POLYNOMIALS NONNEGATIVE ON A GRID AND DISCRETE OPTIMIZATION

JEAN B. LASSERRE

ABSTRACT. We characterize the real-valued polynomials on $\mathbb{R}^n$ that are nonnegative (not necessarily strictly positive) on a grid $K$ of points of $\mathbb{R}^n$, in terms of a weighted sum of squares whose degree is bounded and known in advance. We also show that the minimization of an arbitrary polynomial on $K$ (a discrete optimization problem) reduces to a convex continuous optimization problem of fixed size. The case of concave polynomials is also investigated. The proof is based on a recent result of Curto and Fialkow on the $K$-moment problem.

1. Introduction

This paper is concerned with the characterization of real-valued polynomials on $\mathbb{R}^n$ that are nonnegative (and not necessarily strictly positive) on a pre-defined grid of points of $\mathbb{R}^n$. That is, we consider the polynomials $p(x) : \mathbb{R}^n \to \mathbb{R}$ such that $p(x) \geq 0$ on $K$, where $K$ is the subset of $\mathbb{R}^n$ defined by

$$K := \{ x \in \mathbb{R}^n \mid g_k(x) = 0, k = 1, \ldots, n \},$$

where the polynomials $g_k(x) : \mathbb{R}^n \to \mathbb{R}$ are given by

$$x \mapsto g_k(x) := \prod_{i=1}^{2r_k} (x_k - a_{ki}), \quad k = 1, \ldots, n.$$

$K$ defines a grid in $\mathbb{R}^n$ with $s := \prod_{k=1}^n 2r_k$ points.

We obtain a result in the spirit of the “linear” representation of polynomials positive on a compact semi-algebraic set, obtained by Putinar [10], Jacobi and Prestel [11, 12] in a general framework. Namely, we show that every polynomial $p(x)$ of degree $2r_0$ or $2r_0 - 1$, nonnegative on $K$, can be written as a sum of squares of polynomials weighted by the polynomials $g_k(x)$ defining the set $K$, and whose degree is bounded by $r + v$ with $r := \sum_{k=1}^n (2r_k - 1)$ and $v := \max\{r_0 - r, \max_{k=1}^n r_k\}$, independently of the grid points. The important thing is that in this case, the degree of the polynomials in that representation is bounded and known in advance.

To prove this result, we use a detour and first consider the associated discrete optimization problem

$$\mathbb{P} \to p^* := \min_{x \in K} p(x),$$

Received by the editors November 11, 2000.

2000 Mathematics Subject Classification. Primary 13P99, 30C10; Secondary 90C10, 90C22.

Key words and phrases. Algebraic geometry; positive polynomials; $K$-moment problem; semi-definite programming.

©2001 American Mathematical Society
which is known to be NP-hard in general. However, we show that \( P \) is equivalent to a continuous convex optimization problem whose size depends on the number of points, but not on the points in the grid. The idea is to define a sequence of refined so-called convex *semidefinite (SDP) relaxations* \( \{Q_i\} \) of \( P \). For the general optimization problem where one minimizes a polynomial \( p(x) \) on a compact semi-algebraic set \( K \), we have shown in Lasserre [7] that the SDP relaxations \( Q_i \) “converge”, that is, \( \inf \{Q_i\} \uparrow p^* \) as \( i \to \infty \), provided the set \( K \) has some property that permits a “linear” representation of polynomials positive on \( K \) (a particular case of the more general Schmüdgen’s representation [11]), as proved in Putinar [10], Jacobi and Prestel [4, 5]. This approach is also valid for 0-1 optimization problems (see Lasserre [8]). Other sequences of SDP relaxations have been provided by the more general Schmudgen’s representation [11], as proved in Putinar [10], and also by Schrijver [9] (see also extensions by Kojima and Tuncel [6]) that eventually yields the convex hull of the feasible set of \( P \) whenever a weak separation oracle is available for the homogeneous cone associated with the constraint set. The approach developed in Lasserre [7, 8] is different and directly relates results of algebraic geometry with optimization. In particular, the relaxations \( \{Q_i\} \) and their dual \( \{Q^*_i\} \) perfectly match the duality between the \( K \)-moment problem and the theory of polynomials, positive on a compact semi-algebraic set \( K \).

However, in the present context, we may even further extend the properties of the above relaxations. We use a recent result of Curto and Fialkow [2] on the \( K \)-moment problem, to show that the optimal value \( p^* \) is obtained at most at the relaxation \( Q_{r+1} \). Next, by using a standard “strong duality” result in convex optimization, we show that the dual optimization problem \( Q^*_{r+1} \) is solvable and yields the coefficients of the polynomials in the representation of \( p(x) - p^* \) as a sum of squares, weighted by the polynomials \( g_k(x) \) defining \( K \). In a sense, the order of the relaxation \( Q_i \) for which the optimum value \( p^* \) is obtained, measures the “hardness” of problem \( P \), as it expresses the “effort” (in degree) needed to represent \( p(x) - p^* \) as a weighted sum of squares. In the present case, this “effort” is finite and bounded by \( r + v \). It also follows that any discrete polynomial optimization problem on \( K \), with additional arbitrary polynomial constraints \( h_j(x) \geq 0 \), also reduces to a convex continuous optimization problem of fixed size. Finally, we also obtain a representation of concave polynomials.

The paper is organized as follows. We first introduce the notation as well as some definitions. In Section 3, we present the results on the discrete optimization problem \( P \). In Section 4, we then present the result on the representation of \( p(x) - p^* \) as a weighted sum of squares.

2. Notation and definitions

We adopt the following notation. Given any two real-valued symmetric matrices \( A, B \) let \( \langle A, B \rangle \) denote the usual scalar product \( \text{trace}(AB) \) and let \( A \succeq B \) (resp. \( A \succ B \)) stand for \( A - B \) positive semidefinite (resp. \( A - B \) positive definite). Let \( \{1, x_1, x_2, \ldots, x_n, x_1^2, x_1 x_2, \ldots, x_1 x_n, x_2^2, x_2 x_3, \ldots, x_n^2, x_1, x_1^2, \ldots, x_n \} \) be a basis for the space \( \mathcal{A}_r \) of real-valued polynomials of degree at most \( r \), and let \( s(r) \) be its dimension. Therefore, an \( r \)-degree polynomial \( p(x) : \mathbb{R}^n \to \mathbb{R} \) is written

\[
p(x) = \sum_\alpha p_\alpha x^\alpha, \quad x \in \mathbb{R}^n,
\]
where
\[ x^{\alpha} = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}, \quad \text{with} \sum_{i=1}^{n} \alpha_i = k, \]
is a monomial of degree \( k \) with coefficient \( p_\alpha \). Denote by \( p = \{p_\alpha\} \in \mathbb{R}^{s(r)} \) the coefficients of the polynomial \( p(x) \) in the basis (2.1). Hence, the respective vectors of coefficients of the polynomials \( g_k(x), k = 1, \ldots, n, \) in (1.1), are denoted \( \{(g_k)_\alpha\} = g_k \in \mathbb{R}^{s(2r_k)}, k = 1, \ldots, n. \)

We next define the important notions of moment matrix and localizing matrix already introduced in Curto and Fialkow [2], Berg [1].

2.1. Moment matrix. Given an \( s(2r) \)-sequence \((1, y_1, \ldots)\), let \( M_r(y) \) be the moment matrix of dimension \( s(r) \) (denoted \( M(r) \) in Curto and Fialkow [2]), with rows and columns labelled by (2.1). For instance, for illustration purposes, and for clarity of exposition, consider the 2-dimensional case. The moment matrix \( M_r(y) \) is the block matrix \( \{M_{i,j}(y)\}_{0 \leq i,j \leq r} \) defined by
\[
M_{i,j}(y) = \begin{bmatrix}
y_{i+j,0} & y_{i+j-1,1} & \cdots & y_{i,j} \\
y_{i+j-1,1} & y_{i+j-2,2} & \cdots & y_{i-1,j+1} \\
\vdots & \vdots & \ddots & \vdots \\
y_{j,i} & y_{i-j+1,1} & \cdots & y_{0,i+j}
\end{bmatrix}.
\]

To fix ideas, with \( n = 2 \) and \( r = 2 \), one obtains
\[
M_2(y) = \begin{bmatrix}
1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\
y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\
y_{01} & y_{11} & y_{20} & y_{31} & y_{22} & y_{13} \\
y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\
y_{11} & y_{21} & y_{31} & y_{41} & y_{32} & y_{23} \\
y_{02} & y_{12} & y_{22} & y_{32} & y_{42} & y_{33}
\end{bmatrix}.
\]

Another more intuitive way of constructing \( M_r(y) \) is as follows. If \( M_r(y)(1, i) = y_{\alpha} \) and \( M_r(y)(j, 1) = y_{\beta} \), then
\[
M_r(y)(i, j) = y_{\alpha + \beta}, \quad \text{with} \ \alpha + \beta = (\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n).
\]

\( M_r(y) \) defines a bilinear form \( \langle \cdot, \cdot \rangle_y \) on \( A_r \) by
\[
\langle q(x), v(x) \rangle_y := \langle q, M_r(y)v \rangle, \quad q(x), v(x) \in A_r,
\]
and if \( y \) is a sequence of moments of some measure \( \mu_y \), then
\[
\langle q, M_r(y)q \rangle = \int q(x)^2 \mu_y(dx) \geq 0,
\]
so that \( M_r(y) \geq 0. \)

2.2. Localizing matrix. If the entry \( (i, j) \) of the matrix \( M_r(y) \) is \( y_{\beta} \), let \( \beta(i, j) \) denote the subscript \( \beta \) of \( y_{\beta} \). Next, given a polynomial \( \theta(x) : \mathbb{R}^n \rightarrow \mathbb{R} \) with coefficient vector \( \theta \), we define the matrix \( M_r(\theta y) \) by
\[
M_r(\theta y)(i, j) = \sum_{\alpha} \theta_{\alpha} y_{(\beta(i, j) + \alpha)}.
\]
For instance, with
\[ M_1(y) = \begin{bmatrix} 1 & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{bmatrix} \]
and \( x \mapsto \theta(x) = a - x_1^2 - x_2^2 \),
we obtain
\[ M_1(\theta y) = \begin{bmatrix} a - y_{20} - y_{02} & ay_{10} - y_{30} - y_{12} & ay_{01} - y_{21} - y_{03} \\ ay_{10} - y_{30} - y_{12} & ay_{20} - y_{40} - y_{22} & ay_{11} - y_{31} - y_{13} \\ ay_{01} - y_{21} - y_{03} & ay_{11} - y_{31} - y_{13} & ay_{02} - y_{22} - y_{04} \end{bmatrix}. \]
In a manner similar to what we have in (2.4), if \( y \) is a sequence of moments of some measure \( \mu_y \), then
\[ \langle q, M_v(\theta y)q \rangle = \int \theta(x)q(x)^2 \mu_y(dx), \tag{2.6} \]
for every polynomial \( q(x) : \mathbb{R}^n \rightarrow \mathbb{R} \) with coefficient vector \( q \in \mathbb{R}^{s(r)} \). Therefore, \( M_v(\theta y) \succeq 0 \) whenever \( \mu_y \) has its support contained in the set \( \{ \theta(x) \geq 0 \} \). In Curto and Fialkow [2], \( M_v(\theta y) \) is called a localizing matrix (denoted by \( M_v(r + v) \) if \( \deg \theta = 2v \) or \( 2v-1 \)).

The \( \mathbb{K} \)-moment problem identifies those sequences \( y \) that are moment-sequences of a measure with support contained in the semi-algebraic set \( \mathbb{K} \). In duality with the theory of moments is the theory of representation of positive polynomials, which dates back to Hilbert’s 17th problem. This fact will be reflected in the semidefinite relaxations proposed later. For details and recent results, the interested reader is referred to Curto and Fialkow [2], Jacobi [3], Jacobi and Prestel [4, 5], Simon [12], Schmüdgen [11] and the many references therein.

3. THE ASSOCIATED DISCRETE OPTIMIZATION PROBLEM \( \mathbb{P} \)

Consider the discrete optimization problem \( \mathbb{P} \)
\[ \mathbb{P} \rightarrow p^* := \min_{x \in \mathbb{R}^n} \{ p(x) \mid g_k(x) = 0, \ k = 1, \ldots, n \}, \tag{3.1} \]
where the polynomials \( g_k(x) : \mathbb{R}^n \rightarrow \mathbb{R} \) are defined by
\[ g_k(x) := 2^{r_k} \prod_{i=1}^{2r_k} (x - a_{ki}), \quad k = 1, \ldots, n, \tag{3.2} \]
with \( r_k \in \mathbb{N}, k = 1, \ldots, n, \) and the \( \{a_{ki}\}, i = 1, \ldots, r_k, \) are given real numbers such that for every \( k = 1, \ldots, n, \) \( a_{ki} \neq a_{kj} \) whenever \( i \neq j \). For homogeneity in notation, we have chosen to assume that all the polynomials \( g_k(x) \) have an even degree. The results presented below are also valid when the polynomials have arbitrary degree.

Let
\[ \mathbb{K} := \{ x \in \mathbb{R}^n \mid g_k(x) = 0, \ k = 1, \ldots, n \} \tag{3.3} \]
be the feasible set associated with \( \mathbb{P} \).

As we minimize \( p(x) \) on \( \mathbb{K} \), we could assume that the degree \( 2r_0 \) (or \( 2r_0 - 1 \)) of \( p(x) \) is not larger than \( r := \sum_{k=1}^{n} (2r_k - 1) \), since otherwise, using the equations \( g_k(x) = 0 \), we may replace \( p(x) \) with another polynomial \( \tilde{p}(x) \) of degree not larger than \( r \), and identical to \( p(x) \) on \( \mathbb{K} \). However, as for the representation of \( p(x) \) in [41] we will consider an arbitrary \( r_0 \), we do not make this assumption.
When needed below, for \( i \geq \max_k r_k \), the vectors \( g_k \in \mathbb{R}^{s(2r_k)} \) are extended to vectors of \( \mathbb{R}^{s(2n)} \) by completing with zeros. As we minimize \( p(x) \) we may and will assume that its constant term is zero, that is, \( p(0) = 0 \).

### 3.1. SDP relaxations of \( \mathbb{P} \)

For \( i \geq \max_{k \in \{0, n\}} r_k \), consider the following family \( \{Q_i\} \) of convex positive semidefinite (psd) programs (or semidefinite programming (SDP) relaxations of \( \mathbb{P} \))

\[
Q_i \begin{cases} 
\min_y \sum_{\alpha} p_{\alpha} y_{\alpha} \\
M_i(y) \succeq 0, \\
M_{i-r_k}(g_k y) = 0, \quad k = 1, \ldots, n, 
\end{cases}
\]

with respective dual problems

\[
Q_i^* \begin{cases} 
\max_{X \succeq 0, Z_k} -X(1,1) - \sum_{k=1}^n g_k(0) Z_k(1,1) \\
\langle X, B_{\alpha} \rangle + \sum_{k=1}^n \langle Z_k, C_{\alpha}^k \rangle = p_{\alpha}, \forall \alpha \neq 0 
\end{cases}
\]

where \( X, Z_k \) are real-valued symmetric matrices, the “dual variables” associated with the constraints \( M_i(y) \succeq 0 \) and \( M_{i-r_k}(g_k y) = 0 \) respectively, and where we have written

\[
M_i(y) = \sum_{\alpha} B_{\alpha} y_{\alpha}; \quad M_{i-r_k}(g_k y) = \sum_{\alpha} C_{\alpha}^k y_{\alpha}, \quad k = 1, \ldots, n,
\]

for appropriate real-valued symmetric matrices \( B_{\alpha}, C_{\alpha}^k, k = 1, \ldots, n \).

In the standard terminology, the constraint \( M_i(y) \succeq 0 \) is called a “linear matrix inequality” (LMI) and \( Q_i \) and its dual \( Q_i^* \) are so-called positive semidefinite (psd) programs, the semidefinite (SDP) relaxations of \( \mathbb{P} \). Both are convex optimization problems that can be solved efficiently via interior points methods and nowadays, several software packages (like e.g. the LMI toolbox of MATLAB) are available. The reader interested in more details on semidefinite programming is referred to Vandenberghe and Boyd [14] and the many references therein.

Note that the localizing matrices \( M_{i-r_k}(g_k y) \) are easily obtained from the data \( \{g_k\} \) of the problem \( \mathbb{P} \) by (2.5).

**Interpretation of** \( Q_i \). The linear matrix inequalities (LMI) constraints of \( Q_i \) state (only) necessary conditions for \( y \) to be the vector of moments up to order \( 2i \), of some probability measure \( \mu_y \) with support contained in \( \mathbb{K} \). This clearly implies that \( \inf Q_i \leq p^* \), for all \( i \), since the vector of moments of the Dirac measure at a feasible point of \( \mathbb{P} \) is feasible for \( Q_i \).

**Interpretation of** \( Q_i^* \). Let \( X \succeq 0 \) and \( \{Z_k\}, k = 1, \ldots, n \), be a feasible solution of \( Q_i^* \) with value \( \rho \). From the spectral decomposition of the symmetric matrices \( X \) and \( Z_k \), write

\[
X = \sum_j u_j w_j' \quad \text{and} \quad Z_k = \sum_j v_{kj} v_{kj}' - \sum_l w_{kl} w_{kl}', \quad k = 1, \ldots, n,
\]

where the vectors \( \{u_j\} \) correspond to the positive eigenvalues of \( X \), and the vectors \( \{v_{kj}\} \) (resp. \( \{w_{kl}\} \)) correspond to the positive (resp. negative) eigenvalues of \( Z_k \). Consider the polynomials \( \{u_j(x)\} \) and \( \{v_{kj}(x), w_{kl}(x)\} \) with respective coefficient vectors \( \{u_j\} \) and \( \{v_{kj}, w_{kl}\} \) in the basis (2.1).
Let \( x \in \mathbb{R}^n \) be fixed, arbitrary. Then, from the feasibility of \((X, Z_k)\) in \( \mathbb{Q}_+^n \),
\[
(X, B_\alpha x^\alpha) + \sum_{k=1}^n (Z_k, C_\alpha^k x^\alpha) = p_\alpha x^\alpha, \; \forall \alpha \neq 0,
\]
and
\[
X(1,1) + \sum_{k=1}^n g_k(0)Z_k(1,1) = -\rho.
\]
Using the notation
\[
y^x = (x_1, \ldots, x_n, x_1^2, x_1x_2, \ldots, x_1^i, \ldots, x_n^i),
\]
and summing up over all \( \alpha \), yields
\[
(X, M_1(y^x)) + \sum_{k=1}^n (Z_k, M_{i-r_k} (g_k y^x)) = p(x) - \rho,
\]
or, equivalently,
\[
\sum_j \langle u_j, M_1(y^x)u_j \rangle + \sum_{k=1}^n \left[ \sum_j \langle v_{kj}, M_{i-r_k} (g_k y^x) v_{kj} \rangle - \sum_l \langle w_{kl}, M_{i-v_k} (g_k y^x) w_{kl} \rangle \right] = p(x) - \rho.
\]
Therefore, from the definition of the matrices \( M_1(y), M_{i-r_k} (g_k y) \) (with \( y = y^x \)) and using \((3.7),(2.6)\),
\[
p(x) - \rho = \sum_j u_j(x)^2 + \sum_{k=1}^n g_k(x) \left[ \sum_j v_{kj}(x)^2 - \sum_l w_{kl}(x)^2 \right].
\]
As \( \rho \leq p^* \), \( p(x) - \rho \) is nonnegative on \( \mathbb{R} \) (strictly positive if \( p < p^* \)), and one recognizes in \((3.10)\) a “linear” representation into a weighted sum of squares of the polynomial \( p(x) - p^* \), nonnegative on \( \mathbb{R} \), as in the theory of representation of polynomials, strictly positive on a compact semi-algebraic set \( \mathbb{R} \) (see e.g., Schmüdgen [11], Putinar [10], Jacobi [8], Jacob and Prestel [4,5]). Indeed, when the set \( \mathbb{R} \) has a certain property, the “linear” (in the sense that no product term like \( g_k(x)g_l(x) \) is needed) representation \((3.10)\) holds. This property states that there is some polynomial \( U(x) : \mathbb{R}^n \to \mathbb{R} \) such that \( U(x) \) has the representation \((3.10)\) and \( \{U(x) \geq 0\} \) is compact (see Putinar [10], Jacob and Prestel [4,5]).

It will be shown below that indeed, the polynomial \( p(x) - p^* \), which is only nonnegative (and not strictly positive) on \( \mathbb{R} \), has the representation \((3.10)\).

Hence, both psd programs \( \mathbb{Q}_i \) and \( \mathbb{Q}_i^\ast \) perfectly match the duality between the \( \mathbb{R} \)-moment problem and the theory of polynomials that are positive on \( \mathbb{R} \).

3.2. Simplified SDP relaxations. The SDP relaxation \( \mathbb{Q}_i \) has a much simpler form that can be derived as follows. For convenience, write
\[
g_k(x) := x_k^{2r_k} + \sum_{j=0}^{2r_k-1} \gamma_{kj} x_k^j, \quad k = 1, \ldots, n.
\]
Observe that in view of the construction of the localizing matrices in \((2.4)\), and the form of the polynomials \( g_k(x), k = 1, \ldots, n, \) the constraints \( M_{i-r_k} (g_k y) = 0 \)
generate linear relationships between “moments” that can be translated into the moment matrix $M_i(y)$ as follows.

Given a monomial $x^\alpha$, use (3.11) to replace $x_k^{2r_k}$ by

$$
\sum_{j=0}^{2r_k-1} \gamma_{kj} x_k^j, \quad k = 1, \ldots, n,
$$

every time $\alpha_k \geq 2r_k$, until $x^\alpha$ is expressed as a linear combination $\sum_l h_l x^\beta_l$ of monomials $x^\beta_l$ with $(\beta_l)_k \leq 2r_k$ for all $k = 1, \ldots, n$ and all $l$, that is,

$$
x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n} = \sum_l h_l x^{\beta_l}.
$$

Now, if $M_i(y)(s,t) = y_\alpha = (y_{\alpha_1}, \ldots, y_{\alpha_n})$ is the entry $(s,t)$ of the matrix $M_i(y)$, then formally replace $y_\alpha$ in $M_i(y)(s,t)$ by $\sum_l h_l y_{\beta_l}$. Hence, the LMI constraints $M_i(y) \succeq 0$ and $M_{i-r_k}(g_k y) = 0$, for all $k = 1, \ldots, n$, reduce to the single LMI constraint $\tilde{M}_i(y) \succeq 0$ obtained from $M_i(y) \succeq 0$ by making the above substitution in all the entries of $M_i(y)$. Observe that no matter how large $i$ is, $\tilde{M}_i(y)$ contains only “moments” $y_\beta$ with $\beta_k \leq 2r_k$ for all $k = 1, \ldots, n$.

Therefore, $r := \sum_{k=1}^n (2r_k - 1)$ is the maximum degree of distinct monomials in the sense that whenever $i > r$ a polynomial of degree $i$ can be expressed as a linear combination of monomials of degree less than $r$, after the simplification induced by the constraints $g_k(x) = 0$, for all $k = 1, \ldots, n$.

Thus, with $s := \prod_{k=1}^n 2r_k$, there are no more than $s - 1$ variables $y_\beta$ in all relaxations $Q_i$ (that is, $s$ is the number of monomials in (2.11) of degree less than $r$), and the relaxation $Q_i$ has the simplified form

$$
\begin{align}
\min_y \quad & \sum_\alpha \tilde{p}_\alpha y_\alpha \\
\text{subject to} \quad & \tilde{M}_i(y) \succeq 0,
\end{align}
$$

where we only have variables $y_\alpha$ with $\alpha_k < 2r_k$ for all $k = 1, \ldots, n$, and $\{\tilde{p}_\alpha\}$ is the vector of coefficients of a polynomial $\tilde{p}(x)$ identical to $p(x)$ on $K$.

For instance, in $\mathbb{R}^2$ ($n = 2$), let $g_k(x) = x_k^2 - x_k$ $(r_k = 1)$ so that the set $K$ defines the 4 grid points $(0,0), (0,1), (1,0), (1,1)$ in $\mathbb{R}^2$. Then,

$$
M_2(y) = \begin{bmatrix}
1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\
 y_{10} & y_{10} & y_{11} & y_{30} & y_{21} & y_{12} \\
 y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\
 y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\
 y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\
 y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04}
\end{bmatrix}
$$
is replaced with
\[
\tilde{M}_2(y) = \begin{bmatrix}
1 & y_{10} & y_{01} & y_{10} & y_{11} & y_{01} \\
- & - & - & - & - & - \\
y_{10} & y_{10} & y_{11} & y_{10} & y_{11} & y_{11} \\
y_{01} & y_{11} & y_{01} & y_{11} & y_{11} & y_{11} \\
- & - & - & - & - & - \\
y_{10} & y_{10} & y_{11} & y_{10} & y_{11} & y_{11} \\
y_{11} & y_{11} & y_{11} & y_{11} & y_{11} & y_{11} \\
y_{01} & y_{11} & y_{01} & y_{11} & y_{11} & y_{11} \\
\end{bmatrix}
\]
and only the variables $y_{10}, y_{01}, y_{11}$ appear in all the relaxations $Q_i$.

3.3. Main result. Before stating the main result of this section, begin with the following crucial result:

**Proposition 3.1.** Let $r := \sum_{k=1}^{n} (2r_k - 1)$ and $s := \prod_{k=1}^{n} 2r_k$.
(a) All the simplified relaxations $Q_i$ involve at most $s - 1$ variables $y_\alpha$.
(b) For all the relaxations $Q_i$ with $i > r$, one has
\[
\text{rank } \bar{M}_i(y) = \text{rank } M_r(y).
\]

**Proof.** (a) is just a consequence of the comment preceding Proposition 3.1.
To get (b) observe that with $i > r$, one may write
\[
M_i(y) = \begin{bmatrix}
M_r(y) & B(y) \\
- & - \\
B'(y) & C(y)
\end{bmatrix}
\]
for appropriate matrices $B, C$, and we next prove that each column of \( \begin{bmatrix} B(y) \\ C(y) \end{bmatrix} \) is a linear combination of some columns of \( \begin{bmatrix} M_r(y) \\ B'(y) \end{bmatrix} \).

Indeed, remember from (2.3) how an element $\bar{M}_i(y)(k,p)$ can be obtained. Let $y_\delta = M_i(y)(k,1)$ and $y_\alpha = M_i(y)(1,p)$. Then
\[
M_i(y)(k,p) = y_\eta \text{ with } \eta_i = \delta_i + \alpha_i, i = 1, \ldots, n,
\]
that is, $M_i(y)(k,p)$ is the “moment” which corresponds to the monomial $x^\delta \times x^\alpha$ (denoted $M_i(k,p) \leftrightarrow x^\delta \times x^\alpha$). Now, consider a column \( \begin{bmatrix} B(y)(\cdot,m) \\ C(y)(\cdot,m) \end{bmatrix} \) of \( \begin{bmatrix} B \\ C \end{bmatrix} \), that is, some column $M_i(y)(\cdot,p)$ of $M_i(y)$, with first element $y_\alpha = B(y)(1,m) = M_i(y)(1,p)$. Therefore, the element \( \begin{bmatrix} B(y) \\ C(y) \end{bmatrix}(k,m) \) (or, equivalently, the element $M_i(y)(k,p)$) is the variable $y_\eta$ in (3.10). Note that $\alpha$ corresponds to the monomial $x^{\alpha_1 \cdots \alpha_n}$ in the basis (2.1), of degree larger than $r$. We have seen that in view of the constraint $M_{i-r_k}(g_k y) = 0$, this element satisfies
\[
y_\alpha = \sum_j \gamma_j y_{\beta(j)},
\]
that is, $y_\alpha$ is a linear combination of variables $y_{\beta(j)}$ with
\[
\beta(j) = \beta_1^{(j)} \cdots \beta_n^{(j)} \text{ and } \beta_k^{(j)} < 2r_k, \forall j = 1, \ldots, n.
\]
Now, as $y_{\beta(j)}$ corresponds to a monomial in the basis (2.1) of degree less than $r$, to each $y_{\beta(j)}$ corresponds a column $M_r(y)(\cdot,p_j)$ (more precisely, $M_r(y)(1,p_j)$ ↔
atom probability measure \( x \) and thus, \( B(y)(1,m) \) is a linear combination of the elements \( M_i(y)(1,p_j) \) with coefficients \( \gamma_j \) as in \( (3.17) \). Next, by construction, the element \( M_i(y)(k,p) \) corresponds to the monomial \( x^\delta \times x^\alpha \), and we have the correspondences

\[
y_q = M_i(y)(k,p) \leftrightarrow x^\delta \times x^\alpha = x^\delta \times \sum_j \gamma_j x^{\beta(j)}
\]

\[
= \sum_j \gamma_j (x^\delta \times x^{\beta(j)}) \leftrightarrow \sum_j \gamma_j \left[ \frac{M_r(y)}{B'} \right](k,p_j),
\]

and thus,

\[
M_i(y)(k,p) = \sum_j \gamma_j \left[ \frac{M_r(y)}{B'} \right](k,p_j),
\]

which states that the column \( \left[ \frac{B(y)}{C(y)} \right](.,m) \) is a linear combination of the columns \( \left[ \frac{M_r(y)}{B'} \right](.,p_j) \). From, this it immediately follows that rank \( M_i(y) = \text{rank} \ M_r(y) \) whenever \( i > r \).

We now are in position to state the following result.

**Theorem 3.2.** Let \( \mathcal{P} \) be as in \( (3.1) \) with \( \mathbb{K} \) as in \( (3.3) \) and let \( r := \sum_{k=1}^n (2k-1) \), \( v := \max\{r_0-r, \max_{k=1,...,n} r_k \} \). Then, whenever \( i \geq r + v \),

\[
p^* = \min Q_i,
\]

and every optimal solution \( y^* \) of \( Q_i \) is the vector of moments of a probability measure \( \mu_{y^*} \) supported on \( s = \text{rank} M_r(y) \) optimal solutions of \( \mathcal{P} \).

**Proof.** First let \( i := r+v \). Let \( y \) be any feasible solution of \( Q_i \). From Proposition \( (3.1) \) it follows that rank \( M_{r+1}(y) = \text{rank} M_r(y) \). Therefore, from a deep result of Curto and Fialkow, namely \[2, \text{Th. 1.1, p. 4}] \), \( M_r(y) \) has positive flat extensions \( M_{r+j}(y) \) for all \( j = 1, 2, \ldots \), that is, \( M_{r+j}(y) \geq 0 \) and rank \( M_{r+j}(y) = \text{rank} M_r(y) \) for all \( j = 1, 2, \ldots \). This implies that \( y \) is the vector of moments of some rank \( M_r(y) \) -atomic probability measure \( \mu_y \). Moreover, from \( M_{r-k}(g_k y) = 0 \) (equivalently \( M_{r+k}(g_k y) \geq 0 \) and \( M_{r-k}(-g_k y) \geq 0 \)), it also follows that \( \mu_y \) has its support contained in \( \bigcap_{k=1}^n \{ g_k(x) = 0 \} \), that is, the set \( \mathbb{K} \). In the terminology of Curto and Fialkow \[2, \text{p. 4}] \), \( M_{r+v}(g_k y) \) is well-defined relative to the unique flat extension \( M_{r+v}(y) \) of \( M_r(y) \). Theorem 1.1 in Curto and Fialkow \[2, \text{p. 4}] \) is stated for the complex \( \mathbb{K} \) -moment problem, but is valid for the \( \mathbb{K} \) -moment problem in several real or complex variables (see \[2, \text{p. 2}] \).

But then, for every admissible solution \( y \) of \( Q_i \), we have

\[
\sum_\alpha p_\alpha y_\alpha = \int_\mathbb{R} p(x) \mu_y(dx) \geq p^* \mu_y(\mathbb{K}) = p^*
\]

so that \( p^* \geq \inf Q_i \), which combined with \( \inf Q_i \leq p^* \) for all \( i \), yields \( \inf Q_i = p^* \) when \( i = r+v \). For \( i > r+v \), it suffices to notice that \( p^* \geq \inf Q_{i+1} \geq \inf Q_i \) for all \( i \), to get \( \inf Q_i = p^* \) whenever \( i > r+v \). Moreover, to a global minimizer \( x^* \in \mathbb{K} \) of \( \mathcal{P} \), corresponds the admissible solution

\[
y^* = (x_1^*, \ldots, x_n^*, \ldots, (x_1^*)^{2i}, \ldots, (x_n^*)^{2i})
\]
of $Q_i$, with value $p^*$, which implies $\inf Q_i = \min Q_i = p^*$ whenever $i > r + v$. \hfill \square

In Theorem 3.2 we have $i \geq r + v$, and $v$ depends on $r_0$. Indeed, to be consistent, $Q_i$ must be such that $2i \geq 2r_0$ to contain all the moments up to $2r_0$. However, we can make the following remark.

Remark 3.3. We have seen that $Q_i$ has the simplified equivalent form (3.12) with the single LMI constraint $\tilde{M}_i(y) \succeq 0$ after the simplification induced by the constraints $g_k(x) = 0$. But, since $M_i(y) \succeq 0$ implies that $M_{i+j}(y) \succeq 0$ and rank $M_{i+j}(y) = \text{rank } M_i(y)$ for all $j = 1, \ldots$, it follows that $\tilde{M}_i(y) \succeq 0$ also implies $\tilde{M}_{i+j}(y) \succeq 0$ and rank $\tilde{M}_{i+j}(y) = \text{rank } \tilde{M}_i(y)$ for all $j = 1, \ldots$. Therefore, $p^*$ is obtained by solving the SDP relaxation

\begin{equation}
\min_y \{ \sum_\alpha \tilde{p}_\alpha y_\alpha | \tilde{M}_r(y) \succeq 0 \},
\end{equation}

in which we only have variables $y_\alpha$ such that $\alpha_k < 2r_k$ for all $k = 1, \ldots, n$.

Therefore, no matter its degree, the global minimization of a polynomial $p(x)$ on $\mathbb{R}$ reduces to a **convex continuous optimization problem of fixed dimension**, namely the convex psd program (3.19). This is because even if $p(x)$ has degree larger than $r$, the $\{\tilde{p}_\alpha\}$ are the coefficients of a polynomial $\tilde{p}(x)$ of degree less than $r$, identical to $p(x)$ on $\mathbb{R}$. However, this simplification is formally obtained at the relaxation $Q_{r+v}$ that we will need for the representation of $p(x)$ in (4).

Despite its theoretical interest, Theorem 3.2 is of little value for solving discrete optimization problems for the dimension of the psd program $Q_{r+v}$ (or, equivalently, (3.19)) is exponential in the problem size.

Fortunately, in many cases, the optimal value $p^*$ is obtained at relaxations $Q_i$ with $i \ll r + v$. For instance, this will be the case whenever $p(x) - p^*$ has the representation (3.10) with polynomials $\{u_j(x)\}$ of maximum degree $a_0 < r + v$, and polynomials $\{v_{kl}(x), w_{kl}(x)\}$ of maximum degree $a_k < r + v - r_k$, $k = 1, \ldots, n$. Indeed, let $\{u_j, v_{kj}, w_{kl}\}$ be their vectors of coefficients in $\mathbb{R}^{s(i)}$, $\mathbb{R}^{s(i-r_k)}$ respectively, with $i := \text{max}[a_0, \text{max}_k(a_k + r_k)]$. Then, the real-valued symmetric matrices $X \in \mathbb{R}^{s(a_0) \times s(a_0)}$ and $Z_k \in \mathbb{R}^{s(a_k) \times s(a_k)}$, $k = 1, \ldots, n$, defined by

\begin{align*}
X &:= \sum_j u_j u_j', & Z_k &:= \sum_j v_{kj} v_{kj}' - \sum_l w_{kl} w_{kl}'
\end{align*}

form an admissible solution of $Q_i^*$ with value $p^*$ (just redo backward the derivations (3.6)-(3.9)). Since from weak duality we must have $\sup Q_i^* \leq \inf Q_i \leq p^*$, it follows that $\inf Q_i = p^*$, and thus, $\min Q_i = p^*$ (since the vector $y^*$ of moments of the Dirac measure $\delta_{x^*}$ at a global minimizer $x^*$ of $\mathbb{P}$, is obviously an admissible solution of $Q_i$ with value $p^*$).

For instance, consider the so-called MAX-CUT discrete optimization problem

$$\min_{x \in \{0,1\}^n} x' M x,$$

where $M \in \mathbb{R}^{n \times n}$ is a real-valued symmetric matrix with a null diagonal, which is known to be NP-hard. We have solved a sample of MAX-CUT problems in $\mathbb{R}^{10}$ with the the nondiagonal entries of $M$ randomly generated, uniformly between 0 and 1. In all cases, the SDP relaxation $Q_2$ provided the optimal value $p^*$, with no need to solve $Q_{11}$, as predicted by Theorem 3.2 (since in this case, with $r_k = 1$, $r = \sum_k (2r_k - 1) = 10$). The computational savings are significant since the simplified
form $\mathcal{Q}_2$ of $\mathbb{Q}_2$ is a psd program with 385 variables $y_{i0}$ and a single LMI constraint of dimension $56 \times 56$, whereas $\mathcal{Q}_{11}$ would involve $2^4 - 1 = 1023$ variables and a matrix $M_{10}(y)$ of size $1024 \times 1024$!

**Remark 3.4.** If $p(x) : \mathbb{R}^n \to \mathbb{R}$ is concave, then its minimum $p^*$ on the grid $\mathbb{K}$ is attained at some “corner” of $\mathbb{K}$. Therefore, only the corner points are useful and one may replace the initial grid $\mathbb{K}$, by the new grid $\mathbb{K}'$ defined by the quadratic equality constraints

$$g'_k(x) := (x_k - a_{k1}) \times (x_k - a_{k2r_k}) = 0, \quad k = 1, \ldots, n.$$  

The resulting grid $\mathbb{K}'$ has only $2^n$ points and the optimal value $p^*$ is obtained at most at the $\mathcal{Q}_{n+v}$ relaxation (since $r = n$).

### 3.4. Constrained discrete optimization.** Consider now the problem $\mathcal{P}$ with the additional constraints $h_j(x) \geq 0$, $j = 1, \ldots, m$, for some real-valued polynomials $h_j(x) : \mathbb{R}^n \to \mathbb{R}$ of degree $2v_j$ or $2v_j - 1$ (depending on the parity), that we may assume to be not larger than $r$. The new set $\mathbb{K}$ is now $\mathbb{K}' := \mathbb{K} \cap \{ [h_j(x) \geq 0] \}$ and the new associated relaxations $\mathcal{Q}_i$ now include the additional constraints $M_{1-v_j}(h_jy) \geq 0, \ j = 1, \ldots, m$. Theorem 3.2 also holds, that is, the (new) optimal value $p^*$ is also obtained at the relaxation $\mathcal{Q}_{r+v}$ with new $v := \max\{r_0 - r, \max_j v_j, \max_k r_k\}$. Indeed, as in the proof of Theorem 3.2, we may invoke Theorem 1.1 of Curto and Fialkow [2] to show that $M_r(y)$ admits flat positive extensions $M_{r+l}(y)$, $l = 1, \ldots, n$, and as $M_{r+v-j}(h_jy)$ are all well defined, relative to the unique flat extension $M_{r+v}(y)$, every admissible solution $y$ of $\mathcal{Q}_{r+v}$ is the vector of moments of some probability measure $\mu_y$ supported on $\mathbb{K'}$, and the result follows with the same arguments.

### 4. Representation of polynomials nonnegative on $\mathbb{K}$

We now investigate the representation of polynomials that are nonnegative on $\mathbb{K}$. We will see that the previous results obtained for the discrete optimization problem $\mathcal{P}$ will help us to get the desired representation of $p(x) - p^*$, and thus, the representation of every polynomial $p(x)$, nonnegative on $\mathbb{K}$. We then present results for concave polynomials.

#### 4.1. Representation of polynomials, nonnegative on $\mathbb{K}$

For a polynomial $g(x)$, nonnegative on $\mathbb{K}$, let $\tilde{g}(x) := g(x) - g(0)$ and let $\tilde{p}^* := \min_{x \in \mathbb{K}} \tilde{g}(x)$ and $0 \leq p^* := \min_{x \in \mathbb{K}} g(x)$. Then, $g(x) = \tilde{g}(x) - \tilde{p}^* + p^*$. Therefore, if one obtains a representation of $\tilde{g}(x) - \tilde{p}^*$ as a weighted sum of squares, one obtains the same representation for the polynomial $g(x)$, positive on $\mathbb{K}$, by adding the nonnegative constant term $p^*$. Therefore, with no loss of generality, we may and will assume that the constant term of $p(x)$ vanishes, that is, $p(0) = 0$. Consider the polynomial $p(x) - p^*$, with $p^* := \min_{x \in \mathbb{K}} p(x)$, which is nonnegative on $\mathbb{K}$ and let $2r_0$ (resp. $2r_0 - 1$) be its degree if even (resp. odd). Before proceeding further, we need the following remark.

**Remark 4.1.** We have seen that with

$$v := \max\{r_0 - r, \max_{k=1,\ldots,n} r_k\},$$
we have \(\min Q_{r+v} = p^*\) and \(Q_{r+v}\) is equivalent to solving the psd program
\[
\min_y \left\{ \sum_{\alpha} \tilde{p}_\alpha y_\alpha \mid \tilde{M}_r(y) \succeq 0 \right\},
\]
where \(\tilde{M}_r(y)\) is the “simplified” form of \(M_r(y)\) after the simplifications in \(Q_{r+v}\) induced by the constraints \(g_k(x) = 0, k = 1, \ldots, n\) (cf. (3.19) in Remark 3.3), so that we only have variables \(y_\alpha\) with \(\alpha_k < 2r_k\) for all \(k\).

In fact, \(\tilde{M}_r(y)\) can be further simplified and reduced in size. Indeed, all the columns of \(\tilde{M}_r(y)\) that correspond to monomials \(x^\alpha\) in the basis (2.1), and simplified as a linear combination of monomials \(x^\beta\) with \(\beta_k < 2r_k\) for all \(k\), can be deleted (and the corresponding rows as well). Indeed, by rearranging rows and columns, write
\[
\tilde{M}_r(y) = \begin{bmatrix} H_r(y) & B(y) \\ - & - \\ B'(y) & C(y) \end{bmatrix},
\]
where the elements of \(H_r(y)(1,\) correspond to the monomials \(x^\beta\) in the basis (2.1) with \(\beta_k < 2r_k\) for all \(k\). Thus, \(\begin{bmatrix} B(y) \\ C(y) \end{bmatrix}\) is the submatrix of \(\tilde{M}_r(y)\) whose columns are precisely the columns of \(\tilde{M}_r(y)\) that are linear combinations of the columns of \(\begin{bmatrix} H_r(y) \\ B'(y) \end{bmatrix}\). Therefore, we can write \(B(y) = H_r(y)W\) and \(C(y) = B'(y)W\) for some matrix \(W\). Next, let \(y\) be such that \(H_r(y) \succeq 0\). From \(C(y) = WH_r(y)W \succeq 0\), it follows that \(\tilde{M}_r(y) \succeq 0\) and, in addition, rank \(\tilde{M}_r(y) = \text{rank} H_r(y)\). Therefore, \(y\) is admissible for \(Q_{r+v}\) with same value, so that solving \(Q_{r+v}\) is equivalent to solving the new psd program of reduced size
\[
(4.1) \quad \min \left\{ \sum_{\alpha} \tilde{p}_\alpha y_\alpha \mid H_r(y) \succeq 0 \right\}.
\]
With \(s := \prod_{k=1}^n 2r_k\), \(H_r(y)\) is a matrix of size \(s \times s\), \((s\) is precisely the number of monomials \(x^\beta\) with \(\beta_k < 2r_k\) for all \(k = 1, \ldots, n\)).

To illustrate the above Remark 4.1, consider the matrix \(\tilde{M}_2(y)\) in (5.14) (for the case \(n = 2\) and \(g_k(x) = x_k^2 - x_k\)), which can be written
\[
\tilde{M}_2(y) = \begin{bmatrix} 1 & y_{10} & y_{01} & y_{11} & y_{10} & y_{01} \\ y_{10} & y_{10} & y_{11} & y_{11} & y_{10} & y_{11} \\ y_{01} & y_{11} & y_{01} & y_{11} & y_{11} & y_{01} \\ y_{11} & y_{11} & y_{11} & y_{11} & y_{11} & y_{11} \\ - & - & - & - & - & - \\ y_{10} & y_{10} & y_{11} & y_{11} & y_{10} & y_{11} \\ y_{01} & y_{11} & y_{01} & y_{11} & y_{11} & y_{01} \end{bmatrix},
\]
with the last two columns being identical to the second and third columns respectively. Therefore, in this case, solving \(Q_{2+v}\) reduces to solving the psd program
\[
\begin{array}{c} \min \sum_{\alpha} \tilde{p}_\alpha y_\alpha \\ \min H_2(y) \end{array} \begin{bmatrix} 1 & y_{10} & y_{01} & y_{11} \\ y_{10} & y_{10} & y_{11} & y_{11} \\ y_{01} & y_{11} & y_{01} & y_{11} \\ y_{11} & y_{11} & y_{11} & y_{11} \end{bmatrix} \succeq 0.
\]
We need the simplified psd program $\overline{Q}_r$ to prove the absence of a duality gap between $Q_{r+v}(y)$ and its dual $Q_{r+v}^*$, which in turn, will yield the desired result on the representation of polynomials $p(x)$, nonnegative on $\mathbb{K}$.

**Theorem 4.2.** Let $p(x) : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary polynomial of degree $2r_0$ or $2r_0 - 1$, with $p(0) = 0$. Let $\mathbb{P}$ be as in (3.1) with $\mathbb{K}$ as in (3.3) and let $r := \sum_{k=1}^n (2r_k - 1)$ and $s := \prod_{k=1}^n 2r_k$. Then with $v := \max\{r_0 - r, \max_{k=1}^n r_k\}$.

(i) There is no duality gap between the psd program $Q_{r+v}$ and its dual $Q_{r+v}^*$.

Moreover, $Q_{r+v}^*$ is solvable, that is,

$$\max Q_{r+v}^* = \min Q_{r+v} = p^*.$$  \hspace{1cm} (4.2)

(ii) $p(x) - p^*$ has the following representation:

$$p(x) - p^* = \sum_{j=1}^{m_0} u_j(x)^2 + \sum_{k=1}^{n} q_k(x) \left[ \sum_{j=1}^{m_k} v_{kj}(x)^2 - \sum_{l=1}^{n} w_{kl}(x)^2 \right],$$  \hspace{1cm} (4.3)

for some $m_0 (\leq s(r + v))$ polynomials $\{u_j(x)\}$ of degree at most $r + v$ and some polynomials $\{v_{kj}(x), w_{kl}(x)\}$ of degree at most $r + v - r_k$, with $m_k + n_k \leq s(r + v - r_k), k = 1, \ldots, n$.

**Proof.** From Remark 4.1 we know that $\min \overline{Q}_r = p^*$ and the dimension of the matrix $H_r(y)$ is $s \times s$ with $s := \prod_{k=1}^n 2r_k$. Next, to each of the $s$ (grid) points of $\mathbb{K}$, labelled $x(k), k = 1, \ldots, s$, corresponds an admissible solution $y(k)$ of $Q_{r+v}$ with

$$y(k) = (x(k)_{1, \ldots, n}, \ldots, x(k)^{2(r+v)}, \ldots, x(k)^{2(r+v)}),$$  \hspace{1cm} (4.4)

and with

$$M_{r+v}(y(k)) \succeq 0, \quad k = 1, \ldots, s.$$ 

Let $\tilde{y}(k)$ be the vector obtained from $y(k)$ by keeping the only variables $y(k)_\alpha$ corresponding to the monomials $x^\alpha$ with $\alpha_k < 2r_k$ for all $k$, then the vector $(1, \tilde{y}(k))$ satisfies $H_r(\tilde{y}(k)) \succeq 0$ and $\tilde{y}(k)$ is admissible for $\overline{Q}_r(y)$ in (4.1).

Moreover, the $s$ points $(1, \tilde{y}(k)) \in \mathbb{R}^s$ are linearly independent (see a proof in the Appendix). Consider a convex combination $z \in \mathbb{R}^s$ defined by

$$z := \sum_{k=1}^s z_k (1, \tilde{y}(k)), \text{ with } z_k > 0 \text{ and } \sum_{k=1}^s z_k = 1.$$ 

From the linearity of $H_r(z)$ it follows that

$$H_r(z) = \sum_{k=1}^s z_k H_r(\tilde{y}(k)) \succeq 0.$$ 

Moreover, from the definition of the moment matrix $M_{r+v}(y(k))$, all the matrices $M_{r+v}(y(k))$ are the rank-one matrices

$$M_{r+v}(y(k)) = \begin{bmatrix} 1 \\ y(k) \end{bmatrix} \times [1, y(k)], \quad k = 1, \ldots, s.$$ 

From this and the definition of the simplified moment matrix $H_r(y)$, it also follows that all the matrices $H_r(\tilde{y}(k))$ are the rank-one matrices

$$H_r(\tilde{y}(k)) = \begin{bmatrix} 1 \\ \tilde{y}(k) \end{bmatrix} \times [1, \tilde{y}(k)], \quad k = 1, \ldots, s.$$
As the points \((1, \tilde{y}(k))\) are linearly independent in \(\mathbb{R}^n\), it follows that \(H_r(z) > 0\), that is, \(H_r(z)\) is positive definite, and thus, \(z\) is a strictly admissible solution of \(\mathcal{Q}_r\). In other words, Slater’s interior point condition holds for the convex psd program \(\mathcal{Q}_r\).

As \(\inf \mathcal{Q}_r = p^* > -\infty\), by a standard (strong duality) result in convex optimization, we have that \(\mathcal{Q}_r^*\) is solvable and there there is no duality gap between \(\mathcal{Q}_r\) and \(\mathcal{Q}_r^*\) (see e.g. Sturm [13, Th. 2.24]). That is,

\[
\max \mathcal{Q}_r^* = \min \mathcal{Q}_r = p^*.
\]

If we now remember how \(\mathcal{Q}_r\) was obtained from \(\mathcal{Q}_{r+v}\), it also follows that \(\mathcal{Q}_{r+v}^*\) is also solvable and there is no duality gap between \(\mathcal{Q}_{r+v}\) and \(\mathcal{Q}_{r+v}^*\), which proves the first assertion of Theorem 1.2.

Next, as \(\mathcal{Q}_{r+v}^*\) is solvable with optimal value \(p^*\), let \(X^* \succeq 0\) and \(Z_k^+, k = 1, \ldots, n\), be an optimal solution of \(\mathcal{Q}_{r+v}^*\). From the spectral decomposition of the symmetric matrices \(X^*\) and \(Z_k\), write

\[
X^* = \sum_{j=1}^{m_0} u_j u_j^T \quad \text{and} \quad Z_k^+ = \sum_{j=1}^{m_k} v_{kj} v_{kj}^T - \sum_{l=1}^{n_k} w_{kl} w_{kl}^T, \quad k = 1, \ldots, n,
\]

where the vectors \(\{u_j\}\) correspond to the \(m_0\) positive eigenvalues of \(X^*\), and the vectors \(\{v_{kj}\}\) (resp. \(\{w_{kl}\}\)) correspond to the \(m_k\) positive (resp. the \(n_k\) negative) eigenvalues of \(Z_k^+\). Consider the polynomials \(\{u_j(x)\}\) and \(\{v_{kj}(x), w_{kl}(x)\}\) with respective coefficient vectors \(\{u_j\}\) and \(\{v_{kj}, w_{kl}\}\), in the basis (2.1).

Fix \(x \in \mathbb{R}^n\), arbitrary. From the feasibility of \((X^*, Z_k^+)\) in \(\mathcal{Q}_{r+v}^*\),

\[
\langle X^*, B_{\alpha} x^\alpha \rangle + \sum_{k=1}^{n} \langle Z_k^+, C_{\alpha} x^\alpha \rangle = p_{\alpha} x^\alpha, \forall \alpha \neq 0,
\]

and from \(\max \mathcal{Q}_{r+v}^* = p^*\)

\[
X^*(1, 1) + \sum_{k=1}^{n} g_k(0)Z_k^+(1, 1) = -p^*.
\]

Using the notation \(y^x = (x_1, \ldots, x_n, x_1^2, x_1 x_2, \ldots, x_1^{2(r+v)}, \ldots, x_n^{2(r+v)})\), and summing up over all \(\alpha\), yields

\[
\langle X^*, M_{r+v} (y^x) \rangle + \sum_{k=1}^{n} \langle Z_k^+, M_{r+v-r_k} (g_k y^x) \rangle = p(x) - p^*,
\]

or, equivalently,

\[
p(x) - p^* = \sum_{j=1}^{m_0} \langle u_j, M_{r+v} (y^x) u_j \rangle + \sum_{k=1}^{n} \left( \sum_{j=1}^{m_k} \langle v_{kj}, M_{r+v-r_k} (g_k y^x) v_{kj} \rangle - \sum_{l=1}^{n_k} \langle w_{kl}, M_{r+v-r_k} (g_k y^x) w_{kl} \rangle \right).
\]

Therefore, using (2.4)–(2.6) (with \(y^x\)),

\[
p(x) - p^* = \sum_{j=1}^{m_0} u_j(x)^2 + \sum_{k=1}^{n} g_k(x) \left[ \sum_{j=1}^{m_k} v_{kj}(x)^2 - \sum_{l=1}^{n_k} w_{kl}(x)^2 \right]
\]

which is (1.3).
4.2. Concave polynomials. As a consequence of Theorem 3.2 we obtain a representation of concave polynomials. Remember that the global minimum of a concave polynomial on a grid \( \mathbb{K} \) is attained at some “corner” point of the grid \( \mathbb{K} \) (see Remark 3.3). Therefore, let \( p(x) : \mathbb{R}^n \to \mathbb{R} \) be a concave polynomial of degree \( 2r_0 \) or \( 2r_0 - 1 \), with \( p^* := \min_{x \in \mathbb{K}} p(x) \). Then,

\[
p(x) - p^* = \sum_{j=1}^{m_0} u_j(x)^2 + \sum_{k=1}^{n} (x_k - a_{k1})(a_{k2r_k} - x_k) \left[ \sum_{j=1}^{m_k} v_{kj}(x)^2 - \sum_{l=1}^{n_k} w_{kl}(x)^2 \right]
\]

for some polynomials \( \{u_j(x)\} \) of degree at most \( n + v \), and some polynomials \( \{v_{kj}(x)\} \) and \( \{w_{kl}(x)\} \) of degree at most \( n + v - 1 \).

This is because, we may replace the initial grid \( \mathbb{K} \) with the coarser grid \( \mathbb{K}' \) defined by the \( 2^n \) corner points, so that we have \( r = n \) and \( v = \max[1, r_0 - n] \). Moreover, from the concavity of \( p(x) \),

\[
p^* = \min_{x \in \mathbb{K}} p(x) = \min \{ p(x) \mid a_{k1} \leq x_k \leq a_{k2r_k}, \ x \in \mathbb{K} \},
\]

that is, \( p^* \) is also the global minimum of \( p(x) \) on the box \( S := \prod_{k=1}^{n} [a_{k1}, a_{k2r_k}] \) (equivalently, the convex hull of \( \mathbb{K} \)). But, observe that \( \text{(1.5)} \) is also a “linear” representation with respect to the constraints \( (x - a_{k1})(a_{k2r_k} - x) \geq 0, k = 1, \ldots, n \), that are necessary conditions for \( x \in S \). Therefore, we have

\[
p(x) - \min_{x \in S} p(x) = \sum_{j=1}^{m_0} u_j(x)^2 + \sum_{k=1}^{n} (x_k - a_{k1})(a_{k2r_k} - x_k) \left[ \sum_{j=1}^{m_k} v_{kj}(x)^2 - \sum_{l=1}^{n_k} w_{kl}(x)^2 \right].
\]

Remark 4.3. The reader will easily convince himself that Theorem 3.2 of this paper is also valid if instead of a grid, \( \mathbb{K} \) is now the more general variety \( \prod_{k=1}^{n} \{ g_k(x) = 0 \} \), with polynomials \( g_k(x) \) of the form

\[
x \mapsto g_k(x) := x_{rk} + h_k(x), \quad k = 1, \ldots, n,
\]

and where \( h_k(x) \) is a polynomial of degree not larger than \( r_k - 1 \).

4.3. Concave polynomials on a simplex. We end up this section with the characterization of concave polynomials on a simplex. Let \( \Delta \subset \mathbb{R}^n \) be the canonical simplex

\[
\Delta := \{ x \in \mathbb{R}_+^n \mid \sum_{i=1}^{n} x_i = 1 \}.
\]

Let \( p(x) : \mathbb{R}^n \to \mathbb{R} \) be a concave polynomial of degree \( 2r_0 \) or \( 2r_0 - 1 \). Its minimum on \( \Delta \) is attained at some vertex of \( \Delta \). Therefore,

\[
P \to p^* = \min_{x \in \Delta} p(x) = \min_{x \in \Delta \cap \theta(0,1)^n} p(x)
\]

\[
= \min_{x \in \theta(0,1)^n} \{ p(x) \mid \theta(x) := 1 - \sum_{i=1}^{n} x_i = 0 \}.
\]

The constraints \( x \in \{0,1\}^n \) are modelled with the polynomial constraints

\[
g_k(x) := x_k - x_k^2 = 0, \quad k = 1, 2, \ldots, n.
\]
As we already did before, we also assume with no loss of generality that \( p(0) = 0 \).

In view of Section 3.3 and Remark 3.3 (and as \( r = n, v := \max[1,r_0 - n] \)) the optimal value \( p^* \) is obtained at the \( Q_{n+v} \) SDP relaxation

\[
Q_{n+v} \begin{cases}
\min_{\alpha} \sum_{\alpha} \bar{p}_\alpha y_\alpha \\
\left.\begin{array}{c}
1 \\
y_{10} \\
y_{01} \\
\vdots \\
y_{0^v} \\
\end{array}\right| \left.\begin{array}{c}
y_{10} \\
y_{01} \\
y_{0^v} \\
1 \\
0 \\
\end{array}\right) \geq 0
\end{cases}
\]

The constraints \( M_{n+v-1}(g_{ki}) = 0 \) and \( M_{n+v-1}(\theta y) = 0 \) generate linear relationships between the variables \( y_\alpha \) and in view of those constraints, there are only \( n-1 \) independent variables that may be chosen to be

\[
y_{10\ldots0} y_{01\ldots0} \cdots y_{0^v01},
\]

corresponding to the monomials \( x_1, \ldots, x_{n-1} \). Indeed, for instance, the constraint

\[
M_{n+v-1}(\theta y)(1,1) = 0
\]

which yields \( n-1 \) independent variables among the variables \( y_\alpha \) corresponding to first-order moments, that is, with \( |\alpha| = 1 \) (where \( |\alpha| = \sum_i \alpha_i \)). The constraint

\[
M_{n+v-1}(\theta y)(1,2) = 0
\]

yields \( n-1 \) independent variables among the variables \( y_\alpha \) corresponding to monomials \( x_{i_1} x_{i_2} \), \( i_2 = 2, \ldots, n \), etc. so that finally, all the variables \( y_\alpha \), with \( |\alpha| = 2 \) vanish. From the other constraints induced by \( M_{n+v-1}(\theta y) = 0 \), it follows that all the variables \( y_\alpha \) with \( |\alpha| > 1 \) vanish. This corresponds to the fact that, on \( \Delta \cap \{0,1\}^n \), all the monomials \( x^\alpha \) with \( \alpha_i < 2 \) and \( |\alpha| > 1 \) vanish. Thus, after these substitutions, \( Q_{n+v} \) is the SDP relaxation,

\[
\hat{Q}_{n+v} \begin{cases}
\min \sum_{|\alpha| \leq 1} \tilde{p}_\alpha y_\alpha \\
\left.\begin{array}{c}
1 \\
y_{10} \\
y_{01} \\
\vdots \\
y_{0^v} \\
\end{array}\right| \left.\begin{array}{c}
y_{10} \\
y_{01} \\
y_{0^v} \\
1 \\
0 \\
\end{array}\right) \geq 0
\end{cases}
\]

with

\[
(4.9) \quad y_{10\ldots0} y_{01\ldots0} \cdots y_{0^v01} = 1.
\]

Therefore, using \( (4.9) \) to remove one variable, say \( y_{0^v01} \), the SDP \( \hat{Q}_{n+v} \) is strictly equivalent to the simplified SDP relaxation

\[
\hat{Q}_{n+v} \begin{cases}
\tilde{p}_{0^v01} + \min \sum_{|\alpha| \leq 1} (\tilde{p}_\alpha - \tilde{p}_{0^v01}) y_\alpha \\
\left.\begin{array}{c}
1 \\
y_{10} \\
y_{01} \\
\vdots \\
y_{0^v} \\
\end{array}\right| \left.\begin{array}{c}
y_{10} \\
y_{01} \\
y_{0^v} \\
1 \\
0 \\
\end{array}\right) \geq 0
\end{cases}
\]

\[
H(y) := \begin{cases}
\left.\begin{array}{c}
1 \\
y_{10} \\
y_{01} \\
\vdots \\
y_{0^v} \\
\end{array}\right| \left.\begin{array}{c}
y_{10} \\
y_{01} \\
y_{0^v} \\
1 \\
0 \\
\end{array}\right) \geq 0
\end{cases}
\]
with \( \{y_\alpha\} \in \mathbb{R}^{n-1} \). Next, let \( \{x(k)\} \), \( k = 1, \ldots, n \), be the \( n \) vertices of of \( \Delta \), that is, \( x(k) = e_k \in \mathbb{R}^n \) with \( e_k = \delta_{ki} \). To each \( x(k) \in \mathbb{R}^n \) corresponds the vector \((1, y(k)) \in \mathbb{R}^n \) with \( y(k)_j = \delta_{kj}, \ j = 1, \ldots, n-1 \), for all \( k = 1, \ldots, n-1 \), and \( y(n) = 0 \). Therefore, the \( n \)-vectors \( (1, y(k)) \) are linearly independent, and, in addition,

\[
H(y(k)) = \begin{bmatrix} 1 \\ y(k) \end{bmatrix} \times [1, y(k)] \geq 0, \quad k = 1, \ldots, n,
\]

so that every \( y(k) \) is admissible for \( \mathcal{Q}_{n+v} \). Moreover, let \( z := \sum_{k=1}^{n} z_k (1, y(k)) \) with \( z_k > 0 \) and \( \sum_{k=1}^{n} z_k = 1 \). From the linear independence of the vectors \((1, y(k))\), it follows that \( H(z) \succ 0 \), and thus, \( z \) is a strictly admissible solution of \( \mathcal{Q}_{n+v} \), that is, Slater’s interior point condition holds for \( \mathcal{Q}_{n+v} \). As \( \min \mathcal{Q}_{n+v} = \min \mathcal{Q}_{n+v}^* = p^* \), from a standard result in convex optimization, it follows that there is no duality gap between \( \mathcal{Q}_{n+v} \) and its dual \( \mathcal{Q}_{n+v}^* \), and \( \sup \mathcal{Q}_{n+v}^* = \max \mathcal{Q}_{n+v}^* = p^* \). If we remember how \( \mathcal{Q}_{n+v} \) was obtained from \( \mathcal{Q}_{n+v}^* \), this in turn implies that \( \max \mathcal{Q}_{n+v}^* = p^* \).

Now, from the solvability of the dual \( \mathcal{Q}_{n+v} \) with \( \max \mathcal{Q}_{n+v}^* = p^* \), and proceeding as in the proof of Theorem 4.2 we obtain that

\[
(4.10) \quad p(x) - p^* = \sum_{j=1}^{m} q_j(x)^2 + \sum_{k=1}^{n} (x_k - x_k^2) \left[ \sum_j v_{kj}(x)^2 - \sum_l w_{kl}(x)^2 \right] \\
+ \ (1 - \sum_{k=1}^{n} x_j) \left[ \sum_j y_{kj}(x)^2 - \sum_l z_{kl}(x)^2 \right]
\]

for some polynomials \( \{q_j(x)\} \) of degree at most \( n + v \), and some polynomials \( \{v_{kj}(x)\}, \{w_{kl}(x)\} \) and \( \{y_{kj}(x), z_{kl}(x)\} \) of degree at most \( n + v - 1 \).

Hence, every concave polynomial \( p(x) \), nonnegative on \( \Delta \), has the representation \( P.Y.I.11 \) (write it \( p(x) - p^* + \alpha \) for some scalar \( \alpha > 0 \)).

5. Conclusion

We have provided a representation of polynomials that are nonnegative on a grid \( \mathbb{K} \) of points of \( \mathbb{R}^n \). This representation is a sum of squares “linearly” weighted by the polynomials defining the grid \( \mathbb{K} \) as in Putinar’s representation. However, the degree of the polynomials in this representation is bounded and known in advance and depends on the size of the grid, not on the points of the grid. A related result is that every discrete optimization problem (on a grid) is also equivalent to a continuous convex optimization problem whose size depends on the size of the grid but not on the points of the grid.

6. Appendix

Let \( \{y(k)\}, k = 1, \ldots, s \), be the vectors in \( 4.14 \), that is, \( y(k) \) is the vector \( \{x(k)^\beta\} \) of all monomials \( x^\beta \) with \( \beta_k < 2r_k \), evaluated at the point \( x(k) \) of the grid \( \mathbb{K} \). We show by induction on the dimension of the space \( \mathbb{R}^n \), that the \( s \) points \( \{(1, y(k))\} \) in \( \mathbb{R}^s \) are linearly independent.
Denote by \( \mathbb{K} \) the grid in \( \mathbb{R}^n \) and let \( S_n \in \mathbb{R}^{s_n \times s_n} \) be the matrix formed by the \( s_n \) points \( \{(1, y(k))\} \) in \( \mathbb{R}^n \) where \( s_n = \prod_{k=1}^{2r_k} \), that is,

\[
S_n = \begin{bmatrix}
1 & 1 & \ldots & 1 \\
a_{11} & a_{11} & \ldots & a_{1r_1} \\
a_{21} & a_{21} & \ldots & a_{2r_2} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \ldots & a_{nr_n} \\
a_{11}^2 & a_{11}^2 & \ldots & a_{1r_1}^2 \\
a_{11}a_{21} & a_{11}a_{21} & \ldots & a_{1r_1}a_{2r_2} \\
\vdots & \vdots & \ddots & \vdots \\
\end{bmatrix}.
\]

Assume that the property is true (that is, \( S_n \) is nonsingular) whenever the dimension is \( 1, 2, \ldots, n \). Consider a grid \( \mathbb{K}_{n+1} \) in \( \mathbb{R}^{n+1} \) defined by \( g_k(x) = 0 \) for all \( k = 1, \ldots, n+1 \), and where

\[
g_k(x) := \prod_{j=1}^{2r_k} (x - a_{kj}), \quad k = 1, 2, \ldots, n+1.
\]

With \( s_{n+1} := \prod_{k=1}^{n+1} 2r_k \), we also define the \( s_{n+1} \) points \( \{(1, y(k))\} \) in \( \mathbb{R}^{s_{n+1}} \) and the associated matrix \( S_{n+1} \).

From its definition, and for an easy use of the induction argument, we can rearrange \( S_{n+1} \) to rewrite it as

\[
S_{n+1} = \begin{bmatrix}
S_n & S_n & \ldots & S_n \\
& & & \\
C_1 & C_2 & \ldots & C_{2r_{n+1}}
\end{bmatrix}.
\]

with

\[
C_1 = \begin{bmatrix}
a_{n1+1} & a_{n1+1} & \ldots & a_{n1+1} \\
a_{11}a_{n1+1} & a_{11}a_{n1+1} & \ldots & a_{1r_1}a_{n1+1} \\
a_{11}^2a_{n1+1} & a_{11}^2a_{n1+1} & \ldots & a_{1r_1}^2a_{n1+1} \\
\vdots & \vdots & \ddots & \vdots \\
\end{bmatrix}
\]

that is, the first column \( C_1(:, 1) \) of \( C_1 \) contains all the monomials \( x^\beta \) with \( 1 < \beta_{n+1} < 2r_{n+1} \) evaluated at the point \( \tilde{y}(k) = (a_{11}, a_{21}, \ldots, a_{n1}, a_{n+1}) \) of the grid \( \mathbb{K}_{n+1} \). Similarly, the second column \( C_1(:, 2) \) of \( C_1 \) contains all the monomials \( x^\beta \) with \( 1 < \beta_{n+1} < 2r_{n+1} \) evaluated at the point \( (a_{11}, a_{21}, \ldots, a_{n2}, a_{n+1}) \), etc.

\( C_i \) is a verbatim copy of \( C_1 \) with \( a_{n1+1} \) replaced with \( a_{n+i} \), for all \( i = 1, \ldots, 2r_{n+1} \). Next, by the induction hypothesis, \( S_n \) is full rank, and thus the only way for \( S_{n+1} \) to be singular, is that at least one column of \( \begin{bmatrix} S_n & C_i \end{bmatrix} \) matches exactly one column of \( \begin{bmatrix} S_n \\ C_j \end{bmatrix} \) for some pair \((i, j)\). But this is not possible. Indeed, assume that \( C_i(:, m) = C_j(:, p) \) for some pair \((m, p)\). Note that because of \( S_n \), we must have \( m = p \). An element \( C_i(k, p) \) is of the form \( x^{\beta_1} \cdots x_n^{\beta_n} a_{n+1}^{\beta_{n+1}} \), whereas the element \( C_j(k, p) \) is \( x^{\beta_1} \cdots x_n^{\beta_n} a_{n+1}^{\beta_{n+1}} \). Therefore, it suffices to consider a row \( k \) corresponding to a monomial \( x^\beta \) in the basis (241), with \( \beta_i = 0 \) for all \( i = 1, \ldots, n \), to yield the contradiction \( a_{n+1}^{\beta_{n+1}} a_{n+1}^{\beta_{n+1}} \) since the \( a_{n+1,k}'s \) are distinct.
It remains to prove that the induction hypothesis is true for \( n = 1 \). With \( n = 1 \) and a grid \( K_1 \) of \( 2r_1 \) points, the matrix \( S_1 \) reads

\[
\begin{pmatrix}
1 & 1 & \ldots & 1 \\
x_1 & x_2 & \ldots & x_{2r_1} \\
x_1^2 & x_2^2 & \ldots & x_{2r_1}^2 \\
\vdots & \vdots & \ddots & \vdots \\
x_{2r_1-1}^2 & x_{2r_1-1}^2 & \ldots & x_{2r_1-1}^2 \\
x_1 & x_2 & \ldots & x_{2r_1}
\end{pmatrix}
\]

If \( S_1 \) is singular, then there is a linear combination \( \{ \gamma_k \} \) of the rows such that

\[
\sum_{k=0}^{2r_1-1} \gamma_k x_k = 0, \quad \forall x \in K_1,
\]

that is, the univariate polynomial \( x \mapsto h(x) := \sum_{k=0}^{2r_1-1} \gamma_k x^k \) has \( 2r_1 \) distinct real zeros, which is impossible.

References