ for $v \in S = V \cap \mathbb{R}^n$ and $w_v = 0$ for $v \in V \setminus S$.*

Proof. For any element $v \in V$, its complex conjugate $\overline{v}$ also belongs to V. Hence, the set V can be partitioned into $V = S \cup T \cup \overline{T}$, where $\overline{T} = \{\overline{v} \mid v \in T\}$. Accordingly, w, Z and Z^{-1} have the following block decompositions:

$$w = \begin{array}{c} S \\ T \\ \overline{T} \end{array} \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \quad Z = \begin{array}{c} \begin{array}{ccc} S & T & \overline{T} \end{array} \\ \begin{pmatrix} A & B & \overline{B} \end{pmatrix} \end{array}, \quad Z^{-1} = \begin{array}{c} S \\ T \\ \overline{T} \end{array} \begin{pmatrix} C \\ D \\ \overline{D} \end{pmatrix}$$

where A, C are real matrices. Set $M := Z\text{diag}(w)Z^T$; that is, $\text{diag}(w) = Z^{-1}M(Z^{-1})^T$. We obtain:

$$\begin{array}{c} \\ S \\ T \\ \overline{T} \end{array} \begin{array}{c} \begin{array}{ccc} S & T & \overline{T} \end{array} \\ \begin{pmatrix} \text{diag}(a) & 0 & 0 \\ 0 & \text{diag}(b) & 0 \\ 0 & 0 & \text{diag}(c) \end{pmatrix} \end{array} = \begin{array}{c} \\ S \\ T \\ \overline{T} \end{array} \begin{array}{c} \begin{array}{ccc} S & T & \overline{T} \end{array} \\ \begin{pmatrix} CMC^T & CMD^T & CM\overline{D}^T \\ DMC^T & DMD^T & DM\overline{D}^T \\ \overline{D}MC^T & \overline{D}MD^T & \overline{D}M\overline{D}^T \end{pmatrix} \end{array}.$$

This implies that $c = \overline{b}$ and

$$CMC^T = \text{diag}(a),\ DMD^T = \text{diag}(b),\ CMD^T = CM\overline{D}^T = DM\overline{D}^T = 0.$$

By assumption, $M \succeq 0$. This implies that $a \geq 0$, since $\text{diag}(a) = CMC^T \succeq 0$. We now verify that $b = 0$. For this, write $D = E + iF$ where E is the real part of D and F its imaginary part. Then, $4EME^T = (D + \overline{D})M(D + \overline{D})^T = \text{diag}(b) + \text{diag}(\overline{b})$, and $-4FMF^T = (D - \overline{D})M(D - \overline{D})^T = \text{diag}(b) + \text{diag}(\overline{b})$, implying that $b + \overline{b} = 0$ and, thus, $EME^T = FMF^T = 0$. From this follows that $0 = DM\overline{D}^T = i(EMF^T - FME^T)$ and thus $EMF^T = FME^T$. Moreover, $(E + F)M(E + F)^T = EMF^T + FME^T \succeq 0$, $(E - F)M(E - F)^T = -EMF^T - FME^T \succeq 0$, which implies that $EMF^T = FME^T = 0$ and, therefore, $b = 0$. ∎

THEOREM 8 *Assume that $\mathcal{I}$ is radical and V is finite. Given $\underline{y} \in \mathbb{R}^{\mathcal{B}}$, the following assertions are equivalent.*

(i) *The vector $\underline{y}$ is a conic combination of the columns of the matrix Z indexed by points in F.*

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

(iii) *The vector $\underline{y}$ is the sequence of moments of order $\beta \in \mathcal{B}$ of a nonnegative atomic measure supported by F.*

PROOF. The equivalence (i) $\iff$ (iii) is obvious and (i) $\iff$ (ii) follows using Lemmas 6 and 7. ∎

This shows that problems (1), (20), and (21) are equivalent. Let us now consider the dual semidefinite program of (21). Write $M_{\mathcal{B}}(h_j * \underline{y}) = \sum_{\beta\in\mathcal{B}} C^j_\beta \underline{y}_\beta$; setting $h_0(x) := 1$, $M_{\mathcal{B}}(\underline{y}) = M_{\mathcal{B}}(h_0 * \underline{y})$ holds. The dual semidefinite program of (21) reads:

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

Analogously to the fact that the dual of (5) is equivalent to (7), the dual program of (21) is equivalent to

$$\begin{array}{rl} \rho^* = & \max \rho \\ & \text{s.t. } r(x) - \rho = (\sum_{i_0} q_{0,i_0}^2) + \sum_{j=m+1}^{m+k} h_j (\sum_{i_j} q_{j,i_j}^2) \ + q \\ & \text{where all } q_{j,i_j} \text{ belong to } \mathbb{R}^{\mathcal{B}} \text{ and } q \in \mathcal{I}. \end{array} \tag{25}$$

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

$$\begin{array}{l} r(x) - \rho = \sum_{\beta \neq 0} r_\beta x^\beta + r_0 - \rho = \sum_{\beta \neq 0} x^\beta (\sum_j \langle C_\beta^j, Z_j \rangle) + \sum_j \langle C_0^j, Z_j \rangle \\ = \sum_j \langle Z_j, \sum_\beta C_\beta^j x^\beta \rangle = \sum_j \sum_{i_j} \langle q_{j,i_j} q_{j,i_j}^T, \sum_\beta C_\beta^j x^\beta \rangle, \end{array}$$

which is congruent modulo $\mathcal{I}$ to $\sum_j h_j (\sum_{i_j} q_{j,i_j}^2)$. Indeed, $\sum_\beta C_\beta^j x^\beta = M_{\mathcal{B}}(h_j * \underline{y})$, after setting $\underline{y}_\beta := x^\beta$ for $\beta \in \mathcal{B}$. Moreover, for $\alpha \in \mathbb{Z}_+^n$, we have that $h_j * \underline{y}(\alpha) = \sum_\gamma h_j(\gamma) (r^{(\alpha+\gamma)})^T \underline{y} = \sum_\gamma h_j(\gamma) r^{(\alpha+\gamma)}(x)$, which is congruent modulo $\mathcal{I}$ to $\sum_\gamma h_j(\gamma) x^{\alpha+\gamma} = h_j(x) x^\alpha$. Therefore, for $q \in \mathbb{R}^{\mathcal{B}}$, $\langle q q^T, \sum_\beta C_\beta^j x^\beta \rangle = q^T M_{\mathcal{B}}(h_j * \underline{y}) q$ is, by the above, 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 argumant can be reversed, thus showing equality in (25).

Weak semidefinite duality implies $\rho^* \leq p^*$. The following result of Parrilo [19] implies equality: $\rho^* = p^*$. Hence, there is no duality gap between (21) and its dual (25) (although (21) has no strictly feasible solution if $S \neq V$).

Theorem 9 *([19]) Assume that $\mathcal{I}$ is radical and V is finite. Then, every nonnegative polynomial f on F can be written as $f = (\sum_{i_0} q_{0,i_0}^2) + \sum_{j=m+1}^{m+k} h_j (\sum_{i_j} q_{j,i_j}^2) + q$, 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 10 *When $\mathcal{I}$ is radical and V is finite, the programs (1), (21) and (25) are equivalent.*

Proof. As the polynomial $r(x) - p^*$ is nonnegative on F, we deduce using Theorem 9 that p^* is feasible for (25) and, thus, $p^* \le \rho^*$. ∎

Example 11 *([4], p. 229) Let $\mathcal{I}$ be the ideal in $\mathbb{R}[x,y]$ generated by the polynomials $h_1(x) = x^2 + y - 1$, $h_2(x) = xy - 2y^2 + 2y$. 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 $\mathcal{I}$. The set of standard monomials is $\mathcal{B} = \{1, x, y, y^2\}$ and $V = S = \{v_1 = (1,0), v_2 = (-1,0), v_3 = (0,1), v_4 = (-\frac{1}{2}, \frac{3}{4})\}$. The Zeta matrix of $\mathcal{I}$ 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

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

Therefore, a vector $\underline{u} = (1, a, b, c) \in \mathbb{R}^{\mathcal{B}}$ is the sequence of moments of a nonnegative measure supported by S if and only if $M_{\mathcal{B}}(\underline{u}) \succeq 0$ or, equivalently, $Z^{-1}u \ge 0$; that is,

$$\begin{array}{c} \\ 1 \\ x \\ y \\ y^2 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & x & y & y^2 \end{array} \\ \begin{pmatrix} 1 & a & b & c \\ a & 1-b & 2c-2b & \frac{3}{2}c - \frac{3}{2}b \\ b & 2c-2b & c & \frac{7}{4}c - \frac{3}{4}b \\ c & \frac{3}{2}c - \frac{3}{2}b & \frac{7}{4}c - \frac{3}{4}b & \frac{37}{16}c - \frac{21}{16}b \end{pmatrix} \end{array} \succeq 0 \iff \begin{cases} \frac{1}{2} + \frac{1}{2}a + \frac{1}{6}b - \frac{2}{3}c \ge 0 \\ \frac{1}{2} - \frac{1}{2}a - \frac{5}{2}b + 2c \ge 0 \\ -3b + 4c \ge 0 \\ \frac{16}{3}b - \frac{16}{3}c \ge 0. \end{cases}$$

3 Relation with Other Semidefinite Representations

In this section, we relate our result about the semidefinite representation (21) for problem (1) and the result of Lasserre [12, 13] showing the finite convergence of the programs (5) and (7) in the grid case.

The following notion of flat extension (introduced in [5]) will be useful. Let $X = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$ be a symmetric matrix and A a principal submatrix of X. One says that X is a *flat extension* of A if rank $X =$ rank A or, equivalently, if $B = AW$ and $C = B^T W = W^T AW$ for some matrix W. If X is a flat extension of A, then $X \succeq 0 \Longleftrightarrow A \succeq 0$. Curto and Fialkow [5] show the following result about flat extensions of moment matrices.

THEOREM 12 *[5] 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 13 *[5] 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 12, 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 ([5], 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 [12, 13]; see Proposition 15 and Example 16 below. A strategy for proving equivalence of (5) and (1) is indeed to show that $M_t(y)$ is a flat extension of $M_d(y)$; a way to achieve this is to show that the polynomials $f^{(\alpha)}(x) := x^\alpha - r^{(\alpha)}(x)$ (where $r^{(\alpha)}$ is the residue modulo $\mathcal{I}$ of x^α) belong to the kernel of $M_t(y)$, for all $\alpha \in S_t$. For this, the following property of the kernel of moment matrices will be useful; it comes from [5] and its proof is included for completeness.

LEMMA 14 *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 its α-components are zero 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

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

($v_+ = 0$ if $k = 0$); thus, $v = \max(v_0, v_+)$. Following an analogous reasoning as in [13], we can prove:

PROPOSITION 15 *Assume that $\mathcal{I}$ is radical, $h_1, \ldots, h_m$ form a Groebner basis of $\mathcal{I}$, and $V = S$ is finite. Let $t \geq \max(d+2, 2v_0+1, d+v_+)$. The problems (1) and (5) 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 (5). 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 $f \in \mathcal{I}$ 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 14 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 13 implies that $(y_\alpha)_{\alpha \in S_{2t-2}}$ is the sequence of moments of a nonnegative measure μ. By Lemma 1, the support of μ is contained in the set of zeros of h_j (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 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 16 The grid case. *For $j = 1, \ldots, n$, let $a_1^{(j)}, \ldots, a_{m_j}^{(j)}$ be distinct real numbers and define the polynomial $h_j(x) = \prod_{i=1}^{m_j}(x_j - a_i^{(j)})$ of degree m_j. All common zeros of $h_1, \ldots, h_n$ 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 $h_1, \ldots, h_n$ is radical. (To see it, it suffices to show that the Zeta matrix Z from (18) is nonsingular. This can be done using induction on $n \geq 1$; see the Appendix in [13].) Moreover, $h_1, \ldots, h_n$ form a Groebner basis of $\mathcal{I}$. Thus, the assumptions in Proposition 15 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 [13] and in $\max(d+1, v)$ steps in Proposition 17 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 15, based on Parrilo's result from Theorem 9.

PROPOSITION 17 *Assume that $\mathcal{I}$ is radical, V is finite, and $h_1, \ldots, h_m$ form a Groebner basis of $\mathcal{I}$ with respect to some monomial ordering. Let $t \geq \max(d + v_+ + 1, v_0)$ (resp., $t \geq \max(d + v_+, v_0)$ if $h_1, \ldots, h_m$ have an even degree).*

(i) *Problems (1) and (5) 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)** $M_t(y) \succeq 0$, $M_{t-v_j}(h_j * y) = 0$ *$(j = 1, \ldots, m)$,* $M_{t-v_j}(h_j * y) \succeq 0$ *$(j = m+1, \ldots, m+k)$, then* **(b)** $(y_\alpha)_{\alpha \in S_{2d+2v_+}}$ *is the sequence of moments of a nonnegative measure supported by F (and, thus, its projection $\underline{y} := (y_\beta)_{\beta \in \mathcal{B}}$ is feasible for (21)).*

Proof. (i) The polynomial $f(x) - p^*$ is nonnegative on F. Therefore, by Theorem 9, 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 $h_1, \ldots, h_m$ form a Groebner basis of $\mathcal{I}$, the division of q by $h_1, \ldots, h_m$ 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 (7) when $t \geq \max(d + v_+ + 1, v_0)$. Hence, $p^* \leq \rho_t^*$ and, thus, equality holds in (10). When $\deg(h_j) = 2v_j$ for all $j = 1, \ldots, m$, p^* is feasible for the program (7) with $t \geq \max(d + v_+, v_0)$, yielding equality in (10).
(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 [12].

4 Semidefinite Approximations of Low Order

From a practical point of view, the semidefinite formulation (21) 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 $M_{\mathcal{A}}(\underline{y})$ of the combinatorial moment matrix $M_{\mathcal{B}}(\underline{y})$ 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 8, 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 $M_{\mathcal{B}_t}(\underline{y})$ ensuring that the corresponding semidefinite relaxation solves, in fact, problem (1) exactly. We give our result in the case when

$$S = \{x \in \mathbb{R}^n \mid x_j = a_j, b_j \ \text{ for all } j = 1, \ldots, n\}, \tag{26}$$

where $a_1 \neq b_1, \ldots, a_n \neq b_n$ are given real numbers (thus containing the 0/1 and ± 1 cases). Define the polynomials

$$h_j(x) = (x_j - a_j)(x_j - b_j) = x_j^2 - s_j x_j - t_j, \ \text{ setting } s_j := a_j + b_j, \ t_j = -a_j b_j \tag{27}$$

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 (26). W.r.t. the graded lexicographic order, the set of standard monomials is the set $\mathcal{B}$ (also denoted as $\mathcal{B}(n)$) from (19). 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 N having size $\leq t$. Given $y \in \mathbb{R}^{\mathcal{B}_{2t}}$, we can define its truncated combinatorial moment matrix $M_{\mathcal{B}_t}(\underline{y})$ 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

$$\sum_{C \subseteq A \cap B} \left(\prod_{i \in C} s_i \prod_{i \in (A \cap B) \setminus C} t_i \right) y_{(A \Delta B) \cup C}. \tag{28}$$

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, (28) 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, (28) reads $y_{A \Delta B}$.

THEOREM 18 *Let S and $h_1, \ldots, h_n$ be as in (26) and (27). Let $y \in \mathbb{R}^{\mathcal{B}_{2t}(n)}$ with $y_0 = 1$ and $n \geq t \geq 1$. If $M_{\mathcal{B}_t}(y) \succeq 0$ and* rank $M_{\mathcal{B}_s}(y) \leq \sum_{i=1}^{s} \binom{t}{s}$ *for some $1 \leq s \leq t$, then y is the sequence of moments of a nonnegative measure supported by S. If, moreover,* 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 14.

LEMMA 19 *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) $H = s_i D + t_i F$; (c) $C = s_i E + t_i R$; (d) $T = s_i K + t_i G$. To see this, we use the fact that any entry (A, B) of $M_{\mathcal{B}_t}(y)$ is obtained by computing the residue modulo $\mathcal{I}$ of $x^A x^B$ and linearizing, which means replacing any occurrence of x^C by y_C. In this way, equation: $x^{A_m\setminus i} x^{A_\ell+i} = x^{A_m} x^{A_\ell}$ yields (a). Moreover, $x^{A_m} x^{A_\ell+i} = x^{A_m\setminus i} x^{A_\ell} x_i^2$ is congruent modulo $\mathcal{I}$ to $x^{A_m\setminus i} x^{A_\ell}(s_i x_i + t_i) = s_i x^{A_m} x^{A_\ell} + t_i x^{A_m\setminus i} x^{A_\ell}$, which yields (b). Finally, for $m, m' \in M$, $x^{A_m} x^{A_{m'}} \equiv x^{A_m\setminus i} x^{A_{m'}\setminus i}(s_i x_i + t_i)$ and, for $\ell, \ell' \in L$, $x^{A_\ell+i} x^{A_{\ell'}+i} \equiv x^{A_\ell} x^{A_{\ell'}}(s_i x_i + t_i)$, yielding (c) and (d). We can now show that $X \begin{pmatrix} s_i u \\ t_i u \\ v \end{pmatrix} = 0$. Indeed, $s_i C u + t_i E u + H v = s_i C u + t_i E u + (s_i D + t_i F) v$ (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: $s_i E u + t_i R u + S v = 0$ and $s_i H^T u + t_i S^T u + T v = 0$ follow analogously. ∎

The next result will be used in the proof of Theorem 18.

LEMMA 20 *Given $y \in \mathbb{R}^{\mathcal{B}(n)}$ with $y_0 = 1$ and $\mathcal{B}(n)$ as in (19), 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 $u(x) = \sum_{I\in\mathcal{B}_s} u_I x^I$ 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 $u_K x^N + \sum_{I\in\mathcal{B}_s, I\neq K} u_I x^I x^{N\setminus K}$, whose residue modulo $\mathcal{I}$ belongs to the kernel of $M_{\mathcal{B}}(y)$ (by Lemma 19, 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 $x^I x^{N\setminus K}$ 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 18. 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 $\begin{pmatrix} 1 & y_i \\ y_i & s_i y_i + t_i \end{pmatrix}$; as its rank is 1, this implies that $y_i^2 = s_i y_i + t_i$ 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 $y_{ij} = y_i y_j$. Therefore, $y = (v^A)_{A \in \mathcal{B}_2}$ for some $v \in S$ and the conclusion of Theorem 18 holds.

We can now assume that $t \geq 2$. Given a subset $A \subseteq N$ with $|A| = t$, we deduce from Lemma 20 (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 $\leq t-1$. 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 $M_{\mathcal{B}_t}(y)$. For this, use (23) after noting that, for any $\alpha \in S_{2t}$, the polynomial $r^{(\alpha)}$ uses only monomials in $\mathcal{B}_{2t}$. Therefore, $M_t(y)$ is a flat extension of $M_{t-1}(y)$. Corollary 13 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 1, 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 $L \leq 2^{t-1}$ 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 18.

Claim 21 *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 18 in the case when $S = \{\pm 1\}^n$ and $s = 1$ was proven[1] in [9]. 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 13. We can also give a proof for Theorem 18 not using Curto and Fialkow's result. Indeed, the following combinatorial analogue of Theorem 12 holds.

THEOREM 22 *Let S and $h_1, \ldots, h_n$ be as in (26) and (27). 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 $M_{\mathcal{B}_t}(y)$.*

This result (whose proof is given in the Appendix) permits to conclude that y is the sequence of moments of a nonnegative measure supported by S using Theorem 8 instead of Corollary 13.

The condition about the rank of $M_{\mathcal{B}_s}(y)$ in Theorem 18 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 $S = \{0,1\}^2$ (since $M_{\mathcal{B}_2}(y)$ is not positive semidefinite).

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; then 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 following result has this flavor.

PROPOSITION 23 *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] The paper [9] 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.

$(1,-1,-1)^T$ 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 (11) 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 [3] proved that adding the constraint $\text{rank}(X) \leq 2$ in the above program yields again a formulation of the stable set problem. Thus, Proposition 23 is an analogue of their result. Heuristics for max-cut based on low rank formulations are explored in [2]. 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. It is shown in [10] that the smallest such integer is $t \geq \lceil \frac{n}{2} \rceil$ and equality is conjectured (it holds for $n \leq 7$). In other words, it is conjectured that, for $t = \lceil \frac{n}{2} \rceil$,

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

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. 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.

[3] 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.]

[4] 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.

[5] 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.

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

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

[8] M. Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming. Technical Report PNA-R0108, CWI, Amsterdam, 2001.

[9] M. Laurent. Semidefinite relaxations for Max-Cut. Preprint, 2001. To appear in The Sharpest Cut, Festschrift in Honor of M. Padberg's 60th Birthday, M. Grötschel, ed., SIAM.

[10] M. Laurent. Lower bound for the number of iterations in semidefinite relaxations for the cut polytope. Preprint, July 2002.

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

[12] 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.

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

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

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

[16] 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.

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

[18] P.A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming A*, 2002.

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

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

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

[22] 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.

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

Appendix: Proof of Theorem 22

Let $y \in \mathbb{R}^{\mathcal{B}_{2t}}$, set $M := M_{\mathcal{B}_t}(y)$ and let M_I denotes the I-th 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:

$$\text{For } \lambda \in \mathbb{R}^{\mathcal{B}_t}, \lambda \in \ker M \Longleftrightarrow \left(\sum_{I \in \mathcal{B}_t} \lambda_I M_I\right)_H = 0 \text{ for all } H \in \mathcal{B}_{t-1}. \tag{29}$$

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 $\sum_{I \in \mathcal{B}_{2t}} r_I x^I$. Then, φ denotes 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, meaning that $M(A, B) = \varphi(x^A x^B)$ for $A, B \in \mathcal{B}_t$ (recall (28)). Our objective is to construct an extension of y to $\mathbb{R}^{\mathcal{B}_{2t+2}}$ in such a way that $\widetilde{M} := M_{\mathcal{B}_{t+1}}(y)$ is a flat extension of M.

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 with $|K| = t+1$ and $k \in K$. As M is a flat extension of $M_{\mathcal{B}_{t-1}}(y)$,

$$M_{K\setminus k} = M\lambda \ \text{ for some } \lambda \in \mathbb{R}^{\mathcal{B}_{t-1}}; \tag{30}$$

that is, the polynomial $x^{K\setminus k} - \lambda(x)$ belongs to $\ker M$. Lemma 19 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

$$B_K = Mr, \ \text{ where } r \text{ is the residue modulo } \mathcal{I} \text{ of } x_k\lambda(x). \tag{31}$$

Then,

$$\tilde{M}(K, H) = \varphi(x^H r(x)) \ \text{ for } H \in \mathcal{B}_t. \tag{32}$$

Indeed, $\tilde{M}(K,H) = (Mr)_H = \sum_{I\in\mathcal{B}_t} M(H,I)r_I = \sum_{I\in\mathcal{B}_t} \varphi(x^H x^I) r_I = \varphi(x^H r(x))$. We verify that B_K is well defined, i.e., that the definition in (31) 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 (30), then $\lambda - \lambda' \in \ker M$ which implies that the residue of $x_k(\lambda(x) - \lambda'(x))$ belongs to $\ker M$ (by Lemma 19, since $\lambda - \lambda'$ has degree $\le t-1$). 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 (29), it suffices to show that $(Mr)_H = (Mr')_H$ for $H \in \mathcal{B}_{t-1}$, which follows from the next claim:

$$\tilde{M}(H, K) = \varphi(x^K x^H) \ \text{ for all } H \in \mathcal{B}_t \text{ with } |H \cup K| \le 2t. \tag{33}$$

Indeed, $\tilde{M}(K,H) = \varphi(x^H r(x)) = \varphi(x^H x_k \lambda(x))$ by (32). Suppose first that $H \cap K = \emptyset$. Then, $|H| \le t-1$, $k \notin H$, and $\varphi(x^H x_k\lambda(x)) = \varphi(x^{H+k}\lambda(x)) = \sum_I \lambda_I \varphi(x^{H+k}x^I)$, which is equal to $\sum_I \lambda_I M(H+k, I) = (M\lambda)_{H+k} = M_{K\setminus k}(H+k) = M(K\setminus k, H+k) = \varphi(x^{K\setminus k}x^{H+k}) = \varphi(x^K x^H)$. Suppose now that $H \cap K \neq \emptyset$; then we can choose $k \in H\cap K$. In that case, $\varphi(x^H x_k \lambda(x)) = \varphi(x^{H\setminus k}x_k^2\lambda(x)) = s_k\varphi(x^H\lambda(x)) + t_k\varphi(x^{H\setminus k}\lambda(x))$ which can be verified to be equal to $s_k M(K\setminus k, H) + t_k M(K\setminus k, H\setminus k) = s_k\varphi(x^{K\setminus k}x^H) + t_k\varphi(x^{K\setminus k}x^{H\setminus k}) = \varphi(x^{K\setminus k}(s_k x^H + t_k x^{H\setminus k})) = \varphi(x^{K\setminus k}x^H x_k) = \varphi(x^K x^H)$.

$\tilde{M}$ 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

$$\tilde{M}_{K\setminus k} = \tilde{M}\lambda, \ \tilde{M}_K = \tilde{M}r \ \text{ where } r \text{ is the residue modulo } \mathcal{I} \text{ of } x_k\lambda(x). \tag{34}$$

In order to show that $\tilde{M}$ is a combinatorial moment matrix, we have to prove that, for all H with $|H| = t, t+1$, $\tilde{M}(H,K)$ depends only on the residue modulo $\mathcal{I}$ of $x^H x^K$. In view of (33), the case when $|H| = t$ will be settled if the following holds:

$$\tilde{M}(H,K) = \tilde{M}(H',K') \ \text{ for } |H|, |H'| = t, |K|, |K'| = t+1, \ H\cup K = H'\cup K', \ H\cap K = \emptyset. \tag{35}$$

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 (35) can be reached from (H,K) by a sequence of such local exchanges. By (34), $\tilde{M}(K \setminus k, H+k) = (\tilde{M}\lambda)_{H+k} = \sum_I \lambda_I \tilde{M}(H+k, I) = \sum_I \lambda_I \varphi(x^{H+k}x^I)$ (by (33)), which is equal to $\varphi(x^{H+k}\lambda(x)) = \varphi(x^H r(x)) = \tilde{M}(K,H)$ (by (32)).

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 (33) remains valid for $|H \cup K| = 2t+1$ and, therefore, (32) remains valid for $H \in \mathcal{B}_{t+1}$. This implies:

$$\tilde{M}(H,K) = \varphi(x^H x^K) \text{ for } |H| = |K| = t+1 \text{ with } H \cap K \neq \emptyset. \tag{36}$$

Indeed, choose $k \in H \cap K$. Then, $\tilde{M}(H,K) = \varphi(x^H r(x)) = \varphi(x^H x_k \lambda(x)) = s_k \varphi(x^H \lambda(x)) + t_k \varphi(x^{H\setminus k}\lambda(x)) = s_k \tilde{M}(K \setminus k, H) + t_k \tilde{M}(K \setminus k, H \setminus k) = s_k \varphi(x^{K\setminus k}x^H) + t_k \varphi(x^{K\setminus k}x^{H\setminus k}) = \varphi(x^{K\setminus k}x_k x^H) = \varphi(x^K x^H)$.

The proof will be complete if we can show that

$$\tilde{M}(H,K) = \tilde{M}(H',K') \text{ for } |H|,|H'|,|K|,|K'| = t+1,\ H \cup K = H' \cup K',\ H \cap K = \emptyset. \tag{37}$$

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 (37) can be obtained from (H,K) by a sequence of such local exchanges. Let $\tilde{M}_K = \tilde{M}r$ as in (34) 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 22.