REVISITING TWO THEOREMS OF CURTO AND FIALKOW ON MOMENT MATRICES

MONIQUE LAURENT

ABSTRACT. We revisit two results of Curto and Fialkow on moment matrices. The first result asserts that every sequence $y \in \mathbb{R}^{\mathbb{Z}_+^n}$ whose moment matrix $M(y)$ is positive semidefinite and has finite rank r is the sequence of moments of an r-atomic nonnegative measure μ on $\mathbb{R}^n$. We give an alternative proof for this result, using algebraic tools (the Nullstellensatz) in place of the functional analytic tools used in the original proof of Curto and Fialkow. An easy observation is the existence of interpolation polynomials at the atoms of the measure μ having degree at most t if the principal submatrix $M_t(y)$ of $M(y)$ (indexed by all monomials of degree $\leq t$) has full rank r. This observation enables us to shortcut the proof of the following result. Consider a basic closed semialgebraic set $F = \{x \in \mathbb{R}^n \mid h_1(x) \geq 0, \ldots, h_m(x) \geq 0\}$, where $h_j \in \mathbb{R}[x_1, \ldots, x_n]$ and $d := \max_{j=1}^m \lceil \deg(h_j)/2 \rceil$. If $M_t(y)$ is positive semidefinite and has a flat extension $M_{t+d}(y)$ such that all localizing matrices $M_t(h_j * y)$ are positive semidefinite, then y has an atomic representing measure supported by F. We also review an application of this result to the problem of minimizing a polynomial over the set F.

1. INTRODUCTION

1.1. The moment problem. Throughout the paper, $\mathbb{R}[x_1, \ldots, x_n]$ denotes the ring of real polynomials in n indeterminates. Write a polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ as $p(x) = \sum_{\alpha \in \mathbb{Z}_+^n} p_\alpha x^\alpha$, where x^α denotes the monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ for $\alpha \in \mathbb{Z}_+^n$. Let S_k denote the set of $\alpha \in \mathbb{Z}_+^n$ with $|\alpha| := \sum_{i=1}^n \alpha_i \leq k$. As usual, we identify a polynomial $p(x)$ with the sequence of its coefficients $p = (p_\alpha)_\alpha$. Thus, p can be seen as a vector of $\mathbb{R}^{S_k}$ if the (total) degree of $p(x)$ is at most k, as $p_\alpha = 0$ whenever $|\alpha| \geq k+1$, and $\mathbb{R}[x_1, \ldots, x_n]$ can be identified with the set of sequences $p \in \mathbb{R}^{\mathbb{Z}_+^n}$ having a finite support.

Given a probability measure μ on $\mathbb{R}^n$, the quantity $y_\alpha := \int x^\alpha \mu(dx)$ is called its *moment of order* α. The moment problem concerns the charaterization of the sequences $y = (y_\alpha)_{\alpha \in \mathbb{Z}_+^n}$ that are the sequences of moments of some nonnegative measure μ; in that case one says that μ is a *representing measure* for y and μ is a probability measure if $y_0 = 1$. (See, e.g., [12], [16] for background information.)

The results of Curto and Fialkow that we consider here deal with moment sequences of finite atomic measures, i.e., measures having a finite support. A measure μ having a finite support if it is of the form $\mu = \sum_{i=1}^r \lambda_i \delta_{x_i}$ for some $\lambda_1, \ldots, \lambda_r \neq 0$

Date: January 16, 2004.

1991 *Mathematics Subject Classification.* Primary 44A30, 13J30, 14P10, 90C22.

Key words and phrases. Moment matrix, positive semidefinite matrix, polynomial ideal, variety, polynomial optimization.

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

and distinct $x_1, \ldots, x_r \in \mathbb{R}^n$; the x_i's are the *atoms* of μ and μ is said to be r-*atomic*. Here, δ_x is the *Dirac measure* at $x \in \mathbb{R}^n$ (having mass 1 at x and mass 0 elsewhere), whose moment sequence is $\zeta_x := (x^\alpha)_{\alpha \in \mathbb{Z}^n} \in \mathbb{R}^{\mathbb{Z}^n_+}$, called the *zeta vector* of x.

Given $y \in \mathbb{R}^{\mathbb{Z}^n_+}$, its *moment matrix* is the symmetric matrix $M(y)$ indexed by $\mathbb{Z}^n_+$ whose (α, β)th entry is equal to $y_{\alpha+\beta}$, for $\alpha, \beta \in \mathbb{Z}^n_+$. A well known necessary condition for y to have a representing measure is the positive semidefiniteness of its moment matrix.

Lemma 1.1. *Assume that $y \in \mathbb{R}^{\mathbb{Z}^n_+}$ is the sequence of moments of a nonnegative measure μ. Then, $M(y) \succeq 0$; that is, $p^T M(y) p \geq 0$ for all $p \in \mathbb{R}[x_1, \ldots, x_n]$. If $M(y)p = 0$ for some polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$, then the support of μ is contained in the set of zeros of $p(x)$. Moreover, if μ is r-atomic, then* rank $M(y) \leq r$.

Proof. For $p \in \mathbb{R}[x_1, \ldots, x_n]$, we have:

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

showing that $M(y) \succeq 0$. If $M(y)p = 0$, 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 $p(x)$. Assume that $\mu = \sum_{i=1}^r \lambda_i \delta_{x_i}$ where $\lambda_1, \ldots, \lambda_r > 0$ and $x_1, \ldots, x_r \in \mathbb{R}^n$. Then, $M(y) = \sum_{i=1}^r \lambda_i \zeta_{x_i} (\zeta_{x_i})^T$, which shows that rank $M(y) \leq r$. $\square$

The moment problem can alternatively be formulated in terms of linear functionals. Namely, given $y \in \mathbb{R}^{\mathbb{Z}^n_+}$, consider the linear functional $L_y : \mathbb{R}[x_1, \ldots, x_n] \to \mathbb{R}$ defined by $L_y(p) := p^T y = 1^T M(y) p$ (1 denoting the constant polynomial). Then, y has a representing measure if and only if there is a nonnegative measure μ such that $L_y(p) = \int p(x)\mu(dx)$ for all $p \in \mathbb{R}[x_1, \ldots, x_n]$; $M(y) \succeq 0$ if and only if the linear operator L_y is nonnegative on the cone Σ^2, consisting of all sums of squares of polynomials in $\mathbb{R}[x_1, \ldots, x_n]$. The two cones $\mathcal{M}$ and $\mathcal{P}$, defined, respectively, as the set of $y \in \mathbb{R}^{\mathbb{Z}^n_+}$ having a representing measure and the set of nonnegative polynomials on $\mathbb{R}^n$, are dual of each other (equality: $\mathcal{P} = \mathcal{M}^*$ is easy, equality: $\mathcal{M} = \mathcal{P}^*$ is proved by Haviland [13]). The cone $\mathcal{M}_+$ consisting of the sequences $y \in \mathbb{R}^{\mathbb{Z}^n_+}$ with $M(y) \succeq 0$ and the cone Σ^2 of sums of squares of polynomials are dual of each other (equality: $\mathcal{M}_+ = (\Sigma^2)^*$ is easy, equality: $\Sigma^2 = (\mathcal{M}_+)^*$ is proved by Berg, Christensen and Jensen [2]). Thus the moment problem can be cast - via duality - as the problem of characterizing nonnegative polynomials.

The inclusion: $\Sigma^2 \subseteq \mathcal{P}$ is an equality for $n = 1$ and it is strict for $n \geq 2$. (Hilbert characterized the pairs (n, d) for which every polynomial of degree d in n indeterminates nonnegative on $\mathbb{R}^n$ is a sum of squares; see Reznick [27] for a detailed exposition.) Equivalently, the inclusion: $\mathcal{M} \subseteq \mathcal{M}_+$ is an equality for $n = 1$ (this is Hamburger's theorem) and it is strict for $n \geq 2$ (see [2]).

However, there are some cases when the implication: $y \in \mathcal{M}_+ \Longrightarrow y \in \mathcal{M}$ holds. Berg, Christensen and Ressel [3] (see also Lindahl and Maserick [22]) show that this is true when y is bounded. Berg and Maserick [4] extend this result to *exponentially* bounded sequences. Curto and Fialkow [7] show that this is true when $M(y)$ has a finite rank.

Theorem 1.2. [7] *If $M(y) \succeq 0$ and $M(y)$ has finite rank r, then y has a unique representing measure, which is r-atomic.*

As a direct application of Theorem 1.2, the reverse implication also holds: If y has a r-atomic representing measure, then $M(y) \succeq 0$ and rank $M(y) = r$. Curto and Fialkow's proof is based on functional analytic tools (the spectral theorem and the Riesz representation theorem). The first main contribution of this paper is an alternative more elementary proof for Theorem 1.2. Our proof uses Hilbert's Nullstellensatz and, beside this algebraic result, it uses only elementary linear algebra. A basic observation underlying our proof is that the kernel of a positive semidefinite moment matrix is a *radical* ideal in $\mathbb{R}[x_1, \ldots, x_n]$, a fact which seems not to have been noticed so far.

Curto and Fialkow [7, 8, 9] prove several results about the *truncated* moment problem, which deals with the characterization of the (truncated) sequences $y \in \mathbb{R}^{S_{2t}}$ ($t \geq 1$ integer) having a representing measure μ, i.e., $y_\alpha = \int x^\alpha \mu(dx)$ for all $\alpha \in S_{2t}$. Given $y \in \mathbb{R}^{S_{2t}}$, its *moment matrix of order t* is the matrix $M_t(y)$ indexed by S_t with (α, β)th entry $y_{\alpha+\beta}$, for $\alpha, \beta \in S_t$. In particular, Curto and Fialkow [7] show the following key result about 'flat extensions' of moment matrices. 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; then, $X \succeq 0 \Longleftrightarrow A \succeq 0$.

Theorem 1.3. [7] *Let $y \in \mathbb{R}^{S_{2t}}$ for which $M_t(y) \succeq 0$ and $M_t(y)$ is a flat extension of $M_{t-1}(y)$. Then one can extend y to a (unique) 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 $M_t(y)$.*

The following is a direct consequence of Theorem 1.3 combined with Theorem 1.2.

Corollary 1.4. [7] *Given $y \in \mathbb{R}^{S_{2t}}$, assume that $M_t(y) \succeq 0$ and that $M_t(y)$ is a flat extension of $M_{t-1}(y)$. Then one can extend y to a vector in $\mathbb{R}^{\mathbb{Z}_+^n}$ having a representing measure which is* (rank $M_t(y)$)*-atomic.*

1.2. **The F-moment problem.** This is the problem of characterizing the sequences $y \in \mathbb{R}^{\mathbb{Z}_+^n}$ having a representing measure supported by a given set $F \subseteq \mathbb{R}^n$. When F is a closed subset of $\mathbb{R}^n$, Haviland [13] shows that the cone $\mathcal{M}(F)$ of such sequences and the cone of polynomials nonnegative on F are dual cones. Consider the case when F is a basic closed semialgebraic set, of the form

$$F := \{x \in \mathbb{R}^n \mid h_1(x) \geq 0, \ldots, h_m(x) \geq 0\}, \tag{1.1}$$

where $h_1, \ldots, h_m \in \mathbb{R}[x_1, \ldots, x_n]$ and set

$$d_j = \lceil \deg(h_j)/2 \rceil, \; d := \max_{j=1}^{m} d_j. \tag{1.2}$$

Necessary conditions for membership in $\mathcal{M}(F)$ can be formulated in terms of positive semidefiniteness of the localizing matrices of y. Given $h \in \mathbb{R}[x_1, \ldots, x_n]$, let $h * y$ denote the *shifted vector* in $\mathbb{R}^{\mathbb{Z}_+^n}$ whose αth entry is $(h*y)_\alpha := \sum_\beta h_\beta y_{\alpha+\beta}$, for $\alpha \in \mathbb{Z}_+^n$. Curto and Fialkow call the moment matrix $M(h * y)$ a *localizing matrix.* One can easily verify:

Lemma 1.5. *If y has a representing measure supported by $\{x \mid h(x) \geq 0\}$, then $M(h * y) \succeq 0$.*

When F is compact, Schmüdgen [28] shows that the conditions $M(h_J * y) \succeq 0$ ($J \subseteq \{1, \ldots, m\}$, setting $h_J := \prod_{j \in J} h_j$, $h_\emptyset = 1$) are necessary and sufficient for the existence of a representing measure supported by F. When F is compact and satisfies the condition:

$$\text{(1.3)} \quad \text{There exists } p \in \Sigma^2 + \sum_{j=1}^{m} h_j \Sigma^2 \text{ for which } \{x \in \mathbb{R}^n \mid p(x) \geq 0\} \text{ is compact,}$$

Putinar [26] proves that the conditions $M(y) \succeq 0$, $M(h_j * y) \succeq 0$ $(j = 1, \ldots, m)$ suffice for the existence of a measure supported by F. Curto and Fialkow [9] consider the F-moment problem for truncated sequences. They show that, under certain rank assumptions, the conditions $M_t(y) \succeq 0$, $M_{t-d_j}(h_j * y) \succeq 0$ $(j = 1, \ldots, m)$ are sufficient for the existence of a representing measure supported by F. The following is the main result of [9] (Theorem 1.6 there).

Theorem 1.6. [9] *Let F be the set from (1.1) and let $d_1, \ldots, d_m, d$ be as in (1.2). Let $y \in \mathbb{R}^{S_{2t}}$ and $r := \text{rank } M_t(y)$. The following assertions are equivalent.*

(i) *y has a r-atomic representing measure whose support is contained in F.*
(ii) *$M_t(y) \succeq 0$ and y can be extended to a vector $y \in \mathbb{R}^{S_{2(t+d)}}$ in such a way that $M_{t+d}(y)$ is a flat extension of $M_t(y)$ and $M_t(h_j * y) \succeq 0$ for $j = 1, \ldots, m$.*

In that case, the representing measure μ is unique and, setting $r_j := \text{rank } M_t(h_j * y)$, exactly $r - r_j$ of the atoms in the support of μ belong to the set of roots of the polynomial $h_j(x)$.

The second main contribution of our paper is a very short proof of this result. Indeed, the implication (ii) $\Longrightarrow$ (i) follows directly from Corollary 1.4, after observing the existence of interpolation polynomials at the atoms of the representing measure with degree at most t. Finally, we recall in Section 3 an application of this result to the problem of minimizing a polynomial function over a basic closed semialgebraic set.

2. Alternative Proofs

2.1. **Hilbert's Nullstellensatz.** Let I be an ideal in $\mathbb{R}[x_1, \ldots, x_n]$. The set $V(I) := \{x \in \mathbb{C}^n \mid f(x) = 0 \; \forall f \in I\}$ is the (complex) *variety* associated to I. When $V(I)$ is finite, the ideal is said to be *zero-dimensional*. The two sets: $I(V(I)) := \{f \in \mathbb{R}[x_1, \ldots, x_n] \mid f(x) = 0 \; \forall v \in V(I)\}$ and $\sqrt{I} := \{f \in \mathbb{R}[x_1, \ldots, x_n] \mid f^k \in I \text{ for some integer } k \geq 1\}$ are again ideals in $\mathbb{R}[x_1, \ldots, x_n]$, which obviously contain the ideal I. The ideal I is said to be *radical* when $I = \sqrt{I}$.

Hilbert's Nullstellensatz: $\sqrt{I} = I(V(I))$.

We will use the following corollary of the Nullstellenstaz:

$$\text{(2.1)} \quad I \text{ radical} \Longrightarrow \text{a polynomial vanishing at all points of } V(I) \text{ belongs to } I.$$

The dimension of the quotient vector space $\mathbb{R}[x_1, \ldots, x_n]/I$ and the cardinality of $V(I)$ are related by:

$$\text{(2.2)} \quad |V(I)| \leq \dim \mathbb{R}[x_1, \ldots, x_n]/I, \text{ with equality if and only if } I \text{ is radical.}$$

Detailed information about polynomial ideals and varieties can be found, e.g., in [1], [5].

2.2. **Proof of Theorem 1.2.** We begin with a structural property of moment matrices.

Lemma 2.1. *Given $y \in \mathbb{R}^{\mathbb{Z}_+^n}$ and polynomials $f, g, h \in \mathbb{R}[x_1, \ldots, x_n]$, the following identity holds: $(fg)^T M(y)h = f^T M(y)(gh)$.*

Proof. Direct verification. □

Corollary 2.2. *If $M(y) \succeq 0$, then its kernel $I := \{p \in \mathbb{R}[x_1, \ldots, x_n] \mid M(y)p = 0\}$ is a radical ideal in $\mathbb{R}[x_1, \ldots, x_n]$.*

Proof. Assume $f \in I$ and let $g \in \mathbb{R}[x_1, \ldots, x_n]$. By Lemma 2.1, $(fg)^T M(y)(fg) = (fg^2)^T M(y)f = 0$, which implies that $fg \in I$. Hence, I is an ideal in $\mathbb{R}[x_1, \ldots, x_n]$. We prove that I is radical; that is, for any polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ and any integer $k \geq 1$,

$$f^k \in I \Longrightarrow f \in I. \tag{2.3}$$

If $f^2 \in I$, then $0 = 1^T M(y)(f^2) = f^T M(y)f$ (by Lemma 2.1), which implies that $f \in I$. Hence, (2.3) holds for $k = 2$ and thus for any power of 2 using induction. Finally, if $f^k \in I$, choose r in such a way that $r + k$ is a power of 2; then, $f^r f^k \in I$ implies that $f \in I$. □

Corollary 2.3. *If $M(y) \succeq 0$ and $\operatorname{rank} M(y) < \infty$, then $|V(I)| = \operatorname{rank} M(y)$.*

Proof. Let $\mathcal{B}$ be a set of monomials indexing a maximum nonsingular principal submatrix of $M(y)$. One can easily verify that $\mathcal{B}$ is a basis of $\mathbb{R}[x_1, \ldots, x_n]/I$; that is, for every $p \in \mathbb{R}[x_1, \ldots, x_n]$, there exists a unique set of reals λ_β ($\beta \in \mathcal{B}$) for which $p(x) - \sum_{\beta \in \mathcal{B}} \lambda_\beta x^\beta$ belongs to I. The statement now follows from Corollary 2.2 and (2.2). □

We now prove Theorem 1.2. Assume that $M(y) \succeq 0$ and let $r := \operatorname{rank} M(y) < \infty$. By Corollary 2.3, the variety $V(I)$ has cardinality r; say, $V(I) = \{v_1, \ldots, v_r\}$. As $V(I)$ is the set of common roots of a set of real valued polynomials, a complex point v belongs to $V(I)$ if and only if its conjugate $\overline{v}$ also belongs to $V(I)$. Thus, one can write: $V(I) = S \cup T \cup \overline{T}$, where $S := V(I) \cap \mathbb{R}^n$ and $\overline{T} := \{\overline{v} \mid v \in T\}$.

Let $p_{v_1}, \ldots, p_{v_r} \in \mathbb{C}[x_1, \ldots, x_n]$ be interpolation polynomials at the points of $V(I)$; that is, $p_{v_i}(v_j) = 1$ if $i = j$ and $p_{v_i}(v_j) = 0$ if $i \neq j$, for $i, j = 1, \ldots, r$. One can assume that p_v is real valued for $v \in S$ and that $p_{\overline{v}} = \overline{p_v}$ for $v \in T$.

Let Z be the matrix whose columns are the zeta vectors $\zeta_{v_1}, \ldots, \zeta_{v_r}$, and let $\tilde{Z}$ be the matrix whose rows contain the coefficient vectors of the interpolation polynomials $p_{v_1}, \ldots, p_{v_r}$. Thus, $\tilde{Z}Z = I_r$.

Lemma 2.4. $M(y) = Z\operatorname{diag}(\tilde{Z}y)Z^T$.

Proof. Given $\alpha, \beta \in \mathbb{Z}_+^n$, one has to verify that $y_{\alpha+\beta} = \sum_{i=1}^r v_i^\alpha v_i^\beta \sum_\gamma (p_{v_i})_\gamma y_\gamma$. This follows from the fact that the polynomial $x^{\alpha+\beta} - \sum_{i=1}^r v_i^\alpha v_i^\beta p_{v_i}(x)$ belongs to I, since it vanishes at all points of $V(I)$ (using (2.1)). □

Lemma 2.5. $V(I) \subseteq \mathbb{R}^n$.

Proof. With respect to the partition $V(I) = S \cup T \cup \overline{T}$, the matrix Z and the vector $\tilde{Z}y$ have the block decompositions: $Z = \begin{array}{c} \begin{array}{ccc} S & T & \overline{T} \end{array} \\ \begin{pmatrix} A & B & \overline{B} \end{pmatrix} \end{array}$, $\tilde{Z}y = \begin{array}{c} S \\ T \\ \overline{T} \end{array} \begin{pmatrix} a \\ b \\ \overline{b} \end{pmatrix}$, where A and a are real valued. Hence, $M(y) = A\mathrm{diag}(a)A^T + B\mathrm{diag}(b)B^T + \overline{B\mathrm{diag}(b)B^T}$. The term $A\mathrm{diag}(a)A^T$ can be written as $A_+A_+^T - A_-A_-^T$, where A_+, A_- are real matrices and the number of columns of A_+ (resp., A_-) is the number of roots $v \in S$ with $a_v > 0$ (resp., $a_v < 0$). By decomposing $B\mathrm{diag}(\sqrt{b})$ (setting $\sqrt{b} := (\sqrt{b_v})_{v \in T}$) into its real and imaginary parts, the sum $B\mathrm{diag}(b)B^T + \overline{B\mathrm{diag}(b)B^T}$ can be written as $EE^T - FF^T$, where E, F are real matrices with as many columns as the number of $v \in T$ with $b_v \neq 0$. Therefore, $M(y) = (A_+A_+^T + EE^T) - (A_-A_-^T + FF^T)$ is the difference of two real positive semidefinite matrices. As $M(y) \succeq 0$, this implies that the kernel of $A_+A_+^T + EE^T$ is contained in the kernel of $M(y)$. Hence, rank $M(y) \leq \mathrm{rank}(A_+A_+^T + EE^T) \leq |\{v \in S \mid a_v > 0\}| + |\{v \in T \mid b_v \neq 0\}|$. On the other hand, rank $M(y) = |V(I)| = |S| + 2|T|$. This implies that $T = \emptyset$, i.e., $V(I) \subseteq \mathbb{R}^n$. ☐

Corollary 2.6. *$M(y) = \sum_{i=1}^r p_{v_i}^T M(y) p_{v_i} \zeta_{v_i} \zeta_{v_i}^T$ and $\mu := \sum_{i=1}^r p_{v_i}^T M(y) p_{v_i} \delta_{v_i}$ is the unique measure representing y.*

Proof. As the set $\{p_{v_i} \mid i = 1, \ldots, r\}$ is a basis of $\mathbb{R}[x_1, \ldots, x_n]/I$, the equality: $M(y) = N := \sum_{i=1}^r p_{v_i}^T M(y) p_{v_i} \zeta_{v_i} \zeta_{v_i}^T$ follows from the fact that $p_{v_i}^T M(y) p_{v_j} = p_{v_i}^T N p_{v_j}$ for all $i, j = 1, \ldots, r$. (This is obvious for $i = j$ and, for $i \neq j$, $p_{v_i}^T M(y) p_{v_j} = 1^T M(y)(p_{v_i} p_{v_j}) = 0$, since $p_{v_i} p_{v_j} \in I$.) Thus, μ is a representing measure for y. Finally, if $\mu' := \sum_{i=1}^s \lambda_i \delta_{x_i}$ is another measure representing y, then $r = \mathrm{rank}\ M(y) \leq s = |supp(\mu')|$ and $supp(\mu') \subseteq V(I)$ (by Lemma 1.1), which implies $r = s$, $supp(\mu') = V(I)$, and thus the λ_i's are given by $p_{v_i}^T M(y) p_{v_i}$. ☐

This concludes the proof of Theorem 1.2. We make an observation, which we will use in our short proof of Theorem 1.6.

Lemma 2.7. *Let $y \in \mathbb{R}^{\mathbb{Z}_+^n}$ with $M(y) \succeq 0$, $r = \mathrm{rank}\ M(y) = \mathrm{rank}\ M_t(y)$ (for some $t \geq 1$), and let $\mu = \sum_{i=1}^r \lambda_i \delta_{v_i}$ be the r-atomic measure representing y, with $\lambda_1, \ldots, \lambda_r > 0$ and $v_1, \ldots, v_r \in \mathbb{R}^n$. There exist interpolation polynomials $q_1, \ldots, q_r$ at the points $v_1, \ldots, v_r$ having degree at most t.*

Proof. As above, let $p_{v_1}, \ldots, p_{v_r}$ be interpolation polynomials at $v_1, \ldots, v_r$. As $M_t(y)$ has rank r, one can choose a basis $\mathcal{B}$ of $\mathbb{R}[x_1, \ldots, x_n]/I$ which is contained in S_t. Let $q_1, \ldots, q_r$ denote the respective residues of $p_{v_1}, \ldots, p_{v_r}$ modulo I with respect to the basis $\mathcal{B}$. Then, $q_1, \ldots, q_r$ are again interpolation polynomials at $v_1, \ldots, v_r$ and they use only monomials in $\mathcal{B}$, which implies that they have degree at most t. ☐

Remark 2.8. Let us summarize some links between ideals and moment matrices. If $M(y) \succeq 0$, then its kernel I is a radical ideal. Moreover, I is zero-dimensional if and only if $M(y)$ has finite rank, in which case rank $M(y) = |V(I)|$ and $V(I) \subseteq \mathbb{R}^n$. The inclusion $V(I) \subseteq \mathbb{R}^n$ does not hold in general; e.g., $V(I) = \mathbb{C}^n$ if $M(y)$ is positive definite.

Conversely, if I is a zero-dimensional radical ideal in $\mathbb{R}[x_1, \ldots, x_n]$, define $y \in \mathbb{R}^{\mathbb{Z}_+^n}$ as the sequence of moments of the measure $\mu := \sum_{v \in V(I) \cap \mathbb{R}^n} \delta_v$. Then, $M(y) \succeq$

0 and $I \subseteq \text{Ker } M(y)$, with equality if and only if $V(I) \subseteq \mathbb{R}^n$. Hence, every zero-dimensional radical ideal with $V(I) \subseteq \mathbb{R}^n$ can be realized as the kernel of some positive semidefinite moment matrix.

2.3. **Proof of Theorem 1.6.**

Proof. The implication (i) $\Longrightarrow$ (ii) follows easily using Theorem 1.2 and Lemma 1.5. Conversely, assume that (ii) holds and set $r := \text{rank } M_t(y)$. By Corollary 1.4, y has an r-atomic representing measure, say $\mu = \sum_{i=1}^r \lambda_i \delta_{v_i}$, where $\lambda_i > 0$. It suffices to verify that $v_1, \ldots, v_r$ belong to F, i.e., that $h_j(v_i) \geq 0$ for all $i = 1, \ldots, r$, $j = 1, \ldots, m$. By Lemma 2.7, there exist interpolation polynomials $q_1, \ldots, q_r$ at the points $v_1, \ldots, v_r$ having degree at most t. Then, for $k = 1, \ldots, r$, $q_k^T M_t(h_j * y) q_k = \sum_{i=1}^r (q_k(v_i))^2 h_j(v_i) \lambda_i = h_j(v_k)\lambda_k \geq 0$, since $M_t(h_j * y) \succeq 0$. This implies that $h_j(v_k) \geq 0$ for all $j \leq m$, $k \leq r$; that is, the measure μ is supported by the set F.

Finally, we verify that $r - r_j$ of the points $v_1, \ldots, v_r$ are zeros of the polynomial $h_j(x)$. For this, note that $h_j * y$ can be extended to a vector of $\mathbb{R}^{\mathbb{Z}_+^n}$ by setting

$$(h_j * y)_\alpha := \sum_\gamma (h_j)_\gamma y_{\alpha+\gamma} = \sum_\gamma (h_j)_\gamma \sum_{i=1}^r \lambda_i (v_i)^{\alpha+\gamma} = \sum_{i=1}^r \lambda_i h_j(v_i)(v_i)^\alpha.$$

Denote by s_j the number of v_i's for which $h_j(v_i) > 0$. The measure $\sum_{i=1}^r \lambda_i h_j(v_i) \delta_{v_i}$ is a s_j-atomic representing measure for $h_j * y$. Hence, by Theorem 1.2, the rank of the moment matrix $M(h_j * y)$ is equal to s_j. As $M_t(h_j * y)$ has rank r_j, this implies that $s_j \geq r_j$. In fact, equality holds, since $M_{t+1}(h_j * y)$ (and thus $M(h_j * y)$) is a flat extension of $M_t(h_j * y)$. The latter assertion follows from the fact that (i) $M_{t+1}(y)$ is a flat extension of $M_t(y)$ and that (ii) $\text{Ker } M_{t+1}(y) \subseteq \text{Ker } M_{t+1}(h_j * y)$. Condition (i) holds by assumption. Let us verify (ii). Indeed, if $f \in \text{Ker } M_{t+1}(y)$, then $0 = f^T M_{t+1}(y) f = \sum_{i=1}^r (f(v_i))^2 \lambda_i$, which implies that $f(v_1) = \ldots = f(v_r) = 0$. Then, $f^T M_{t+1}(h_j * y) f = \sum_{i=1}^r (f(v_i))^2 h_j(v_i) \lambda_i = 0$, showing that $f \in \text{Ker } M_{t+1}(h_j * y)$. Thus we have shown that the measure representing $h_j * y$ is r_j-atomic, i.e., exactly $r - r_j$ of the atoms of μ are zeros of $h_j(x)$. □

3. Application to Optimization

Consider the problem:

$$p^* := \inf \ p(x) \ \text{ subject to } h_1(x) \geq 0, \ldots, h_m(x) \geq 0 \tag{3.1}$$

of minimizing a polynomial $p(x)$ over the semialgebraic set F from (1.1). Set $d_0 := \lceil \deg(p)/2 \rceil$ and let $d_1, \ldots, d_m, d$ be as in (1.2). One can formulate the following hierarchy of lower bounds for problem (3.1):

$$p_t^* := \inf \ p^T y \ \text{ subject to } y_0 = 1, \ M_t(y) \succeq 0, \ M_{t-d_j}(y) \succeq 0 \ (h = 1, \ldots, m) \tag{3.2}$$

for $t \geq \max(d_0, d)$ (see Lasserre [17]). Then, $p_t^* \leq p_{t+1}^* \leq p^*$. (The inequality $p_t^* \leq p^*$ follows from the fact that the truncated zeta vector $\zeta_{2t,x} := (x^\alpha)_{\alpha \in S_{2t}}$ of any $x \in F$ is feasible for the program (3.2) with objective value $p^T \zeta_{2t,x} = p(x)$.)

The dual semidefinite program of (3.2) is of the form:

$$
\begin{aligned}
\rho_t^* := \quad & \max \quad \rho \\
& \text{s.t.} \quad p(x) - \rho = u_0(x) + \sum_{j=1}^{m} u_j(x) h_j(x) \\
& \qquad\quad \text{where } u_0, u_1, \ldots, u_m \text{ are sums of squares of polynomials} \\
& \qquad\quad \text{with } \deg(u_0), deg(u_1 h_1), \ldots, \deg(u_m h_m) \leq 2t.
\end{aligned} \tag{3.3}
$$

(See Lasserre [17].) Weak semidefinite duality implies that $\rho_t^* \leq p_t^*$. Moreover, there is no duality gap, i.e., $\rho_t^* = p_t^*$, if F has a nonempty interior (Lasserre [17], see also Schweighofer [30]).

When F is compact and satisfies (1.3), the bounds ρ_t^* (and thus the bounds p_t^*) converge to p^* as $t \to \infty$ (Lasserre [17]). This follows directly from a result of Putinar [26] asserting that every positive polynomial on F belongs to $\Sigma^2 + \sum_{j=1}^m h_j \Sigma^2$.

Hence, the semidefinite programs (3.2) and (3.3) can be used for approximating the minimum value of a polynomial over a semialgebraic set. Two software packages based on this approach have been developped for this problem: *GloptiPoly*, developped by Henrion and Lasserre [14], which relies on the program (3.2) involving moment matrices, and *SOSTOOLS*, developped by Prajna et al. [25], which relies on the sums of squares approach studied by Parrilo [23],[24].

The results of Curto and Fialkow mentioned earlier have important applications to optimization. They indeed permit to formulate stopping criterions for the hierarchies (3.2) of semidefinite relaxations for problem (3.1).

For instance, they have been used by Lasserre [18, 19] for proving the *finite* convergence of the bounds p_t^* to p^* in the 0/1 and grid cases. More precisely, the 0/1 case is the case when the equations $x_i^2 = x_i$ $(i = 1, \ldots, n)$ are present in the description of the semialgebraic set F, i.e., when F is the set of 0/1 points satifying certain additionnal polynomial (in)equalities. Then, using Theorem 1.6, Lasserre [18] shows the finite convergence (in n steps); that is, $p_t^* = p^*$ for $t \geq n$. The grid case considered in [19] is the case when F is of the form $I_1 \times \ldots \times I_n$, where each $I_i \subseteq \mathbb{R}$ is a finite set. Then, Lasserre [19] shows the finite convergence in $t = d - n + \sum_{i=1}^n |I_i|$ steps. Laurent ([20],[21]) gives an alternative proof for these convergence results. The paper [20] contains a simple proof for the finite convergence result in the 0/1 case. Extending the idea from the 0/1 case, the paper [21] shows a finite convergence result in a more general setting and the proof is again elementary; in particular, it does not use the results by Curto and Fialkow. This finite convergence result applies to the case when the equations present in the description of F include a set of polynomial equations defining a radical zero-dimensional ideal (thus including the 0/1 and grid cases).

The result of Curto and Fialkow from Theorem 1.6 is used by Henrion and Lasserre [15] for producing a certificate that the relaxation (3.2) in fact solves the original problem (3.1) at optimality. Namely,

Proposition 3.1. *Let $y \in \mathbb{R}^{S_{2t}}$ be an optimum solution to the program (3.2). If*

$$\text{rank } M_t(y) = \text{rank } M_{t-d}(y), \tag{3.4}$$

then $p_t^* = p^*$. As before, $d = \max(\lceil \deg(h_j)/2 \rceil \mid j = 1, \ldots, m)$.

Proof. By assumption, $M_t(y) \succeq 0$, $M_{t-d}(h_j * y) \succeq 0$ for all j, and rank $M_t(y) =$ rank $M_{t-d}(y) =: r$. Therefore, by Theorem 1.6, we can conclude that y has an r-atomic representing measure $\mu = \sum_{i=1}^r \lambda_i \delta_{v_i}$, where $v_i \in F$, $\lambda_i > 0$ and $\sum_{i=1}^r \lambda_i = 1$ (since $y_0 = 1$). Hence, $p_t^* = p^T y = \sum_{i=1}^r \lambda_i p(v_i) \geq p^*$, as $p(v_i) \geq p^*$ for all i. On the other hand, $p^* \geq p_t^*$. This implies that $p^* = p_t^*$ and that each v_i is a minimizer of $p(x)$ over the set F. □

Assume that $y \in \mathbb{R}^{S_{2t}}$ is an optimum solution to the program (3.2) satisfying the rank condition (3.4). Thus, $p_t^* = p^*$. Moreover, as the proof of Proposition 3.1 shows, y has an r-atomic representing measure $\mu = \sum_{i=1}^r \lambda_i \delta_{v_i}$, where $\lambda_i > 0$, $\sum_{i=1}^r \lambda_i = 1$, $r =$ rank $M_t(y) =$ rank $M_{t-d}(y)$, and the points $v_1, \ldots, v_r$ are global minimizers of $p(x)$ over the set F. Henrion and Lasserre [15] propose the following procedure for computing $v_1, \ldots, v_r$, whose details fit nicely within our algebraic setting.

Let $\mathcal{B} \subseteq S_{t-d}$ be a set of monomials indexing a maximum nonsingular principal submatrix of $M_{t-d}(y)$. For each $\beta \in S_t \setminus \mathcal{B}$, there exists a polynomial $g^{(\beta)} \in \mathbb{R}^{S_t}$ belonging to the kernel of $M_t(y)$, of the form:

$$g^{(\beta)}(x) = x^\beta + r^{(\beta)}(x), \text{ where } r^{(\beta)} \in \mathbb{R}^{\mathcal{B}}. \tag{3.5}$$

Let I denote the ideal in $\mathbb{R}[x_1, \ldots, x_n]$ generated by $g^{(\beta)}(x)$ ($\beta \in S_t \setminus \mathcal{B}$), with variety $V(I)$. Obviously, $\{v_1, \ldots, v_r\} \subseteq V(I)$.

Lemma 3.2. $V(I) = \{v_1, \ldots, v_r\}$.

Proof. Observe first that, for every $\beta \in \mathbb{Z}_+^n \setminus \mathcal{B}$, there exists $r \in \mathbb{R}^{\mathcal{B}}$ for which $x^\beta + r(x) \in I$. (True for $|\beta| \leq t$ by (3.5) and in general using induction.) As $M_t(y)$ is a flat extension of $M_{t-d}(y)$, y has an extension to $y \in \mathbb{R}^{\mathbb{Z}_+^n}$ (namely, the sequence of moments of the measure μ) such that $M(y)$ has rank r. Moreover, $I =$ Ker $M(y)$. (The inclusion $I \subseteq$ Ker $M(y)$ is obvious and the reverse inclusion follows using our preliminary observation.) Therefore, $|V(I)| = r$ and thus $V(I) = \{v_1, \ldots, v_r\}$. □

Thus we are left with the task of finding the common roots of a system of polynomial equations, a problem which has received considerable attention in the litterature. A classic method is the so-called eigenvalue method which consists of computing the eigenvalues of the multiplication matrices. (See, e.g., [1], [5], [6]).

4. Concluding Remarks

Curto and Fialkow's proof for Theorem 1.2 is along the following lines. (See chapter 4 in [7].) Assume $M(y) \succeq 0$ and rank $M(y) = r$. As the kernel $I := \{p \in \mathbb{R}[x_1, \ldots, x_n] \mid M(y)p = 0\}$ of $M(y)$ is an ideal in $\mathbb{R}[x_1, \ldots, x_n]$, one can consider the quotient vector space $A := \mathbb{R}[x_1, \ldots, x_n]/I$. Define an inner product on A by setting $\langle p, q \rangle := p^T M(y) q$. In this way, A is a Hilbert space of finite dimension r. For $q \in \mathbb{R}[x_1, \ldots, x_n]$, consider the multiplication operator $\varphi_q : A \to A$ defined by $\varphi_q(p) = pq$. Obviously, the operators $\varphi_{x_1}, \ldots, \varphi_{x_n}$ pairwise commute. Curto and Fialkow use then the spectral theorem and the Riesz representation theorem for proving the existence of a representation measure for y. This type of proof based on functional analytic tools is often used for proving results about the moment problem. See, e.g., Fuglede [12], Schmüdgen [28].

We have given in this paper an alternative algebraic proof based on the Nullstellensatz. In fact, the original proof of Curto and Fialkow can be modified in such a way that only the spectral theorem is used (and not the Riesz representation theorem). This other proof follows an argument used by Freedman, Lovász and Schrijver [11] (for a more general problem) and goes as follows. With respect to an orthonormal basis of the space A (equipped with the inner product $\langle ., . \rangle$), the multiplication operator φ_{x_i} is represented by a real symmetric matrix M_i. As the matrices $M_1, \ldots, M_n$ pairwise commute, they have a common orthogonal basis $p_1, \ldots, p_r$ of real eigenvectors. For $i \neq j$, $p_i p_j = \varphi_{p_i}(p_j) = \lambda p_j$ and $p_i p_j = \varphi_{p_j}(p_i) = \lambda' p_i$ (for some scalars λ, λ'), which implies that $p_i p_j = 0$. Up to rescaling, $p_i p_i = p_i$ for all i. In other words, $\{p_1, \ldots, p_r\}$ is a basis of idempotents for $\mathbb{R}[x_1, \ldots, x_n]/I$ (corresponding to the interpolation polynomials of the points in $V(I)$). For $k = 1, \ldots, n$, decompose the monomial x_k into the basis $\{p_1, \ldots, p_r\}$ as $x_k = \sum_{i=1}^r \beta_i^{(k)} p_i$, for some reals $\beta_i^{(k)}$. Then, for any $\alpha \in \mathbb{Z}_+^n$, we have that $x^\alpha = \sum_{i=1}^r (\beta_i^{(1)})^{\alpha_1} \cdots (\beta_i^{(n)})^{\alpha_n} p_i$. This implies that $y_\alpha = \langle 1, x^\alpha \rangle = \sum_{i=1}^r \langle 1, p_i \rangle \beta_i^\alpha$, after setting $\beta_i := (\beta_i^{(1)}, \ldots, \beta_i^{(n)}) \in \mathbb{R}^n$. That is, the measure $\mu := \sum_{i=1}^r \langle 1, p_i \rangle \delta_{\beta_i}$ represents y (note that $\langle 1, p_i \rangle = \langle p_i, p_i \rangle > 0$).

The original results of Curto and Fialkow [7, 9] are formulated for the complex moment problem. However, as explained e.g. in [10], the n-dimensional complex m oment problem is equivalent to the $2n$-dimensional real moment problem. Hence, our alternative proof also implies a proof in the complex case. Anyway, the above proof also applies in the complex setting (applying appropriately conjugation).

Let us finally observe that the above proof also extends to the moment problem in semigroups (the paper [11] considers a class of semigroups). That is, if S is an abelian semigroup, $y \in \mathbb{R}^S$ such that its moment matrix $M(y) = (y_{s+t})_{s,t \in S}$ is positive semidefinite with finite rank, then there exists a representing measure for y supported by the set of characters on S. Note that the result of Berg, Christensen and Ressel [3] about bounded sequences (quoted before Theorem 1.2) is proved for abelian semigroups.

References

1. S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in Real Algebraic Geometry.* Springer, Berlin, 2003.
2. C. Berg, J.P.R. Christensen, and C.U. Jensen. A remark on the multidimensional moment problem. *Math. Ann.* **243** (1979), 163–169.
3. C. Berg, J.P.R. Christensen, and P. Ressel. Positive definite functions on Abelian semigroups. *Math. Ann.* **223** (1976), 253–272.
4. C. Berg and P.H. Maserick. Exponentially bounded positive definite functions. *Ill. J. Math.* **28** (1984), 162–179.
5. D. A. Cox, J. B. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra.* 2nd edition, Springer, New York, 1997.
6. D. A. Cox, J. B. Little, and D. O'Shea. *Using Algebraic Geometry.* Graduate Texts in Mathematics, N. 185, Springer, New York, 1998.
7. R.E. Curto and L.A. Fialkow. Solution of the truncated complex moment problem for flat data. *Mem. Amer. Math. Soc.* vol. 568, Amer. Math. Soc., Providence, RI, 1996.
8. R.E. Curto and L.A. Fialkow. Flat extensions of positive moment matrices: recursively generated relations. *Mem. Amer. Math. Soc.* vol. 648, Amer. Math. Soc., Providence, RI, 1998.
9. R.E. Curto and L.A. Fialkow. The truncated complex K-moment problem. *Trans. Amer. Math. Soc.* **352** (2000), 2825–2855.
10. R.E. Curto and L.A. Fialkow. Solution of the singular quartic moment problem. *J. Operator Theory* **48** (2002), 315–354.

11. M. Freedman, L. Lovász, and A. Schrijver. Reflection positivity, rank connectivity and homomorphism of graphs. In preparation.
12. B. Fuglede. The multidimensional moment problem. *Expo. Math.* **1** (1983), 47–65.
13. E.K. Haviland. On the momentum problem for distributions in more than one dimension. *Amer. J. Math.* **57** (1935), 562–568.
14. D. Henrion and J.-B. Lasserre. GloptiPoly: Global optimization over polynomials with Matlab and SeDuMi. *ACM Trans. Math. Soft.* **29** (2003), 165–194.
15. D. Henrion and J.-B. Lasserre. Detecting global optimality and extracting solutions in GloptiPoly. Preprint, 2003.
16. H. Landau. Classic background of the moment problem. In *Moments in Mathematics*, Proceedings of Symposia in Applied Mathematics, vol. 37, pages 1–15, Amer. Math. Soc., Providence, RI, 1987.
17. J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.* **11** (2001), 796–817.
18. 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** (2001), 293–303.
19. J.B. Lasserre. Polynomials nonnegative on a grid and discrete representations. *Trans. Amer. Math. Soc.* **354** (2001), 631–649.
20. M. Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming. *Math. Oper. Res.* **28** (2003), 470–496.
21. M. Laurent. Semidefinite representations for finite varieties. Preprint, 2002.
22. R.J. Lindahl and P.H. Maserick. Positive-definite functions on involution semigroups. *Duke Math. J.* **38** (1971), 771–782.
23. P.A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization.* PhD thesis, California Institute of Technology, May 2000.
24. P.A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Program., Ser. B* **96** (2003), 293–320.
25. S. Prajna, A. Papachristodoulou, and P.A. Parrilo. SOSTOOLS (Sums of squares optimization toolbox for MATLAB) User's guide. Available at `http://control.ee.ethz.ch/~parrilo/sostools/index.html`
26. M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana Univ. Math. J.* **42** (1993), 969–984.
27. B. Reznick. Some concrete aspects of Hilbert's 17th problem. *Contemp. Math.* **253** (2000), 251–272, Amer. Math. Soc., Providence, RI, 2000.
28. K. Schmüdgen. The K-moment problem for compact semi-algebraic sets. *Math. Ann.* **289** (1991), 203–206.
29. M. Schweighofer. An algorithmic approach to Schmdgen's Positivstellensatz. *J. Pure Appl. Algebra* **166** (2002), 307–319.
30. M. Schweighofer. Optimization of polynomials on compact semialgebraic sets. Preprint, 2003.

CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands
E-mail address: `M.Laurent@cwi.nl`