CONVERGENT RELAXATIONS OF POLYNOMIAL OPTIMIZATION PROBLEMS WITH NONCOMMUTING VARIABLES[*]

S. PIRONIO[†], M. NAVASCUÉS[‡], AND A. ACÍN[§]

Abstract. We consider optimization problems with polynomial inequality constraints in noncommuting variables. These noncommuting variables are viewed as bounded operators on a Hilbert space whose dimension is not fixed and the associated polynomial inequalities as semidefinite positivity constraints. Such problems arise naturally in quantum theory and quantum information science. To solve them, we introduce a hierarchy of semidefinite programming relaxations which generates a monotone sequence of lower bounds that converges to the optimal solution. We also introduce a criterion to detect whether the global optimum is reached at a given relaxation step and show how to extract a global optimizer from the solution of the corresponding semidefinite programming problem.

Key words. global optimization, semidefinite programming, positive polynomials, noncommutative variables

AMS subject classifications. 90C22, 90C99, 81Q99

DOI. 10.1137/090760155

1. Introduction. A standard problem in optimization theory is to find the global minimum of a polynomial on a set constrained by polynomial inequalities, that is, to solve the program

$$
(1) \qquad \begin{aligned} p^\star = \min_{x \in \mathbb{R}^n} & \quad p(x) \\ \text{subject to (s.t.)} & \quad q_i(x) \geq 0, \qquad i = 1, \ldots, m\,, \end{aligned}
$$

where $p(x)$ and $q_i(x)$ are real-valued polynomials in the variable $x \in \mathbb{R}^n$. To deal with such nonconvex problems, Lasserre [12] introduced a sequence of semidefinite programming (SDP)[1] relaxations of increasing size, whose optima converge monotonically to the global optimum $p^\star$; a similar approach has been proposed by Parrilo [21]. This paper presents a generalization of Lasserre's method for a noncommutative version of the optimization problem (1). That is, we consider a polynomial optimization problem, where the variables $x = (x_1, \ldots, x_n)$ are not simply real numbers but noncommuting (NC) variables for which, in general, $x_i x_j \neq x_j x_i$. Our motivation comes from quantum theory, whose basic objects are matrices and operators that do not commute. But our approach might also find applications in other fields that involve optimization over matrices or operators, such as in systems engineering [6].

To write down the noncommutative version of (1), let $p(x)$ and $q_i(x)$ be polynomial expressions in the noncommuting variables $x = (x_1, \ldots, x_n)$. Given a Hilbert space H and a set $X = (X_1, \ldots, X_n)$ of bounded operators on H, we define operators $p(X)$ and $q_i(X)$ by substituting the variables x by the operators X in the expressions $p(x)$ and $q_i(x)$. Given in addition a normalized vector ϕ in H, we evaluate the polynomial $p(X)$ as $\langle \phi, p(X)\phi \rangle$. The noncommutative version of the optimization problem (1) considered here is then

$$
(2) \qquad
\begin{aligned}
p^\star \quad = \quad &\min_{(H, X, \phi)} \quad \langle \phi, p(X)\phi \rangle \\
&\text{s.t.} \quad\quad q_i(X) \succeq 0, \qquad i = 1, \ldots, m \,,
\end{aligned}
$$

where $q_i(X) \succeq 0$ means that the operator $q_i(X)$ should be positive semidefinite. In other words, given the input data $p(x)$ and $q_i(x)$, we look for the combination (H, X, ϕ) of Hilbert space H, operators X, and normalized state ϕ (both defined on H) that minimizes $\langle \phi, p(X)\phi \rangle$ subject to the constraints $q_i(X) \succeq 0$. It is important to note that the dimension of the Hilbert space H is not fixed but subject to optimization as well.

Taking inspiration from Lasserre's method [12] and from the papers [18, 19], we introduce a hierarchy of SDP relaxations for the above optimization problem. The optimal solutions of these relaxations form a monotonically increasing sequence of lower bounds on the global minimum $p^\star$. We prove that this sequence converges to the optimum $p^\star$ when the set of constraints $q_i(X) \succeq 0$ is such that every tuple of operators $X = (X_1, \ldots, X_n)$ satisfying them is bounded, i.e., such that they satisfy $C^2 - (X_1 + \cdots + X_n) \succeq 0$ for some real constant $C > 0$. Our proof is constructive: from the sequence of optimal solutions of the SDP relaxations, we build an explicit global minimizer $(H^\star, X^\star, \phi^\star)$ for (2), where $H^\star$ is, in general, infinite-dimensional. In some cases, the SDP relaxation at a given finite step in the hierarchy may already yield the global minimum $p^\star$. We introduce a criterion to detect such events and show in this case how to extract the global minimizer $(H^\star, X^\star, \phi^\star)$ from the solution of this particular SDP relaxation. The resulting Hilbert space $H^\star$ is then finite-dimensional, with its dimension determined by the rank of the matrices involved in the solution of the SDP relaxation.

Our method can find direct applications in quantum information science, e.g., to compute upper bounds on the maximal violation of Bell inequalities, and in quantum chemistry to compute atomic and molecular ground-state energies. Practice reveals that convergence is usually fast and finite (up to machine precision).

1.1. Relation to other works. Unconstrained NC polynomial optimization problems (i.e., the minimization of a single polynomial $p(X)$ with no constraints of the form $q_i(X) \succeq 0$) were considered in [10]. Such problems can also be solved using SDP, as implemented in the MATLAB toolbox NCSOStools [1]. Unlike constrained NC optimization (2), which requires a sequence of SDPs to compute the minimum, for unconstrained NC optimization a *single* SDP is sufficient by a theorem of Helton that a symmetric NC polynomial is positive if and only if it admits a sum of square decomposition [4]. This single SDP corresponds actually to the first step of our hierarchy when neglecting constraints coming from the conditions $q_i(X) \succeq 0$.

In [18], a subclass of the general constrained NC problem (2) which is of interest in quantum information (see (67) later in the text) was considered and a sequence of SDP relaxations introduced for it. The convergence of this SDP sequence was established in [19, 3, 9]. Our work can be seen as a generalization of these results to arbitrary NC polynomial optimization.

In the commutative case, the convergence of the relaxations introduced by Lasserre is based on a sum of squares representation theorem of Putinar [23] for positive polynomials. The connection to Putinar's representation arises when considering the dual problems of the SDP relaxations. A noncommutative analogue of Putinar's result, the Positivstellensatz for noncommutative positive polynomials, has been introduced by Helton and McCullough [5]. Although we first prove the convergence of the hierarchy introduced here through the primal version of our SDP relaxations (in the spirit of [19]), we also provide an alternative proof through the duals, which exploits Helton and McCullough's result (as used in [3] and [9]).

Note that problem (2) can also account for equality constraints $q_i(X) = 0$, which can be enforced through the inequalities $q_i(X) \succeq 0$ and $-q_i(X) \succeq 0$. When constraints of the form $x_i x_j - x_j x_i$ are explicitly added to (2), that is, when we require that the variables x commute, our method reduces to the one introduced by Lasserre. It is interesting to note that the results presented here, such as the convergence of the hierarchy or the criterion to detect optimality, are easier to establish in the general noncommutative framework than they are in the specialized commutative case. This commutative setting has generated quite a large literature, and the properties of the corresponding SDP relaxations have been thoroughly investigated. We refer to [13] for a review. Our work provides an NC analogue only of the most basic results in the commutative case. It would be interesting to reexamine from an NC perspective other topics in this subject.

1.2. Organization of the paper. In section 2, we define some notation and introduce in more detail the class of problems that we consider here. Section 3 contains our main results: we introduce our hierarchy of SDP relaxations, prove its convergence, show how to detect optimality at a finite step in the hierarchy, and show how to extract a global optimizer. We then explain the relation between our approach and the works of Helton and McCullough. We proceed by mentioning briefly how to modify our method to deal efficiently with equality constraints. In particular, we discuss how it can be simplified when dealing with Hermitian variables and how it reduces to Lasserre's method in the case of commuting variables. We end section 3 by showing how our method can be extended to solve a slightly more general class of optimization problems. In section 4, we illustrate our method on concrete examples. Finally, we briefly discuss practical applications of our method in the quantum setting in section 5.

2. Notation and definitions. Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ denote the field of real or complex numbers. We consider the algebra $\mathbb{K}[x, x^*]$ of polynomials in the $2n$ noncommuting variables $x = (x_1, \ldots, x_n)$ and $x^* = (x_1^*, \ldots, x_n^*)$ with coefficients from $\mathbb{K}$. That is, an element $p \in \mathbb{K}[x, x^*]$ is a linear combination

$$(3) \qquad p = \sum_w p_w \, w$$

of words w in the $2n$ letters x and x^*, where the sum is finite and $p_w \in \mathbb{K}$. We interpret $*$ as an involution (that is, loosely speaking, a conjugate transpose) defined as follows: on letters, $(x_i)^* = x_i^*$ and $(x_i^*)^* = x_i$; on a word, $w = w_1 \ldots w_n$, $w^* = w_n^* \ldots w_1^*$; and on a polynomial, $p^* = \sum_w p_w^* w^*$, where p_w^* is the complex conjugate of p_w. Thus $\mathbb{K}[x, x^*]$ is the free $*$-algebra generated by the $2n$ variables $x_1, \ldots, x_n, x_1^*, \ldots, x_n^*$. In the following, we will often view these $2n$ variables as $x_1, \ldots, x_n, x_{n+1}, \ldots, x_{2n}$ by identifying x_{n+i} with x_i^*.

Throughout this paper, the symbols u, v, w always denote words, and we denote the empty word by 1. We use the notation $\mathcal{W}_d$ for the set of all words of length $|w|$

at most d and $\mathcal{W}_\infty$ for the set of all words (of unrestricted length). The number of words in $\mathcal{W}_d$ is $|\mathcal{W}_d| = \left((2n)^{d+1} - 1\right)/(2n - 1)$. The degree of a polynomial p is the length of the longest word in p and is denoted $\deg(p)$. We let $\mathbb{K}[x, x^*]_d$ denote the set of polynomials $p = \sum_{|w| \leq d} p_w\, w$ of degree $\leq d$. If necessary, a polynomial of degree d can be viewed as a polynomial of higher degree d' by setting to zero the coefficients associated with words of length larger than d. A polynomial p is said to be Hermitian if $p^* = p$ or, in term of its coefficients, if $p_w^* = p_{w^*}$. Note that words can be interpreted as monomials, and we will sometimes use the two terms interchangeably. We will then also refer to the length $|w|$ of a word as the degree of the monomial w and to $\mathcal{W}_d$ as a monomial basis for $\mathbb{K}[x, x^*]_d$.

Let $\mathcal{B}(H)$ denote the set of bounded operators on a Hilbert space H defined on the field $\mathbb{K}$. Consider a set of operators $X = (X_1, \ldots, X_n)$ from $\mathcal{B}(H)$. Given the polynomial $p \in \mathbb{K}[x, x^*]$, we define the operator $p(X) \in \mathcal{B}(H)$ by substituting every variable x_i by the operator X_i and every variable x_i^* by X_i^*, where $*$ denotes the adjoint operation on H. If $p^* = p$ is a Hermitian polynomial, then $p(X) = p^*(X)$ is a Hermitian operator and the quantity $\langle \phi, p(X)\phi \rangle$ is real for every vector ϕ in H. A Hermitian operator O is said to be positive semidefinite, a fact that we denote by $O \succeq 0$, if $\langle \phi, O\phi \rangle \geq 0$ for all $\phi \in H$.

2.1. Formulation of the optimization problem. Let p and q_i $(i = 1, \ldots, m)$ be Hermitian polynomials in $\mathbb{K}[x, x^*]$. We are interested in the following optimization problem:

$$
(4) \qquad \mathbf{P}: \qquad
\begin{aligned}
p^\star = \min_{(H, X, \phi)} \quad & \langle \phi, p(X)\phi \rangle \\
\text{s.t.} \quad & q_i(X) \succeq 0, \qquad i = 1, \ldots, m,
\end{aligned}
$$

where the optimization should be understood over all Hilbert spaces H (of arbitrary dimension), all sets of operators $X = (X_1, \ldots, X_n)$ in $\mathcal{B}(H)$, and all normalized vectors ϕ in H. We assume throughout the remainder of the paper that this problem admits a feasible solution, that is, that there exists a triple (H, X, ϕ) such that $\langle \phi, \phi \rangle = 1$ and $q_i(X) \succeq 0$ for $i = 1, \ldots, m$.

Let $Q = \{q_i : i = 1, \ldots, m\}$ be the set of polynomials determining the positivity constraints in (4). The following definitions follow those used in [11]. The *positivity domain* $\mathbf{S}_Q$ associated to Q is the class of tuples $X = (X_1, \ldots, X_n)$ of bounded operators on a Hilbert space making each $q_i(X)$ a positive semidefinite operator. The *quadratic module* $\mathbf{M}_Q$ is the set of all elements of the form $\sum_i f_i^* f_i + \sum_i \sum_j g_{ij}^* q_i g_{ij}$, where f_i and g_{ij} are polynomials in $\mathbb{K}[x, x^*]$. We say that $\mathbf{M}_Q$ is Archimedean if there exists a real constant C such that $C^2 - (x_1^* x_1 + \cdots + x_{2n}^* x_{2n}) \in \mathbf{M}_Q$. In this case, the positivity domain $\mathbf{S}_Q$ is bounded: for all $X \in \mathbf{S}_Q$, $C^2 - (X_1^* X_1 + \cdots + X_{2n}^* X_{2n}) \succeq 0$. Note that if $\mathbf{S}_Q$ is bounded, we can always add $C^2 - (x_1^* x_1 + \cdots + x_{2n}^* x_{2n})$ to Q for a sufficiently large C to make $\mathbf{M}_Q$ Archimedean without changing $\mathbf{S}_Q$. In the following we will always assume that the constraints in Q are such that $\mathbf{M}_Q$ is Archimedean.

3. Main results.

3.1. Moment and localizing matrices. Let $y = (y_w)_{|w| \leq d} \in \mathbb{K}^{|\mathcal{W}_d|}$ be a sequence of real or complex numbers indexed in $\mathcal{W}_d$, i.e., to each word $w \in \mathcal{W}_d$ corresponds a number $y_w \in \mathbb{K}$. We define the linear mapping $L_y : \mathbb{K}[x, x^*]_d \mapsto \mathbb{K}$ as

$$
(5) \qquad\qquad p \mapsto L_y(p) = \sum_{|w| \leq d} p_w y_w\,.
$$

By analogy with [12], given a sequence $y = (y_w)_{|w| \leq 2k}$ indexed in W_{2k}, we define the *moment matrix* $M_k(y)$ of order k as a matrix with rows and columns indexed in W_k,

and whose entry (v, w) is given by

$$(6) \qquad M_k(y)(v, w) = L_y(v^*w) = y_{v^*w}.$$

If $q = \sum_{|u| \leq d} q_u u$ is a polynomial of degree d and $y = (y_w)_{|w| \leq 2k+d}$ a sequence indexed in $\mathcal{W}_{2k+d}$, we define the *localizing matrix* $M_k(qy)$ as the matrix with rows and columns indexed in $\mathcal{W}_k$ and whose entry (v, w) is

$$(7) \qquad M_k(qy)(v, w) = L_y(v^*qw) = \sum_{|u| \leq d} q_u y_{v^*uw}.$$

We say that a sequence $y = (y_w)_{|w| \leq 2k}$ admits a *moment representation* if there exists a triple (H, X, ϕ) with a normalized ϕ such that

$$(8) \qquad y_w = \langle \phi, w(X)\phi \rangle$$

for all $|w| \leq 2k$.

LEMMA 1. *Let* $y = (y_w)_{|w| \leq 2k}$ *be a sequence admitting a moment representation. Then* $y_1 = 1$ *and* $M_k(y) \succeq 0$. *If the moment representation* (8) *is such that* $q(X) \succeq 0$ *for some* $q \in \mathbb{K}[x, x^*]$, *then in addition* $M_{k-d}(qy) \succeq 0$, *where* $d = \lceil \deg(q)/2 \rceil$.

Proof. Equation (8) immediately implies $y_1 = 1$ since $\langle \phi, \phi \rangle = 1$. The positivity of the moment matrix $M_k(y)$ follows from the fact that for any vector $z \in \mathbb{K}^{|\mathcal{W}_k|}$,

$$z^*M_k(y)z = \sum_{v,w} z_v^* M_k(y)(v, w) z_w = \sum_{v,w} z_v^* y_{v^*w} z_w$$

$$(9) \qquad = \left\langle \phi, \sum_v z_v^* v^*(X) \sum_w z_w w(X)\, \phi \right\rangle = \langle \phi, z^*(X)z(X)\, \phi \rangle \geq 0,$$

where we have defined the operator $z(X) = \sum_w z_w w(X)$.

Suppose now that y admits a moment representation (8) by a triple (H, X, ϕ) such that $q(X) \succeq 0$. Then $M_{k-d}(qy) \succeq 0$ since for all vectors $z \in \mathbb{K}^{|\mathcal{W}_{k-d}|}$,

$$z^*M_{k-d}(qy)z = \sum_{v,w} z_v^* M_{k-d}(qy)(v, w) z_w = \sum_{v,w,u} z_v^* q_u y_{v^*uw} z_w$$

$$= \left\langle \phi, \sum_v z_v^* v^*(X) \sum_u q_u u(X) \sum_w z_w w(X)\, \phi \right\rangle$$

$$(10) \qquad = \langle \phi, z^*(X)q(X)z(X)\, \phi \rangle \geq 0,$$

where the last inequality follows from the fact that $q(X) \succeq 0$. $\qquad \square$

3.2. Convergent SDP relaxations. For $2k \geq \max\{\deg(p), \max_i \deg(q_i)\}$, consider the SDP problem

$$
\begin{aligned}
p^k = \min_y \quad & \sum_w p_w y_w \\
(11) \qquad \mathbf{R_k}: \qquad & \text{s.t.} \quad y_1 = 1, \\
& \qquad M_k(y) \succeq 0, \\
& \qquad M_{k-d_i}(q_i y) \succeq 0, \qquad i = 1, \dots, m,
\end{aligned}
$$

where $d_i = \lceil \deg(q_i)/2 \rceil$ and the optimization is over $y = (y_w)_{|w| \leq 2k} \in \mathbb{K}^{|\mathcal{W}_{2k}|}$. The optimum p^k provides a lower bound on the global optimum $p^\star$ of the original problem

P since any feasible solution (H, X, ϕ) of **P** yields a feasible solution y of $\mathbf{R_k}$ through (8) and Lemma 1. We refer to $\mathbf{R_k}$ as the SDP relaxation of order k of **P**. Since the positivity of the moment and the localizing matrices of a given order k' implies the positivity of the moment and the localizing matrices of lower orders k, the sequences of SDP relaxations form a hierarchy in the sense that $p^k \leq p^{k'}$ when $k \leq k'$.

THEOREM 1. *If $\mathbf{M}_Q$ is Archimedean, $\lim_{k \to \infty} p^k = p^\star$.*

Remember that if $\mathbf{M}_Q$ is Archimedean, there exist polynomials f_i and g_{ij} and a constant C such that $C^2 - (x_1^* x_1 + \cdots + x_{2n}^* x_{2n}) = \sum_i f_i^* f_i + \sum_i \sum_j g_{ij}^* q_i g_{ij}$. In the following, we write $d_\mathbf{M} = \max_{ij}\{\deg(f_i), \deg(g_{ij}) + d_i\}$. Note that $d_\mathbf{M} \geq 1$, with $d_\mathbf{M} = 1$ when $C^2 - (x_1^* x_1 + \cdots + x_{2n}^* x_{2n})$ is contained in Q. Although the asymptotic behavior of the hierarchy of SDP relaxations depends only on the quadratic module being Archimedean, it may be advantageous in practice to add the constraint $C^2 - (x_1^* x_1 + \cdots + x_{2n}^* x_{2n})$ to Q. This will guarantee in particular that the first step of the hierarchy has a bounded solution (see Lemma 3).

The proof of Theorem 1 is based on the following four lemmas.

LEMMA 2. *Let $c = C^2 - (x_1^* x_1 + \cdots + x_{2n}^* x_{2n})$, and let y be a sequence satisfying $y_1 = 1$, $M_k(y) \succeq 0$, and $M_{k-1}(cy) \succeq 0$. Then $|y_w| \leq C^{|w|}$ for all $|w| \leq 2k$.*

Proof. The diagonal elements of $M_{k-1}(cy)$ are of the form $C^2 y_{w^* w} - \sum_{i=1}^{2n} y_{w^* x_i^* x_i w}$ with $|w| \leq k - 1$. Since the localizing matrix $M_{k-1}(cy)$ is positive semidefinite, these diagonal entries must be positive, that is, $\sum_{i=1}^{2n} y_{w^* x_i^* x_i w} \leq C^2 y_{w^* w}$. In addition, it also holds that $y_{w^* x_i^* x_i w} \geq 0$ since these numbers are diagonal entries of the moment matrix $M_k(y)$. It thus follows that $y_{w^* x_i^* x_i w} \leq C^2 y_{w^* w}$ for all $|w| \leq k - 1$ and all $i = 1, \ldots, 2n$. Given that $y_1 = 1$, we deduce by induction that $y_{w^* w} \leq C^{2|w|}$ for all $|w| \leq k$.

The moment matrix $M_k(y)$ admits the matrix

$$(12) \qquad \begin{pmatrix} y_{w^* w} & y_{w^* v} \\ y_{v^* w} & y_{v^* v} \end{pmatrix}$$

as a submatrix, where $|w|, |v| \leq k$. Since $M_k(y) \succeq 0$, the above submatrix must also be positive semidefinite, which is equivalent to the condition that $y_{w^* v} y_{v^* w} \leq y_{w^* w} y_{v^* v}$. Combining this relation with the previous bound on $y_{w^* w}$ and the fact that $y_{v^* w} = y_{w^* v}^*$ which follows from the hermicity of $M_k(y)$, we deduce that $|y_w| \leq C^{|w|}$ for all $|w| \leq 2k$. ☐

LEMMA 3. *Let $2k \geq \max\{\deg(p), \max_i \deg(q_i)\}$, and let M_Q be Archimedean. Let y be a feasible solution of the relaxation $\mathbf{R_{k-1+d_M}}$. Then $|y_w| \leq C^{|w|}$ for all $|w| \leq 2k$.*

Proof. First note that if $f \in \mathbb{K}[x, x^*]_d$ is a polynomial of degree d and y is a sequence such that $M_{k+d}(y) \succeq 0$, then $M_k(f^* f y) \succeq 0$. This follows from the fact that $M_{k+d}(y) \succeq 0$ implies $\sum_{|v|,|w| \leq k} \sum_{|\tilde{v}|,|\tilde{w}| \leq d} z_v^* f_{\tilde{v}}^* L_y(v^* \tilde{v}^* \tilde{w} w) z_w f_{\tilde{w}} \geq 0$ for all $z \in \mathbb{K}^{|\mathcal{W}_k|}$ and from the identity $\sum_{|v|,|w| \leq k} \sum_{|\tilde{v}|,|\tilde{w}| \leq d} z_v^* f_{\tilde{v}}^* L_y(v^* \tilde{v}^* \tilde{w} w) z_w f_{\tilde{w}} = \sum_{|v|,|w| \leq k} z_v^* L_y(v^* f^* f w) z_w = z^* M_k(f^* f y) z$. Similarly, if $g \in \mathbb{K}[x, x^*]_d$ is a polynomial of degree d and y is a sequence such that $M_{k+d}(q_i y) \succeq 0$, then $M_k(g^* q_i g y) \succeq 0$. Indeed, from $M_{k+d}(q_i y) \succeq 0$ we deduce that for all $z \in \mathbb{K}^{|\mathcal{W}_k|}$, $\sum_{|v|,|w| \leq k} \sum_{|\tilde{v}|,|\tilde{w}| \leq d} z_v^* g_{\tilde{v}}^* L_y(v^* \tilde{v}^* q_i \tilde{w} w) z_w g_{\tilde{w}} \geq 0$, and the left-hand side of this last inequality is equal to $\sum_{|v|,|w| \leq k} z_v^* L_y(v^* g^* q_i g w) z_w = z^* M_k(g^* q_i g y) z$.

Now, let y be the optimal solution of the relaxation $\mathbf{R_{k-1+d_M}}$ as in the statement of the lemma, and let $c = C^2 - (x_1^* x_1 + \cdots + x_{2n}^* x_{2n})$. As $\mathbf{M}_Q$ is Archimedean, we can write $c = \sum_i f_i^* f_i + \sum_{ij} g_{ij}^* q_i g_{ij}$, and thus $M_{k-1}(cy) = \sum_i M_{k-1}(f_i^* f_i y) +$

$\sum_{ij} M_{k-1}(g_{ij}^* q_i g_{ij} y)$. Since $M_{k-1+d_{\mathbf{M}}}(y) \succeq 0$ and $M_{k-1+d_{\mathbf{M}}-d_i}(q_i y) \succeq 0$, the argument outlined above implies that $M_{k-1}(f_i^* f_i y) \succeq 0$ and $M_{k-1}(g_{ij}^* q_i g_{ij} y) \succeq 0$. This in turn implies $M_{k-1}(cy) \succeq 0$. From Lemma 2, we then deduce that $|y_w| \leq C^{|w|}$ for all $|w| \leq 2k$. $\quad\square$

LEMMA 4. *If M_Q is Archimedean, the optima p^k of the relaxations $\mathbf{R_k}$ form, for k large enough, a monotically increasing bounded sequence. Therefore, the limit $\hat{p} = \lim_{k\to\infty} p^k$ exists.*

Proof. Let $l = l' - 1 + d_{\mathbf{M}}$ with $2l' \geq \max\{\deg(p), \max_i \deg(q_i)\}$, and let y be the solution of the relaxation $\mathbf{R_l}$ with objective value p^l. From Lemma 3, the entries y_w with $|w| \leq 2l'$ are bounded, i.e., $|y_w| \leq C^{|w|}$. Thus the solution $p^l = \sum_{|w| \leq l'} p_w y_w$ is bounded as well. We also have that $p^\star$ is bounded since for $\mathbf{M_Q}$ Archimedean, the positivity domain $\mathbf{S_Q}$ is bounded. For all $k \geq l$, $p^l \leq p^k \leq p^{k+1} \leq p^\star$. Thus the $(p^k)_{k \geq l}$ form a monotonically increasing bounded sequence, and the limit $\hat{p} = \lim_{k\to\infty} p^k$ exists. $\quad\square$

LEMMA 5. *Let $\mathbf{M}_Q$ be Archimedean, and let $\hat{p} = \lim_{k\to\infty} p^k$ be the limit of the optimal solutions p^k of the relaxations $\mathbf{R_k}$. Then there exists an infinite sequence $\hat{y} = (\hat{y}_w)_{|w|=0,1,\ldots}$ indexed in $\mathcal{W}_\infty$ such that $|y_w| \leq C^{|w|}$,*

$$\sum_w p_w \hat{y}_w = \hat{p},$$
$$(13) \qquad \hat{y}_1 = 1,$$

and

$$M_k(\hat{y}) \succeq 0,$$
$$(14) \qquad M_{k-d_i}(q_i \hat{y}) \succeq 0, \quad i = 1, \ldots, m$$

for all k large enough.

Proof. For any k such that $2k \geq \max\{\deg(p), \deg(q_i)\}$, let $y^{k-1+d_{\mathbf{M}}}$ be a feasible solution of the relaxation $\mathbf{R_{k-1+d_M}}$ with objective value $\hat{p}$. Such a solution always exists because the problem $\mathbf{R_{k-1+d_M}}$ is convex and there exist feasible points of $\mathbf{R_{k-1+d_M}}$ with optimal values p_1 and p_2 satisfying $p_1 \leq \hat{p} \leq p_2$ (take, for instance, $p_1 = p^{k-1+d_{\mathbf{M}}}$ and $p_2 = p^\star$). By Lemma 3, the entries $y_w^{k-1+d_{\mathbf{M}}}$ with $|w| \leq 2k$ are bounded, i.e., $|y_w^{k-1+d_{\mathbf{M}}}| \leq C^{|w|}$. Let $\tilde{y}^k$ be the restriction of the solution $y^{k-1+d_{\mathbf{M}}}$ to the $|w| \leq 2k$. That is, $\tilde{y}^k = (y_w^{k-1+d_{\mathbf{M}}})_{|w| \leq 2k}$ is the subsequence of $y^{k-1+d_{\mathbf{M}}}$ composed of the entries $y_w^{k-1+d_{\mathbf{M}}}$ with $|w| \leq 2k$. Complete $\tilde{y}^k$ with zeros to make it an infinite vector y^k in l_∞, and perform the renormalization $y_w^k \to z_w^k = y_w^k / C^{|w|}$. Each vector z^k thus belongs to the unit ball of l_∞, and the sequence $(z^k)_{k \geq l}$ admits by the Banach–Alaoglu theorem a subsequence $(z^{k_i})_{i=1,2,\ldots}$ that converges in the weak-$*$ topology to a limit $\lim_{i\to\infty} z^{k_i} = \hat{z}$ [24]. This implies, in particular, pointwise convergence, i.e., $\lim_{i\to\infty} z_w^{k_i} = \hat{z}_w$ for all w. Define the infinite vector $\hat{y}$ through $\hat{y}_w = \hat{z}_w C^{|w|}$. The pointwise convergence $z^{k_i} \to \hat{z}$ implies the pointwise convergence of $y^{k_i} \to \hat{y}$, i.e., $\lim_{i\to\infty} y_w^{k_i} = \hat{y}_w$ for all w. Since $\sum_w p_w y_w^k = \hat{p}$, $y_1^k = 1$, $M_k(y^{k'}) \succeq 0$, and $M_{k-d_i}(q_i y^{k'}) \succeq 0$ $(i = 1, \ldots, m_q)$ for all k, k' with $k' \geq k$, we deduce (13) and (14) from the pointwise convergence of $y^{k_i} \to \hat{y}$. $\quad\square$

Proof of Theorem 1. By Lemma 4, the limit $\hat{p} = \lim_{k\to\infty} p^k$ exists. We obviously have that $\hat{p} \leq p^\star$. We now show that there exist a set of operators $\hat{X}$ and a vector $\hat{\phi}$ in a Hilbert space $\hat{H}$ (possibly of infinite dimension) that yield a feasible solution of $\mathbf{P}$ with objective value $\hat{p}$. Thus, we also have that $\hat{p} \geq p^\star$, and therefore $\hat{p} = p^\star$.

Incidentally, this implies that the minimum appearing in (4) is well defined, i.e., it is not an infimum, as one would have expected in general.

To build $(\hat{H}, \hat{X}, \hat{\phi})$, we perform a Gelfand–Naimark–Segal-like construction. Let $\hat{y}$ be the infinite sequence defined in Lemma 5. Consider the linear functional $L_{\hat{y}}$: $\mathbb{K}[x, x^*] \mapsto \mathbb{K}$, $p \mapsto L_{\hat{y}}(p) = \sum_w p_w \hat{y}_w$. Since $M_k(\hat{y}) \succeq 0$ for all k, this linear functional is positive in the sense that $L_{\hat{y}}(p^*p) = \sum_{v,w} p_v^* L_{\hat{y}}(v^*w) p_w \geq 0$ for all p. It thus defines a semi-inner product on $\mathbb{K}[x, x^*]$ through

$$(15) \qquad \langle p, q \rangle = L_{\hat{y}}(p^*q) \,.$$

Define the set

$$(16) \qquad I = \{p \in \mathbb{K}[x, x^*] \, : \, \langle p, p \rangle = 0\} \,.$$

By the Cauchy–Schwarz inequality (which is valid for semi-inner products), the set I is a linear subspace of $\mathbb{K}[x, x^*]$. Moreover, it is a left ideal of $\mathbb{K}[x, x^*]$. To show that I is a left ideal of $\mathbb{K}[x, x^*]$, it is sufficient to show that $x_i I \subseteq I$ for all $i = 1, \ldots, 2n$. Since $\mathbf{M}_Q$ is Archimedean, there is some C such that $c = C^2 - \sum_i x_i^* x_i \in \mathbf{M}_Q$, and, as in the proof of Lemma 3, one can show that $M_k(c\hat{y}) \succeq 0$, from which it follows that

$$(17) \qquad 0 \leq L_{\hat{y}}(p^*cp) = C^2 L_{\hat{y}}(p^*p) - \sum_i L_{\hat{y}}(p^*x_i^*x_ip) \,.$$

Since $L_{\hat{y}}(p^*x_i^*x_ip) \geq 0$ for all i, (17) implies that

$$(18) \qquad 0 \leq L_{\hat{y}}(p^*x_i^*x_ip) \leq C^2 L_{\hat{y}}(p^*p) \,.$$

For all $p \in I$, we thus have that $L_{\hat{y}}(p^*x_i^*x_ip) = 0$, that is, $x_ip \in I$.

The definition (15) of $\langle \cdot, \cdot \rangle$ induces a well-defined inner product on the quotient $\mathbb{K}[x, x^*]/I$. Let $\hat{H}$ denote the Hilbert space corresponding to the completion of $\mathbb{K}[x, x^*]/I$ with respect to this scalar product. We will now construct operators $\hat{X}$ on $\hat{H}$. For every x_i, let $\hat{X}_i$ be the operator of left multiplication by x_i on $\mathbb{K}[x, x^*]/I$, i.e.,

$$(19) \qquad \hat{X}_i(p + I) = x_ip + I \,.$$

Since I is a left ideal, this map is well defined for every x_i. It is linear, and by (18) it is bounded. Thus it extends uniquely to a bounded operator on H, which we denote by the same symbol $\hat{X}_i$. Note that the map is also consistent with the involution on $\mathbb{K}[x, x^*]$, i.e., it satisfies $\hat{X}_i^* = \hat{X}_{i+n}$, since $\langle p, X_i^*q \rangle = \langle X_ip, q \rangle = \langle x_ip, q \rangle = L_{\hat{y}}(p^*x_i^*q) = L_{\hat{y}}(p^*x_{i+n}q) = \langle p, X_{i+n}q \rangle$.

Let $\hat{\phi}$ be the vector of $\mathbb{K}[x, x^*]/I$ corresponding to the identity polynomial 1. The fact that $\hat{y}_1 = 1$ implies that the vector $\hat{\phi}$ is normalized: $\langle \hat{\phi}, \hat{\phi} \rangle = 1$. From (13), it follows that

$$(20) \qquad \langle \hat{\phi}, p(\hat{X})\hat{\phi} \rangle = \sum_w p_w \langle \hat{\phi}, w(\hat{X})\hat{\phi} \rangle = \sum_w p_w \langle 1, w \rangle = \sum_w p_w \hat{y}_w = \hat{p} \,.$$

To show that $(\hat{H}, \hat{X}, \hat{\phi})$ yields a feasible solution to $\mathbf{P}$ with objective value $\hat{p}$, it remains to show that the operators $\hat{X}$ satisfy $q_i(\hat{X}) \succeq 0$ ($i = 1, \ldots, m$), i.e., that $\langle r, q_i(\hat{X})r \rangle \geq 0$ for all $r \in \hat{H}$. But since any $r \in \hat{H}$ can be approximated to arbitrary precision by elements of the pre-Hilbert space $\mathbb{K}[x, x^*]/I$, it is sufficient to show that $\langle p, q_i(\hat{X})p \rangle \geq 0$ for all $p \in \mathbb{K}[x, x^*]$. This follows from

$$(21) \qquad \langle p, q_i(\hat{X})p \rangle = \langle p, q_ip \rangle = L_{\hat{y}}(p^*q_ip) = \sum_{v,w} p_v^* L_{\hat{y}}(v^*q_iw) p_w \geq 0$$

since $M_{k-d_i}(q_i\hat{y}) \succeq 0$ for all k. $\quad \square$

3.3. Optimality detection and extraction of optimizers. In this subsection, we introduce a criterion that allows us to detect whether the relaxation of order k already yields the optimal value $p^\star$. If so, it is possible to extract a global optimizer $(H^\star, X^\star, \phi^\star)$ from the optimal solution of this relaxation. The procedure to extract this optimizer is described in the proof of the following theorem.

THEOREM 2. *Assume that the optimal solution y^k of the relaxation of order k satisfies*

$$\text{(22)} \qquad \qquad \text{rank}\, M_k(y^k) = \text{rank}\, M_{k-d}(y^k),$$

where $d = \max_i d_i \geq 1$. Then $p^k = p^\star$, i.e., the optimum of the relaxation of order k is the global optimum of the original problem (4). Moreover, there exists a global optimizer $(H^\star, X^\star, \phi^\star)$ of (4) with $\dim H^\star = \text{rank}\, M_{k-d}(y^k)$.

Proof. We show that when (22) holds, we can find a solution (H, X, ϕ) to (4) with objective value p^k. This implies that $p^k \geq p^\star$, and thus $p^k = p^\star$ since we also have $p^k \leq p^\star$.

Let $r = \text{rank}\, M_k(y^k) = \text{rank}\, M_{k-d}(y^k)$. Since the moment matrix $M_k(y^k)$ is positive semidefinite, it admits a Gram decomposition. That is, to each row (and column), indexed by a w with $|w| \leq k$, can be associated a vector $\overline{w} \in \mathbb{K}^r$ such that $M_k(y^k)(w, v) = y^k_{w^* v} = \langle \overline{w}, \overline{v} \rangle$. We define the Hilbert space H as $H = \text{span}\{\overline{w} : |w| \leq k\}$, with dimension $\dim H = r$. Note that (22) implies that

$$\text{(23)} \qquad H = \text{span}\{\overline{w} : |w| \leq k\} = \text{span}\{\overline{w} : |w| \leq k - d\}.$$

We now define $2n$ linear operators X_i through their actions on the $\overline{w}$'s with $|w| \leq k - 1$ in the following way:

$$\text{(24)} \qquad \qquad X_i \overline{w} = \overline{x_i w}.$$

Note that when $d \geq 1$, the operators are well defined on the whole space H since by (23) the set of vectors $\overline{w}$ with $|w| \leq k - d \leq k - 1$ span H. This definition is also consistent in the sense that if $\overline{f} \in H$ admits two different decompositions $\overline{f} = \sum a_w \overline{w} = \sum b_w \overline{w}$ as a linear combination of the vectors $\{\overline{w} : |w| \leq k - 1\}$, then $\sum a_w \overline{x_i w} = \sum b_w \overline{x_i w}$. Indeed, the equality

$$\left\langle \overline{v}, \sum_w (a_w - b_w) \overline{x_i w} \right\rangle = \sum_w (a_w - b_w) y_{v^* x_i w} = \sum_w (a_w - b_w) y_{(x_i^* v)^* w}$$

$$\text{(25)} \qquad \qquad = \left\langle \overline{x_i^* v}, \sum_w (a_w - b_w) \overline{w} \right\rangle = \langle \overline{x_i^* v}, 0 \rangle = 0$$

holds for all $\overline{v}$, with $|v| \leq k - d \leq k - 1$. Since these vectors span the Hilbert space H, this implies that both vectors $\sum a_w \overline{x_i w}$ and $\sum b_w \overline{x_i w}$ are identical. The definition (24) is also consistent with the involution on $\mathbb{K}\langle x \rangle$, i.e., it satisfies $\hat{X}_i^* = \hat{X}_{i+n}$. Indeed, for all v, w of length $|v|, |w| \leq k - 1$,

$$\text{(26)} \qquad \langle \overline{v}, X_i^* \overline{w} \rangle = \langle X_i \overline{v}, \overline{w} \rangle = \langle \overline{x_i v}, \overline{w} \rangle = y^k_{v^* x_i^* w} = y^k_{v^* x_{i+n} w} = \langle \overline{v}, X_{i+n} \overline{w} \rangle.$$

Let us now define $\phi = \overline{1}$. Let w be of length $|w| \leq 2k$, and write $w = w_1 w_2$ with $|w_1|, |w_2| \leq k$. Then $\langle \phi, w(X) \phi \rangle = \langle \overline{w_1^*}, \overline{w_2} \rangle = y^k_{w_1 w_2} = y^k_w$. This implies that $\langle \phi, p(X) \phi \rangle = \sum_{|w| \leq 2k} p_w \langle \phi, w(X) \phi \rangle = \sum_{|w| \leq 2k} p_w y^k_w = p^k$. It remains to check that

the operators X satisfy $q_i(X) \succeq 0$. To verify this it is necessary, because of (23), to show only that the matrix A with entries $A(v, w) = \langle \overline{v}, q_i(X)\overline{w} \rangle$ with $|v|, |w| \leq k - d$ is a positive semidefinite matrix. This is the case since A is equal to $M_{k-d}(q_i y^k)$ and is thus a submatrix of $M_{k-d_i}(q_i y^k) \succeq 0$, which is itself positive semidefinite because y^k is a solution of the relaxation of order k. $\quad\square$

Note that there exists a related optimality detection criterion in the commutative case, which is based on the flat extension theorem of Curto and Fialkow [2, 13]. The matrix $M_k(y^k)$ is said to be a *flat extension* of $M_{k-d}(y^k)$ if $\operatorname{rank} M_k(y^k) = \operatorname{rank} M_{k-d}(y^k)$. When this condition holds, the flat extension theorem permits (in the commutative case) to extend the finite sequence y^k to an infinite sequence $\hat{y}$ satisfying $\operatorname{rank} M_{k'}(\hat{y}) = \operatorname{rank} M_k(y^k)$ for all $k' \geq k$. The proof of Theorem 2 yields an NC analogue of this important result. (Simply define the infinite sequence $\hat{y}$ through $\hat{y}_w = \langle \phi, w(X)\phi \rangle$, where ϕ and X are the vectors and operators defined in the proof of Theorem 2.)

3.4. Relation to the Positivstellensatz for noncommutative polynomials. We now explain the link between the convergence of the SDP relaxations and the Positivstellensatz for noncommutative polynomials introduced by Helton and McCullough [5]. We proceed by analogy with the link that exists in the commutative case between the convergence of Lasserre's relaxations [12] and Putinar's Positivstellensatz [23].

Consider the problem

$$
\begin{aligned}
\lambda^k = \max_{\lambda, b_i, c_{ij}} \quad & \lambda \\
\text{s.t.} \quad & p - \lambda = \sum_j b_j^* b_j + \sum_{i=1}^m \sum_j c_{ij}^* q_i c_{ij}, \\
& \max_j \deg(b_j) \leq k, \\
& \max_j \deg(c_{ij}) \leq k - d_i,
\end{aligned}
$$

(27)

where b_j and c_{ij} are polynomials. The expression $\sum_i b_i^* b_i$ is known as a sum of squares (SOS), and the above problem is a polynomial SOS problem. As shown in Appendix B, this polynomial SOS problem can be formulated as an SDP problem, which turns out to be the dual of $\mathbf{R_k}$. This implies that the optimal solution of (27) provides a lower bound on the solution of $\mathbf{R_k}$, i.e.,

$$
\lambda^k \leq p^k.
$$
(28)

Alternatively, this last relation can be established as follows. Let λ, b_j, c_{ij} be a feasible solution of (27) and y be a feasible solution of (11). We show that $L_y(p - \lambda) = \sum_w p_w y_w - \lambda \geq 0$, which implies (28). As $L_y(p - \lambda) = \sum_j L_y(b_j^* b_j) + \sum_i \sum_j L_y(c_{ij}^* q_i c_{ij})$, it is sufficient to show that $L_y(b_j^* b_j) \geq 0$ and that $L_y(c_{ij}^* q_i c_{ij}) \geq 0$. Writing $b_j = \sum_w b_{j,w} w$, we find

$$
\begin{aligned}
L_y(b_j^* b_j) &= \sum_{v,w} b_{j,v}^* L_y(v^* w) b_{j,w} \\
&= \sum_{v,w} b_{j,v}^* M_k(y)(v, w) b_{j,w} \geq 0,
\end{aligned}
$$
(29)

where we have used the definition (6) of the moment matrix $M_k(y)$ in the second

equality and the property that $M_k(y) \succeq 0$ to deduce the last inequality. Similarly,

$$L_y(c_{ij}^* q_i c_{ij}) = \sum_{v,w} c_{ij,v}^* \sum_u q_{i,u} L_y(v^* uw) c_{ij,w}$$

(30)
$$= \sum_{v,w} c_{ij,v}^* M_k(q_i y)(v,w) c_{ij,w} \geq 0 \,,$$

where we have used the definition (7) of the localizing matrix $M_k(q_i y)$ and the property $M_k(q_i y) \succeq 0$.

So far, we thus have that $\lambda^k \leq p^k \leq p^\star$ for all k. We note now from the definition (4) that for any $\epsilon > 0$, the polynomial $p(X) - (p^\star - \epsilon)$ is strictly positive on $\mathbf{S}_Q$. It then follows from the Positivstellensatz representation theorem of Helton and McCullough[2] [5] that

(31)
$$p - p^\star + \epsilon = \sum_j b_j^* b_j + \sum_i \sum_j c_{ij}^* q_i c_{ij}$$

for some polynomials b_j and c_{ij}. Let $k \geq \max_{ij} \{\deg(b_j), \deg(c_{ij}) + d_i\}$. Then (λ, b_i, c_{ij}) is a feasible solution of (27) with objective value $p^\star - \epsilon$, and therefore $\lambda^k \geq p^\star - \epsilon$. It follows that $p^\star - \epsilon \leq \lambda^k \leq p^k \leq p^\star$, which implies $p^k \to p^\star$ since $\epsilon > 0$ is arbitrary.

We thus have just shown that the convergence of the relaxations $\mathbf{R_k}$ can be proved, alternatively to the proof given in section 3.2, using the Positivstellensatz for noncommutative polynomials. In fact, both proofs are somewhat equivalent, and the proof presented in section 3.2 can itself be viewed as an indirect proof of the Positivstellensatz for noncommutative polynomials. The advantage of the proof given in section 3.2 is that it is more constructive in spirit and it inspired the proof of Theorem 2, where a procedure is given to build an optimizer $(H^\star, X^\star, \phi^\star)$. The proof that we have just given, on the other hand, connects with the fascinating theory of positive polynomials. We see, for instance, that an a priori bound on the maximal degree k necessary in the SOS decomposition (31) would yield information on the speed of convergence of the relaxations $\mathbf{R_k}$.

3.5. Dealing with equality constraints. The problem **P** can contain a set of equality constraints $e_i(X) = 0$ $(i = 1, \ldots, m_e)$, which can be enforced through the pairs of inequalities $e_i(X) \succeq 0$ and $-e_i(X) \succeq 0$. Rather than writing down directly the corresponding relaxations $\mathbf{R}_k$, it can be advantageous to exploit these equalities to reduce the complexity of the problem.

The set of equalities

(32)
$$E = \{e_i \,:\, i = 1, \ldots, m_e\} \subseteq \mathbb{K}[x, x^*]$$

generates the ideal

(33)
$$I = \left\{ \sum_i f_i e_i g_i \,:\, f_i, g_i \in \mathbb{K}[x, x^*] \right\} \,,$$

which is such that any $p \in I$ satisfies $p(X) = 0$ for operators X such that $e_i(X) = 0$ $(i = 1, \ldots, m_e)$. It is therefore sufficient to express every polynomial $p \in \mathbb{K}[x, x^*]$

---

[2]The proof given by Helton and McCullough covers only the case of polynomials with real coefficients, but it is straightforward to generalize it to the complex case (see, for instance, [3]).

modulo I, that is, to work in the quotient ring $\mathbb{K}[x, x^*]/I$. Let B denote a monomial basis for $\mathbb{K}[x, x^*]/I$. Then we need to consider only polynomial expressions of the form $q = \sum_{w \in B} q_w w$ since for every polynomial $p \in \mathbb{K}[x, x^*]$, there exists a unique $q = \sum_{w \in B} q_w w$ such that $p - q \in I$. It is readily seen that all the results presented so far still hold when we work at relaxation step k with the reduced monomial basis $B_k = B \cap W_k$. The relaxation $\mathbf{R_k}$ then corresponds to an optimization over the set of variables $(y_w)_{w \in B_{2k}}$ and involves matrices $M_k(y)$ and $M_{k-d_i}(g_i y)$ of sizes $|B_k| \times |B_k|$ and $|B_{k-d_i}| \times |B_{k-d_i}|$, respectively. This represents a reduction in the complexity of the original problem.

All of the problem, of course, consists in building a monomial basis B for the quotient ring $\mathbb{K}[x, x^*]/I$. This can be done, e.g., if a finite Gröbner basis exists and can be computed efficiently for the ideal I [17]. Below we give two examples where such a reduced monomial basis B is readily obtained.

3.5.1. Hermitian variables. Polynomials in Hermitian variables are elements of the free $*$-algebra $\mathbb{K}[x]$ with generators $x = (x_1, \ldots, x_n)$ and anti-involution $*$ defined on letters as $x_i^* = x_i$. Our previous results carry over to this situation if words are now viewed as built on the n letters $x_1, \ldots, x_n$ rather than the $2n$ letters $x_1, \ldots, x_n, x_{n+1}, \ldots, x_{2n}$ and if the anti-involution $*$ is reinterpreted accordingly. Since the algebra is now based on n generators, the set of words of length d has $|\mathcal{W}_d| = (n^{k+1} - 1)/(n - 1)$ elements, compared to $((2n)^{k+1} - 1)/(2n - 1)$ for the general case in $2n$ variables. The size of the optimization variables y and the dimension of the moment and localizing matrices in the SDP problem $\mathbf{R_k}$ are reduced accordingly.

3.5.2. Commuting variables and link with Lasserre's results. The method that we have presented to solve optimization problems in noncommuting variables also contains, as a particular case, the commutative version (1) considered by Lasserre since constraints of the type $X_i X_j - X_j X_i = 0$ can explicitly be imposed on the operators X_i. More precisely, the problem

$$
(34) \quad
\begin{aligned}
p^c = \min_{(H, X, \phi)} \quad & \langle \phi, p(X) \phi \rangle \\
\text{s.t.} \quad & q_i(X) \succeq 0, && i = 1, \ldots, m, \\
& X_i X_j - X_j X_i = 0, && i, j = 1, \ldots, n,
\end{aligned}
$$

where the variables X_i are assumed to be Hermitian and all polynomials are expressed in terms of real coefficients, is identical to (1). To show that (34) and (1) are equivalent, note that the operators X in any feasible solution (H, X, ϕ) of (34) generate an Abelian algebra. Hence the Hilbert space H (or at least the part of H on which the operators X and the state ϕ have support) is isomorphic to a direct integral $\int^{\oplus} H_x \, d\mu(x)$ of one-dimensional Hilbert spaces H_x, and the operators X_i are decomposable as $X_i = \int^{\oplus} x_i \, d\mu(x)$, where each x_i is a scalar operator that acts only on H_x [27]. A priori, any point $x \in \mathbb{R}^n$ defines a possible n-uple of operators $(x_1, \ldots, x_n)$ and can be associated with a factor H_x, but to satisfy (34) the measure $d\mu(x)$ should be such that $\int_S d\mu(x) = 1$ and $\int_{\mathbb{R}^n \setminus S} d\mu(x) = 0$, where $S = \{x \in \mathbb{R}^n : q_i(x) \geq 0, \ i = 1, \ldots, m\}$. Thus (34) is equivalent to

$$
(35) \quad
\begin{aligned}
p^c = \min_{\mu} \quad & \int p(x) d\mu(x) \\
\text{s.t.} \quad & \int_S d\mu(x) = 1, \ \int_{\mathbb{R}^n \setminus S} d\mu(x) = 0,
\end{aligned}
$$

where the minimum is taken over all measures μ on $\mathbb{R}^n$. As shown by Lasserre [12], the problems (35) and (1) are equivalent. Indeed, as $p(x) \geq p^\star$ on S, $\int p d\mu \geq p^\star$ and thus $p^c \geq p^\star$. On the other hand, if $x^\star$ is a global minimizer of (1), then the measure $\mu^\star = \delta_{x^\star}$ is admissible for (35), and thus $p^c \leq p^\star$.

The relaxations $\mathbf{R_k}$ are constructed on the canonical basis of noncommutative monomials, for instance, for $n = 2$, $\mathcal{W}_2 = \{1, x_1, x_2, x_1^2, x_1 x_2, x_2 x_1, x_2^2\}$. Simplifying these relaxations using the constraints $x_i x_j - x_j x_i = 0$ amounts to considering only the canonical basis of commutative monomials, e.g., $\mathcal{W}_2^c = \{1, x_1, x_2, x_1^2, x_1 x_2, x_2^2\}$, which leads to the exact same construction as the one introduced by Lasserre. In particular, the criterion for detecting global optimality presented in section 3.3 coincides with the detection criterion introduced in the commutative situation [7]. If we apply the procedure outlined in the proof of Theorem 2 to extract optimal solutions from the solution of a finite order relaxation $\mathbf{R_k}$, we end up with a set of operators $X = (X_1, \ldots, X_n)$ which are matrices each of dimension $r = \operatorname{rank} M_k(y^k)$. As these matrices all commute, they can be simultaneously diagonalized, with each set of common eigenvalues $(x_1(j), \ldots, x_n(j))$ $(j = 1, \ldots, r)$ corresponding to one optimal solution of (1). We thus see that if the rank of the moment matrix $r = \operatorname{rank} M_k(y^{(k)})$ is related to the Hilbert space dimension of the global optimal solution in the noncommutative case, it is related to the number of global solutions extracted by the algorithm in the commutative case.

It is interesting to note that most of our results, such as the convergence of the hierarchy or the criterion to detect optimality, are easier to establish in the general noncommutative framework than they are in the specialized commutative case. Note also that it may be easier, from a computational point of view, to solve the noncommutative version of a problem than it is to solve the commutative one. In particular, the speed of convergence of the SDP relaxations may be faster in the noncommutative case than in the commutative one. This is dramatically illustrated in the following example.

Let p be a polynomial of degree 2, and consider the quadratic problem

$$(36) \qquad \begin{aligned} p^\star = \min_{(H, X, \phi)} & \quad \langle \phi, p(X)\phi \rangle \\ \text{s.t.} & \quad X_i^2 - X_i = 0, \qquad i = 1, \ldots, n \,, \end{aligned}$$

where the variables X_i are assumed to be Hermitian. Its first order relaxation is

$$(37) \qquad \begin{aligned} p^1 = \min_y & \quad \sum_\alpha p_w y_w \\ \text{s.t.} & \quad y_1 = 1, \\ & \quad M_1(y) \succeq 0, \\ & \quad y_{ii} - y_i = 0, \quad i = 1, \ldots, n \,. \end{aligned}$$

Any feasible point y of the above SDP problem with objective value $p(y) = \sum_w p_w y_w$ defines a feasible point of (36) with objective value $\langle \phi, p(X)\phi \rangle = p(y)$, and therefore $p^1 = p^\star$, i.e., the first order relaxation already yields the global optimum of the original problem. To see this, perform a Gram decomposition of the matrix $M_1(y)$: $M_1(y)(v, w) = y_{vw} = \langle \overline{v}, \overline{w} \rangle$, where $|v|, |w| \leq 1$, i.e., $v, w \in \{1, x_1, \ldots, x_n\}$. Define the vector $\phi = \overline{1}$, which is normalized since $\langle \overline{1}, \overline{1} \rangle = y_1 = 1$, and the operator X_i $(i = 1, \ldots, n)$ as the projector on $\overline{x_i}$. Obviously, $X_i^2 = X_i$. Moreover, $X_i\phi = X_i\overline{x_i} + X_i(\phi - \overline{x_i}) = \overline{x_i}$, where the last equality follows from the fact that $\langle \overline{x_i}, \phi - \overline{x_i} \rangle = y_i - y_{ii} = 0$. This implies that $y_{vw} = \langle \phi, v(X)w(X)\phi \rangle$ for $v, w \in \{1, x_1, \ldots, x_n\}$ and

therefore that $p(y) = \langle \phi, p(X)\phi \rangle$ since p is of degree 2. Using similar arguments, one can actually show that the minimization of a polynomial of arbitrary degree evaluated over projection operators can always be determined from the first relaxation of the problem.

The commutative version of (36) is the quadratically constrained quadratic program

$$
(38) \qquad p^\star = \min_{x \in \mathbb{R}^n} \quad p(x) \\
\text{s.t.} \quad x_i^2 - x_i = 0, \qquad i = 1, \ldots, n \, .
$$

Since 0-1 integer programming can be formulated in this form, it is NP-hard to solve a general instance of (38). Thus, contrary to the noncommutative case, it is highly unlikely that considering relaxations up to some bounded order might be sufficient to solve this problem.

3.6. Generalization. In this subsection, we introduce a slight generalization of the problem (4) to which our method readily extends. We state the results without entering in the details of the proofs.

In addition to the polynomials p and $\{q_i \, : \, i = 1, \ldots, m_q\}$ defined in (4), consider the sets of polynomials $\{r_i \, : \, i = 1, \ldots, m_r\}$ and $\{s_i \, : \, i = 1, \ldots, m_s\}$, where the s_i's are Hermitian. The problem that we consider is

$$
(39) \qquad \mathbf{\tilde{P}} : \qquad
\begin{aligned}
\tilde{p}^\star = \min_{(H,X,\phi)} \quad & \langle \phi, p(X)\phi \rangle \\
\text{s.t.} \quad & q_i(X) \succeq 0, & i = 1, \ldots, m_q \, , \\
& r_i(X)\phi = 0, & i = 1, \ldots, m_r \, , \\
& \langle \phi, s_i(X)\phi \rangle \geq 0, & i = 1, \ldots, m_s \, .
\end{aligned}
$$

We thus not only require that the operators X satisfy $q_i(X) \succeq 0$, but we also require that $r_i(X)$ acting on ϕ yield the null vector and that the average value of $s_i(X)$ be positive. As before, we assume that $Q = \{q_i \, : \, i = 1, \ldots, m_q\}$ is such that the quadratic module $\mathbf{M}_Q$ is Archimedean.

For $r \in \mathbb{K}[x, x^*]_d$ and $y = \{y_w\}_{|w| \leq k+d}$, a sequence indexed in $\mathcal{W}_{k+d}$, define the vector $\mathrm{m}_k(ry)$ as the vector with components indexed in $\mathcal{W}_k$ and whose component w is equal to

$$
(40) \qquad \mathrm{m}_k(ry)(w) = L_y(wr) = \sum_{|v| \leq d} r_v y_{wv} \, .
$$

If y admits a moment representation (8) such that $r(X)\phi = 0$, then $\mathrm{m}_k(ry) = 0$ since

$$
(41) \qquad \mathrm{m}_k(fy)(w) = \sum_v r_v y_{wv} = \sum_v r_v \langle \phi, w(X)v(X)\phi \rangle = \langle \phi, w(X)r(X)\phi \rangle = 0.
$$

If, in addition, y admits a moment representation such that $\langle \phi, s(X)\phi \rangle \geq 0$, then obviously $\sum_w s_w y_w \geq 0$. These observations motivate the following definition.

For $2k \geq \max \{\deg(p), \deg(q_i), \deg(r_i), \deg(s_i)\}$, we define the relaxation of order k associated to the problem $\mathbf{\tilde{P}}$ as the SDP problem

$$
(42) \qquad \mathbf{\tilde{R}_k} : \qquad
\begin{aligned}
\tilde{p}^k = \min_{y} \quad & \sum_w p_w y_w \\
\text{s.t.} \quad & M_k(y) \succeq 0, \\
& y_1 = 1, \\
& M_{k-d_i}(q_i y) \succeq 0, & i = 1, \ldots, m_q, \\
& \mathrm{m}_{2k-d_i'}(r_i y) = 0, & i = 1, \ldots, m_r, \\
& \sum_w s_{i,w} y_w \geq 0, & i = 1, \ldots, m_s,
\end{aligned}
$$

where $d_i = \lceil \deg(q_i)/2 \rceil$, $d'_i = \deg(r_i)$, and the optimization is over $y \in \mathbb{K}^{|\mathcal{W}_{2k}|}$. It is easily verified that $\tilde{p}^k \geq \tilde{p}^{k'}$ when $k \leq k'$ and that $\tilde{p}^k \leq \tilde{p}^\star$ for all k.

The results obtained in sections 3.2 and 3.3 can easily be adapted to the above situation.

THEOREM 3. *If* $\mathbf{M}_Q$ *is Archimedean,* $\lim_{k\to\infty} \tilde{p}^k = \tilde{p}^\star$.

THEOREM 4. *Assume that the optimal solution* y^k *of the relaxation* $\tilde{\mathbf{R}}_\mathbf{k}$ *of order* k *satisfies*

$$\text{(43)} \qquad\qquad \operatorname{rank} M_k(y^k) = \operatorname{rank} M_{k-d}(y^k),$$

where $d = \max_i d_i \geq 1$ *and*

$$\text{(44)} \qquad\qquad d'_i - d \leq k$$

for all $i = 1, \ldots, m_r$. *Then* $\tilde{p}^k = \tilde{p}^*$, *i.e., the optimum of the relaxation of order* k *is the global optimum of the original problem* $\tilde{\mathbf{P}}$. *Moreover, there exists a global optimizer* $(H^\star, X^\star, \phi^\star)$ *of* $\tilde{\mathbf{P}}$ *with* $\dim H^\star = \operatorname{rank} M_{k-d}(y^k)$.

The proof of both of these theorems follows along the same line as the proofs of Theorems 1 and 2, respectively. One has simply to show that the reconstructed operators $\hat{X}$ and the state $\hat{\phi}$ satisfy the additional properties $r_i(\hat{X})\hat{\phi} = 0$ and $\langle \hat{\phi}, s_i(\hat{X})\hat{\phi} \rangle \geq 0$. This can be established given the conditions $\mathrm{m}_{2k-d'_i}(r_i y) = 0$ and $\sum_w s_{i,w} y_w \geq 0$ present in $\tilde{\mathbf{R}}_\mathbf{k}$. The additional constraint (44) with respect to Theorem 2 comes from the fact that to show that $r_i(X)\phi = 0$, we need to show, because of (23), that $\langle \overline{w}, r_i(X)\phi \rangle = 0$ for all $|w| \leq k - d$. This is implied by $\mathrm{m}_{2k-d'_i}(r_i y) = 0$ when $2k - d'_i \geq k - d$, i.e., when (44) is satisfied.

The duals of the relaxations $\tilde{\mathbf{R}}_\mathbf{k}$ can be shown to be equivalent to the problems

(45)
$$\tilde{\lambda}^k = \max_{\lambda, b_i, c_{ij}, f_i, g_i} \quad \lambda$$
$$\text{s.t.} \qquad p - \lambda = \sum_j b_j^* b_j + \sum_{i=1}^{m_q} \sum_j c_{ij}^* q_i c_{ij} + \sum_{i=1}^{m_r} (f_i r_i + r_i^* f_i^*) + \sum_{i=1}^{m_s} g_i s_i,$$
$$\max_j \deg(b_j) \leq k,$$
$$\max_j \deg(c_{ij}) \leq k - d_i,$$
$$\deg(f_i) \leq 2k - d'_i,$$
$$g_i \geq 0,$$

where b_j, c_{ij}, f_i are polynomials and g_i are real numbers. From the decomposition of $p - \tilde{\lambda}^k$ appearing in (45), it clearly follows that $p(X) - \tilde{\lambda}^k \geq 0$ for any (H, X, ϕ) satisfying the constraints in $\tilde{\mathbf{P}}$. Thus the solution of the dual (45) provides a certificate that the optimal solution $\tilde{p}^\star$ of $\tilde{\mathbf{P}}$ cannot be lower than $\tilde{\lambda}^k$.

Finally, we mention that it is possible, taking inspiration from [8], to generalize the problem (4) and the results associated to it to the case of matrix-valued polynomials, that is, polynomials $\sum_w p_w w$, where each coefficient p_w is now an $a \times b$ matrix with entries from $\mathbb{K}$. A Positivstellensatz also exists in this case [5].

4. Illustration of the method. For the sake of illustration, we now apply our approach on simple examples. To simplify the notation, through all this section we label monomials (i.e., words) by the indices of the ordered noncommutative variables of which they are composed. For instance, the word $w = x_2 x_1 x_2 x_2$ will be referred to

as 2122. The empty word 1 corresponding to the identity element of the algebra will be labeled by the symbol $\emptyset$.

Our first example involves two Hermitian variables $X_1 = X_1^*$ and $X_2 = X_2^*$ and has the form

$$
\begin{aligned}
p^\star = \min_{(H,X,\phi)} \quad & \langle \phi, X_1 X_2 + X_2 X_1 \phi \rangle \\
\text{s.t.} \quad & X_1^2 - X_1 = 0, \\
& -X_2^2 + X_2 + 1/2 \succeq 0 .
\end{aligned}
\tag{46}
$$

Since all constraint and objective variables are at most of degree 2, the first order relaxation is associated with the monomial basis $\mathcal{W}_2 = \{1, x_1, x_2, x_1 x_2, x_2 x_1, x_2^2\}$, where, following the approach of section 3.5, we used that $x_1^2 = x_1$. The first relaxation step thus involves the relaxed variables $\{y_\emptyset, y_1, y_2, y_{12}, y_{21}, y_{22}\}$ and corresponds to the SDP problem

$$
\begin{aligned}
p^1 = \min_y \quad & y_{12} + y_{21} \\
\text{s.t.} \quad & \begin{bmatrix} 1 & y_1 & y_2 \\ \hline y_1 & y_1 & y_{12} \\ y_2 & y_{21} & y_{22} \end{bmatrix} \succeq 0, \\
& -y_{22} + y_2 + 1/2 \geq 0.
\end{aligned}
\tag{47}
$$

We solved this SDP problem using the Matlab toolboxes YALMIP [14] and SeDuMi [25]. After rounding, we obtain the solution $p^1 = -3/4$, achieved for the moment matrix

$$
M_1 = \begin{bmatrix} 1 & 3/4 & -1/4 \\ \hline 3/4 & 3/4 & -3/8 \\ -1/4 & -3/8 & 1/4 \end{bmatrix}
\tag{48}
$$

with eigenvalues 0, $1 \pm \sqrt{37}/8$. The second order relaxation is

(49)
$$
p^2 = \min_y \quad y_{12} + y_{21}
$$

$$
\text{s.t.} \quad \begin{bmatrix} 1 & y_1 & y_2 & y_{12} & y_{21} & y_{22} \\ \hline y_1 & y_1 & y_{12} & y_{12} & y_{121} & y_{122} \\ y_2 & y_{21} & y_{22} & y_{212} & y_{221} & y_{222} \\ \hline y_{21} & y_{21} & y_{212} & y_{212} & y_{2121} & y_{2122} \\ y_{12} & y_{121} & y_{122} & y_{1212} & y_{1221} & y_{1222} \\ y_{22} & y_{221} & y_{222} & y_{2212} & y_{2221} & y_{2222} \end{bmatrix} \succeq 0,
$$

$$
\begin{bmatrix} -y_{22} + y_2 + \frac{1}{2} & -y_{221} + y_{21} + \frac{1}{2} y_1 & -y_{222} + y_{22} + \frac{1}{2} y_2 \\ \hline -y_{221} + y_{21} + \frac{1}{2} y_1 & -y_{1221} + y_{121} + \frac{1}{2} y_1 & -y_{1222} + y_{122} + \frac{1}{2} y_{12} \\ -y_{222} + y_{22} + \frac{1}{2} y_2 & -y_{1222} + y_{122} + \frac{1}{2} y_{12} & -y_{2222} + y_{222} + \frac{1}{2} y_{22} \end{bmatrix} \succeq 0
$$

with solution $p^2 = -3/4$. The moment matrix associated to this solution is

$$
M_2 = \begin{bmatrix} 1 & 3/4 & -1/4 & -3/8 & -3/8 & 1/4 \\ \hline 3/4 & 3/4 & -3/8 & -3/8 & -3/16 & 0 \\ -1/4 & -3/8 & 1/4 & 3/16 & 0 & 1/8 \\ \hline -3/8 & -3/8 & 3/16 & 3/16 & 3/32 & 0 \\ -3/8 & -3/16 & 0 & 3/32 & 3/16 & -3/16 \\ 1/4 & 0 & 1/8 & 0 & -3/16 & 1/4 \end{bmatrix},
\tag{50}
$$

which as two nonzero eigenvalues $3/32 \times (14 \pm \sqrt{61})$.

Optimality criterion and extraction of optimizers. Since the matrix M_2 has two nonzero eigenvalues, it has rank 2. Let $M_1(y^2)$ be the upper right 3×3 submatrix of $M_2 = M_2(y^2)$. This submatrix is, in fact, equal to (48) and thus also has rank 2. The matrices $M_1(y^2)$ and $M_2(y^2)$ have thus the same rank, and condition (22) of Theorem 2 is satisfied. It follows that $p^\star = p^2 = -3/4$. It also follows that we can extract a global optimizer for (47), which will be realized in a space of dimension 2. For this, write down the Gram decomposition $M_2 = R^T R$, where

$$(51) \qquad R = \left[\begin{array}{cccccc} 1 & 3/4 & -1/4 & -3/8 & -3/8 & 1/4 \\ 0 & \sqrt{3}/4 & -\sqrt{3}/4 & -\sqrt{3}/8 & \sqrt{3}/8 & -\sqrt{3}/4 \end{array} \right] .$$

Following the procedure specified in the proof of Theorem 2, we find the optimal solutions

$$(52) \quad X_1^\star = \left[\begin{array}{cc} 3/4 & \sqrt{3}/4 \\ \sqrt{3}/4 & 1/4 \end{array} \right] , \qquad X_2^\star = \left[\begin{array}{cc} -1/4 & -\sqrt{3}/4 \\ -\sqrt{3}/4 & 5/4 \end{array} \right] , \qquad \phi^\star = \left[\begin{array}{c} 1 \\ 0 \end{array} \right] .$$

Dual. Solving the dual of the first order relaxation (47) yields, in the notation of Appendix B, the solutions

$$\lambda = -3/4,$$
$$V = \left[\begin{array}{ccc} 1/4 & -1/2 & -1/2 \\ -1/2 & 1 & 1 \\ -1/2 & 1 & 1 \end{array} \right] ,$$
$$(53) \qquad W = 1 .$$

The matrix V has only one nonzero eigenvalue and can be written as $V = aa^T$, where $a = [-1/2, 1, 1]$. In the formlation of (27), this corresponds to an SOS decomposition for $x_1 x_2 + x_2 x_1$ of the form

$$(54) \qquad x_1 x_2 + x_2 x_1 - \left(-\frac{3}{4} \right) = \left(-\frac{1}{2} + x_1 + x_2 \right)^2 + \left(-x_2^2 + x_2 + \frac{1}{2} \right) .$$

It immediately follows that $\langle \phi, X_1 X_2 + X_2 X_1 \phi \rangle \geq -3/4$ for every (H, X, ϕ) satisfying $X_1^2 = X_1$ and $-X_2^2 + X_2 + \frac{1}{2} \succeq 0$. Thus the decomposition (54) certifies that the solution (52) is optimal.

Comparison with the commutative case. To illustrate the differences and similarities between the noncommutative and commutative cases, let

$$(55) \qquad \begin{aligned} p^\star = \min_{x \in \mathbb{R}^2} \quad & 2x_1 x_2 \\ \text{s.t.} \quad & x_1^2 - x_1 = 0, \\ & -x_2^2 + x_2 + 1/2 \geq 0 \end{aligned}$$

be the commutative version of (46). The first relaxation step associated to this problem involves the monomial basis $\mathcal{W}_2^c = \{1, x_1, x_2, x_1 x_2, x_2^2\}$ (we used $x_1 x_2 = x_2 x_1$) and the corresponding relaxation variables $\{y_\emptyset, y_1, y_2, y_{12}, y_{22}\}$. This should be compared to $\mathcal{W}_2 = \{1, x_1, x_2, x_1 x_2, x_2 x_1 x_2^2\}$ and $\{y_\emptyset, y_1, y_2, y_{12}, y_{21}, y_{22}\}$ in the noncommutative case. The first order relaxation associated to (55) is thus

$$(56) \qquad \begin{aligned} p^1 = \min_y \quad & 2\, y_{12} \\ \text{s.t.} \quad & \left[\begin{array}{c|cc} 1 & y_1 & y_2 \\ \hline y_1 & y_1 & y_{12} \\ y_2 & y_{12} & y_{22} \end{array} \right] \succeq 0, \\ & -y_{22} + y_2 + 1/2 \geq 0. \end{aligned}$$

Note that (47) and (56) are, in fact, identical because the hermicity of the moment matrix in (47) implies that $y_{12} = y_{21}$. In general, it always happens that the first order relaxations of the commutative and noncommutative versions of a problem coincide. We thus find as before that $p^1 = -3/4$. The relaxation of order two of (56), however, is

(57)
$$p^2 = \min_y \quad 2y_{12}$$

$$\text{s.t.} \quad \begin{bmatrix} 1 & y_1 & y_2 & y_{12} & y_{22} \\ y_1 & y_1 & y_{12} & y_{12} & y_{122} \\ y_2 & y_{12} & y_{22} & y_{122} & y_{222} \\ y_{12} & y_{12} & y_{122} & y_{122} & y_{1222} \\ y_{22} & y_{122} & y_{222} & y_{1222} & y_{2222} \end{bmatrix} \succeq 0,$$

$$\begin{bmatrix} -y_{22} + y_2 + \frac{1}{2} & -y_{122} + y_{12} + \frac{1}{2}y_1 & -y_{222} + y_{22} + \frac{1}{2}y_2 \\ -y_{122} + y_{12} + \frac{1}{2}y_1 & -y_{122} + y_{12} + \frac{1}{2}y_1 & -y_{1222} + y_{122} + \frac{1}{2}y_{12} \\ -y_{222} + y_{22} + \frac{1}{2}y_2 & -y_{1222} + y_{122} + \frac{1}{2}y_{12} & -y_{2222} + y_{222} + \frac{1}{2}y_{22} \end{bmatrix} \succeq 0 .$$

Solving it, we obtain $p^2 = 1 - \sqrt{3} \simeq -0.7321$. Again, it can be verified that the rank condition (22) of Theorem 2 is satisfied, so that this solution is optimal, and the following optimizer can be reconstructed:

(58)
$$x_1^\star = 1, \qquad x_2^\star = (1 - \sqrt{3})/2 .$$

As expected, the global minimum of (55) is higher than the one of (46) as the commutative case is more constrained than the noncommutative one.

Additional constraints. We now consider a problem of the form (39) by adding two constraints to (46):

(59)
$$\begin{aligned} p^\star = \min_{(H,X,\phi)} \quad & \langle \phi, X_1 X_2 + X_2 X_1 \phi \rangle \\ \text{s.t.} \quad & X_1^2 - X_1 = 0, \\ & -X_2^2 + X_2 + 1/2 \succeq 0, \\ & (3X_1 + 2X_2 - 1)\,\phi = 0, \\ & -\langle \phi, X_1 \phi \rangle + 1/3 \geq 0 . \end{aligned}$$

Following (42), the corresponding first order relaxation is

(60)
$$\begin{aligned} p^1 = \min_y \quad & y_{12} + y_{21} \\ \text{s.t.} \quad & \begin{bmatrix} 1 & y_1 & y_2 \\ y_1 & y_1 & y_{12} \\ y_2 & y_{21} & y_{22} \end{bmatrix} \succeq 0, \\ & -y_{22} + y_2 + 1/2 \geq 0, \\ & 3y_\alpha + 2y_v - y_u = 0, \qquad (w, v, u) \in J_1, \\ & -y_1 + 1/3 \geq 0 , \end{aligned}$$

where $J_1 = \{(1, 2, \emptyset), (1, 12, 1), (2, 22, 2)\}$. This problem admits the solution $p^1 = -2/3$, achieved for the moment matrix

(61)
$$M_1 = \begin{bmatrix} 1 & 1/3 & 0 \\ 1/3 & 1/3 & -1/3 \\ 0 & -1/3 & 1/2 \end{bmatrix}$$

with eigenvalues 0, 2/3, and 7/6. The solution $p^1 = -2/3$ thus yields a lower bound on $p^\star$, which is already higher, as expected, than the optimal solution of (46). The second order relaxation is

$$p^2 = \min_y \ y_{12} + y_{21}$$

$$(62) \ \text{s.t.} \ \begin{bmatrix} 1 & y_1 & y_2 & y_{12} & y_{21} & y_{22} \\ y_1 & y_1 & y_{12} & y_{12} & y_{121} & y_{122} \\ y_2 & y_{21} & y_{22} & y_{212} & y_{221} & y_{222} \\ \hline y_{21} & y_{21} & y_{212} & y_{212} & y_{2121} & y_{2122} \\ y_{12} & y_{121} & y_{122} & y_{1212} & y_{1221} & y_{1222} \\ y_{22} & y_{221} & y_{222} & y_{2212} & y_{2221} & y_{2222} \end{bmatrix} \succeq 0,$$

$$\begin{bmatrix} -y_{22} + y_2 + \frac{1}{2} & -y_{221} + y_{21} + \frac{1}{2}y_1 & -y_{222} + y_{22} + \frac{1}{2}y_2 \\ \hline -y_{221} + y_{21} + \frac{1}{2}y_1 & -y_{1221} + y_{121} + \frac{1}{2}y_1 & -y_{1222} + y_{122} + \frac{1}{2}y_{12} \\ -y_{222} + y_{22} + \frac{1}{2}y_2 & -y_{1222} + y_{122} + \frac{1}{2}y_{12} & -y_{2222} + y_{222} + \frac{1}{2}y_{22} \end{bmatrix} \succeq 0,$$

$$3y_w + 2y_v - y_u = 0, \qquad (w, v, u) \in J_2,$$

$$-y_1 + \frac{1}{3} \geq 0,$$

where $J_2 = \{(1, 2, \emptyset), (1, 12, 1), (21, 22, 2), (121, 122, 12), (21, 22, 21), (221, 222, 22),$
$(121, 1212, 121), (1221, 1222, 122), (2121, 2122, 212), (221, 2212, 221), (2221, 2222, 222)\}$.
It admits the solution $p^2 = -2/3$ with

$$(63) \qquad M_2 = \begin{bmatrix} 1 & 1/3 & 0 & -1/3 & -1/3 & 1/2 \\ \hline 1/3 & 1/3 & -1/3 & -1/3 & 0 & -1/6 \\ 0 & -1/3 & 1/2 & 1/3 & -1/6 & 1/2 \\ \hline -1/3 & -1/3 & 1/3 & 1/3 & 0 & 1/6 \\ -1/3 & 0 & -1/6 & -0 & 1/6 & -1/3 \\ 1/2 & -1/6 & 1/2 & 1/6 & -1/3 & 3/4 \end{bmatrix},$$

which has two nonzero eigenvalues 17/12 and 5/3. As in the previous examples, it is easily verified that the rank condition (22) is satisfied, and we thus deduce that $p^\star = p^2 = -2/3$. From the Gram decomposition $M_2 = R^T R$, with

$$(64) \qquad R = \begin{bmatrix} 1 & 1/3 & 0 & -1/3 & -1/3 & 1/2 \\ 0 & \sqrt{2}/3 & -\sqrt{2}/2 & -\sqrt{2}/3 & \sqrt{2}/6 & -\sqrt{2}/2 \end{bmatrix},$$

one obtains the global optimizer

$$(65) \quad X_1^\star = \begin{bmatrix} 1/3 & \sqrt{2}/3 \\ \sqrt{2}/3 & 2/3 \end{bmatrix}, \qquad X_2^\star = \begin{bmatrix} 0 & -\sqrt{2}/2 \\ -\sqrt{2}/2 & 1 \end{bmatrix}, \quad \phi = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Finally, the dual of the first order relaxation (60) yields the SOS decomposition

$$x_1 x_2 + x_2 x_1 - \left(-\frac{2}{3}\right) = \frac{1}{9}\left(-1 + 3x_1 + 2x_2\right)^2 + \frac{4}{9}\left(-x_2^2 + x_2 + \frac{1}{2}\right) + \left(\frac{1}{3} - x_1\right)$$

$$(66) \qquad \qquad + \frac{1}{6}x_1\left(3x_1 + 2x_2 - 1\right) + \frac{1}{6}\left(3x_1 + 2x_2 - 1\right)x_1,$$

which clearly implies $p^\star \geq -2/3$.

5. Applications. The results presented so far have immediate applications in quantum theory and quantum information science. Since the dimension of the underlying Hilbert space is not specified in optimization problems (4) or (39), they are well adapted to situations where we want to optimize a quantity over all its possible physical realizations, that is to say, over Hilbert spaces of arbitrary dimensions. Computing the maximal quantum violation of a Bell inequality is an example of this sort.

Let $S_1, \ldots, S_N$ be a collection of finite disjoint sets. Each S_k represents a measurement that can be performed on a given system, and the elements $i \in S_k$ are the possible outcomes of the k-measurement. We suppose that the system is composed of two noninteracting subsystems and that measurements $S_1, \ldots, S_n$ are performed on the first system and measurements $S_{n+1}, \ldots, S_N$ on the second. We put $A = S_1 \cup \ldots \cup S_n$ and $B = S_{n+1} \cup \ldots \cup S_N$, and we denote by $P(ij)$ as the joint probability to obtain outcome $i \in A$ and outcome $j \in B$ when measurements associated to these outcomes are made on the first and second subsystems, respectively. In quantum theory, these probabilities are given by $P(ij) = \langle \phi, E_i E_j \phi \rangle$, where ϕ describes the state of the system under observation and the self-adjoint operators E_i describe the measurements performed on ϕ. The measurement operators $\{E_i : i \in S_k\}$ associated to the measurement S_k form an orthogonal resolution of the identity, and operators corresponding to different subsystems commute, i.e., $[E_i, E_j] = 0$ when $i \in A$ and $j \in B$.

For our purposes, a Bell inequality is simply a linear expression $\sum_{ij} c_{ij} P(ij)$ in the joint probabilities. We are interested in the maximal value that this quantity can take over all probabilities $P(ij)$ that admit a quantum representation. This amounts to solving the problem

$$
\min_{(H,E,\phi)} \quad \left\langle \phi, \sum_{ij} c_{ij} E_i E_j \phi \right\rangle
$$

$$
\text{s.t.} \quad E_i E_j = \delta_{ij} E_i \quad \text{for all } S_k \text{ and for all } i, j \in S_k,
$$

(67)
$$
\sum_{i \in S_k} E_i = 1 \quad \text{for all } S_k,
$$

$$
[E_i, E_j] = 0 \quad \text{for all } i \in A \text{ and for all } j \in B,
$$

which is a particular instance of the noncommutative optimization problem (4) and involves polynomials of degree at most 2. Note that $1 - \sum_{i \in \mathcal{S}_k} E_i^2 = (1 - \sum_{i \in \mathcal{S}_k} E_i) + \sum_{i \in \mathcal{S}_k} (E_i - E_i^2) = 0$, and thus the quadratic module associated to the constraints in (67) is Archimedean. The sequence of SDP relaxations associated to (67) thus converges to the optimal solution. This particular sequence of SDP relaxations is the one already introduced in [18, 19] and the source of inspiration for the present work. It represents the unique tool that is currently available to compute the maximal violation of a generic Bell inequality. It has been applied up to the third order in [20] to derive upper bounds on the maximal violation of 241 Bell inequalities. The resulting upper bounds are tight for all but 20 of these inequalities; for the remaining 20 inequalities the gap between the upper bound and the best known lower bound is small.

The sequence of SDP relaxations introduced here can also be used to decide if a given set of probabilities $P(ij)$ admits a quantum representation [18, 19]. More generally, it has the potential to find other applications in quantum information science (see, for instance, [3, 9, 22]).

Besides applications where the dimension of the underlying Hilbert space is not fixed, optimization problems (4) and (39) are also well suited to problems where the

Hilbert space is the unique irreducible representation space of a set of operators satisfying algebraic constraints. Consider, for instance, a system of N electrons that can occupy M orbitals, each orbital being associated with annihilation and creation operators a_i and $a_i^\dagger$, $i = 1, \ldots, M$. (We use the common physics notation $\dagger$ for the conjugate transpose.) Since electrons interact pairwise, the Hamiltonian for such a system involves only two-body interactions, and its ground-state energy can be computed as

$$
\begin{aligned}
\min \quad & \left\langle \phi, \sum_{ijkl} h_{ijkl} a_i^\dagger a_j^\dagger a_k a_l \, \phi \right\rangle \\
\text{s.t.} \quad & \{a_i, a_j\} = 0, \\
& \{a_i^\dagger, a_j^\dagger\} = 0, \\
& \{a_i^\dagger, a_j\} = \delta_{ij}, \\
& \left(\sum_i a_i^\dagger a_i - N \right) \phi = 0 \,.
\end{aligned}
$$

(68)

The first three constraints represent the usual anticommutation fermionic relations, while the last constraint fixes the number of electrons to N. This problem is a particular case of (39), and it involves polynomials of degree 4. Note that the algebra of operators generated by (68) has a unique irreducible representation of dimension 2^M. Since a product (in normal order) of more than N of the operators $\{a_i, a_i^\dagger\}$ vanishes, the sequence of SDP relaxations halts at order N, and thus $p^N = p^\star$.

The hierarchy of SDP relaxations associated to problem (68) can be used, for instance, to compute the ground-state electronic energy of atoms or molecules. In the last years, very successful SDP methods based on the N-representability problem have been independently introduced in quantum chemistry to compute these electronic energies [15, 16]. Our hierarchy of SDP relaxations actually reduces to these existing SDP techniques. But our approach is more general and can be used to compute the ground-state energy of other many-body systems, such as spin systems or systems described by *unbounded* operators satisfying the canonical relations $[x, p] = i$ (in which case it has to be slightly adapted). These applications will be presented in a forthcoming paper.

Finally, the method presented here might also prove useful for problems where the Hilbert space dimension is fixed in advance. Consider, for instance, a polynomial optimization problem of the form (4), where $\dim H = r$, i.e, where the operators X are $r \times r$ matrices. We may, in principle, solve such a problem by introducing an explicit parameterization of the matrices X and by using Lasserre's method for polynomial scalar optimization [12] or its extension, taking into account polynomial matrix inequalities [8]. This would necessitate, however, introduction of the order of r^2 scalar variables for each operator X_i. This renders this approach impractical, even for small problems. In comparison, the method presented here treats each matrix as a single variable. Although it represents only a relaxation of the original problem since the Hilbert space dimension is not fixed (in particular, we have no guarantee that the sequence of relaxations will converge to a solution with $\dim H = r$), it may nevertheless provide a cheap way to compute the lower bound on the optimal solutions of these problems when it is too costly to introduce an explicit parameterization.

Appendix A. Basics of semidefinite programming. Semidefinite programming [26] is a subfield of convex optimization concerned with the following optimization problem, known as the *primal problem*:

$$\text{minimize} \quad c^T x$$

$$(69) \qquad \text{s.t.} \quad F(x) = \sum_{i=1}^m x_i F_i - G \succeq 0 \,.$$

The problem variable is the vector x with m components x_i, and the problem parameters are the $n \times n$ matrices G, F_i and the scalars c_i. A vector x is said to be *primal feasible* when $F(x) \geq 0$.

For each primal problem there is an associated *dual problem*, which is a maximization problem of the form

$$\text{maximize} \quad \text{tr}(GZ)$$

$$(70) \qquad \text{s.t.} \quad \text{tr} F_i Z = c_i, \quad i = 1, \ldots, m,$$

$$Z \succeq 0,$$

where the optimization variable is the $n \times n$ matrix Z. The dual problem is also a semidefinite program, i.e., it can be put in the same form as (70). A matrix Z is said to be *dual feasible* if it satisfies the conditions in (70).

The key property of the dual program is that it yields bounds on the optimal value of the primal program. To see this, take a primal feasible point x and a dual feasible point Z. Then $c^T x - \text{tr}(GZ) = \sum_{i=1}^m \text{tr}(ZF_i)x_i - \text{tr}(GZ) = \text{tr}(ZF(x)) \geq 0$. This proves that the optimal primal value p^* and the optimal dual value d^* satisfy $d^* \leq p^*$. In fact, it usually happens that $d^* = p^*$. A sufficient condition for this to hold is that the dual (primal) problem admits a strict feasible point, that is, that there exists a matrix $Z \succ 0$ ($F(x) \succ 0$) that is dual (primal) feasible [26]. We refer the reader to the review of Vandenberghe and Boyd [26] for further information on SDP.

There exist many available numerical packages to solve SDPs, for instance, for Matlab, the toolboxes SeDuMi [25] and YALMIP [14]. These algorithms solve both the primal and the dual at the same time and thus yield bounds on the accuracy of the solution that is obtained.

Appendix B. Duals of the SDP relaxations. Here we show that the duals of the relaxations $\mathbf{R_k}$ defined in (11) correspond to problem (27). To simplify the presentation, we do this explicitly only in the case where we are dealing with polynomials defined in the real free $*-$algebra $\mathbb{R}[x, x^*]$ and where the SDP relaxations (11) involve only real quantities. The more general case of complex SDP relaxations can be treated similarly by decomposing them in real and imaginary parts.

Write $M_k(y) = \sum_w B_w y_w$ and $M_{k-d_i}(q_i y) = \sum_w C_w^i y_w$ for appropriate symmetric matrices B_w and C_w^i. The SDP relaxation (11) is then expressed as an SDP problem in primal form (69), and its dual is

$$\lambda^k = \max_{\lambda, V, W_i} \quad \lambda$$

$$\text{s.t.} \quad p_1 = \lambda + \text{tr}\left(B_\emptyset V\right) + \sum_{i=1}^m \text{tr}(C_\emptyset^i W_i),$$

$$(71) \qquad p_w = \text{tr}\left(B_w V\right) + \sum_{i=1}^m \text{tr}(C_w^i W_i) \qquad \text{for all } 0 < |w| \leq 2k,$$

$$V \succeq 0,$$

$$W_i \succeq 0, \quad i = 1, \ldots, m,$$

where $\lambda \in \mathbb{R}$, $V \in \mathbb{R}^{|\mathcal{W}_k|} \times \mathbb{R}^{|\mathcal{W}_k|}$ and $W_i \in \mathbb{R}^{|\mathcal{W}_{k-d_i}|} \times \mathbb{R}^{|\mathcal{W}_{k-d_i}|}$.

The terms on the left-hand sides of the above equality constraints are the coefficients in the canonical basis of monomials $\mathcal{W}_{2k} = \{w \,:\, |w| \leq 2k\}$ of the polynomial p. The quantities $\mathrm{tr}(B_w V)$ on the right-hand side are the coefficients of a polynomial of the form $\sum_j b_j^* b_j$, where each b_j is a polynomial of degree k. Indeed, it is easily seen from the definition of the moment matrix $M_k(y)$ that the entries of the matrices B_w satisfy $B_w(u,v) = 1$ if $w = u^*v$ or $w = v^*u$ and $B_w(u,v) = 0$ otherwise. It follows that $\sum_{|w| \leq 2k} \mathrm{tr}(B_w V) w = \sum_{|u|,|v| \leq k} V_{vu}\, u^*v$, where we used that V is symmetric. As V is positive semidefinite, we can write $V = \sum_j \mu_j a_j a_j^T$, where $\mu_j \geq 0$ are the eigenvalues of V and a_j the corresponding eigenvectors. Using this expression for V, we obtain that $\sum_{|w| \leq 2k} \mathrm{tr}(B_w V) w = \sum_j \mu_j a_j^* a_j$, which is of the announced form with $b_j = \sqrt{\mu_i} a_j$. In a similar way, it can be shown that $\sum_{|w| \leq 2k} \mathrm{tr}(C_w^i W_i) w = \sum_j c_{ij}^* q_i c_{ij}$. Putting it all together, we find that problem (71) is equivalent to

$$
\begin{aligned}
\lambda^k = \max_{\lambda, b_i, c_{ij}} \quad & \lambda \\
\text{s.t.} \quad & p - \lambda = \sum_j b_j^* b_j + \sum_{i=1}^m \sum_j c_{ij}^* q_i c_{ij}, \\
& \max_j \deg(b_j) \leq k, \\
& \max_j \deg(c_{ij}) \leq k - d_i\,.
\end{aligned}
\tag{72}
$$

In the case of polynomials defined on $\mathbb{C}[x, x^*]$, the dual of (11) has the same form as above, but now all polynomials are allowed to take complex coefficients.

A similar analysis can be carried out to show that problems (42) and (45) are dual to each other.

REFERENCES

[1] K. Cafuta, I. Klep, and J. Povh. *NCSOStools: A computer algebra system for symbolic and numerical computation with noncommutative polynomials*, http://ncsostools.fis.unm.si.

[2] R. E. Curto and L. A. Fialkow, *Solution of the truncated complex moment problem with flat data*, Mem. Amer. Math. Soc. 568, 119 (1996).

[3] A. C. Doherty, Y. C. Liang, B. Toner, and S. Wehner, *The quantum moment problem and bounds on entangled multi-prover games*, in Proceedings of IEEE Conference on Computational Complexity, 2008, pp. 199–210.

[4] J. W. Helton, *"Positive" noncommutative polynomials are sums of squares*, Ann. of Math., 156 (2002), pp. 675–694.

[5] J. W. Helton and S. A. McCullough, *A Positivstellensatz for non-commutative polynomials*, Trans. Amer. Math. Soc., 356 (2004), pp. 3721–3737.

[6] J. W. Helton and M. Putinar, *Positive polynomials in scalar and matrix variables, the spectral theorem, and optimization*, in Operator Theory, Structured Matrices, and Dilations: Tiberiu Constantinescu Memorial Volume, International Book Series of Mathematical Texts, Theta, Bucharest, 2007, pp. 229–306.

[7] D. Henrion and J. B. Lasserre, *Detecting global optimality and extracting solutions in GloptiPoly*, in Positive Polynomials in Control, Lecture Notes in Control and Information Sciences 312, D. Henrion and A. Garulli, eds., Springer, Berlin, 2005, pp. 293–310.

[8] D. Henrion and J. B. Lasserre, *Convergent relaxations of polynomial matrix inequalities and static output feedback*, IEEE Trans. Automat. Control, 51 (2006), pp. 192–202.

[9] T. Ito, H. Kobayashi, and K. Matsumoto, *Quantum multi-prover interactive proofs and decidability*, in preparation.

[10] I. KLEP AND J. POVH, *Semidefinite programming and sums of Hermitian squares of noncommutative polynomials*, J. Pure Appl. Algebra, 214 (2010), pp. 740–749.

[11] I. KLEP AND M. SCHWEIGHOFER, *A Nichtnegativstellensatz for polynomials in noncommuting variables*, Israel J. Math., 161 (2007), pp. 17–27.

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

[13] M. LAURENT, *Sums of squares, moment matrices and optimization over polynomials*, in Emerging Applications of Algebraic Geometry, M. Putinar and S. Sullivant, eds., IMA Vol. Math. Appl. 149, Springer, New York, 2009, pp. 157–270.

[14] J. LÖFBERG, *YALMIP: A toolbox for modeling and optimization in MATLAB*, available from http://users.isy.liu.se/johanl/yalmip.

[15] D. A. MAZZIOTTI, *Variational minimization of atomic and molecular ground-state energies via the two-particle reduced density matrix*, Phys. Rev. A, 65 (2002), 062511.

[16] D. A. MAZZIOTTI, *Realization of quantum chemistry without wave functions through first-order semidefinite programming*, Phys. Rev. Lett., 93 (2004), 213001.

[17] T. MORA, *An introduction to commutative and noncommutative Gröbner bases*, Theoret. Comput. Sci, 134 (1994), pp. 131–173.

[18] M. NAVASCUES, S. PIRONIO, AND A. ACIN, *Bounding the set of quantum correlations*, Phys. Rev. Lett., 98 (2007), 010401.

[19] M. NAVASCUES, S. PIRONIO, AND A. ACIN, *A convergent hierarchy of semdefinite programs characterizing the set of quantum correlations*, New J. Phys., 10 (2008), 073013.

[20] K. F. PÁL AND T. VÉRTESI, *Quantum bounds on Bell inequalities*, Phys. Rev. A, 79 (2009), 022120.

[21] P. A. PARRILO, *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robstness and Optimization*, Ph.D. thesis, California Institute of Technology, Pasadena, CA, 2000.

[22] S. PIRONIO, A. ACIN, S. MASSAR, A. BOYER DE LA GIRODAY, D. N. MATSUKEVICH, P. MAUNZ, S. OLMSCHENK, D. HAYES, L. LUO, T. A. MANNING, AND C. MONROE, *Random numbers certified by Bell's theorem*, arXiv:0911.3427.

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

[24] M. REED AND B. SIMON, *Functional Analysis*, Vol. 1, Methods of Modern Mathematical Physics, Academic Press, New York, 1972.

[25] J. F. STURM, *SeDuMi, a MATLAB toolbox for optimization over symmetric cones*, http://sedumi.mcmaster.ca.

[26] L. VANDENBERGHE AND S. BOYD, *Semidefinite programming*, SIAM Rev., 38 (1996), pp. 49–95.

[27] J. VON NEUMANN, *On rings of operators. Reduction theory*, Ann. of Math., 50 (1949), pp. 401–485.