CONVERGENT SDP-RELAXATIONS IN POLYNOMIAL OPTIMIZATION WITH SPARSITY

JEAN B. LASSERRE

Abstract. We consider a polynomial programming problem $\mathbf{P}$ on a compact semi-algebraic set $\mathbf{K} \subset \mathbb{R}^n$, described by m polynomial inequalities $g_j(X) \geq 0$, and with criterion $f \in \mathbb{R}[X]$. We propose a hierarchy of semidefinite relaxations in the spirit those of Waki et al. [9]. In particular, the SDP-relaxation of order r has the following two features:

(a) The number of variables is $O(\kappa^{2r})$ where $\kappa = \max[\kappa_1, \kappa_2]$ with κ_1 (resp. κ_2) being the maximum number of variables appearing the monomials of f (resp. appearing in a single constraint $g_j(X) \geq 0$).

(b) The largest size of the LMI's (Linear Matrix Inequalities) is $O(\kappa^r)$.

This is to compare with the respective number of variables $O(n^{2r})$ and LMI size $O(n^r)$ in the original SDP-relaxations defined in [11]. Therefore, great computational savings are expected in case of sparsity in the data $\{g_j, f\}$, i.e. when κ is small, a frequent case in practical applications of interest. The novelty with respect to [9] is that we prove convergence to the global optimum of $\mathbf{P}$ when the sparsity pattern satisfies a condition often encountered in large size problems of practical applications, and known as the *running intersection property* in graph theory. In such cases, and as a by-product, we also obtain a new representation result for polynomials positive on a basic closed semi-algebraic set, a *sparse* version of Putinar's Positivstellensatz [16].

1. Introduction

In this paper we consider the polynomial programming problem

$$\mathbf{P}: \quad \inf_{x \in \mathbb{R}^n} \{ f(x) \mid \quad x \in \mathbf{K}\}, \tag{1.1}$$

where $f \in \mathbb{R}[X]$, and $\mathbf{K} \subset \mathbb{R}^n$ is the basic closed semi-algebraic set defined by

$$\mathbf{K} := \{x \in \mathbb{R}^n \mid \quad g_j(x) \geq 0, \quad j = 1, \ldots, m \}, \tag{1.2}$$

for some polynomials $\{g_j\}_{j=1}^m \subset \mathbb{R}[X]$.

The hierarchy of semidefinite programming (SDP) relaxations introduced in Lasserre [11] provides a sequence of SDPs of increasing size, whose associated sequence of optimal values converges to the global minimum of $\mathbf{P}$. Moreover, as proved in Schweighofer [17], convergence to a global minimizer of $\mathbf{P}$ (if unique) also holds. For more details, the reader is referred to [5, 11, 17] and the many references therein. In addition, practice reveals that convergence is usually fast, and often *finite* (up to machine precision); see e.g. Henrion and Lasserre [5].

However, despite these nice features, the size of the SDP-relaxations grows rapidly with the size of the original problem. Typically, the k^{th} SDP-relaxation

Date: April 12, 2006.

1991 *Mathematics Subject Classification.* 90C22 90C25.

Key words and phrases. Polynomial programming; semidefinite relaxations; measures; moments.

has to handle at least one LMI of size $\binom{n+k}{n}$ and $\binom{n+2k}{n}$ variables, which clearly limits the applicability of the methodology to problems with small to medium size only. Therefore, validation of the above methodology for larger size problems (and even more, for large scale problems) is a real challenge of practical importance.

One way to extend the applicability of the methodology to problems of larger size, is to take into account *sparsity* in the original data, frequently encountered in practical cases. Indeed, as typical in many applications of interest, f as well as the polynomials $\{g_j\}$ that describe $\mathbf{K}$, are sparse, i.e., each monomial of f and each polynomial g_j are only concerned with a small subset of variables. This is the approach taken in Waki et al. [9] (extending Kim et al. [7] and Kojima et al. [8]), where the authors have built up a hierarchy of SDP-relaxations in the spirit of those in [11], but where sparsity is taken into account. Sometimes, a sparsity pattern can be "read" from the data of $\mathbf{P}$ but not always, and in [9], the authors have proposed a systematic procedure to detect and structure sparsity in $\mathbf{P}$, via the so-called *chordal extension* of the *correlation sparsity pattern graph* (csp graph); the csp graph has as many nodes as variables, and a link beween two nodes (i.e., variables) means that these two variables both appear in a monomial of the objective function or in some inequality constraint $g_j \geq 0$ of $\mathbf{P}$. Once a sparsity pattern has been detected, they define a simplified "sparse" version of the SDP-relaxations of [11]; briefly, in the dual, the sum of squares (s.o.s.) multiplier associated with a constraint is now a polynomial in only those variables appearing in that constraint. In doing so, they have obtained impressive gains in the size of the resulting SDP-relaxations, as well as in the computational time needed for obtaining an optimal solution. As a matter of fact, they were even able to solve problems that could not be handled with the original SDP-relaxations. However, and despite good approximations are obtained in most problems in their sample of experiments, convergence to the global minimum is *not* guaranteed.

Contribution. Our contribution is twofold: We first propose a hierarchy of SDP-relaxations $\{\mathbf{Q}_r\}$ in the spirit of the original SDP-relaxations [11] and close to those defined in [9]. They are valid for arbitrary polynomial programming problems, and have the following three appealing features:

(a) In the SDP-relaxation $\mathbf{Q}_r$ of order r, the number of variables is $O(\kappa^{2r})$ where $\kappa = \max[\kappa_1, \kappa_2]$ witth κ_1 (resp. κ_2) being the maximum number of variables appearing in f (resp. in a single constraint $g_j(X) \geq 0$).

(b) The largest size of the LMI's (Linear Matrix Inequalities) is $O(\kappa^r)$.

This is to compare with the respective number of variables $O(n^{2r})$ and LMI size $O(n^r)$ in the original SDP-relaxations defined in [11].

(c) Under a certain condition on the sparsity pattern, the resulting sequence of their optimal value *converges* to the global minimum of $\mathbf{P}$.

So in view of (a) and (b), and when κ is small ($\kappa \ll n$), i.e., when sparsity is present, dramatic computational savings can be expected. In other words, these new SDP-relaxations are inherently exploiting sparsity in the data $\{f, g_j\}$ when present. Moreover, the size of the SDP-relaxation $\mathbf{Q}_r$ is in a sense *minimal*, at least when considering such types of SDP-relaxations, because one should at least handle moments involving κ variables, whenever some monomial of κ variables appears in the data $\{f, g_j\}$.

The condition under which such SDP-relaxations converge to the global minimum of $\mathbf{P}$ is easy to describe, and reflects a sparsity pattern frequently encountered in large scale problems. Namely, let $\{1,\ldots,n\}$ be the union $\bigcup_{k=1}^{p} I_k$ of subsets $I_k \subset \{1,\ldots,n\}$. Every polynomial g_j in the definition (1.2) of $\mathbf{K}$, is only concerned with variables $\{X_i \,|\, i \in I_k\}$ for some k. Next, $f \in \mathbb{R}[X]$ can be written $f = f_1 + \cdots + f_p$ where each f_k uses only variables $\{X_i \,|\, i \in I_k\}$. In cases where the subsets $\{I_k\}$ are not so easy to detect, one may use the procedure of Waki et al. [9] via the chordal extension of the csp graph.

Finally, the collection $\{I_1,\ldots,I_p\}$ should obey the following condition: For every $k = 1,\ldots,p-1$,

$$I_{k+1} \cap \bigcup_{j=1}^{k} I_j \subseteq I_s \quad \text{for some } s \leq k. \tag{1.3}$$

Notice that (1.3) is always satisfied when $p = 2$. Property (1.3) depends on the ordering and so, can be satisfied possibly after some relabelling of the I_k's. Moreover, if not satisfied, one may enforce (1.3) but at the price of enlarging some of the sets I_k. If $I_1,\ldots,I_p$ are the maximal cliques of a chordal graph then (1.3) is satisfied possibly after some reordering of the cliques, and is known as the *running intersection property*; for more details on chordal graphs, the reader is referred to Fukuda et al. [4] and Nakata et al. [15].

In particular, (1.3) is naturally satisfied in a number of applications, in particular, in what we call *strong* and *weak* coupling. In the former, we have $I_k \cap I_{k+j} = \emptyset$ whenever $j > 1$, so that (1.3) holds. In the latter, there is a set of *coupling variables* with index set $I_0' \subset \{1,\ldots,n\}$, and a partition of $\{1,\ldots,n\} \setminus I_0'$ into p disjoint subsets of *independent variables* I_k', $k = 1,\ldots,p$. In this case one has $I_k := I_0' \cup I_k'$, $k = 1,\ldots,p$, and so $I_k \cap I_j = I_0'$ for all $j \neq k$, which in turn implies that (1.3) holds.

At last, and as a by-product of the property (1.3) of the sparsity pattern, we also obtain a new *sparse representation* result for polynomials, nonnegative on a basic closed semi-algebraic set, a *sparse* version of Putinar's Positivstellensatz [16].

Link with related literature. As already mentioned, our work is closely related to the recent work of Kojima et al. [8] and Waki et al. [9], in which they were the first to exploit sparsity of data and modify (or simplify) in an appropriate way the original SDP-relaxations defined in [11]. Our SDP-relaxations are very close to those defined in [9], but handle p additional quadratic constraints. These p additional constraints together with condition (1.3), are crucial to prove our convergence result. To summarize, our result implies that by a slight modification of the SDP-relaxations defined in [9], convergence is now guaranteed when the sparsity pattern satisfies (1.3)

The paper is organized as follows. After introducing notation and definitions, our main result is presented in section 3, and for clarity of exposition, some proofs are postponed to section 4, whereas auxiliary results needed in some proofs are postponed to an appendix section.

2. Notation and definitions

As common in algebra, variables of polynomials are denoted with capitals (e.g. X) whereas points in $\mathbb{R}^n$ are denoted with small letters (e.g. x). For a real symmetric matrix $A \in \mathbb{R}^{n\times n}$, the notation $A \succeq 0$ (resp. $A \succ 0$) stands for A is positive definite (resp. semidefinite), and for a vector x, let x' denote its transpose.

Let $\mathbb{R}[X]$ denote the ring of real polynomials in the variables $X_1,\dots,X_n$. In the usual canonical basis $v_\infty(X) = \{X^\alpha \,|\, \alpha \in \mathbb{N}^n\}$ of monomials, a polynomial $g \in \mathbb{R}[X]$ is written

$$g(X) \,=\, \sum_{\alpha\in\mathbb{N}^n} g_\alpha X^\alpha, \tag{2.1}$$

for some real vector $\mathbf{g} = \{g_\alpha\}$ with finitely many non zero coefficients.

With $\alpha \in \mathbb{N}^n$, let $|\alpha| := \sum_i \alpha_i$, and let $\mathbb{R}_r[X] \subset \mathbb{R}[X]$ be the $\mathbb{R}$-vector space of polynomials of degree at most r, with usual canonical basis of monomials $v_r(X) = \{X^\alpha \,|\, \alpha \in \mathbb{N}^n;\, |\alpha| \leq r\}$.

Let $I_0 := \{1,\dots,n\}$ be the union $\cup_{k=1}^p I_k$ of p subsets I_k, $k = 1,\dots,p$, with cardinal denoted n_k. Let $\mathbb{R}[X(I_k)]$ denote the ring of polynomials in the n_k variables $X(I_k) = \{X_i \,|\, i \in I_k\}$, and so $\mathbb{R}[X(I_0)] = \mathbb{R}[X]$.

For each $k = 0,1,\dots,p$, let $\mathcal{I}_k$ be the set of all subsets of I_k. Next, for every $\alpha \in \mathbb{N}^n$, let $\mathrm{supp}\,(\alpha) \in \mathcal{I}_0$ be the support of α, i.e.,

$$\mathrm{supp}\,(\alpha) \,:=\, \{\, i \in \{1,\dots,n\} \;:\quad \alpha_i \neq 0 \,\}, \qquad \alpha \in \mathbb{N}^n.$$

For instance, with $n = 6$ and $\alpha := (004020)$, $\mathrm{supp}\,(\alpha) = \{3,5\}$. Next, define

$$S_k := \{\, \alpha \in \mathbb{N}^n \;:\quad \mathrm{supp}\,(\alpha) \,\in\, \mathcal{I}_k \,\}, \qquad k = 1,\dots,p. \tag{2.2}$$

A polynomial $h \in \mathbb{R}[X(I_k)]$ can be viewed as a member of $\mathbb{R}[X]$, and is written

$$h(X) \,=\, h(X(I_k)) \,=\, \sum_{\alpha\in S_k} h_\alpha\, X^\alpha \tag{2.3}$$

for some real vector $\mathbf{h} = \{h_\alpha\}$ with finitely many non zero coefficients.

2.1. **Moment matrix.** Let $y = (y_\alpha)_{\alpha\in\mathbb{N}^n}$ (i.e. a sequence indexed in the canonical basis $v_\infty(X)$), and define the linear functional $L_y : \mathbb{R}[X] \to \mathbb{R}$ to be:

$$g \mapsto L_y(g) \,:=\, \sum_{\alpha\in\mathbb{N}^n} g_\alpha\, y_\alpha, \tag{2.4}$$

whenever g is as in (2.1).

As already presented in [11], given a sequence $y = (y_\alpha)_{\alpha\in\mathbb{N}^n}$, the *moment* matrix $M_r(y)$ associated with y, is the matrix with rows and columns indexed in $v_r(X)$, and such that

$$M_r(y)(\alpha,\beta) \,:=\, L_y(X^\alpha X^\beta) \,=\, y_{\alpha+\beta}, \quad \forall\, \alpha,\beta \in \mathbb{N}^n \text{ with } |\alpha|,|\beta| \leq r.$$

A sequence y is said to have a representing measure μ on $\mathbb{R}^n$ if

$$y_\alpha \,=\, \int_{\mathbb{R}^n} X^\alpha\, \mu(dX), \qquad \forall\, \alpha \in \mathbb{N}^n.$$

Let $s(r) := \binom{n+r}{r}$ be the dimension of vector space $\mathbb{R}_r[X]$. For a vector $\mathbf{u} \in \mathbb{R}^{s(r)}$, let $u \in \mathbb{R}[X]$ be the polynomial $u(X) = \langle \mathbf{u}, v_r(X)\rangle$. Then, one has

$$\langle \mathbf{u}, M_r(y)\mathbf{u}\rangle \,=\, L_y(u^2), \qquad \forall\, \mathbf{u} \in \mathbb{R}^{s(r)}.$$

Therefore, if y has a representing measure μ, then

$$\langle \mathbf{u}, M_r(y)\mathbf{u}\rangle \,=\, L_y(u^2) \,=\, \int_{\mathbb{R}^n} u(X)^2\, \mu(dX) \,\geq 0,$$

which implies $M_r(y) \succeq 0$ (as $\mathbf{u} \in \mathbb{R}^{s(r)}$ was arbitrary).

Of course, in general, not every sequence y such that $M_r(y) \succeq 0$ for all $r \in \mathbb{N}$, has a representing measure. The $\mathbf{K}$-moment problem is precisely concerned with finding conditions on the sequence y, to ensure it is the moment sequence of some measure μ, with support contained in $\mathbf{K} \subset \mathbb{R}^n$.

2.2. **Localizing matrix.** Let $h \in \mathbb{R}[X]$ be a given polynomial

$$h(X) \,=\, \sum_{\gamma\in\mathbb{N}^n} h_\gamma\, X^\gamma,$$

and let $y = (y_\alpha)_{\alpha\in\mathbb{N}^n}$ be given. The *localizing* matrix $M_r(h\,y)$ associated with h and y, is the matrix with rows and columns indexed in $v_r(X)$, obtained from the moment matrix $M_r(y)$ by:

$$M_r(h\,y)(\alpha,\beta) \,:=\, L_y(h(X)\,X^\alpha X^\beta) \,=\, \sum_{\gamma\in\mathbb{N}^n} h_\gamma\, y_{\gamma+\alpha+\beta},$$

for all $\alpha,\beta \in \mathbb{N}^n$, with $|\alpha|, |\beta| \leq r$.

As before, let $\mathbf{u} \in \mathbb{R}^{s(r)}$, and let $u := \langle \mathbf{u}, v_r(X)\rangle \in \mathbb{R}_r[X]$. Then

$$\langle \mathbf{u}, M_r(h\,y)\mathbf{u}\rangle \,=\, L_y(h\,u^2), \qquad \forall\, \mathbf{u} \in \mathbb{R}^{s(r)},$$

and if y has a representing measure μ with support contained in the set $\{x \in \mathbb{R}^n : h(x) \geq 0\}$, then

$$\langle \mathbf{u}, M_r(h\,y)\mathbf{u}\rangle \,=\, L_y(h\,u^2) \,=\, \int_{\mathbb{R}^n} h(X)\,u(X)^2\,\mu(dX) \,\geq 0,$$

which implies $M_r(h\,y) \succeq 0$ (as $\mathbf{u} \in \mathbb{R}^{s(r)}$ was arbitrary).

Next, with $k \in \{1,\ldots,p\}$ fixed, and $h \in \mathbb{R}[X(I_k)]$, let $M_r(y, I_k)$ (resp. $M_r(h\,y, I_k)$) be the moment (resp. localizing) submatrix obtained from $M_r(y)$ (resp. $M_r(h\,y)$) by retaining only those rows (and columns) $\alpha \in \mathbb{N}^n$ of $M_r(y)$ (resp. $M_r(h\,y)$) with $\mathrm{supp}\,(\alpha) \in \mathcal{I}_k$.

In doing so, $M_r(y, I_k)$ and $M_r(h\,y, I_k)$ can be viewed as moment and localizing matrices with rows and columns indexed in the canonical basis $v_r(X(I_k))$ of $\mathbb{R}_r[X(I_k)]$. Indeed, $M_r(y, I_k)$ contain only variables y_α with $\mathrm{supp}\,(\alpha) \in \mathcal{I}_k$, and so does $M_r(h\,y, I_k)$ because $h \in \mathbb{R}[X(I_k)]$. And for every polynomial $u \in \mathbb{R}_r[X(I_k)]$, with coefficient vector $\mathbf{u}$ in the basis $v_r(X(I_k))$, we also have

$$\begin{aligned}\langle \mathbf{u}, M_r(y, I_k)\mathbf{u}\rangle &= L_y(u^2), \qquad \forall\, u \in \mathbb{R}_r[X(I_k)]\\ \langle \mathbf{u}, M_r(h\,y, I_k)\mathbf{u}\rangle &= L_y(h\,u^2), \qquad \forall\, u \in \mathbb{R}_r[X(I_k)],\end{aligned}$$

and therefore,

$$M_r(y, I_k) \succeq 0 \;\Leftrightarrow\; L_y(u^2) \geq 0, \qquad \forall\, u \in \mathbb{R}_r[X(I_k)] \tag{2.5}$$

$$M_r(h\,y, I_k) \succeq 0 \;\Leftrightarrow\; L_y(h\,u^2) \geq 0, \qquad \forall\, u \in \mathbb{R}_r[X(I_k)]. \tag{2.6}$$

3. Main result

Consider problem $\mathbf{P}$ as defined in (1.1), and recall that $I_0 = \{1,\ldots,n\} = \bigcup_{k=1}^p I_k$ for some subsets $I_k \subset \{1,\ldots,n\}$, $k = 1,\ldots,p$. The subsets $\{I_k\}$ may be read directly from the data or may have been obtained by some procedure, e.g. the one described in Waki et al. [9].

With $\|x\|_\infty$ (resp. $\|x\|$) denoting the usual sup-norm (resp. euclidean norm) of a vector $x \in \mathbb{R}^n$, we make the following assumption.

Assumption 3.1. *Let $\mathbf{K} \subset \mathbb{R}^n$ be as in (1.2). Then, there is $M > 0$ such that $\|x\|_\infty < M$ for all $x \in \mathbf{K}$.*

In view of Assumption 3.1, one has $\|X(I_k)\|^2 \leq n_k M^2$, $k = 1, \ldots, p$, and therefore, in the definition (1.2) of $\mathbf{K}$, we add the p redundant quadratic constraints

$$g_{m+k}(X) := n_k M^2 - \|X(I_k)\|^2 \geq 0, \quad k = 1, \ldots, p, \tag{3.1}$$

and set $m' = m + p$, so that $\mathbf{K}$ is now defined by:

$$\mathbf{K} := \{x \in \mathbb{R}^n \mid \quad g_j(x) \geq 0, \quad j = 1, \ldots, m' \}. \tag{3.2}$$

Notice that $g_{m+k} \in \mathbb{R}[X(I_k)]$, for every all $k = 1, \ldots, p$.

Assumption 3.2. *Let $\mathbf{K} \subset \mathbb{R}^n$ be as in (3.2). The index set $J = \{1, \ldots, m'\}$ is partitioned into p disjoint sets J_k, $k = 1, \ldots, p$, and the collections $\{I_k\}$ and $\{J_k\}$ satisfy:*

(i) *For every $j \in J_k$, $g_j \in \mathbb{R}[X(I_k)]$, that is, for every $j \in J_k$, the constraint $g_j(X) \geq 0$ is only concerned with the variables $X(I_k) = \{X_i \mid i \in I_k\}$. Equivalently, viewing g_j as a polynomial in $\mathbb{R}[X]$, $g_{j\alpha} \neq 0 \Rightarrow \mathrm{supp}\,(\alpha) \in \mathcal{I}_k$.*

(ii) *The objective function $f \in \mathbb{R}[X]$ can be written*

$$f = \sum_{k=1}^{p} f_k, \quad \text{with } f_k \in \mathbb{R}[X(I_k)], \quad k = 1, \ldots, p. \tag{3.3}$$

Equivalently, $f_\alpha \neq 0 \Rightarrow \mathrm{supp}\,(\alpha) \in \cup_{k=1}^{p} \mathcal{I}_k$.

(iii) *(1.3) holds.*

As already mentioned, (1.3) always holds when $p \leq 2$.

Example 3.3. With $n = 6$, and $m = 6$, let

$$g_1(X) = X_1X_2 - 1; \quad g_2(X) = X_1^2 + X_2X_3 - 1; \quad g_3(X) = X_2 + X_3^2X_4,$$

and

$$g_4(X) = X_3 + X_5; \quad g_5(X) = X_3X_6; \quad g_6(X) = X_2X_3,$$

Then one may choose $p = 4$ with

$$I_1 = \{1, 2, 3\};\ I_2 = \{2, 3, 4\};\ I_3 = \{3, 5\};\ I_4 = \{3, 6\},$$

and $J_1 = \{1, 2, 6\}, J_2 = \{3\}$, $J_3 = \{4\}$, $J_4 = \{4\}$.

So in Example 3.3, the objective function $f \in \mathbb{R}[X]$ should be a sum of polynomials in $\mathbb{R}[X_1, X_2, X_3]$, $\mathbb{R}[X_2, X_3, X_4]$, $\mathbb{R}[X_3, X_5]$ and $\mathbb{R}[X_3, X_6]$ (also considered as polynomials in $\mathbb{R}[X_1, \ldots, X_6]$).

Remark 3.4. For every $k = 1, \ldots, p$, let

$$\mathbf{K}_k := \{x \in \mathbb{R}^{n_k} \ : \ g_j(x) \geq 0, \quad \forall\, j \in J_k\}. \tag{3.4}$$

For every $k = 1, \ldots, p$, the set $\mathbf{K}_k \subset \mathbb{R}^{n_k}$ satisfies Putinar's condition, that is, there exists $u \in \mathbb{R}[X(I_k)]$ which can be written $u = u_0 + \sum_{l \in J_k} u_l\, g_l$ for some s.o.s. polynomials $\{u_0, u_l\} \subset \mathbb{R}[X(I_k)]$, and such that the level set $\{x \in \mathbb{R}^{n_k} \ : \ u \geq 0\}$ is compact. (Take $u = g_{m+k}$.) When satisfied, Putinar's condition has the important consequences stated in Theorem 4.1.

3.1. **Convergent SDP-relaxations.** For each $j = 1, \ldots, m'$, and depending on its parity, write $\deg g_j = 2r_j - 1$ or $2r_j$. Next, with $2r \geq 2r_0 := \max[\deg f, \max_j 2r_j]$, consider the following semidefinite program:

$$
\text{(3.5)} \qquad \mathbf{Q}_r : \quad \left\{ \begin{array}{lrl} \inf\limits_y & L_y(f) & \\ \text{s.t.} & M_r(y, I_k) \succeq 0, & k = 1, \ldots, p \\ & M_{r-r_j}(g_j\, y, I_k) \succeq 0, & j \in J_k; \quad k = 1, \ldots, p \\ & y_0 = 1 & \end{array} \right. ,
$$

where the moment and localizing matrices $M_r(y, I_k)$, $M_r(g_j\, y, I_k)$ have been defined at the end of §2.2. Denote the optimal value of $\mathbf{Q}_r$ by $\inf \mathbf{Q}_r$, and $\min \mathbf{Q}_r$ if the infimum is attained.

Notice that $\mathbf{Q}_r$ is well-defined under Assumption 3.2(i)-(ii). Assumption 3.2(iii) is only useful to show convergence in Theorem 3.6 below.

The semidefinite program $\mathbf{Q}_r$ is a relaxation of $\mathbf{P}$. Indeed, with $x \in \mathbb{R}^n$ being a feasible solution of $\mathbf{P}$, the moment vector $y = \{y_\alpha\}$ of the Dirac measure $\mu = \delta_x$ at x, is feasible for $\mathbf{Q}_r$, with value $L_y(f) = \int f d\mu = f(x)$.

Under Assumption 3.2, and from the definition of $M_r(y, k)$ and $M_r(g_j\, y, k)$ in §2.2, the SDP-relaxation $\mathbf{Q}_r$ contains only variables y_α with α in the set

$$
\text{(3.6)} \qquad \Gamma_r \,:=\, \{\, \alpha \in \mathbb{N}^n \;:\; \quad \text{supp}\,(\alpha) \in \bigcup_{k=1}^{p} \mathcal{I}_k\,; \quad |\alpha| \leq 2r\,\}.
$$

Remark 3.5. (i) Maximality of the I_k's is not required, i.e., one may have $I_j \subset I_k$ for some pair (j, k). In this case, the LMI constraint $M_r(y, I_j) \succeq 0$ is redundant. However, if non desirable in theory, in practice it may be more convenient to allow for non maximality.

(ii) Comparing with the SDP-relaxations of Waki et al. [9]. When the sets $\{I_k\}$ are just the *cliques* $\{C_k\}$ obtained from the chordal extension of the csp graph as defined in [9], then the SDP-relaxations (3.5) are basically the same as those defined in (32) in [9]. The only difference is in the definition of the feasible set $\mathbf{K}$ of $\mathbf{P}$, where we have now included the p redundant quadratic constraints (3.1). In this case, the SDP-relaxations (3.5) are thus stronger than (32) in [9], because they are more constrained.

In view of the definition of the moment matrix $M_r(y, I_k)$, write

$$
M_r(y, I_k) \,=\, \sum_{\alpha \in \mathbb{N}^n} y_\alpha B_\alpha^k, \qquad k = 1, \ldots, p,
$$

for appropriate symmetric matrices $\{B_\alpha^k\}$, and notice that for every $k = 1, \ldots, p$, one has $B_\alpha^k = 0$ whenever $\text{supp}\,(\alpha) \notin \mathcal{I}_k$. Similarly, for every $k = 1, \ldots, p$, and $j \in J_k$, write

$$
M_{r-r_j}(g_j\, y, I_k) \,=\, \sum_{\alpha \in \mathbb{N}^n} y_\alpha C_\alpha^{jk},
$$

for appropriate symmetric matrices $\{C_\alpha^{jk}\}$, and notice that $C_\alpha^{jk} = 0$ whenever $\text{supp}\,(\alpha) \notin \mathcal{I}_k$.

The dual SDP $\mathbf{Q}_r^*$ of $\mathbf{Q}_r$, reads

$$(3.7)\quad \left\{\begin{array}{ll} \sup\limits_{\Omega_k,Z_{jk},\lambda} & \lambda \\ \text{s.t.} & \displaystyle\sum_{k:\,\mathrm{supp}\,(\alpha)\in\mathcal{I}_k} [\,\langle \Omega_k, B_\alpha^k\rangle + \sum_{j\in J_k} \langle Z_{jk}, C_\alpha^{jk}\rangle\,] + \lambda\,\delta_{\alpha 0} \;=\; f_\alpha \\ & \text{for all } \alpha\in\Gamma_r \\ & \Omega_k, Z_{jk}\succeq 0, \qquad j\in J_k, \quad k=1,\ldots,p \end{array}\right. ,$$

where Γ_r is defined in (3.6) and $\delta_{\alpha 0}$ is the usual Kronecker symbol. From an arbitrary feasible solution $(\lambda, \Omega_k, Z_{jk})$ of $\mathbf{Q}_r^*$, multiplying each side of the constraint in (3.7) with X^α, for all $\alpha\in\Gamma_r$, and summing up, yields

$$\sum_{\alpha\in\Gamma_r}\left[\sum_{k:\,\mathrm{supp}\,(\alpha)\in\mathcal{I}_k}\left(\langle \Omega_k, B_\alpha^k X^\alpha\rangle + \sum_{j\in J_k}\langle Z_{jk}, C_\alpha^{jk}X^\alpha\rangle\right)\right] \;=\; f(X)-\lambda,$$

which, denoting $\Gamma_{kr} := \{\alpha\in\mathbb{N}^n \,:\, \mathrm{supp}\,(\alpha)\in\mathcal{I}_k;\ |\alpha|\leq 2r\}$, can be rewritten

$$(3.8)\qquad \sum_{k=1}^p\left[\langle \Omega_k, \sum_{\alpha\in\Gamma_{kr}} B_\alpha^k X^\alpha\rangle + \sum_{j\in J_k}\langle Z_{jk}, \sum_{\alpha\in\Gamma_{kr}} C_\alpha^{jk}X^\alpha\rangle\right] \;=\; f(X)-\lambda.$$

Proceding as in Lasserre [11], and using the spectral decomposition of matrices $\Omega_k, Z_{jk}\succeq 0$, write

$$\Omega_k = \sum_l \mathbf{q}_{kl}\mathbf{q}_{kl}', \quad Z_{jk} = \sum_t \mathbf{q}_{jkt}\mathbf{q}_{jkt}', \quad j\in J_k, \quad k=1,\ldots,p,$$

for some vectors $\{\mathbf{q}_{kl}, \mathbf{q}_{jkt}\}$. Next, notice that

$$(3.9)\qquad \sum_{\alpha\in\Gamma_{kr}} B_\alpha^k X^\alpha \;=\; v_r(X(I_k))\,v_r(X(I_k))', \qquad k=1,\ldots,p$$

(recall that $v_r(X(I_k))$ is the canonical basis of $\mathbb{R}_r[X(I_k)]$). Similarly, for every $k=1,\ldots,p$, and $j\in J_k$,

$$(3.10)\qquad \sum_{\alpha\in\Gamma_{kr}} C_\alpha^{jk} X^\alpha \;=\; g_j(X)\,v_{r-r_j}(X(I_k))\,v_{r-r_j}(X(I_k))'.$$

In view of the dimension of the matrix Ω_k (resp. Z_{jk}), one may identify $\mathbf{q}_{kl}$ (resp. $\mathbf{q}_{jkt}$) with the vector of coefficients of a polynomial $q_{kl}\in\mathbb{R}_r[X(I_k)]$ (resp. $q_{jkt}\in\mathbb{R}_{r-r_j}[X(I_k)]$), and so for every l, t

$$\langle v_r(X(I_k)), \mathbf{q}_{kl}\rangle \;=\; q_{kl}(X), \quad k=1,\ldots,p,$$
$$\langle v_{r-r_j}(X(I_k)), \mathbf{q}_{jkt}\rangle \;=\; q_{jkt}(X), \quad j\in J_k, \quad k=1,\ldots,p.$$

Combining the latter with (3.8)-(3.10), one may rewrite (3.8) as

$$\sum_{k=1}^p\left[\sum_l q_{kl}(X)^2 + + \sum_{j\in J_k} g_j(X)\sum_t q_{jkt}(X)^2\right] \;=\; f(X)-\lambda.$$

In other words,

$$(3.11)\qquad f-\lambda \;=\; \sum_{k=1}^p (\,q_k + \sum_{j\in J_k} q_{jk}\,g_j\,),$$

for some s.o.s. polynomials $q_k, q_{jk} \in \mathbb{R}[X(I_k)]$, $k = 1, \ldots, p$, a *sparse* version of Putinar's representation [16] for the polynomial $f - \lambda$, nonnegative on $\mathbf{K}$.

Finally, in view of what precedes, the dual $\mathbf{Q}_r^*$ also reads:

$$
(3.12) \quad \left\{ \begin{array}{lll}
\sup\limits_{q_k, q_{jk}, \lambda} & \lambda & \\
\text{s.t.} & f - \lambda = \displaystyle\sum_{k=1}^{p} (q_k + \sum_{j \in J_k} q_{jk}\, g_j\,) & \\
& q_k,\, q_{jk} \in \mathbb{R}[X(I_k)] \text{ and s.o.s.}, \qquad j \in J_k, \quad k = 1, \ldots, p & \\
& \deg q_k,\, \deg q_{jk} g_j \leq 2r, \qquad j \in J_k, \quad k = 1, \ldots, p, &
\end{array} \right.
$$

Theorem 3.6. *Let* $\mathbf{P}$ *be as defined in (1.1), with global minimum denoted* $\min \mathbf{P}$*, and let Assumption 3.1 and 3.2 hold. Let* $\{\mathbf{Q}_r\}$ *be the hierarchy of SDP-relaxations defined in (3.5). Then:*

(a) $\inf \mathbf{Q}_r \uparrow \min \mathbf{P}$ *as* $r \to \infty$.

(b) *If* $\mathbf{K}$ *has a nonempty interior, then there is no duality gap between* $\mathbf{Q}_r$ *and its dual* $\mathbf{Q}_r^*$*, and* $\mathbf{Q}_r^*$ *is solvable for sufficiently large* r*, i.e.,* $\inf \mathbf{Q}_r = \max \mathbf{Q}_r^*$.

(c) *Let* y^r *be a nearly optimal solution of* $\mathbf{Q}_r$*, with e.g.*

$$L_{y^r}(f) \leq \inf \mathbf{Q}_r + \frac{1}{r}, \qquad \forall\, r \geq r_0,$$

and let $\widehat{y}^r := \{y^r_\alpha : |\alpha| = 1\}$*. If* $\mathbf{P}$ *has a unique global minimizer* $x^* \in \mathbf{K}$*, then*

$$(3.13) \qquad \widehat{y}^r \to x^* \qquad \text{as } r \to \infty.$$

For a proof see §4.1. Theorem 3.6 establishes convergence of the hierarchy of SDP-relaxations to the global minimum $\min \mathbf{P}$, as well as convergence to a global minimizer $x^* \in \mathbf{K}$ (if unique).

3.2. **Computational complexity.** The number of variables for the SDP-relaxation $\mathbf{Q}_r$ defined in (3.5) is bounded by $\sum_{k=1}^{p} \binom{n_k+2r}{2r}$, and so, if all n_k's are *close* to each other, say $n_k \approx n/p$ for all k, then one has one has at most $O(p(\frac{n}{p})^{2r})$ variables, a big saving when compared with $O(n^{2r})$ in the original SDP-relaxations defined in [11] and implemented in [5].

In addition, one also has p LMI constraints of size $O((\frac{n}{p})^r)$ and $m + p$ LMI constraints of size $O((\frac{n}{p})^{r-r'})$ (where $2r'$ is the largest degree of the polynomials g_j's), to be compared with a single LMI constraint of size $O(n^r)$ and m LMI constraints of size $O(n^{r-r'})$ in [5, 11]. So for instance, when using an interior point method, it is definitely better to handle p LMIs, each of size $(n/p)^r$, rather than a single LMI of size n^r.

Example: For illustration purposes, consider the following elementary example.

Let $n=4$, and consider the optimization problem:

$$\mathbf{P}: \quad \left\{ \begin{array}{lrl} \inf\limits_x & x_1x_2+x_1x_3+x_1x_4 & \\ \text{s.t.} & x_1^2+x_2^2 & \leq a_{12} \\ & x_1^2+x_3^2 & \leq a_{13} \\ & x_1^2+x_4^2 & \leq a_{14} \end{array} \right.$$

Hence, $I_1=\{1,2\}$, $I_2=\{1,3\}$, $I_3=\{1,4\}$. The first SDP-relaxation $\mathbf{Q}_1$ in the hierarchy is obtained with $r=1$, and reads

$$\inf_y \; y_{1100}+y_{1010}+y_{1001}$$

$$\left[\begin{array}{ccc} 1 & y_{1000} & y_{0100} \\ y_{1000} & y_{2000} & y_{1100} \\ y_{0100} & y_{1100} & y_{0200} \end{array}\right], \left[\begin{array}{ccc} 1 & y_{1000} & y_{0010} \\ y_{1000} & y_{2000} & y_{1010} \\ y_{0010} & y_{1010} & y_{0020} \end{array}\right], \left[\begin{array}{ccc} 1 & y_{1000} & y_{0001} \\ y_{1000} & y_{2000} & y_{1001} \\ y_{0001} & y_{1001} & y_{0002} \end{array}\right] \succeq 0$$

$$a_{12}-y_{2000}-y_{0200} \geq 0; \; a_{13}-y_{2000}-y_{0020} \geq 0; \; a_{14}-y_{2000}-y_{0002} \geq 0.$$

3.3. **Extraction of solutions.** As for the standard SDP-relaxations of [11], one may also detect global optimality, i.e., when $\min \mathbf{Q}_{s_0} = \min \mathbf{P}$ for some s_0, in which case *finite* convergence occurs, and the SDP-relaxation $\mathbf{Q}_{s_0}$ is said to be *exact*. Recall that for the standard SDP-relaxations [11], one has defined a rank-test to detect finite convergence (see e.g. Lasserre [12]), as well as an *extraction procedure* (applied to the moment matrix of an exact SDP-relaxation) to obtain one or several global minimizers $x^*\in\mathbb{R}^n$ of $\mathbf{P}$; for more details, see Henrion and Lasserre [5, 6].

For all j,k with $I_{jk}:=I_j\cap I_k\neq\emptyset$, denote by $\mathcal{I}_{jk}$ the set of subsets of I_{jk}. Let $M_r(y,I_{jk})$ be the submatrix obtained from $M_r(y,I_j)$ or $M_r(y,I_k)$, by selecting only those rows and columns $\alpha\in\mathbb{N}^n$, with $\mathrm{supp}\,(\alpha)\in\mathcal{I}_{jk}$ and $|\alpha|\leq r$.

Theorem 3.7. *Let Assumption 3.2(i)-(ii) hold, and let $\{\mathbf{Q}_r\}$ be the hierarchy of SDP-relaxations defined in (3.5). Let $a_k:=\max_{j\in J_k}[r_j]$, for all $k=1,\ldots,p$, and assume that y is an optimal solution of $\mathbf{Q}_{s_0}$ for some s_0.*

The SDP-relaxation $\mathbf{Q}_{s_0}$ is exact, i.e., $\min \mathbf{Q}_{s_0}=\min\mathbf{P}$, if

$$\operatorname{rank} M_{s_0}(y,I_k) \;=\; \operatorname{rank} M_{s_0-a_k}(y,I_k), \qquad k=1,\ldots,p, \tag{3.14}$$

and if $\operatorname{rank} M_{s_0}(y,I_{jk})=1$, for all pairs (j,k) with $I_j\cap I_k\neq\emptyset$.

Moreover, let $\Delta_k:=\{x^*(k)\}\subset\mathbb{R}^{n_k}$ be a set of solutions obtained from the extraction procedure applied to each moment matrix $M_{s_0}(y,I_k)$, $k=1,\ldots,p$. Then every $x^*\in\mathbb{R}^n$ obtained by $(x_i^*)_{i\in I_k}=x^*(k)$ for some $x^*(k)\in\Delta_k$, is an optimal solution of $\mathbf{P}$.

For a proof see §4.2.

Remark 3.8. In Theorem 3.7 Assumption 3.2(iii) is not needed. In addition, it also holds even if the SDP-relaxations are defined with the original set $\mathbf{K}$ defined in (1.2) instead of (3.2), i.e., without the additional quadratic constraints (3.1). And so, Theorem 3.7 is also valid for non compact sets $\mathbf{K}$, provided Assumption 3.2(i)-(ii) hold true.

3.4. **A sparse representation result.** As a by-product of Theorem 3.6, we obtain the following representation result[1].

Corollary 3.9. *Let* $\mathbf{K}$ *be as in (3.2) with the additional quadratic constraints (3.1), and with nonempty interior. Let Assumption 3.2 hold. If* $f \in \mathbb{R}[X]$ *is strictly positive on* $\mathbf{K}$ *then*

$$f \;=\; \sum_{k=1}^{p} (\, q_k + \sum_{j\in J_k} q_{jk}\, g_j \,), \tag{3.15}$$

for some s.o.s. polynomials $q_k, q_{jk} \in \mathbb{R}[X(I_k)]$, $k = 1, \ldots, p$.

Proof. Let $f \in \mathbb{R}[X]$ be strictly positive on $\mathbf{K}$, and let $f^* > 0$ be its global minimum on $\mathbf{K}$. From Theorem 3.6(a)-(b), we have $\inf \mathbf{Q}_r = \max \mathbf{Q}_r^* \uparrow f^*$, as $r \to \infty$. Therefore, let $r \in \mathbb{N}$ be such that $\max \mathbf{Q}_r^* \geq f^*/2 > 0$, and as $\mathbf{Q}_r^*$ is solvable, let (q_k, q_{jk}, λ) be an arbitrary optimal solution, so that $\max \mathbf{Q}_r^* = \lambda > 0$. From that solution, one obtains (3.11), i.e.,

$$f - \lambda \;=\; \sum_{k=1}^{p} (\, q_k + \sum_{j\in J_k} q_{jk}\, g_j \,),$$

for some s.o.s. polynomials $q_k, q_{jk} \in \mathbb{R}[X(I_k)]$, $k = 1, \ldots, p$ (associated with the optimal solution (q_k, q_{jk}, λ) of $\mathbf{Q}_r^*$. But then,

$$f \;=\; \lambda + \sum_{k=1}^{p} (\, q_k + \sum_{j\in J_k} q_{jk}\, g_j \,),$$

the desired result (by adding $\lambda > 0$ to one of the s.o.s. polynomials q_k). □

Observe that (3.15) is a sparse version of Putinar's representation for polynomials strictly positive on $\mathbf{K}$; see Theorem 4.1. Indeed, (3.15) is a certificate of nonnegativity of f on $\mathbf{K}$. Finally, Corollary 3.9 also holds if $\mathbf{K}$ is such that for every $k = 1, \ldots, p$, $\mathbf{K}_k$ satisfies Putinar's condition (so that there is no need of the quadratic constraints (3.1)).

3.5. **Examples.** We here provide some examples considered in Waki et al. [9].

Example 3.10. The chained singular function. With n a multiple of 4,

$$I_k = \{k, k+1, k+2, k+3\}, \quad k = 1, \ldots, n-3,$$

and the sparsity pattern satisfies (1.3). One has $\kappa = 4$.

Example 3.11. The Broyden banded function. In this case,

$$I_k = \{k, k+1, \ldots, \min[k+6, n]\}, \quad k = 1, \ldots, n,$$

and the sparsity pattern also satisfies (1.3). One has $\kappa = 7$;

Example 3.12. The Broyden tridiagonal function. In this case

$$I_k = \{k, k+1, \min[n, k+2]\}, \quad k = 1, \ldots, n,$$

and the sparsity pattern also satisfies (1.3). One has $\kappa = 3$.

[1] In the recent note [10], Kojima and Maramatsu have improved Corollary 3.9 and show the same result without assuming that $\mathbf{K}$ has a nonempty interior.

Example 3.13. The chained Wood function. In this case, with n a multiple of 4,

$$I_k = \{k, k+1, k+2, k+3]\}, \quad k = 1, \ldots, n-3,$$

and the sparsity pattern also satisfies (1.3). One has $\kappa = 2$.

Example 3.14. The generalized Rosenbrock function. In this case,

$$I_k = \{k, k-1\}, \quad k = 2, \ldots, n,$$

and the sparsity pattern also satisfies (1.3).

Example 3.15. The optimal control problem (38) considered in [9]. In this case,

$$I_k = \{\{y_{k,j}\}_{j=1}^{n_y}, \{x_{k,l}\}_{l=1}^{n_x}\}, \quad k = 1, \ldots, M-1,$$

$I_M = \{\{y_{M,j}\}_{j=1}^{n_y}\}$, and the sparsity pattern also satisfies (1.3). One has $\kappa = n_x \times n_y$.

Example 3.15 is typical of what we call *strong coupling*, always the case in discrete-time optimal control problems. Indeed, the *control* variables at each period are *independent*, whereas the coupling of periods is done through the *state* equations (i.e. the dynamics) and via the *state* variables.

In view of Remark 3.5, the SDP-relaxations (3.5) are stronger than (32) in [9], when the sets $\{I_k\}$ are the same as the cliques $\{C_k\}$ in [9], which is the case in all the above examples, for which Waki et al. [9] report excellent numerical results; in particular, problems of large size that could not be handled via the standard SDP-relaxations of [11], have been solved relatively easily.

Indeed, for instance, in Examples 3.12, 3.13, and 3.14, they have solved problems with up to $n = 500$ variables, a remarkable result! For the interested reader, more details and numerical results can be found in [9].

4. Proofs

We first restate Putinar's theorem that is crucial in the proof of Theorem 3.6 below.

Theorem 4.1 (Putinar [16])**.** *Let $\mathbf{K} \subset \mathbb{R}^n$ be a compact basic semi-algebraic set as defined in (1.2), and let $y = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ be given. Let $M_r(y)$ and $M_r(g_j\, y)$ be the moment and localizing matrices defined in §2. Assume that there exists $u \in \mathbb{R}[X]$ such that $u = u_0 + \sum_{j=1}^m u_j\, g_j$ for some s.o.s. polynomials $\{u_j\}_{j=0}^m \subset \Sigma^2$, and such that the level set $\{x \,:\, u(x) \geq 0\}$ is compact.*

(a) *If $h \in \mathbb{R}[X]$ is strictly positive on $\mathbf{K}$ then $h = h_0 + \sum_{j=1}^m h_j\, g_j$ for some s.o.s. polynomials $\{h_j\}_{j=0}^m \subset \Sigma^2$.*

(b) *If $M_r(y) \succeq 0$ and $M_r(g_j\, y) \succeq 0$ for all $j = 1, \ldots, m$, and all $r = 0, 1, \ldots,$ then y has a representing measure μ with support contained in $\mathbf{K}$.*

4.1. **Proof of Theorem 3.6.** (a) We first prove that $\mathbf{Q}_r$ has a feasible solution. Recall the definitions

$$\begin{aligned}\Gamma_{kr} &:= \{\,\alpha\in\mathbb{N}^n\,:\,\mathrm{supp}\,(\alpha)\in\mathcal{I}_k;\quad |\alpha|\leq 2r\,\},\quad k=1,\ldots,p.\\ \Gamma_r &:= \bigcup_{k=1}^p \Gamma_{kr} \,=\, \{\,\alpha\in\mathbb{N}^n\,:\,\mathrm{supp}\,(\alpha)\in\bigcup_{k=1}^p\mathcal{I}_k;\quad |\alpha|\leq 2r\,\}.\\ \Gamma &:= \bigcup_{r\in\mathbb{N}}\Gamma_r \,=\, \{\,\alpha\in\mathbb{N}^n\,:\,\mathrm{supp}\,(\alpha)\in\bigcup_{k=1}^p\mathcal{I}_k\,\}.\end{aligned}$$

Let $\nu:=\delta_x$ be the Dirac measure at a feasible solution $x\in\mathbf{K}$ of $\mathbf{P}$, and let

$$y_\alpha=\int X^\alpha\,d\nu,\qquad \forall\,\alpha\in\Gamma_r.$$

Recalling the definition of $M_r(y,I_k)$ and $M_{r-r_j}(g_j\,y,I_k)$ in §2.2, one has $M_r(y,I_k)\succeq 0$ and $M_{r-r_j}(g_j\,y,I_k)\succeq 0$; therefore, y is an obvious feasible solution of $\mathbf{Q}_r$. Next we prove that $\inf\mathbf{Q}_r>-\infty$ for all sufficiently large r.

Recall that $2r_0\geq\max[\deg f,\max_j\deg r_j]$. In view of Assumption 3.1 and from the definition of the set $\mathbf{K}_k$ in (3.4), there exists N such that $N\pm X^\alpha>0$ on $\mathbf{K}_k$, for all $\alpha\in\Gamma_{kr_0}$, and all $k=1,\ldots,p$. Therefore, for every $k=1,\ldots,p$, and $\alpha\in\Gamma_{kr_0}$, the polynomial $N\pm X^\alpha$ belongs to the quadratic module $Q_k\subset\mathbb{R}[X(I_k)]$ generated by $\{g_j\}_{j\in J_k}\subset\mathbb{R}[X(I_k)]$, i.e.,

$$Q_k\,:=\,\{\sigma_0+\sum_{j\in J_k}\sigma_j\,g_j\;:\quad \sigma_j \text{ s.o.s. in } \in\mathbb{R}[X(I_k)]\quad\forall\,j\in\{0\}\cup J_k\}.$$

But there is even some $l(r_0)$ such that $N\pm X^\alpha\in Q_k(l(r_0))$ for all $\alpha\in\Gamma_{kr_0}$ and $k=1,\ldots,p$, where $Q_k(t)\subset Q_k$ is the set of elements of Q_k which have a representation $\sigma_0+\sum_{j\in J_k}\sigma_j\,g_j$ for some s.o.s. $\{\sigma_j\}\subset\mathbb{R}[X(I_k)]$ with $\deg\sigma_0\leq 2t$ and $\deg\sigma_j g_j\leq 2t$ for all $j\in J_k$. Of course we also have $N\pm X^\alpha\in Q_k(l)$ for all $\alpha\in\Gamma_{kr_0}$, whenever $l\geq l(r_0)$. Therefore, let us take $l(r_0)\geq r_0$.

For every feasible solution y of $\mathbf{Q}_{l(r_0)}$ one has

$$|\,L_y(X^\alpha)\,|\leq N,\qquad \alpha\in\Gamma_{kr_0};\quad k=1,\ldots,p.$$

This follows from $y_0=1$, $M_{l(r_0)}(y,I_k)\succeq 0$ and $M_{l(r_0)-r_j}(g_j\,y,I_k)\succeq 0$, which implies

$$L_y(N\pm X^\alpha)\,=\,L_y(\sigma_0)+\sum_{j\in J_k}L_y(\sigma_j\,g_j)\geq 0$$

because the σ_j's are s.o.s. (see (2.5) and (2.6)).

As $2r_0\geq\deg f$, it follows that $L_y(f)\geq -N\sum_\alpha|f_\alpha|$. This is because by Assumption 3.2(ii), $f_\alpha\neq 0\Rightarrow\alpha\in\Gamma_{r_0}$. Hence $\inf\mathbf{Q}_{l(r_0)}>-\infty$.

So from what precedes, and with $s\in\mathbb{N}$ arbitrary, let $l(s)\geq s$ be such that

$$N_s\pm X^\alpha\in Q_k(l(s)),\qquad\forall\,\alpha\in\Gamma_{ks};\quad k=1,\ldots,p, \tag{4.1}$$

for some N_s. Next, let $r\geq l(r_0)$ (so that $\inf\mathbf{Q}_r>-\infty$), and let y^r be a nearly optimal solution of $\mathbf{Q}_r$ with value

$$\inf\mathbf{Q}_r\,\leq\,L_{y^r}(f)\,\leq\,\inf\mathbf{Q}_r+\frac{1}{r}\quad(\leq\min\mathbf{P}+\frac{1}{r}). \tag{4.2}$$

Fix $s \in \mathbb{N}$. Notice that from (4.1), for all $r \geq l(s)$, one has

$$|\, L_{y^r}(X^\alpha)\,| \;\leq\; N_s, \qquad \forall\, \alpha \in \Gamma_s.$$

Therefore, for all $r \geq r_0$,

$$|y^r_\alpha| \;=\; |\, L_{y^r}(X^\alpha)\,| \;\leq\; N'_s, \quad \forall\, \alpha \in \Gamma_s, \tag{4.3}$$

where $N'_s = \max[N_s, V_s]$, with

$$V_s := \max\,\{\, |y^r_\alpha| \;:\quad \alpha \in \Gamma_s;\; r_0 \leq r \;<\; l(s)\,\}.$$

Complete each y^r with zeros to make it an infinite vector in l_∞, indexed in the canonical basis $v_\infty(X)$ of $\mathbb{R}[X]$. Notice that $y^r_\alpha \neq 0$ only if $\alpha \in \Gamma$.

In view of (4.3), one has

$$|y^r_\alpha| \;\leq N'_s, \qquad \forall\, \alpha \in \Gamma; \quad 2s-1 \leq |\alpha| \leq 2s, \tag{4.4}$$

for all $s = 1, 2, \ldots$.

Hence, define the new sequence $\widehat{y}^r \in l_\infty$ defined by $\widehat{y}_0 := 1$, and

$$\widehat{y}^r_\alpha \;:=\; \frac{y^r_\alpha}{N'_s}, \qquad \forall\, \alpha \in \Gamma, \quad 2s-1 \leq |\alpha| \leq 2s,$$

for all $s = 1, 2, \ldots$, and in l_∞, consider the sequence $\{\widehat{y}^r\}$ as $r \to \infty$.

Obviously, the sequence $\{\widehat{y}^r\}$ is in the unit ball B_1 of l_∞, and so, by Banach-Alaoglu theorem (see e.g. Ash [1, Theor. 3.5.16]), there exists $\widehat{y} \in B_1$, and a subsequence $\{r_i\}$, such that $\widehat{y}^{r_i} \to \widehat{y}$ as $i \to \infty$, for the weak $\star$ topology $\sigma(l_\infty, l_1)$ of l_∞. In particular, pointwise convergence holds, that is,

$$\lim_{i\to\infty} \widehat{y}^{r_i}_\alpha \;\to\; \widehat{y}_\alpha, \qquad \alpha \in \mathbb{N}^n.$$

Notice that $\widehat{y}_\alpha \neq 0$ only if $\alpha \in \Gamma$. Next, define $y_0 := 1$ and

$$y_\alpha \;:=\; \widehat{y}_\alpha \times N'_s, \quad 2s-1 \leq |\alpha| \leq 2s, \quad s = 1, 2, \ldots,$$

The pointwise convergence $\widehat{y}^{r_i} \to \widehat{y}$ implies the pointwise convergence $y^{r_i} \to y$, i.e.,

$$\lim_{i\to\infty} y^{r_i}_\alpha \;\to\; y_\alpha \qquad \forall\, \alpha \in \Gamma. \tag{4.5}$$

Let $s \in \mathbb{N}$ be fixed. From the pointwise convergence (4.5), we deduce that

$$\lim_{i\to\infty} M_s(y^{r_i}, I_k) \;=\; M_s(y, I_k) \succeq 0, \qquad k = 1, \ldots, p.$$

Similarly

$$\lim_{i\to\infty} M_s(g_j\, y^{r_i}, I_k) \;=\; M_s(g_j\, y, I_k) \succeq 0, \quad j \in J_k, \qquad k = 1, \ldots, p.$$

As s was arbitrary, we obtain that for all $k = 1, \ldots, p$,

$$M_r(y, I_k) \succeq 0; \quad M_r(g_j\, y, I_k) \succeq 0, \quad j \in J_k; \quad r = 0, 1, 2, \ldots \tag{4.6}$$

Introduce the subsequence y^k obtained from y by

$$y^k \;:=\; \{\, y_\alpha \;:\; \mathrm{supp}\,(\alpha) \;\in\; \mathcal{I}_k \,\}, \qquad \forall\, k = 1, \ldots, p. \tag{4.7}$$

Recall that $M_r(y, I_k)$ (resp. $M_r(g_j\, y, I_k)$) is also the moment matrix $M_r(y^k)$ (resp. the localizing matrix $M_r(g_j\, y^k)$) for the sequence y^k indexed in the canonical basis $v_\infty(X(I_k))$ of $\mathbb{R}[X(I_k)]$; see §2.2.

Therefore, by Remark 3.4, (4.6) implies that y^k has a representing measure ν_k with support contained in $\mathbf{K}_k$, $k = 1, \ldots, p$; see Theorem 4.1. As $y^k_0 = 1$, ν_k is a probability measure on $\mathbf{K}_k$ for all $k = 1, \ldots, p$.

Next, let j,k be such that $I_{jk} := I_j \cap I_k \neq \emptyset$, and recall that $\mathcal{I}_{jk}$ is the set of all subsets of I_{jk}. Let $m_{jk} := \mathrm{card}\,(I_j \cup I_k)$ and let $n_{jk} := \mathrm{card}\,(I_j \cap I_k)$. Define $\pi_j : \mathbb{R}^{m_{jk}} \to \mathbb{R}^{n_j}$, $\pi_k : \mathbb{R}^{m_{jk}} \to \mathbb{R}^{n_k}$, and $\pi_{jk} : \mathbb{R}^{m_{jk}} \to \mathbb{R}^{n_{jk}}$, the natural projections with respect to the variables $\{X_i \,|\, i \in I_j\}$, $\{X_i \,|\, i \in I_k\}$, and $\{X_i \,|\, i \in I_j \cap I_k\}$ respectively. Let $\mathbf{K}_{j\vee k} \subset \mathbb{R}^{m_{jk}}$ and $\mathbf{K}_{j\wedge k} \subset \mathbf{K}_{j\vee k}$ be the compact sets

$$\mathbf{K}_{j\vee k} := \{\, x \in \mathbb{R}^{m_{jk}} \;:\; \pi_j(x) \in \mathbf{K}_j; \quad \pi_k(x) \in \mathbf{K}_k \,\}; \quad \mathbf{K}_{j\wedge k} := \pi_{jk}(\mathbf{K}_{j\vee k}).$$

The probability measures ν_j and ν_k can be understood as probability measures on $\mathbf{K}_{j\vee k}$, supported on $\mathbf{K}_j = \pi_j(\mathbf{K}_{j\vee k})$ and $\mathbf{K}_k = \pi_k(\mathbf{K}_{j\vee k})$, respectively.

Observe that from the definition (4.7) of y^j and y^k, one has

$$y^j_\alpha \;=\; y^k_\alpha \qquad \forall\, \alpha \text{ with } \mathrm{supp}\,(\alpha) \in \mathcal{I}_{jk},$$

and as measures on compact sets are moment determinate, it follows that the marginal probability measures of ν_j and ν_k on $\mathbf{K}_{j\wedge k}$ (i.e. with respect to the variables $X = \{X_i \,|\, i \in I_{jk}\}$), are the *same* probability measure, denoted ν_{jk}. That is,

$$y^k_\alpha \;=\; y^j_\alpha \;=\; \int X^\alpha \, d\nu_{jk}, \qquad \forall\, \alpha \;\text{ with } \mathrm{supp}\,(\alpha) \in \mathcal{I}_{jk}.$$

From Lemma 6.4, there exists a probability measure μ on $\mathbf{K}$, constructed from the ν_k's, and with marginal ν_k on $\mathbf{K}_k$, for all $k = 1, \ldots, p$. In particular, this implies

$$y_\alpha \;=\; \int X^\alpha \, d\mu \qquad \forall\, \alpha \in \Gamma. \tag{4.8}$$

Recall that by Assumption 3.2, $f_\alpha \neq 0 \Rightarrow \alpha \in \Gamma$, and so $L_y(f) = \int f\, d\mu$. On the other hand, from (4.2) and the pointwise convergence (4.5),

$$\min \mathbf{P} \;\geq\; \lim_{i\to\infty} \inf \mathbf{Q}_{r_i} \;=\; \lim_{i\to\infty} L_{y^{r_i}}(f) \;=\; L_y(f) \;=\; \int f\, d\mu.$$

But as μ is supported on $\mathbf{K}$, we necessarily have $\int f\, d\mu \;\geq\; f^* \;=\; \min \mathbf{P}$, and so $\min \mathbf{P} = \int f\, d\mu$. Therefore, we have proved that $\lim_{i\to\infty} \inf \mathbf{Q}_{r_i} = \min \mathbf{P}$, and so $\inf \mathbf{Q}_r \uparrow \min \mathbf{P}$ follows because the sequence $\{\inf \mathbf{Q}_r\}$ is monotone nondecreasing. This completes the proof of (a).

(b) In the feasible solution ν that we have constructed at the beginning of the proof of (a), choose now ν to be *uniform* on $\mathbf{K}$, and let $y = \{y_\alpha\}_{\alpha\in\mathbb{N}^n}$ be the vector of all its moments, well defined because $\mathbf{K}$ is compact. As $\mathbf{K}$ has a nonempty interior, the probability measure ν satisfies $M_r(y) \succ 0$ and $M_r(g_j y) \succ 0$, for all all $j = 1, \ldots, m$, and all $r = 0, 1, \ldots$

Then, obviously, $M_r(y, I_k) \succ 0$ (resp. $M_r(g_j\, y, I_k) \succ 0$, $j \in J_k$) as a submatrix of $M_r(y) \succ 0$ (resp. $M_r(g_j\, y) \succ 0$), for all $k = 1, \ldots, p$.

Hence, the feasible solution y is now strictly feasible, i.e., Slater's condition holds for $\mathbf{Q}_r$. This implies the absence of a duality gap between $\mathbf{Q}_r$ and its dual $\mathbf{Q}^*_r$, and as $\inf \mathbf{Q}_r > -\infty$ for sufficiently large r, $\mathbf{Q}^*_r$ is solvable, i.e., $\inf \mathbf{Q}_r = \sup \mathbf{Q}^*_r = \max \mathbf{Q}^*_r$. This completes the proof of (b).

(c) Finally, let $x^* \in \mathbf{K}$ be the unique global minimizer of $\mathbf{P}$, and let y^r be as in Theorem 3.6(c). From (a) there exists a subsequence y^{r_i} for which we have the pointwise convergence $y^{r_i} \to y$ (see (4.5)), where y is the moment sequence of a probability measure μ on $\mathbf{K}$. In particular, (4.8) holds and $\min \mathbf{P} = \int f\, d\mu$. From uniqueness of the global minimizer $x^* \in \mathbf{K}$, it follows that $\mu = \delta_{x^*}$ (the Dirac measure at $x^*\in\mathbf{K}$). But then (4.8) yields

$$\lim_{i\to\infty} y_\alpha^{r_i} \,=\, y_\alpha \,=\, \int X^\alpha\, d\mu \,=\, (x^*)^\alpha, \qquad \forall \alpha\in\Gamma.$$

Taking $\alpha\in\Gamma$ with $|\alpha|=1$ yields $\widehat{y}^{r_i}\to x^*$, and as the converging subsequence was arbitrary, it follows that the whole sequence $\widehat{y}^r$ converges to $x^*\in\mathbf{K}$, the desired result. $\square$

4.2. **Proof of Theorem 3.7.** Let $\gamma_k := \mathrm{rank}\, M_{s_0}(y, I_k)$, $k=1,\ldots,p$. From (3.14) the vector $y^k=\{y^k_\alpha\}$ defined in (4.7) (with $|\alpha|\leq 2s_0$) is the vector of moments (up to order $2s_0$) of a γ_k-atomic probability measure ν_k supported on $\mathbf{K}_k\subset\mathbb{R}^{n_k}$, with $\mathbf{K}_k$ being defined in (3.4), $k=1,\ldots,p$. This follows from a result of Curto and Fialkow [3, Theor. 1.6] already used in Lasserre [12] to prove finite convergence of SDP-relaxations for 0-1 programs; see also Laurent [14] for a shorter proof of Theorem 1.6 in [3], and related comments.

Therefore, when applying the extraction procedure defined in [6] to the moment matrix $M_{s_0}(y^k)$ $(=M_{s_0}(y,I_k))$, $k=1,\ldots,p$, one obtains sets of vectors $\Delta_k := \{x^l(k)\}_{l=1}^{\gamma_k}\subset\mathbf{K}_k$, for all $k=1,\ldots,p$.

With $\delta_\bullet$ denoting the Dirac measure at $\bullet$, one may thus write

$$\nu_k \,=\, \sum_{l=1}^{\gamma_k} p_{kl}\,\delta_{x^l(k)}, \qquad \text{for some} \quad p_{kl}>0 \quad \forall\, l; \quad \sum_{l=1}^{\gamma_k} p_{kl}=1,$$

for all $k=1,\ldots,p$.

But then, pick *any* solution $x^{l_k}(k)\in\Delta_k$, for some l_k, $k=1,\ldots,p$, and define $x^*\in\mathbb{R}^n$ to be the vector such that

$$x^*(k) \,:=\, \{x_i^*\}_{i\in I_k} \,=\, x^{l_k}(k); \quad k=1,\ldots,p. \tag{4.9}$$

There is no ambiguity for x_i^* when $i\in I_j\cap I_k\neq\emptyset$ for some $j,k\in\{1,\ldots,p\}$, because in this case, from $\mathrm{rank}\, M_{s_0}(y,j,I_{jk})=1$, we deduce that $y^{jk}=\{y_\alpha\}$ with $\mathrm{supp}\,(\alpha)\in\mathcal{I}_j\cap\mathcal{I}_k$, is the vector of moments (up to order $2s_0$) of some Dirac measure ν_{jk}. As in the proof of (a), ν_{jk} is the marginal of ν_k and ν_j on $\mathbf{K}_{j\wedge k}$ (i.e. with respect to the variables $\{X_i\,:\, i\in I_j\cap I_k\}$), and so the Dirac measure at some point denoted $x(j\wedge k)\in\mathbf{K}_{j\wedge k}$.

Hence, for any two choices $x^{l_j}(j)\in\Delta_j$ and $x^{l_k}(k)\in\Delta_k$, the point $x^*\in\mathbb{R}^n$ defined in (4.9) is in $\mathbf{K}$. We can thus construct $s:=\prod_{k=1}^p\gamma_k$ solutions $\{x^\omega\}_{\omega=1}^s\subset\mathbf{K}$, each associated with the probability $p_\omega := \prod_{k=1}^p p_{kl_k}$ if $x^\omega(k)=x^{l_k}(k)\in\Delta_k$, for some $l_k\in\{1,\ldots,\gamma_k\}$, $k=1,\ldots,p$. But then, by construction, the probability measure μ on $\mathbb{R}^n$, defined by

$$\mu := \sum_{\omega=1}^s p_\omega\delta_{x^\omega},$$

is supported on $\mathbf{K}$, and its marginal probability measure on $\mathbf{K}_k$, is ν_k, for all $k=1,\ldots,p$. Therefore,

$$\min\mathbf{P} \,\geq\, \min\mathbf{Q}_{s_0} \,=\, L_y(f) \,=\, \int f\, d\mu \,=\, \sum_{\omega=1}^s p_\omega f(x^\omega),$$

which implies that $f(x^\omega)=\min\mathbf{P}$, for all $\omega=1,\ldots,s$, because $x^\omega\in\mathbf{K}$ for all $\omega=1,\ldots,s$. Therefore, we have proved that $\min\mathbf{P}=\min\mathbf{Q}_{s_0}$. In addition, each $x^\omega\in\mathbf{K}$ is an optimal solution of $\mathbf{P}$. $\square$

5. Conclusion

We have provided a hierarchy of SDP-relaxations when the polynomial optimization problem **P** has some structured sparsity (which can be detected as in Waki et al. [9]). This hierarchy is of the same flavor (in fact a minor modification) as that in Waki et al. [9], for which excellent numerical results have been reported. Our contribution was to prove convergence of the optimal values to the global minimum of **P** when the sparsity pattern satisfies the condition (1.3), called the *running intersection property* in graph theory, and frequently encountered in practice. Therefore, this result together with [9], opens the door for the applicability of the general approach of SDP-relaxations to medium (and even large) scale polynomial optimization problems, at least when a certain sparsity pattern is present.

Acknowledgements. The author is indebted to Prof. M. Kojima for very interesting and helpful discussions on the topic of sparse SDP-relaxations. He also wishes to thank T. Netzer and M. Schweighofer from Konstanz University (Germany), who indicated a way to simplify the original SDP-rexations of the author in an earlier version, so as to yield the SDP-relaxations of this paper. Finally, the author wishes to thank anonymous referees for helpful remarks and suggestions to improve the initial version of the paper.

6. Appendix

We state some auxiliary results needed in the proof of Theorem 3.6 in §4.1.

For a topological space Y let $\mathcal{B}(Y)$ denote the usual Borel σ-algebra associated with Y, and let $P(Y)$ denote the space of probability measures on Y. A Borel space is a Borel subset of a complete separable metric space. Let Y, Z be two Borel spaces. A stochastic kernel $q(dy\,|\,z)$ on Y given Z is defined by:
- $q(dy\,|\,z) \in P(Y)$ for all $z \in Z$.
- The function $z \mapsto q(B\,|\,z)$ is $\mathcal{B}(Z)$-measurable for all $B \in \mathcal{B}(Y)$.

6.1. **Disintegration of a Borel probability measure.** The following result states that one may decompose or *disintegrate* a probability measure on a product of Borel spaces into a marginal and a stochastic kernel (also called *conditional probability* when dealing with distributions of random variables).

Proposition 6.1. *Let Y, Z be two Borel spaces, and let μ be a probability measure on $Y \times Z$. Then there exists a probability measure $\nu \in P(Z)$ and a stochastic kernel $q(dy\,|\,z)$ on Y given Z, such that*

$$\mu(A \times B) = \int_B q(A\,|\,z)\,\nu(dz), \qquad \forall\, A \in \mathcal{B}(Y),\; B \in \mathcal{B}(Z). \tag{6.1}$$

(Proposition 6.1 can be extended to the cartesian product of an arbitrary number of Borel spaces.) The probability measure ν is called the *marginal* of μ on Z. One also has the converse.

Proposition 6.2. *Let Y, Z be two Borel spaces, and let ν be a probability measure on Z and $q(dy\,|\,z)$ a stochastic kernel on Y given Z. Then there exists a unique probability measure μ on $Y \times Z$ such that*

$$\mu(A \times B) = \int_B q(A\,|\,z)\,\nu(dz), \qquad \forall\, A \in \mathcal{B}(Y),\; B \in \mathcal{B}(Z). \tag{6.2}$$

(See e.g. Ash [1, §6] and Bertsekas and Schreve [2, p. 139-141].)

Let μ (resp. ν) be a finite Borel probability measure on $\mathbb{R}^n\times\mathbb{R}^m$ (resp. $\mathbb{R}^m\times\mathbb{R}^p$) with all moments $y=(y_{\alpha\beta})_{\alpha\in\mathbb{N}^n,\beta\in\mathbb{N}^m}$ (resp. $z=(z_{\beta\gamma})_{\beta\in\mathbb{N}^m,\gamma\in\mathbb{N}^p}$) finite. Let μ_1 and ν_1 be the respective marginals of μ and ν on $\mathbb{R}^m$, hence with moments

$$\int X^\beta\,d\mu_1(X)\,=\,\int Y^0X^\beta\,d\mu(Y,X)\,=\,y_{0\beta}\qquad\forall\beta\in\mathbb{N}^m,$$

$$\int X^\beta\,d\nu_1(X)\,=\,\int X^\beta Z^0\,d\mu(X,Z)\,=\,z_{\beta0}\qquad\forall\beta\in\mathbb{N}^m.$$

If both μ and ν have compact support and $y_{0\beta}=z_{\beta0}$ for all $\beta\in\mathbb{N}^m$, then $\mu_1=\nu_1$. This is because measures with compact support are *moment determinate*, i.e., if two measures on a compact subset of $\mathbb{R}^m$ have all same moments, they must coincide.

6.2. **Probability measures with given marginals. Case** $p=2$**.** Let $I_0:=\{1,\ldots,n\}$, and let $I_0=I_1\cup I_2$ with $I_1\cap I_2\neq\emptyset$. Let $n_k=\mathrm{card}\,I_k$, for $k=1,2$, and $n_{12}=\mathrm{card}\,I_1\cap I_2$. For $k=1,2$, let $\pi_k:\mathbb{R}^n\to\mathbb{R}^{n_k}$ be the natural projection with respect to I_k, that is,

$$x\,\mapsto\,\pi_k(x)\,=\,\{\,x_i\,:\,i\in I_k\,\},\qquad x\in\mathbb{R}^n,$$

and let $\pi_{12}:\mathbb{R}^{n_1}\to\mathbb{R}^{n_{12}}$, $\pi_{21}:\mathbb{R}^{n_2}\to\mathbb{R}^{n_{12}}$ be the projections with respect to $I_1\cap I_2$, that is,

$$\begin{aligned}x\,\mapsto\,\pi_{12}(x)\;&=\;\{\,x_i\,:\,i\in I_1\cap I_2\,\},\qquad x\in\mathbb{R}^{n_1}\\ x\,\mapsto\,\pi_{21}(x)\;&=\;\{\,x_i\,:\,i\in I_1\cap I_2\,\},\qquad x\in\mathbb{R}^{n_2}.\end{aligned}$$

and one also extends π_{12} and π_{21} to $\mathbb{R}^n$ in the obvious way.

Next, for $k=1,2$, let $\mathbf{K}_k\in\mathcal{B}(\mathbb{R}^{n_k})$ be given, and let $\nu_k\in P(\mathbf{K}_k)$. Denote by ν_{12} and ν_{21} the respective marginals of ν_1 and ν_2 on $\mathbb{R}^{n_{12}}$ (i.e., with respect to the variables $\{X_i,\,i\in I_1\cap I_2\}$). That is, letting $Z:=\mathbb{R}^{n_{12}}$,

$$\begin{aligned}\nu_{12}(B)\;&=\;\nu_1(\pi_{12}^{-1}(B)\cap\mathbf{K}_1),\qquad\forall\,B\in\mathcal{B}(Z)\\ \nu_{21}(B)\;&=\;\nu_2(\pi_{21}^{-1}(B)\cap\mathbf{K}_2),\qquad\forall\,B\in\mathcal{B}(Z),\end{aligned}$$

and we have

$$\nu_{12}(\pi_{12}(\mathbf{K}_1))\,=\,\nu_{21}(\pi_{21}(\mathbf{K}_2))\,=\,1.\tag{6.3}$$

Let $\mathbf{K}\subset\mathbb{R}^n$ be the set

$$\mathbf{K}\,:=\,\{\,x\in\mathbb{R}^n:\quad\pi_k(x)\in\mathbf{K}_k,\qquad k=1,2\,\},\tag{6.4}$$

and view the sets $\mathbf{K}_k$, $k=1,2$ as naturally embedded in $\mathbb{R}^n$, with $\mathbf{K}_k=\pi_k(\mathbf{K})$, for every $k=1,2$.

Lemma 6.3. *For $k=1,2$, let $\mathbf{K}_k\in\mathcal{B}(\mathbb{R}^{n_k})$ be given, and let $\nu_k\in P(\mathbf{K}_k)$ be such that $\nu_{12}=\nu_{21}=:\nu$. Then there exists a probability measure μ on $\mathbf{K}$ with marginals ν_k on $\mathbf{K}_k=\pi_k(\mathbf{K})$, $k=1,2$, and marginal ν on $\pi_{12}(\mathbf{K})$.*

Proof. For $k=1,2$, let π'_k be the natural projection with respect to $I_k\setminus I_1\cap I_2$, i.e.,

$$x\,\mapsto\,\pi'_k(x)\,=\,\{x_i\,:\,i\in I_k\setminus I_1\cap I_2\,\},\quad x\in\mathbb{R}^{n_k},\;k=1,2,$$

and define $Y_k\in\mathcal{B}(\mathbb{R}^{n_k-n_{12}})$ to be the Borel set $\{\pi'_k(x)\,:\,x\in\mathbf{K}_k\}$, $k=1,2$.

Then, for $k = 1, 2$, one may view ν_k as a probability measure on the cartesian product $Y_k \times Z$. By Proposition 6.1, and from $\nu_{12} = \nu_{21} =: \nu$, for $k = 1, 2$, one may disintegrate ν_k as

$$\nu_k(A \times B) \,=\, \int_B q_k(A \,|\, z)\, \nu(dz), \qquad \forall\, A \in \mathcal{B}(Y_k),\, B \in \mathcal{B}(Z),$$

for some stochastic kernels q_k, $k = 1, 2$. Next, let μ be the measure on $Y_1 \times Z \times Y_2$, defined by

$$\mu(A \times B \times C) \,=\, \int_B q_1(A \,|\, z)\, q_2(C \,|\, z)\, \nu(dz),$$

for every Borel rectangle

$$A \times B \times C \,\in\, \mathcal{B}(Y_1) \times \mathcal{B}(Z) \times \mathcal{B}(Y_2).$$

Taking $A = Y_1$ yields $q_1(A \,|\, z) = 1$, ν-a.e. and so

$$\mu(Y_1 \times B \times C) \,=\, \int_B q_2(C \,|\, z)\, \nu_{12}(dz) \,=\, \nu_2(B \times C).$$

Therefore, ν_2 is the marginal of μ on $Z \times Y_2$ (and so on $\mathbf{K}_2$). With similar argument, ν_1 is the marginal of μ on $Y_1 \times Z$ (and so on $\mathbf{K}_1$). Finally, taking $A = Y_1$, $C = Y_2$ and using $q_k(Y_k \,|\, z) = 1$, ν-a.e., yields

$$\mu(Y_1 \times B \times Y_2) \,=\, \int_B \nu(dz) \,=\, \nu(B),$$

which shows that ν is the marginal of μ on Z, i.e. with respect to the variables X_i, $i \in I_1 \cap I_2$. It remains to prove that $\mu(\mathbf{K}) = 1$. But notice that from the definitions of $\mathbf{K}_1, \mathbf{K}_2$ and ν,

$$q_1(\{\, y \,:\, (y, z) \in \mathbf{K}_1\} \,|\, z) \,=\, q_2(\{\, y' \,:\, (z, y') \in \mathbf{K}_2\} \,|\, z) = 1, \qquad \nu\text{-a.e.}$$

So, writing (6.4) as

$$\mathbf{K} \,=\, \{(y, z, y') \,\in\, \mathbb{R}^n \,:\, \quad (y, z) \,\in\, \mathbf{K}_1\,;\, (z, y') \,\in\, \mathbf{K}_2\,\},$$

yields

$$\mu(\mathbf{K}) \,=\, \int_Z q_1(\{y \,:\, (y, z) \in \mathbf{K}_1\} \,|\, z)\, q_2(\{y' \,:\, (z, y') \in \mathbf{K}_2\} \,|\, z)\, \nu(dz) \,=\, 1.$$

Therefore, ν_k is the marginal of μ on $\mathbf{K}_k = \pi_k(\mathbf{K})$ for $k = 1, 2$, and ν is the marginal of μ on $\pi_{12}(\mathbf{K})$. □

6.3. **Probability measures with given marginals. General case.** Let I_k, J_k, $k = 1, \ldots, p$, be as in §2, and let $\mathbf{K} \subset \mathbb{R}^n$ be as defined in (1.2), with $\mathbf{K}_k \subset \mathbb{R}^{n_k}$ as in (3.4), $k = 1, \ldots, p$. Let ν_k be a given probability measure on $\mathbf{K}_k$, $k = 1, \ldots, p$.

Given a set $I \subset I_k$ denote by $X(I)$ the vector of variables $\{X_i\}_{i \in I} \in \mathbb{R}^{|I|}$, and denote by ν_{kI} the marginal of ν_k on $\mathbb{R}^{|I|}$ (i.e., with respect to the variables X_i, $i \in I$), so that ν_k can be disintegrated into $q_k(. \,|\, z) d\nu_{kI}(dz)$ for a stochastic kernel q on $\mathbb{R}^{n_k - |I|}$ given $\mathbb{R}^{|I|}$ (see Proposition 6.1)

We say that the family of probability measures $\{\nu_k\}_{k=1}^p$ is *consistent* with respect to marginals, if whenever $l, k \in \{1, \ldots, p\}$ and $I_k \cap I_l \neq \emptyset$,

$$I \subseteq I_k \cap I_l \Rightarrow \nu_{kI} \,=\, \nu_{lI}.$$

Equivalently, when ν_k and ν_l have compact support,

$$\int X^\alpha\, d\nu_k \,=\, \int X^\alpha\, d\nu_l \qquad \forall\, \alpha \,:\, \sup\,(\alpha) \,\subseteq\, I_k \cap I_l.$$

For every $k=1,\ldots,p$, let $W_k:=\bigcup_{l=1}^k I_l$, $s_k:=\vert W_k\vert$, and

$$\Omega_k\,:=\,\{x\in\mathbb{R}^{s_k}\mid\quad g_j(x)\geq 0,\quad j\in\bigcup_{l=1}^k J_l\}. \tag{6.5}$$

Notice that $\Omega_n\equiv\mathbf{K}\subset\mathbb{R}^n$.

Lemma 6.4. *Let ν_k be a probability measure on $\mathbf{K}_k\subset\mathbb{R}^{n_k}$, $k=1,\ldots,p$, and assume that the family $\{\nu_k\}_{k=1}^p$ is consistent with respect to marginals. If (1.3) holds then :*

(a) *There exists a probability measure μ on $\mathbb{R}^n$ such that ν_k is the marginal of μ with respect to I_k, for all $k=1,\ldots,p$.*

(b) *μ is supported on $\mathbf{K}\subset\mathbb{R}^n$.*

Proof. The proof is by induction on p. With $p=1$ it is trivial. Let $p=2$. Observe that the condition (1.3) is automatically satisfied. If $I_1\cap I_2=\emptyset$ just let $\mu:=\nu_1\otimes\nu_2$, the product measure on $\mathbf{K}_1\times\mathbf{K}_2$, i.e.,

$$\mu(A\times B)\,=:\,\nu_1(A)\,\nu_2(B),\qquad\forall\,(A,B)\in\mathcal{B}(\mathbf{K}_1)\times\mathcal{B}(\mathbf{K}_1).$$

If $I_1\cap I_2\neq\emptyset$ then the result follows from Lemma 6.3.

Next, suppose that the results holds for $1\leq m<p$. That is, let Ω_m be as in (6.5), and let ν_k be given probability measures on $\mathbf{K}_k$, $k=1,\ldots,m$, consistent with marginals, i.e., whenever $l,k\in\{1,\ldots,m\}$, and $I_l\cap I_k\neq\emptyset$,

$$I\subseteq I_k\cap I_l\Rightarrow\nu_{lI}\,=\,\nu_{kI}.$$

Then there exists a probability measure μ_m on Ω_m, such that ν_k is the marginal of μ_m on $\mathbf{K}_k$ (i.e., with respect to the variables X_i, $i\in I_k$), for every $k=1,\ldots,m$. We next whow that it holds true for $m+1$.

Set $\Delta:=I_{m+1}\cap W_m$. If $\Delta=\emptyset$ then just take $\mu_{m+1}:=\mu_m\otimes\nu_{m+1}$, the product measure on $\Omega_m\times\mathbf{K}_{m+1}$, and the induction is trivially satisfied for $m+1$. (As $\Delta=\emptyset$, one has $\Omega_{m+1}=\Omega_m\times\mathbf{K}_{m+1}$.)

Consider the case $\Delta\neq\emptyset$, and let $\delta:=\vert\Delta\vert$, $s_{m+1}:=\vert W_{m+1}\vert$. Let $\pi_\Delta:\Omega_m\to\mathbb{R}^\delta$, and $\pi'_\Delta:\mathbf{K}_{m+1}\to\mathbb{R}^\delta$ be the natural projection with respect to the variables X_i, $i\in\Delta$. Similarly, let $\pi_{\Delta^c}:\Omega_m\to\mathbb{R}^{s_m-\delta}$, and $\pi'_{\Delta^c}:\mathbf{K}_{m+1}\to\mathbb{R}^{n_{m+1}-\delta}$ be the natural projections with respect to the variables X_i, $i\in W_m\setminus\Delta$, and X_i, $i\in I_{m+1}\setminus\Delta$, respectively. So consider μ_m and ν_{m+1} as probability measures on the Borel spaces

$$Y\times Z\,:=\,\pi_{\Delta^c}(\Omega_m)\times\pi_\Delta(\Omega_m),\quad\text{and}\quad Z'\times Y'\,:=\,\pi'_\Delta(\mathbf{K}_{m+1})\times\pi'_{\Delta^c}(\mathbf{K}_{m+1}),$$

respectively. Next, consider the marginals $\mu_{m\Delta}$ and $\nu_{(m+1)\Delta}$ of μ_m and ν_{m+1} on Z and Z' respectively, and the corresponding disintegrations

$$\mu_m\,=\,q_m(.\,\vert\,z)\,\mu_{m\Delta}(dz)\,;\quad\nu_{m+1}\,=\,q'_m(.\,\vert\,z)\,\nu_{(m+1)\Delta}(dz).$$

From (1.3), $\Delta\subseteq I_s$ for some $s\in\{1,\ldots,m\}$. Therefore, $\nu_{(m+1)\Delta}=\nu_{s\Delta}$ because $\{\nu_k\}_{k=1}^{m+1}$ are consistent with marginals, and $\mu_{m\Delta}=\nu_{s\Delta}=:\nu$ by the induction hypothesis. Hence, one may take $Z=Z'$, and notice that

$$q_m(Y\,\vert\,z)\,=\,q'_m(Y'\,\vert\,z)\,=\,1,\quad\nu\text{-a.e.} \tag{6.6}$$

Then define the probability measure μ_{m+1} on $Y\times Z\times Y'\subset\mathbb{R}^{s_{m+1}}$ by:

$$\mu_{m+1}(A\times B\times C)\,:=\,\int_B q_m(A\,\vert z)\,q'_m(C\,\vert z)\,\nu(dz), \tag{6.7}$$

for all Borel rectangles $A \times B \times C \in \mathcal{B}(Y) \times \mathcal{B}(Z) \times \mathcal{B}(Y')$.

We claim that μ_{m+1} has the required properties of the induction hypothesis. First consider the marginal $\mu_{(m+1)I_{m+1}}$ of μ_{m+1} on $Z \times Y'$. It is obtained from (6.7) with $A = Y$. But from (6.6),

$$\begin{aligned}\mu_{(m+1)I_{m+1}}(B \times C) \,=\, \mu_{m+1}(Y \times B \times C) &= \int_B q'_m(C\,|z)\,\nu(dz)\\ &= \int_B q'_m(C\,|z)\,\nu_{(m+1)\Delta}(dz)\\ &= \nu_{m+1}(B \times C),\end{aligned}$$

for all $B \times C$ in $\mathcal{B}(Z) \times \mathcal{B}(Y')$, which proves that $\mu_{(m+1)I_{m+1}} = \nu_{m+1}$, the desired result. Next, consider the marginal $\mu_{(m+1)W_m}$ of μ_{m+1} with respect to the variables X_i, $i \in W_m$. It is obtained from (6.7) with $C = Y'$. So, using (6.6) again,

$$\begin{aligned}\mu_{(m+1)W_m}(A \times B) \,=\, \mu_{m+1}(A \times B \times Y') &= \int_B q_m(A\,|z)\,\nu(dz)\\ &= \int_B q_m(A\,|z)\,\mu_{m\Delta}(dz)\\ &= \mu_m(A \times B),\end{aligned}$$

for all $A \times B$ in $\mathcal{B}(Y) \times \mathcal{B}(Z)$, which proves that $\mu_{(m+1)W_m} = \mu_m$. But then, $\mu_{(m+1)I_k} = \mu_{mI_k}$ for all $k \leq m$, and so, by the induction hypothesis, $\mu_{(m+1)I_k} = \mu_{mI_k} = \nu_k$ for all $k \leq m$.

Hence, we have constructed a probability measure μ_{m+1} on $Y \times Z \times Y'$, such that for all $k = 1, \ldots, m+1$, ν_k is the marginal of $\mu_{(m+1)I_k}$ with respect to the variables X_i, $i \in I_k$. It remains to show that $\mu_{m+1}(\Omega_{m+1}) = 1$.

But from the definition of $\mathbf{K}_{m+1}$, Y', ν and $\nu_{m+1}(\mathbf{K}_{m+1}) = 1$,

$$q'_m(B(z)\,|\,z) = 1 \quad \nu\text{-a.e. with } B(z) := \{y \;:\; g_j(z,y) \geq 0,\ \forall\, j \in J_{m+1}\}.$$

Similarly, from the definitions of Ω_m, Y, ν, and $\mu_m(\Omega_m) = 1$,

$$q_m(A(z)\,|\,z) = 1 \quad \nu\text{-a.e. with } A(z) := \{y \;:\; g_j(y,z) \geq 0,\ \forall\, j \in \cup_{k=1}^m J_k\}.$$

Therefore, (6.7) together with the definition (6.5) of Ω_{m+1}, yields

$$\mu_{m+1}(\Omega_{m+1}) \,=\, \int_Z q_m(A(z)\,|\,z)\,q'_m(B(z)\,|\,z) \times \nu(dz) \,=\, 1.$$

Therefore, the induction hypothesis is also true for $m+1$.

(b) From $\mu(\Omega_n) = 1$, and $\Omega_n = \mathbf{K}$, we obtain $\mu(\mathbf{K}) = 1$, the desired result. $\square$

References

[1] R.B. Ash. *Real Analysis and Probability*, Academic Press Inc, Boston, 1972.

[2] D.P. Bertsekas, S.E. Schreve. *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York, 1978.

[3] R.E. Curto and L. A. Fialkow. The truncated complex K-moment problem, *Trans. Amer. Math. Soc.* **352** (2000), 2825–2855.

[4] M. Fukuda, M. Kojima, K. Murota, K. Nakata. Exploiting Sparsity in Semidefinite Programming via Matrix Completion I: General Framework, *SIAM J. Optim.* **11** (2001), 647–674.

[5] D. Henrion, J.B. Lasserre. GloptiPoly : Global Optimization over Polynomials with Matlab and SeDuMi, *ACM Trans. Math. Soft.* **29** (2003), 165–194.

[6] D. Henrion, J.B. Lasserre. Detecting global optimality and extracting solutions in GloptiPoly, in: *Positive Polynomials in Control*. D. Henrion, A. Garulli (Editors). Lecture Notes on Control and Information Sciences, Vol. 312, Springer Verlag, Berlin, January 2005.

[7] S.Kim, M. Kojima, H. Waki. Generalized Lagrangian Duals and Sums of Squares Relaxations of Sparse Polynomial Optimization Problems, *SIAM J. Optim.* **15** (2005), 697–719.

[8] M. Kojima, S. Kim, H. Waki. Sparsity in sums of squares of polynomials, *Math. Prog.*, to appear.

[9] H. Waki, S. Kim, M. Kojima, M. Maramatsu. Sums of squares and semidefinite programming relaxations for polynomial optimization problems witth structured sparsity, Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo, 2004.

[10] M. Kojima, M. Maramatsu. A note on sparse SOS and SDP-relaxations for polynomial optimization problems over symmetric cones, Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo, 2006.

[11] J.B. Lasserre. Global optimization with polynomials and the problem of moments, *SIAM J. Optim.* **11** (2001), 796–817.

[12] J.B. Lasserre. An explicit equivalent positive semidefinite program for nonliner 0-1 programs, *SIAM J. Optim.* **12** (2002), 756–769.

[13] J.B. Lasserre. A sum of squares approximation of nonnegative polynomials, *SIAM J. Optim.* **16** (2006), 751–765.

[14] M. Laurent. Revisiting two theorems of Curto and Fialkow on moment matrices. *Proc. Amer. Math. Soc.* **133** (2005), 2965–2976.

[15] K. Nakata, K. Fujisawa, M. Fukuda, M. Kojima, K. Murota. Exploiting Sparsity in Semidefinite Programming via Matrix Completion II: Implementation and Numerical Results, *Math. Progr.* **95** (2003), 303–327.

[16] M. Putinar. Positive polynomials on compact semi-algebraic sets, *Indiana Univ. Math. J.* **42** (1993), 969–984.

[17] M. Schweighofer. Optimization of polynomials on compact semialgebraic sets, *SIAM J. Optim*, **15** (2005), 805–825.

LAAS-CNRS and Institute of Mathematics, LAAS 7 Avenue du Colonel Roche, 31077 Toulouse Cédex, France.

E-mail address: lasserre@laas.fr