Universität Konstanz

The tracial moment problem and trace-optimization of polynomials

Sabine Burgdorf
Kristijan Cafuta
Igor Klep
Janez Povh

Konstanzer Schriften in Mathematik

Nr. 287, August 2011

ISSN 1430-3558

THE TRACIAL MOMENT PROBLEM AND TRACE-OPTIMIZATION OF POLYNOMIALS

SABINE BURGDORF[1,3], KRISTIJAN CAFUTA, IGOR KLEP[2,3], AND JANEZ POVH[4]

Abstract. The main topic addressed in this paper is *trace-optimization of polynomials* in noncommuting (nc) variables: given an nc polynomial f, what is the smallest trace $f(\underline{A})$ can attain for a tuple of matrices $\underline{A}$? A relaxation using semidefinite programming (SDP) based on sums of hermitian squares and commutators is proposed. While this relaxation is not always exact, it gives effectively computable bounds on the optima. To test for exactness, the solution of the dual SDP is investigated. If it satisfies a certain condition called *flatness*, then the relaxation *is exact*. In this case it is shown how to extract global trace-optimizers with a procedure based on two ingredients. The first is the solution to the truncated *tracial moment problem*, and the other crucial component is the numerical implementation of the Artin-Wedderburn theorem for matrix $*$-algebras due to Murota, Kanno, Kojima, Kojima, and Maehara.

Trace-optimization of nc polynomials is a nontrivial extension of polynomial optimization in commuting variables on one side and eigenvalue optimization of nc polynomials on the other side – two topics with many applications, the most prominent being to linear systems engineering and quantum physics. The optimization problems discussed here facilitate new possibilities for applications, e.g. in operator algebras and statistical physics.

1. Introduction

A matrix has nonnegative trace if and only if it is a sum of a positive semidefinite matrix (a hermitian square) and a trace zero matrix (a commutator).

In this article we propose a method for finding and proving trace inequalities involving symmetric matrices. Our procedure provides certificates holding irrespective of the size of the matrices involved. Following Helton and his school [dOHMP08] we call such situations *dimension-free*. The algorithm is based on sum of squares and commutators certificates for noncommutative (nc) polynomials which can be obtained using semidefinite programming and has been implemented in the open source Matlab toolbox `NCSOStools` written by the second, third and fourth author [CKP+]. We refer the reader to [KP10, PNA10] for a similar treatment of dimension-free matrix inequalities given via positive semidefiniteness, and to GloptiPoly [HLL09], SparsePOP [WKKMS09], YALMIP [Löf04], and SOSTOOLS [PPSP05] for

Date: 22 April 2011.

2010 *Mathematics Subject Classification.* Primary 90C22, 13J30; Secondary 47A57, 11E25, 08B20.

Key words and phrases. sum of squares, noncommutative polynomial, semidefinite programming, tracial moment problem, flat extension, free positivity, real algebraic geometry.

[1]Partially supported by the Zukunftskolleg Konstanz.

[2]Partially supported by the Slovenian Research Agency (project no. J1-3608 and program no. P1-0222). Part of this research was done while the author held a visiting professorship at the University Konstanz supported by the program "free spaces for creativity".

[3]Partially supported by the French-Slovene partnership project Proteus 20208ZM.

[4]Supported by the Slovenian Research Agency under program no. P1-0297(B).

optimization software for polynomials in commuting variables based on sum of squares methods. Readers interested in symbolic computation with noncommuting variables are advised to see `NCAlgebra` [HdOMS] under Mathematica.

1.1. **Motivation.** Starting with Helton's seminal paper [Hel02], *free real algebraic geometry* (including *free positivity*, the study of positivity of polynomials in noncommutating variables) is being established. In this article we focus on *trace-positive* polynomials. These are nc polynomials all of whose evaluations at tuples of matrices have nonnegative trace.

Much of today's interest in real algebraic geometry is due to its powerful applications. For instance, the use of sum of squares and the truncated moment problem for polynomial optimization on $\mathbb{R}^n$ established by Lasserre and Parrilo [Las01, Las09, PS03, Par03] is nowadays a common fact in real algebraic geometry with applications to control theory, mathematical finance or operations research. In the free context there are many facets of applications as well. A nice survey on connections to control theory, systems engineering and optimization is given by Helton, McCullough, de Oliveira, Putinar [dOHMP08]. Another interesting use of nc sum of squares is given by Cimprič [Cim10], who investigates PDEs and eigenvalues of polynomial partial differential operators. Applications to quantum physics are explained by Pironio, Navascués, Acín [PNA10] who also consider computational aspects related to nc sum of squares. Furthermore, optimization of nc polynomials has direct applications in quantum information science (to compute upper bounds on the maximal violation of a generic Bell inequality [PV09]), and also in quantum chemistry (e.g. to compute the ground-state electronic energy of atoms or molecules [Maz04]). Another application in quantum physics is presented by Doherty, Liang, Toner, Wehner [DLTW08], who use free real algebraic geometry to consider the quantum moment problem and multi-player quantum games. Certificates of positivity via sums of squares are often used in the theoretical physics literature to place very general bounds on quantum correlations (cf. [Gla63]). These applications of free real algebraic geometry in quantum physics are based on finding lower bounds or estimates for the smallest eigenvalue of a given system represented by an nc polynomial.

Considering quantum mechanical many particle systems one often investigates the *statistical means* of the system instead of the system itself. Hence one is interested in bounds or estimates of the *trace* of a quantum statistical system. This brings us to the consideration of trace-positive nc polynomials, the main topic of this article. Trace-positive polynomials also arise in the Lieb-Seiringer reformulation of the important Bessis-Moussa-Villani (BMV) conjecture [BMV75] from statistical quantum mechanics. This reformulation states on the polynomial level that the nc polynomials $S_{m,k}(X^2, Y^2)$ that describe the coefficient of t^k in $(X^2 + tY^2)^m \in \mathbb{R}[t]$ are trace-positive for all $m, k \in \mathbb{N}$. In addition, trace-positive polynomials (and the tracial moment problem we discuss) occur naturally in von Neumann algebras and functional analysis. For instance, Connes' embedding problem [Con76] on finite II_1-factors is a question about the existence of a certain type of sum of hermitian squares (sohs) certificates for trace-positive polynomials [KS08a]. It is widely believed that Connes' conjecture is false and our results will enable us to look for a counterexample using a computer algebra system.

We developed `NCSOStools` [CKP+] as a consequence of this surge of interest in free real algebraic geometry and sums of (hermitian) squares of nc polynomials. `NCSOStools` is an open source Matlab toolbox for solving sohs problems using *semidefinite programming* (SDP). As a side product our toolbox implements symbolic computation with noncommuting variables in Matlab.

For a precise statement of our contribution we need a bit of notation. We start by explaining the gist of the idea on an example.

Example 1.1. For symmetric matrices A, B of the same size we have

$$\operatorname{tr}(A^2B^2 + AB^2A + ABAB + BA^2B + BABA + B^2A^2) \geq 0, \tag{1}$$

where tr stands for trace. In fact,

$$\begin{aligned}&\operatorname{tr}(A^2B^2 + AB^2A + ABAB + BA^2B + BABA + B^2A^2)\\ &\qquad = \operatorname{tr}(ABAB + BABA + AB^2A + BA^2B) + 2\operatorname{tr}(AB^2A)\\ &\qquad\qquad = \operatorname{tr}((AB+BA)^t(AB+BA)) + 2\operatorname{tr}((BA)^t(BA)) \geq 0\end{aligned}$$

since $(AB+BA)^t(AB+BA)$ and $(BA)^t(BA)$ are positive semidefinite matrices.

1.2. **Words and nc polynomials.** Fix $n \in \mathbb{N}$ and let $\langle \underline{X} \rangle$ be the monoid freely generated by $\underline{X} := (X_1, \dots, X_n)$, i.e., $\langle \underline{X} \rangle$ consists of *words* in the n noncommuting letters $X_1, \dots, X_n$ (including the empty word denoted by 1). We consider the free algebra $\mathbb{R}\langle \underline{X} \rangle$. The elements of $\mathbb{R}\langle \underline{X} \rangle$ are linear combinations of words in the n letters $\underline{X}$ and are called *nc polynomials*. An element of the form aw where $a \in \mathbb{R} \setminus \{0\}$ and $w \in \langle \underline{X} \rangle$ is called a *monomial* and a its *coefficient*. Words are monomials with coefficient 1. The length of the longest word in an nc polynomial $f \in \mathbb{R}\langle \underline{X} \rangle$ is the *degree* of f and is denoted by $\deg f$. The set of all nc polynomials of degree $\leq d$ will be denoted by $\mathbb{R}\langle \underline{X} \rangle_{\leq d}$. If an nc polynomial f involves only two variables, we write $f \in \mathbb{R}\langle X, Y \rangle$.

1.3. **Sums of hermitian squares.** We equip $\mathbb{R}\langle \underline{X} \rangle$ with the *involution* $*$ that fixes $\mathbb{R} \cup \{\underline{X}\}$ pointwise and thus reverses words, e.g. $(X_1X_2^2X_3 - 2X_3^3)^* = X_3X_2^2X_1 - 2X_3^3$. Hence $\mathbb{R}\langle \underline{X} \rangle$ is the $*$-algebra freely generated by n symmetric letters. Let $\operatorname{Sym}\mathbb{R}\langle \underline{X} \rangle$ denote the set of all *symmetric elements*, that is,

$$\operatorname{Sym}\mathbb{R}\langle \underline{X} \rangle := \{f \in \mathbb{R}\langle \underline{X} \rangle \mid f = f^*\}.$$

An nc polynomial of the form g^*g is called a *hermitian square* and the set of all sums of hermitian squares will be denoted by Σ^2. Clearly, $\Sigma^2 \subsetneq \operatorname{Sym}\mathbb{R}\langle \underline{X} \rangle$. The involution $*$ extends naturally to matrices (in particular, to vectors) over $\mathbb{R}\langle \underline{X} \rangle$. For instance, if $V = (v_i)$ is a (column) vector of nc polynomials $v_i \in \mathbb{R}\langle \underline{X} \rangle$, then V^* is the row vector with components v_i^*. We use V^t to denote the row vector with components v_i.

The main idea in systematizing the verification of inequalities as in Example 1.1 is to look for certificates at the level of nc polynomials. In particular, we propose a relaxation for finding the trace-optimum based on sums of hermitian squares and commutators.

1.4. **Contribution and reader's guide.** To verify the trace-inequality of Example 1.1 via sums of hermitian squares and commutators at the level of nc polynomials consider

$$f = X^2Y^2 + XY^2X + XYXY + YX^2Y + YXYX + Y^2X^2 \in \mathbb{R}\langle X, Y \rangle.$$

This f is of the form

$$\begin{aligned} f &= (XYXY + YXYX + XY^2X + YX^2Y) + 2XY^2X \\ &\qquad\qquad + (X^2Y^2 - XY^2X) + (Y^2X^2 - XY^2X) \\ &= (XY+YX)^*(XY+YX) + 2(YX)^*(YX) + (\text{sum of commutators}). \end{aligned}$$

Note that the two differences in the brackets are commutators, e.g. $X^2Y^2 - XY^2X = X \cdot XY^2 - XY^2 \cdot X$. Hence $f(A, B)$ is a sum of hermitian squares and commutators for all symmetric matrices A, B of the same size, and so has nonnegative trace.

The purpose of this paper is threefold.

First, we present how to systematize the search for sum of hermitian squares (sohs) and commutators certificates using a computer algebra system. This is done via a variant of the classical Gram matrix method. It is purely symbolic and constructs an SDP whose feasibility is equivalent to the existence of such a certificate. In order to find the *best* possible bound (equivalently, what is the greatest lower bound for the trace an nc polynomial can attain), we study a closely related instance of a semidefinite programming problem. From the solution of this SDP we extract the desired bound and the corresponding polynomial sohs certificate.

Second, to investigate exactness of the obtained bound and the corresponding certificate, we consider the dual SDP, giving rise to the *tracial moment problem*. Loosely speaking, it asks which linear functionals on $\mathbb{R}\langle \underline{X} \rangle$ are integration of the trace of an nc polynomial. In Section 3 we continue the investigation of the tracial moment problem started in [BK+] by the first and the third author. Motivated by optimization problems, our main focus is on the *truncated* tracial moment problem, like in the classical case of polynomial optimization on $\mathbb{R}^n$ [Las01, Las09, PS03, Par03]. We define a seemingly more general version of the tracial moment problem by considering integrals over Borel measures on tuples of matrices as opposed to finite atomic measures as is done in [BK+]. In the truncated case both definitions are equivalent by the tracial version of the Bayer-Teichmann theorem [BT06] presented in Theorem 3.8 below. We emphasize that the truncated version is more general than the full tracial moment problem. In fact, solving the truncated moment problems solves the full moment problem. This is the topic of Section 3.2.

Third, the solution of the truncated tracial moment problem is utilized to give a condition for the exactness of the sohs certificate for trace-optimization of polynomials. If the solution to the dual SDP satisfies a condition called *flatness*, then our sohs relaxation is exact (Theorem 3.12). While this resembles the classical case of polynomial optimization on $\mathbb{R}^n$, the extraction of optimizers is more involved and is explained in detail in Section 3.3. First of all, the Gelfand-Naimark-Segal (GNS) construction gives rise to a set of symmetric matrices $\hat{X}_j$, one for each of the noncommuting variables. Unlike in the commutative [Las01] or the free noncommutative setting [PNA10], an additional step is needed to recover trace-optimizers. We consider the matrix $*$-algebra generated by the $\hat{X}_j$ and compute its Artin-Wedderburn decomposition. This is done with the aid of the algorithm of Murota, Kanno, Kojima, and Kojima [MKKK10], and Maehara and Murota [MM10]. It produces a simultaneous block diagonalization of the $\hat{X}_j$, and each of these blocks yields a trace-optimizer.

2. Sums of hermitian squares and commutators

In this section we present the main notions we exploit in the sequel, namely sums of hermitian squares and commutators of nc polynomials. Via the so-called Gram matrix method they relate naturally to semidefinite programming.

2.1. **Matrix-positive polynomials and sums of hermitian squares.** Every positive semidefinite matrix A has a square root, i.e., A is a hermitian square. On the polynomial level we have the following:

Definition 2.1. An nc polynomial $f \in \mathbb{R}\langle \underline{X} \rangle$ is called *matrix-positive* if

$$f(\underline{A}) \succeq 0 \text{ for all tuples of symmetric matrices } \underline{A} \text{ of the same size.} \tag{2}$$

If $f \in \mathbb{R}\langle \underline{X} \rangle$ is a sum of hermitian squares, i.e., $f \in \Sigma^2$, then f is matrix-positive. Helton [Hel02] (and independently, McCullough [McC01]) proved the converse of this easy observation: if $f \in \mathbb{R}\langle \underline{X} \rangle$ is matrix-positive, then $f \in \Sigma^2$.

2.2. **Trace zero polynomials and cyclic equivalence.** It is well-known and easy to see that trace zero matrices are (sums of) commutators. To mimic this property for nc polynomials, we introduce cyclic equivalence [KS08a]:

Definition 2.2. An element of the form $[p, q] := pq - qp$ for $p, q \in \mathbb{R}\langle \underline{X} \rangle$ is called a *commutator*. nc polynomials $f, g \in \mathbb{R}\langle \underline{X} \rangle$ are called *cyclically equivalent* ($f \overset{\text{cyc}}{\sim} g$) if $f - g$ is a sum of commutators:

$$f - g = \sum_{i=1}^{k} [p_i, q_i] = \sum_{i=1}^{k} (p_i q_i - q_i p_i) \text{ for some } k \in \mathbb{N} \text{ and } p_i, q_i \in \mathbb{R}\langle \underline{X} \rangle.$$

Example 2.3. We have $2X^2Y^2X^3 + XY^2X^2 + XY^2X^4 \overset{\text{cyc}}{\sim} 3YX^5Y + YX^3Y$ as

$$\begin{aligned} 2X^2Y^2X^3 + XY^2X^2 + XY^2X^4 - (3YX^5Y + YX^3Y) = \\ = [2X^2Y, YX^3] + [XY, YX^4] + [XY, YX^2]. \end{aligned}$$

It is clear that $\overset{\text{cyc}}{\sim}$ is an equivalence relation. The following remark shows that it can be easily tested and motivates its name.

Remark 2.4.

(a) For $v, w \in \langle \underline{X} \rangle$, we have $v \overset{\text{cyc}}{\sim} w$ if and only if there are $v_1, v_2 \in \langle \underline{X} \rangle$ such that $v = v_1 v_2$ and $w = v_2 v_1$. That is, $v \overset{\text{cyc}}{\sim} w$ if and only if w is a cyclic permutation of v.

(b) nc polynomials $f = \sum_{w \in \langle \underline{X} \rangle} a_w w$ and $g = \sum_{w \in \langle \underline{X} \rangle} b_w w$ ($a_w, b_w \in \mathbb{R}$) are cyclically equivalent if and only if for each $v \in \langle \underline{X} \rangle$,

$$\sum_{\substack{w \in \langle \underline{X} \rangle \\ w \overset{\text{cyc}}{\sim} v}} a_w = \sum_{\substack{w \in \langle \underline{X} \rangle \\ w \overset{\text{cyc}}{\sim} v}} b_w. \tag{3}$$

This notion is important for us because trace zero nc polynomials are exactly sums of commutators:

Theorem 2.5 (Klep-Schweighofer [KS08a])**.** *Let $s \in \mathbb{N}$ and $f \in \operatorname{Sym} \mathbb{R}\langle \underline{X} \rangle_{\leq s}$. Then $f \overset{\text{cyc}}{\sim} 0$ if and only if $\operatorname{tr}(f(\underline{A})) = 0$ for all n-tuples $\underline{A} = (A_1, \ldots, A_n)$ of symmetric $s \times s$-matrices.*

2.3. **Trace-positive polynomials, cyclic equivalence and sums of hermitian squares.** A matrix has nonnegative trace if and only if it is a sum of a positive semidefinite matrix and a trace zero matrix.

Definition 2.6. An nc polynomial $f \in \mathbb{R}\langle \underline{X} \rangle$ is called *trace-positive* if

$$\operatorname{tr} f(\underline{A}) \geq 0 \text{ for all tuples of symmetric matrices } \underline{A} \text{ of the same size.} \tag{4}$$

Clearly, every matrix-positive $f \in \mathbb{R}\langle \underline{X} \rangle$ is trace-positive and the same is true for every nc polynomial cyclically equivalent to a sum of hermitian squares.

Definition 2.7. Let
$$\Theta^2 := \{f \in \mathbb{R}\langle \underline{X}\rangle \mid \exists g \in \Sigma^2 : \ f \overset{\text{cyc}}{\sim} g\}$$
denote the convex cone of all nc polynomials cyclically equivalent to a sum of hermitian squares. By definition, the elements in Θ^2 are exactly nc polynomials which can be written as sums of hermitian squares and commutators.

Unlike in the matrix-positive case, there are trace-positive polynomials which are *not* members of Θ^2. The easiest example is the noncommutative Motzkin polynomial, $f = X_1X_2^4X_1 + X_2X_1^4X_2 - 3X_1X_2^2X_1 + 1$ [KS08a, Example 4.4]. We also refer the reader to [KS08b, Example 3.5] for more sophisticated examples obtained by considering the BMV conjecture. Nevertheless, this obvious certificate for trace-positivity turns out to be useful in optimization, so merits a further systematic investigation here.

2.4. **Gram matrix method.** Testing whether a given $f \in \mathbb{R}\langle \underline{X}\rangle$ is an element of Θ^2 can be done using semidefinite programming as first observed in [KS08b, Section 3]. This is based on the *Gram matrix method*. The core of the method is given by the following proposition, an extension of the results for sums of hermitian squares (cf. [Hel02, Section 2.2] or [KP10, Theorem 3.1 and Algorithm 1]), which are in turn variants of the classical result for polynomials in commuting variables due to Choi, Lam and Reznick ([CLR95, Section 2]; see also Parrilo [Par03], and Parrilo and Sturmfels [PS03]).

Proposition 2.8. *Suppose $f \in \mathbb{R}\langle \underline{X}\rangle$. Then $f \in \Theta^2$ if and only if there exists a positive semidefinite matrix G such that*
$$f \overset{\text{cyc}}{\sim} W^*GW, \tag{5}$$
where W is a vector consisting of all words $w \in \langle \underline{X}\rangle$ satisfying $2\deg(w) \leq \deg(f)$. Conversely, given such a positive semidefinite matrix G of rank r, one can construct nc polynomials $g_1, \dots, g_r \in \mathbb{R}\langle \underline{X}\rangle$ with
$$f \overset{\text{cyc}}{\sim} \sum_{i=1}^{r} g_i^* g_i. \tag{6}$$

The matrix G is called *a (tracial) Gram matrix* for f. More generally, given a vector of words V, every symmetric matrix G satisfying $f \overset{\text{cyc}}{\sim} V^*GV$ is called a Gram matrix. If $f = V^*GV$, then G is an *exact* Gram matrix. The proof of Proposition 2.8 is straightforward as in the commutative case.

For an nc polynomial $f \in \mathbb{R}\langle \underline{X}\rangle$ the tracial Gram matrix is *not* unique, hence determining whether $f \in \Theta^2$ amounts to finding *a* positive semidefinite Gram matrix from the affine set of all Gram matrices for f. Problems like this can be (in theory) solved *exactly* using quantifier elimination. However, this only works for problems of small size, so a *numerical* approach is needed in practice. Thus we turn to semidefinite programming.

2.5. **Semidefinite programming.** Semidefinite programming (SDP) is a subfield of convex optimization concerned with the optimization of a linear objective function over the intersection of the cone of positive semidefinite matrices with an affine space. More precisely, given symmetric matrices C, $A_1, \dots, A_m \in \mathbb{R}^{s\times s}$ and a vector $b \in \mathbb{R}^m$, we formulate a *semidefinite program in standard primal form* (in the sequel we refer to problems of this type by PSDP) as follows:
$$\begin{array}{lrcl} \inf & \langle C, G\rangle & & \\ \text{s.t.} & \langle A_i, G\rangle & = & b_i, \quad i = 1, \dots, m \\ & G & \succeq & 0. \end{array} \tag{PSDP}$$

Here $\langle\_\,,\,\_\rangle$ stands for the standard scalar product of matrices: $\langle A, B\rangle = \operatorname{tr}(B^t A)$. The dual problem to (PSDP) is the *semidefinite program in the standard dual form*

$$\begin{array}{ll} \sup & \langle b, y\rangle \\ \text{s.t.} & \sum_i y_i A_i \preceq C. \end{array} \tag{DSDP}$$

Here $y \in \mathbb{R}^m$, and the difference $C - \sum_i y_i A_i$ is usually denoted by Z.

The relevance of SDPs increased with the ability to solve these problems efficiently in theory and in practice. Given an $\varepsilon > 0$ we can extend most interior point methods for linear programming to polynomial time algorithms giving an ε-optimal solution for SDPs [NN94] (provided that both (PSDP) and (DSDP) have non-empty interiors of feasible sets and we have good initial points). The variables appearing in these polynomial bounds are the size s of the matrix variable, the number m of linear constraints in (PSDP) and $\log \varepsilon$ (cf. [WSV00, Ch. 10.4.4] and [BTN01] for details). However, the complexity to obtain exact solutions of an SDP is still an open question in semidefinite optimization, see e.g. [Ram97]. Nevertheless, there exist several general purpose open source packages (cf. SeDuMi [Stu99], SDPA [YFK03], SDPT3 [TTT99]) which can efficiently find ε-optimal solutions in practice. If the problem is of medium size (i.e., $s \leq 1000$ and $m \leq 10.000$), these packages are based on interior point methods, while packages for larger semidefinite programs use some variant of the first order methods (see [Mit03] for a comprehensive list of state-of-the-art SDP solvers and also [MPRW09]). However, once $s \geq 3000$ or $m \geq 250000$, the problem must share some special property otherwise state-of-the-art solvers will fail to solve it for complexity reasons.

3. Trace-optimization of nc polynomials

One of the main features of our freely available Matlab software package `NCSOStools` [CKP+] is `NCcycMin` which uses a sum of hermitian squares and commutators relaxation to approximate a trace-minimum of a given nc polynomial. The purpose of this section is three-fold. The first subsection presents our relaxation as an SDP and states its duality properties. We then recall the tracial moment problem (Section 3.2) introduced and studied by the first and third author in [BK+], needed in Section 3.3 where we show how to use the solution to the tracial moment problem to test for exactness of our Θ^2-relaxation and to extract trace-optimizers. This part is influenced by the method of Henrion and Lasserre [HL05] for the commutative case, which has been implemented in GloptiPoly [HLL09]. For a similar investigation in the free noncommutative setting see [PNA10].

Let $\mathbb{S}^{s\times s}$ denote the set of symmetric matrices of size s, for some $s \in \mathbb{N}$, and let Tr denote the *normalized* trace.

3.1. SDP relaxation and its duality properties.

Let $f \in \mathbb{R}\langle \underline{X}\rangle$ be given. We are interested in the *trace-minimum* of f, that is,

$$f_* := \inf\{\operatorname{Tr}\big(f(\underline{A})\big) \mid d \in \mathbb{N},\ \underline{A} \in (\mathbb{S}^{d\times d})^n\}. \tag{7}$$

This is a hard problem. For instance, a good understanding of trace-positive polynomials is likely to lead to a solution of two outstanding open problems: Connes' embedding conjecture [Con76] from operator algebras, and the BMV conjecture [BMV75] from quantum statistical mechanics; see [KS08b, KS08a]. In fact, our computational advances will make it possible to look for a *counterexample* to Connes' conjecture using our software.

We propose the following relaxation of trace-minimization of nc polynomials:

$$f_{\text{sos}} := \sup\{a \mid f - a \in \Theta^2\}. \tag{8}$$

Remark 3.1. Since we are only interested in the trace of the values of $f \in \mathbb{R}\langle \underline{X} \rangle$, we may use that $\operatorname{tr}(f(\underline{A})) = \operatorname{tr}(f^*(\underline{A}))$ for all real $\underline{A}$; hence there is no harm in replacing f by its symmetrization $\frac{1}{2}(f + f^*)$. Thus we will mostly focus on *symmetric* nc polynomials.

Lemma 3.2. *Let $f \in \operatorname{Sym} \mathbb{R}\langle \underline{X} \rangle$. Then $f_{\text{sos}} \leq f_*$.* ■

In general we do not have equality in Lemma 3.2. For instance, the Motzkin polynomial f satisfies $f_* = 0$ and $f_{\text{sos}} = \sup \varnothing := -\infty$, see [KS08a]. Nevertheless, f_{sos} gives a solid approximation of f_* for most of the examples and is easier to compute. It is obtained by solving the SDP

$$\begin{array}{ll} \sup & a \\ \text{s.t.} & f - a \in \Theta^2. \end{array} \tag{SDP_{min}}$$

Suppose $f \in \operatorname{Sym} \mathbb{R}\langle \underline{X} \rangle$ is of degree $\leq 2d$ (with constant term f_1). Let W be a vector of all words up to degree d with first entry equal to 1. Then (SDP_{min}) rewrites into

$$\begin{array}{lrcl} \sup & \multicolumn{3}{l}{f_1 - \langle E_{11}, G \rangle} \\ \text{s.t.} & f - f_1 & \overset{\text{cyc}}{\sim} & W^t(G - g_{11}E_{11})W \\ & G & \succeq & 0. \end{array} \tag{$\text{SDP}_{\text{min}'}$}$$

Here E_{11} is the matrix with all entries 0 except for the $(1,1)$-entry which is 1, and g_{11} denotes the $(1,1)$-entry of G. The cyclic equivalence translates into a set of linear constraints, cf. Remark 2.4.

In general (SDP_{min}) does not satisfy the Slater condition. Nevertheless:

Theorem 3.3. (SDP_{min}) *satisfies strong duality.*

Proof. The proof is essentially the same as that of [KP10, Theorem 5.1] so is omitted. We only mention an important ingredient is the closedness of the cone Θ^2 established in [BK+, Lemma 4.5]. ■

The dual problem to the (SDP_{min}) can be written as

$$\begin{array}{ll} \inf & L(f) \\ \text{s.t.} & L : \mathbb{R}\langle \underline{X} \rangle_{\leq 2d} \to \mathbb{R} \text{ is a linear } *\text{-map} \\ & L(1) = 1 \\ & L(p) \geq 0 \text{ for all } p \in \Theta^2 \cap \mathbb{R}\langle \underline{X} \rangle_{\leq 2d}. \end{array} \tag{DSDP_{min}}$$

(L is a $*$-*map* means $L(p^*) = L(p)$ for all p. Note the last constraint enforces $L(pq - qp) = 0$ for all $p, q \in \mathbb{R}\langle \underline{X} \rangle_{\leq d}$, i.e., L is *tracial*.) Let f^{sos} denote the optimal value of (DSDP_{min}). By Theorem 3.3, we have $f_{\text{sos}} = f^{\text{sos}}$. The question is, does $f_{\text{sos}} = f^{\text{sos}} = f_*$ hold? And if so, can we detect this using the above SDP? If the dual optimizer L_* satisfies an easy to check condition called *flatness* (see Subsection 3.3.1 for a definition), then the answer to both questions is affirmative. In particular, the proposed Θ^2-relaxation is then exact. Furthermore, in this case we can even extract global trace-minimizers of f. This is based on the solution to the truncated tracial moment problem, uses the Gelfand-Naimark-Segal construction and the Artin-Wedderburn theorem; see Section 3.3.

3.2. **Tracial moment problem.** The moment problem is a classical question in functional analysis, well studied because of its importance and applications [Akh65, CF96, Lau09]. For the free noncommutative moment problem see McCullough [McC01]. In this section we recall the tracial moment problem from [BK+], which is essentially the study of feasible points of ($\mathrm{DSDP}_{\min}$). In fact, we define a seemingly more general version using integrals over Borel measures as opposed to finite atomic measures as is done in [BK+]. However, in the truncated case both versions are equivalent by the tracial version of the Bayer-Teichmann theorem [BT06] presented in Theorem 3.8 below. Our emphasis on the truncated tracial moment problem is justified for two reasons. First of all, this is what is needed for the application to trace-optimization of nc polynomials. Second, by Theorem 3.6, a tracial analog of the classical result of Stochel [Sto01], solving the truncated tracial moment problems solves the full tracial moment problem.

Definition 3.4. A sequence of real numbers (y_w) indexed by words $w \in \langle \underline{X} \rangle$ satisfying

$$y_w = y_u \text{ whenever } w \overset{\mathrm{cyc}}{\sim} u, \quad y_w = y_{w^*} \text{ for all } w, \tag{9}$$

and $y_1 = 1$, is called a (normalized) *tracial sequence*.

Example 3.5.

(a) Given $s \in \mathbb{N}$ and a probability measure μ on $(\mathbb{S}^{s\times s})^n$, the sequence given by

$$y_w := \int \mathrm{Tr}(w(\underline{A}))\, d\mu(\underline{A}) \tag{10}$$

is a tracial sequence since the traces of cyclically equivalent words coincide.

(b) Every feasible point L of ($\mathrm{DSDP}_{\min}$) induces a truncated tracial sequence $y_L := (L(w))_w$, where $w \in \langle \underline{X} \rangle$ are constrained by $\deg w \leq 2d$. Conversely, every finite tracial sequence $(y_w)_{\leq 2d}$ yields a linear $*$-map (often called the Riesz functional) $L_y : \mathbb{R}\langle \underline{X} \rangle_{\leq 2d} \to \mathbb{R}$, $w \mapsto y_w$.

For us the converse of Example 3.5(a) (the *tracial moment problem*) is of importance: for which sequences (y_w) do there exist an $s \in \mathbb{N}$ and a probability measure μ on $(\mathbb{S}^{s\times s})^n$ such that (10) holds? We then say that (y_w) has a *tracial moment representation* and call it a *tracial moment sequence*. The *truncated tracial moment problem* is the study of (finite) tracial sequences $(y_w)_{\leq k}$ where w is constrained by $\deg w \leq k$ for some $k \in \mathbb{N}$, and properties (9) hold for these w. For instance, which sequences $(y_w)_{\leq k}$ have a tracial moment representation, i.e., when does there exist a representation of the values y_w as in (10) for $\deg w \leq k$? If this is the case, the sequence $(y_w)_{\leq k}$ is called a *truncated tracial moment sequence*.

3.2.1. *Stochel's theorem.* The truncated tracial moment problem is more general than the full tracial moment problem in the sense explained in Theorem 3.6.

Theorem 3.6. *Suppose $y = (y_w)_w$ is a tracial sequence. If there is an $s \in \mathbb{N}$ such that for all $k \in \mathbb{N}$ there is a probability measure μ_k on $(\mathbb{S}^{s\times s})^n$ satisfying (10) for all $w \in \langle \underline{X} \rangle$ with $\deg w \leq k$, then y is a tracial moment sequence. Furthermore, there is a probability measure μ on $(\mathbb{S}^{s\times s})^n$ such that (10) holds for all $w \in \langle \underline{X} \rangle$.*

We start by a preliminary lemma showing that a specific function needed in the proof of Theorem 3.6 vanishes at infinity.

Lemma 3.7. *Let $s \in \mathbb{N}$ be fixed. For $u \in \langle \underline{X} \rangle$ the map $\varphi_u : (\mathbb{S}^{s\times s})^n \to \mathbb{R}$ defined by*

$$\varphi_u(\underline{A}) := \frac{\operatorname{Tr}\big(u(\underline{A})\big)}{1 + \sum_{i=1}^n \operatorname{Tr}\big(A_i^{2\deg(u)+2}\big)}$$

lies in $\mathcal{C}_0\big((\mathbb{S}^{s\times s})^n, \mathbb{R}\big)$.

Proof. Let $u \in \mathbb{R}\langle \underline{X} \rangle$ be fixed with $\deg(u) =: d$ and let $\underline{A} \in (\mathbb{S}^{s\times s})^n$ be such that $\sum_{i=1}^n \operatorname{Tr}(A_i^2) > \ell^2$ for some $\ell \in \mathbb{N}$. Choose the index $i_A \in \{1\dots,n\}$ such that $\operatorname{Tr}(A_{i_A}^2) \geq \operatorname{Tr}(A_i^2)$ for all $i = 1, \dots, n$. Then

$$\operatorname{Tr}(A_{i_A}^2) \geq \frac{\sum_i \operatorname{Tr}(A_i^2)}{n} > \frac{\ell^2}{n}.$$

Since the matrices A_i^2 are positive semidefinite we have $\operatorname{Tr}(A_i^{2d+2}) = \|A_i^2\|_{d+1}^{d+1}$, where $\|\_\|_p$ denotes the normalized p-Schatten norm on $\mathbb{S}^{s\times s}$, which generalizes the Hilbert-Schmidt norm $(p = 2)$ and is given by

$$\|T\|_p^p = \operatorname{Tr}(|T|^p) \text{ with } |T| = \sqrt{T^2} \text{ for } T \in \mathbb{S}^{s\times s}.$$

Since $\mathbb{S}^{s\times s}$ is finite dimensional, the $(d + 1)$-Schatten norm is equivalent to the 1-Schatten norm, also known as the trace-norm, on $\mathbb{S}^{s\times s}$. Hence there is a $c \in \mathbb{R}_{>0}$ such that

$$c \operatorname{Tr}(A_i^2)^{d+1} = c\|A_i^2\|_1^{d+1} \leq \|A_i^2\|_{d+1}^{d+1} = \operatorname{Tr}(A_i^{2d+2})$$

for all $A_i \in \mathbb{S}^{s\times s}$. Further, for the numerator of φ_u we have

$$(\operatorname{Tr}(u(\underline{A})))^2 \leq s^{d-2} u(\operatorname{Tr}(A_1^2), \dots, \operatorname{Tr}(A_n^2)) \leq s^{d-2}(\operatorname{Tr}(A_{i_A}^2))^d$$

by induction on d and the Cauchy-Schwarz inequality. All together this implies

$$\begin{aligned}
\varphi_u(\underline{A})^2 = \frac{\big(\operatorname{Tr}(u(\underline{A}))\big)^2}{\big(1 + \sum_{i=1}^n \operatorname{Tr}(A_i^{2d+2})\big)^2} &\leq \frac{s^{d-2}(\operatorname{Tr}(A_{i_A}^2))^d}{\big(1 + \sum_{i=1}^n \operatorname{Tr}(A_i^{2d+2})\big)^2} \\
\leq \frac{s^{d-2}(\operatorname{Tr}(A_{i_A}^2))^d}{\big(1 + c\sum_{i=1}^n (\operatorname{Tr}(A_i^2))^{d+1}\big)^2} &< \frac{s^{d-2}(\operatorname{Tr}(A_{i_A}^2))^d}{c^2(\operatorname{Tr}(A_{i_A}^2))^{2d+2}} \\
\leq \frac{s^{d-2}}{c^2 \operatorname{Tr}(A_{i_A}^2)^{d+2}} &< \frac{s^{d-2} n^{d+2}}{c^2 \ell^{2d+4}}
\end{aligned}$$

which goes to zero for large ℓ. Hence $\varphi_u \in \mathcal{C}_0\big((\mathbb{S}^{s\times s})^n, \mathbb{R}\big)$. ∎

Proof of Theorem 3.6. Endow $\mathcal{C}_0 := \mathcal{C}_0\big((\mathbb{S}^{s\times s})^n, \mathbb{R}\big)$ with the maximum norm $\|\_\|_\infty$. To every finite measure η on $(\mathbb{S}^{s\times s})^n$ we associate the linear functional $\widehat{\eta} : \mathcal{C}_0 \to \mathbb{R}$,

$$\widehat{\eta}(f) := \int f(\underline{A})\, d\eta(\underline{A}).$$

Due to our normalization, for all $k \in \mathbb{N}$ we have

$$|\widehat{\mu}_k(f)| \leq \int \|f\|_\infty \, d\mu_k = \|f\|_\infty \quad \text{for all } f \in \mathcal{C}_0,$$

so all the $\widehat{\mu}_k$ belong to $\mathbb{B}$, the closed unit ball in the dual space $\mathcal{C}_0^\vee = \mathcal{C}_0\big((\mathbb{S}^{s\times s})^n, \mathbb{R}\big)^\vee$.

By the Banach-Alaoglu theorem, there is a subsequence $(\widehat{\mu}_{k_\ell})_\ell$ of $(\widehat{\mu}_k)_k$ converging to some $\psi \in \mathbb{B}$. For simplicity of notation, we omit the subindex ℓ in the sequel and assume that $(\widehat{\mu}_k)_k$ converges to ψ. If $f \in \mathcal{C}_0$ and $f \geq 0$, then

$$\psi(f) = \lim_{k\to\infty} \widehat{\mu}_k(f) \geq 0.$$

Hence by the Riesz representation theorem, there is a finite positive Borel measure μ on $(\mathbb{S}^{s\times s})^n$ with $\widehat{\mu} = \psi$. Since $\widehat{\mu}(1) = 1$, μ is a probability measure.

Let $u \in \langle \underline{X} \rangle$ be fixed with $\deg(u) =: d$ and $\varrho_u(\underline{A}) := 1 + \sum_{i=1}^n \operatorname{Tr}\left(A_i^{2d+2}\right)$. The assumption that $(y_w)_{\leq 2k}$ is a truncated tracial moment sequence with corresponding measure μ_k, implies

$$\int \varrho_u \, d\mu_k = \int \Big(1 + \sum_{i=1}^n \operatorname{Tr}\left(A_i^{2d+2}\right)\Big) \, d\mu_k(\underline{A}) = 1 + \sum_{i=1}^n y_{X_i^{2d+2}}. \quad \text{for all} \quad k \geq 2d+2.$$

Thus the sequence $(\widehat{\nu}_k)_k$ of linear functionals associated to the Borel measures ν_k on $(\mathbb{S}^{s\times s})^n$ which are defined by

$$d\nu_k(\underline{A}) = \varrho_u(\underline{A}) \, d\mu_k(\underline{A}),$$

is uniformly bounded. We now proceed to show that the Borel measure ν, given by

$$d\nu(\underline{A}) = \varrho_u(\underline{A}) \, d\mu(\underline{A}),$$

is finite. Let $(X_\ell)_\ell$ be an increasing sequence of compact subsets of $\left(\mathbb{S}^{s\times s}\right)^n$ with $\bigcup_{\ell=1}^\infty X_\ell = \left(\mathbb{S}^{s\times s}\right)^n$. For each $\ell \geq 1$ there is a continuous function $\tau_\ell : \left(\mathbb{S}^{s\times s}\right)^n \to \mathbb{R}$ with compact support such that $0 \leq \tau_\ell \leq 1$ and $\tau_\ell = 1$ on X_ℓ. Then,

$$\int d\nu = \int \varrho_u \, d\mu = \lim_{\ell\to\infty} \int_{X_\ell} \varrho_u \, d\mu \leq \limsup_{\ell\to\infty} \lim_{k\to\infty} \int \tau_\ell \varrho_u \, d\mu_k \leq \limsup_{k\to\infty} \int \varrho_u \, d\mu_k < \infty.$$

The finiteness of ν yields that $(\widehat{\nu}_k)_k$ converges pointwise to $\widehat{\nu} \in \mathcal{C}_0^\vee$ in the $\sigma(\mathcal{C}_0^\vee, \mathcal{C}_0)$-topology. Since $\varphi_u : (\mathbb{S}^{s\times s})^n \to \mathbb{R}$,

$$\varphi_u(\underline{A}) := \frac{\operatorname{Tr}\left(u(\underline{A})\right)}{1 + \sum_{i=1}^n \operatorname{Tr}\left(A_i^{2\deg(u)+2}\right)}$$

lies in $\mathcal{C}_0$ by Lemma 3.7, we get the desired conclusion

$$y_u = \lim_{k\to\infty} \int \operatorname{Tr}(u(\underline{A})) \, d\mu_k(\underline{A}) = \lim_{k\to\infty} \int \varphi_u \varrho_u \, d\mu_k = \int \varphi_u \varrho_u \, d\mu = \int \operatorname{Tr}\left(u(\underline{A})\right) d\mu(\underline{A}). \qquad \blacksquare$$

3.2.2. *Bayer-Teichmann theorem.* Our next theorem is a tracial version of the classical result of Bayer and Teichmann [BT06] stating that every truncated moment sequence y that admits a representing measure, admits a finite atomic representing measure. That is, the corresponding linear map L_y is given by a cubature formula. Our proof is an easy modification of the Schweighofer adaptation of the original proof as presented by Laurent in [Lau09, Section 5.2].

Theorem 3.8. *If $y = (y_w)_{\leq k}$ is a truncated tracial moment sequence with probability measure μ on $(\mathbb{S}^{s\times s})^n$ for some $s \in \mathbb{N}$, then there exist $N \in \mathbb{N}$, $\lambda_i \in \mathbb{R}_{>0}$ with $\sum_i^N \lambda_i = 1$ and n-tuples $\underline{A}^{(i)} = (A_1^{(i)}, \ldots, A_n^{(i)}) \in (\mathbb{S}^{s\times s})^n$, such that for all w with $\deg w \leq k$:*

$$y_w = \sum_{i=1}^N \lambda_i \operatorname{Tr}(w(\underline{A}^{(i)})). \tag{11}$$

Proof. Let $S = \operatorname{supp} \mu \subseteq (\mathbb{S}^{s\times s})^n$ and

$$C = \operatorname{conv}\operatorname{cone}\{y^{\underline{A}} = ({y^{\underline{A}}}_w)_{\leq k} \mid {y^{\underline{A}}}_w = \operatorname{Tr}(w(\underline{A})) \text{ for some } \underline{A} \in \operatorname{supp}\mu\}.$$

The closure of C can be written as the intersection of supporting halfspaces H, that is,

$$\overline{C} = \{z = (z_w)_{\leq k} \mid \forall c \in H: \; c^t z \geq 0\}.$$

Thus $y \in \overline{C}$. We now proceed to show that $y \in \operatorname{rel\,int} \overline{C}$. For this, consider a supporting hyperplane $\{z = (z_w)_{\leq k} \mid c^t z = 0\}$ that does not contain $\overline{C}$ and assume $c^t y = 0$. Let

$$X = \{\underline{A} \in S \mid c^t y^{\underline{A}} > 0\} \text{ and } X_\ell = \{\underline{A} \in S \mid c^t y^{\underline{A}} \geq \frac{1}{\ell}\}.$$

Then $X \neq \varnothing$ and $X = \bigcup_\ell X_\ell$, hence there is some ℓ with $\mu(X_\ell) > 0$. We have

$$0 = c^t y = \int_X c^t y^{\underline{A}} d\mu(\underline{A}) \geq \int_{X_\ell} c^t y^{\underline{A}} d\mu(\underline{A}) \geq \frac{1}{\ell} \int_{X_\ell} d\mu = \frac{1}{\ell}\mu(X_\ell) > 0,$$

a contradiction. This shows $c^t y > 0$ thus $y \in \operatorname{rel\,int} \overline{C} = \operatorname{rel\,int} C$. Whence $y \in C$, as desired. ■

Remark 3.9. Using Carathéodory's theorem, we deduce that y from Theorem 3.8 can be written as a convex combination of at most $N \leq 1 + \sum_{\ell=1}^{k} B_n(\ell)$ tracial sequences $y^{\underline{A}}$, where

$$B_n(\ell) = \begin{cases} \frac{1}{2} N_n(\ell) + \frac{1}{4}(n+1)n^{\ell/2}; & \text{if } \ell \text{ even} \\ \frac{1}{2} N_n(\ell) + \frac{1}{2} n^{(\ell+1)/2}; & \text{if } \ell \text{ odd} \end{cases}$$

is the bracelet number,

$$N_n(\ell) = \frac{1}{\ell} \sum_{d|\ell} \phi\left(\frac{\ell}{d}\right) n^d$$

is the necklace number, and ϕ is the Euler function.

3.3. **Exactness of the Θ^2-relaxation and extraction of trace-optimizers.** In this subsection we shall use our results on the truncated tracial moment problem and flat extensions of tracial moment matrices to detect exactness of the Θ^2-relaxation and to extract global trace-optimizers.

3.3.1. *The flatness condition.* The *tracial moment matrix* $M_k(y)$ of a truncated tracial sequence $y = (y_w)_{\leq 2k}$ is

$$M_k(y) = (y_{u^* v})_{u,v},$$

a matrix indexed by words u, v with $\deg u, \deg v \leq k$. The tracial moment matrix represents the bilinear form on $\mathbb{R}\langle \underline{X} \rangle_{\leq k} \times \mathbb{R}\langle \underline{X} \rangle_{\leq k}$ given by $(f, g) \mapsto L_y(f^* g)$, cf. Example 3.5(b). Hence if y is a truncated tracial moment sequence, then $M_k(y)$ is positive semidefinite.

Example 3.10. A feasible point L of ($\mathrm{DSDP}_{\min}$) with corresponding tracial sequence y_L has a tracial moment matrix $M_L = M_d(y_L)$. Since $L(p^* p) \geq 0$ for all $p \in \mathbb{R}\langle \underline{X} \rangle_{\leq d}$ the tracial moment matrix M_L is positive semidefinite.

Definition 3.11. Let $A \in \mathbb{S}^{s \times s}$ be given. A (symmetric) extension of A is a matrix $\tilde{A} \in \mathbb{S}^{(s+\ell)\times(s+\ell)}$ of the form

$$\tilde{A} = \begin{bmatrix} A & B \\ B^t & C \end{bmatrix}$$

for some $B \in \mathbb{R}^{s \times \ell}$ and $C \in \mathbb{R}^{\ell \times \ell}$. Such an extension is *flat* if $\operatorname{rank} A = \operatorname{rank} \tilde{A}$, or, equivalently, if $B = AZ$ and $C = Z^t A Z$ for some matrix Z.

The property we use is that a truncated tracial sequence $y = (y_w)_{\leq 2k}$ with a positive semidefinite tracial moment matrix $M_k(y)$ which is a flat extension of $M_{k-1}(y)$, is a truncated tracial moment sequence [BK+, Corollary 3.19]. How the finite atomic measure as in (11) can be explicitly constructed we explain in Subsections 3.3.2 and 3.3.3 below.

Theorem 3.12. *If the optimizer L_* of* (DSDP$_{\min}$) *satisfies the* flatness condition, *i.e.,* $M_{L_*} = M_d(y_{L_*})$ *is flat over* $M_{d-1}(y_{L_*})$, *then the Θ^2-relaxation is exact:* $f_{\rm sos} = f^{\rm sos} = f_*$.

Proof. By assumption the tracial moment matrix M_{L_*} is a flat extension of $M_{d-1}(y_{L_*})$. From $L_*(\Theta^2 \cap \mathbb{R}\langle \underline{X}\rangle_{\leq 2d}) \subseteq [0,\infty)$ it follows that M_{L_*} is positive semidefinite. Then, by [BK+, Theorem 3.18], there exists a unique (infinite) tracial extension $\tilde{y}$ of y_{L_*} with tracial moment matrix $M(\tilde{y})$ being a flat extension of M_{L_*}. Thus y_{L_*} is a truncated tracial moment sequence [BK+, Corollary 3.19], and has a finite representation (11). Hence there exist $N \in \mathbb{N}$, $\lambda_i \in \mathbb{R}_{>0}$ with $\sum_i^N \lambda_i = 1$ and tuples $\underline{A}^{(i)} \in (\mathbb{S}^{s\times s})^n$, such that

$$L_*(f) = \sum_{i=1}^N \lambda_i \operatorname{Tr}(f(\underline{A}^{(i)})).$$

Since L_* is the optimizer of (DSDP$_{\min}$), we have $L_*(f) = f^{\rm sos} = f_{\rm sos}$. Further,

$$\operatorname{Tr}(f(\underline{A}^{(i)})) \geq f_{\rm sos}$$

for each $i = 1, \ldots, N$. Hence

$$f_* \leq \operatorname{Tr}(f(\underline{A}^{(i)})) = f_{\rm sos} \leq f_*.$$

Thus the minimum $f_* = f_{\rm sos}$ is attained at each of the $\underline{A}^{(i)}$. ∎

For the rest of this section assume $f \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}$ is such that the optimizer L of (DSDP$_{\min}$) is flat. By Theorem 3.12, $f_* = f_{\rm sos} = f^{\rm sos}$. In the next two subsections we explain how to *construct* the trace-minimizing tuples $\underline{A}^{(i)}$ for f.

3.3.2. *GNS construction.* In this subsection we use the Gelfand-Naimark-Segal (GNS) construction to associate a matrix $*$-algebra $\mathcal{A}$ to L.

Since $M_d = M_d(y_L)$ is flat over $M_{d-1} = M_{d-1}(y_L)$, there exist $s = \operatorname{rank} M_d$ linear independent columns of M_{d-1} labeled by words $w \in \langle \underline{X}\rangle$ with $\deg w \leq d-1$ which form a basis $\mathcal{B}$ of $E = \operatorname{Ran} M_d$, the range of M_d. Now L (or M_d) induces a positive definite bilinear form (i.e., a scalar product) $\langle \_,\_ \rangle_E$ on E.

Let $\hat{X}_i$ be the right multiplication with X_i on E, i.e., if $\overline{w}$ denotes the column of M_d labeled by $w \in \langle \underline{X}\rangle_{\leq d}$, then $\hat{X}_i \overline{u} := \overline{uX_i}$ for $u \in \langle \underline{X}\rangle_{\leq d-1}$. The operator $\hat{X}_i$ is well defined and symmetric by the tracial property of L:

$$\langle \hat{X}_i \overline{p}, \overline{q}\rangle_E = L(X_i p^* q) = L(p^* q X_i) = \langle \overline{p}, \hat{X}_i \overline{q}\rangle_E.$$

Therefore we can construct matrix representations $A_i \in \mathbb{S}^{s\times s}$ of these multiplication operators $\hat{X}_i$ by calculating their image according to our chosen basis $\mathcal{B}$. To be more specific, $\hat{X}_i \overline{u}_1$ for $u_1 \in \langle \underline{X}\rangle_{\leq d-1}$ being the first label in $\mathcal{B}$, can be written as a unique linear combination $\sum_{j=1}^s \lambda_j \overline{u}_j$ with words u_j labeling $\mathcal{B}$ such that $L\big((u_1X_i - \sum \lambda_j u_j)^*(u_1X_i - \sum \lambda_j u_j)\big) = 0$. Then $\begin{bmatrix}\lambda_1 & \ldots & \lambda_s\end{bmatrix}^t$ will be the first column of A_i.

Remark 3.13. We note there is an alternative and more abstract approach to the construction of the $\hat{X}_i$ based upon properties of flat moment matrices. Let $\tilde{L} : \mathbb{R}\langle \underline{X}\rangle \to \mathbb{R}$ be the linear functional corresponding to the unique flat extension $\tilde{y}$ of y_L [BK+, Theorem 3.18]. Since $\tilde{L}|_{\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}} = L$ we write L instead of $\tilde{L}$. Equip $\mathbb{R}\langle \underline{X}\rangle$ with the bilinear form given by

$$\langle p, q\rangle := L(p^* q).$$

Let $I = \{p \in \mathbb{R}\langle \underline{X}\rangle \mid L(p^*p) = 0\}$. By [BK+, Proposition 3.7], I is an ideal of $\mathbb{R}\langle \underline{X}\rangle$. Thus $E := \mathbb{R}\langle \underline{X}\rangle / I$ with the induced scalar product is a Hilbert space of dimension $\operatorname{rank} M_d(y) < \infty$.

Let $\hat{X}_i$ be the right regular representation of X_i on E, i.e., $\hat{X}_i\overline{p} := \overline{pX_i}$ for $\overline{p} = p + I \in E$. The operator $\hat{X}_i$ is well defined and symmetric with respect to the scalar product induced by L. The construction of the matrices A_i is now similar as above.

Let $\mathcal{A}$ denote the unital ($*$-)subalgebra of $\mathbb{R}^{s\times s}$ generated by $A_1, \dots, A_n$.

3.3.3. *Artin-Wedderburn block decomposition.* The matrix $*$-algebra $\mathcal{A}$ is semisimple and thus admits an Artin-Wedderburn block decomposition [Lam91, (3.5)]. In this subsection we employ this block decomposition of $\mathcal{A}$; each of the blocks obtained will yield a trace-minimizer of f.

Elements of $\mathcal{A}$ can be presented as $\hat{p} := p(A_1, \dots, A_n)$ for $p \in \mathbb{R}\langle \underline{X} \rangle$. Let $\hat{L} : \mathcal{A} \to \mathbb{R}$ be the induced linear functional given by $\hat{L}(\hat{p}) = L(p)$. By construction, $\hat{L}$ is a tracial state, that is, $\hat{L}$ maps positive semidefinite matrices to nonnegative scalars, $\hat{L}(1) = 1$, and $\hat{L}$ vanishes on commutators.

By [BK+, Proposition 3.13], the tracial state $\hat{L}$ is given by a conic combination of normalized traces on the Artin-Wedderburn blocks of $\mathcal{A}$. More precisely, there exist unital $*$-subalgebras $\mathcal{A}^{(i)}$ of $\mathbb{R}^{s\times s}$, each isomorphic to a full matrix algebra over $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$, a $*$-isomorphism

$$\mathcal{A} \to \bigoplus_{i=1}^{N} \mathcal{A}^{(i)}, \tag{12}$$

and $\lambda_1, \dots, \lambda_N \in \mathbb{R}_{>0}$ with $\sum_i \lambda_i = 1$, such that for all $A \in \mathcal{A}$,

$$\hat{L}(A) = \sum_{i=1}^{N} \lambda_i \operatorname{Tr}(A^{(i)}).$$

Here, $\bigoplus_i A^{(i)}$ denotes the image of A under the isomorphism (12). In particular,

$$L(p) = \hat{L}(\hat{p}) = \sum_{i=1}^{N} \lambda_i \operatorname{Tr}(p(A_1^{(i)}, \dots, A_n^{(i)})) \quad \text{for} \quad p \in \mathbb{R}\langle \underline{X} \rangle. \tag{13}$$

As $\operatorname{Tr}(f(A_1^{(i)}, \dots, A_n^{(i)})) \geq f_* \geq L(f)$ for all i, (13) implies $L(f) = \operatorname{Tr}(f(A_1^{(i)}, \dots, A_n^{(i)}))$. That is, each of the tuples $(A_1^{(i)}, \dots, A_n^{(i)})$ is a trace-minimizer for f.

3.3.4. *Implementation.* All steps in our algorithm to extract trace-minimizers are straightforward with the possible exception of the last one where one has to *construct* for given matrices $A_j \in \mathbb{S}^{s\times s}$, the matrices $A_j^{(i)}$ as in Subsection 3.3.3, i.e. one has to implement the decomposition of $\mathcal{A}$ into simple components. The first efficient algorithm to decompose a semisimple algebra over a number field into simple components goes back to Friedl and Rónyai [FR85]. Later, Eberly and Giesbrecht [EG04] modified their method to obtain an efficient algorithm to find the simple components of a separable algebra over an infinite field by decomposing its center. In particular, their algorithm works for semisimple algebras over a field of characteristic 0. One can also employ the Murota, Kanno, Kojima, Kojima, and Maehara probabilistic method [MKKK10, MM10] which produces an orthogonal change of basis U for $\mathbb{R}^s$ so that the matrix $*$-algebra $\mathcal{A} \subseteq \mathbb{R}^{s\times s}$ decomposes into a direct sum of simple matrix algebras $\mathcal{A}^{(i)}$ which cannot be further decomposed. Then $U^t A_j U = \oplus_i A_j^{(i)}$.

The entire algorithm using the probabilistic method of Murota et al. has been implemented in `NCSOStools` [CKP+]. We conclude by an example.

Example 3.14. Let

$$f = 3 + X_1^2 + 2X_1^3 + 2X_1^4 + X_1^6 - 4X_1^4X_2 + X_1^4X_2^2 + 4X_1^3X_2 + 2X_1^3X_2^2 - 2X_1^3X_2^3 \\ + 2X_1^2X_2 - X_1^2X_2^2 + 8X_1X_2X_1X_2 + 2X_1^2X_2^3 - 4X_1X_2 + 4X_1X_2^2 + 6X_1X_2^4 - 2X_2 \\ + X_2^2 - 4X_2^3 + 2X_2^4 + 2X_2^6.$$

The minimum of f on $\mathbb{R}^2$ is 1.0797. Using `NCcycMin` we obtain the floating-point trace-minimum $f_{\text{sos}} = 0.2842$ for f which is different from the commutative minimum. In particular, the minimizers will not be scalar matrices. The tracial moment matrix M_{L_*} of the optimizer L_* in (DSDP_{min}) is of rank 4 and flat over $M_2(y_{L_*})$. Thus the matrix representation of the multiplication operators $\hat{X}_i$ is given by 4×4 matrices:

$$\hat{X}_1 = \begin{bmatrix} -1.0761 & 0.1802 & 0.5107 & 0.2590 \\ 0.1802 & -0.3393 & -0.1920 & 0.9428 \\ 0.5107 & -0.1920 & 0.5094 & 0.0600 \\ 0.2590 & 0.9428 & 0.0600 & -0.3020 \end{bmatrix},$$

$$\hat{X}_2 = \begin{bmatrix} 0.7108 & 0.7328 & 0.1043 & 0.4415 \\ 0.7328 & -0.3706 & 0.4757 & -0.2147 \\ 0.1043 & 0.4757 & 0.0776 & -0.9102 \\ 0.4415 & -0.2147 & -0.9102 & 0.1393 \end{bmatrix}.$$

The Artin-Wedderburn decomposition for the matrix $*$-algebra $\mathcal{A}$ generated by $\hat{X}_1, \hat{X}_2$ gives in this case only one block. Using `NCcycOpt` leads to the trace-minimizer

$$A_1 = \begin{bmatrix} -1.1843 & 0 & -0.2095 & 0.3705 \\ 0 & -1.1843 & 0.3705 & 0.2095 \\ -0.2095 & 0.3705 & 0.5803 & 0 \\ 0.3705 & 0.2095 & 0 & 0.5803 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} -0.1743 & 0 & 0.4851 & -0.8577 \\ 0 & -0.1743 & -0.8577 & -0.4851 \\ 0.4851 & -0.8577 & 0.4529 & 0 \\ -0.8577 & -0.4851 & 0 & 0.4529 \end{bmatrix}.$$

The reader can easily verify that $\operatorname{Tr}(f(A_1, A_2)) = 0.2842$.

Note that $\mathcal{A}$ is (as a real $*$-algebra) isomorphic to $M_2(\mathbb{C})$. For instance,

$$A_1 = \begin{bmatrix} -1.1843 & 0.3705 - 0.2095\mathbb{i} \\ 0.3705 + 0.2095\mathbb{i} & 0.5803 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -0.1743 & -0.8577 + 0.4851\mathbb{i} \\ -0.8577 - 0.4851\mathbb{i} & 0.4529 \end{bmatrix}.$$

In this case it is possible to find a unitary matrix $U \in \mathbb{C}^{2\times 2}$ with $A_j' = U^* A_j U \in \mathbb{R}^{2\times 2}$, e.g.

$$U = \begin{bmatrix} 0.180122 - 0.0473861\mathbb{i} & 0.950143 - 0.250076\mathbb{i} \\ 0.950143 + 0.250076\mathbb{i} & -0.180122 - 0.0473861\mathbb{i} \end{bmatrix},$$

$$A_1' = \begin{bmatrix} 0.674861 & 0.0731923 \\ 0.0731923 & -1.27886 \end{bmatrix}, \quad A_2' = \begin{bmatrix} 0.0705101 & -1.03179 \\ -1.03179 & 0.20809 \end{bmatrix}.$$

Then $(A_1', A_2') \in \left(\mathbb{S}^{2\times 2}\right)^2$ is also a trace-minimizer for f.

Acknowledgments. The authors thank Markus Schweighofer, Scott McCullough, and Marko Kandić for sharing their expertise.

References

[Akh65] N.I. Akhiezer. *The classical moment problem and some related questions in analysis.* Translated by N. Kemmer. Hafner Publishing Co., New York, 1965. 9

[BK+] S. Burgdorf and I. Klep. The truncated tracial moment problem. To appear in *J. Operator Theory*, http://arxiv.org/abs/1001.3679. 4, 7, 8, 9, 12, 13, 14

[BMV75] D. Bessis, P. Moussa, and M. Villani. Monotonic converging variational approximations to the functional integrals in quantum statistical mechanics. *J. Math. Phys.*, 16(11):2318–2325, 1975. 2, 7

[BT06] C. Bayer and J. Teichmann. The proof of Tchakaloff's theorem. *Proc. Amer. Math. Soc.*, 134(10):3035–3040, 2006. 4, 9, 11

[BTN01] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization.* MPS/SIAM Series on Optimization. SIAM, Philadelphia, PA, 2001. 7

[Bur11] S. Burgdorf. Sums of Hermitian squares as an approach to the BMV conjecture. *Linear Multilinear Algebra* 59(1):1–9, 2011.

[CF96] R.E. Curto and L.A. Fialkow. Solution of the truncated complex moment problem for flat data. *Mem. Amer. Math. Soc.*, 119(568):x+52, 1996. 9

[Cim10] J. Cimprič. A method for computing lowest eigenvalues of symmetric polynomial differential operators by semidefinite programming. *J. Math. Anal. Appl.*, 369(2):443–452, 2010. 2

[CKP+] K. Cafuta, I. Klep, and J. Povh. NCSOStools: a computer algebra system for symbolic and numerical computation with noncommutative polynomials. To appear in *Optim. Methods Softw.* http://ncsostools.fis.unm.si 1, 2, 7, 14

[CKP10] K. Cafuta, I. Klep, and J. Povh. On the nonexistence of sum of squares certificates for the BMV conjecture. *J. Math. Phys.*, 51:083521, 2010.

[CLR95] M.D. Choi, T.Y. Lam, and B. Reznick. Sums of squares of real polynomials. In *K-theory and algebraic geometry: connections with quadratic forms and division algebras*, volume 58 of *Proc. Sympos. Pure Math.*, pages 103–126. AMS, Providence, RI, 1995. 6

[Con76] A. Connes. Classification of injective factors. Cases II_1, II_∞, III_λ, $\lambda \neq 1$. *Ann. of Math. (2)*, 104:73–115, 1976. 2, 7

[dOHMP08] M.C. de Oliveira, J.W. Helton, S. McCullough, and M. Putinar. Engineering systems and free semi-algebraic geometry. In *Emerging Applications of Algebraic Geometry*, volume 149 of *IMA Vol. Math. Appl.*, pages 17–62. Springer, 2008. 1, 2

[DLTW08] A.C. Doherty, Y.-C. Liang, B. Toner, and S. Wehner. The quantum moment problem and bounds on entangled multi-prover games. In *Twenty-Third Annual IEEE Conference on Computational Complexity*, pages 199-210. IEEE Computer Soc., Los Alamitos, CA, 2008. 2

[EG04] W. Eberly and M. Giesbrecht. Efficient decomposition of separable algebras. *J. Symbolic Comput.*, 37: 35–81, 2004 14

[FR85] K. Friedl and L. Rónyai. Polynomial time solutions of some problems in computational algebra. *Symp. on Theory of Computing*, Amer. Math. Soc., 17:153–162, 1985 14

[Gla63] R.J. Glauber. The quantum theory of optical coherence. *Phys. Rev.*, 130(6):2529–2539, 1963. 2

[HdOMS] J.W. Helton, M. de Oliveira, R.L. Miller, and M. Stankus. NCAlgebra: A Mathematica package for doing non commuting algebra. http://www.math.ucsd.edu/~ncalg/. 2

[Hel02] J.W. Helton. "Positive" noncommutative polynomials are sums of squares. *Ann. of Math. (2)*, 156(2):675–694, 2002. 2, 5, 6

[HL05] D. Henrion and J.-B. Lasserre. Detecting global optimality and extracting solutions in GloptiPoly. In *Positive polynomials in control*, volume 312 of *Lecture Notes in Control and Inform. Sci.*, pages 293–310. Springer, Berlin, 2005. 7

[HLL09] D. Henrion, J.-B. Lasserre, and J. Löfberg. GloptiPoly 3: moments, optimization and semidefinite programming. *Optim. Methods Softw.* 24(4-5): 761–779, 2009. http://www.laas.fr/~henrion/software/gloptipoly3/ 1, 7

[KP10] I. Klep and J. Povh. Semidefinite programming and sums of hermitian squares of noncommutative polynomials. *J. Pure Appl. Algebra*, 214:740–749, 2010. 1, 6, 8

[KS08a] I. Klep and M. Schweighofer. Connes' embedding conjecture and sums of Hermitian squares. *Adv. Math.*, 217(4):1816–1837, 2008. 2, 5, 6, 7, 8

[KS08b] I. Klep and M. Schweighofer. Sums of Hermitian squares and the BMV conjecture. *J. Stat. Phys*, 133(4):739–760, 2008. 6, 7

[Lam91] T.Y. Lam. *A first course in noncommutative rings*, volume 131 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1991. 14

[Las01] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2000/01. 2, 4

[Las09] J.B. Lasserre. *Moments, Positive Polynomials and Their Applications*, volume 1. Imperial College Press, 2009. 2, 4

[Lau09] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, volume 149 of *IMA Vol. Math. Appl.*, pages 157–270. Springer, New York, 2009. 9, 11

[Löf04] J. Löfberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004. http://control.ee.ethz.ch/~joloef/yalmip.php. 1

[Maz04] D.A. Mazziotti. Realization of quantum chemistry without wave functions through first-order semidefinite programming. *Phys. Rev. Lett.*, 93(21):213001, 4, 2004. 2

[McC01] S. McCullough. Factorization of operator-valued polynomials in several non-commuting variables. *Linear Algebra Appl.*, 326(1-3):193–203, 2001. 5, 9

[Mit03] D. Mittelmann. An independent benchmarking of SDP and SOCP solvers. *Math. Program.*, 95(2, Ser. B):407–430, 2003. http://plato.asu.edu/sub/pns.html. 7

[MKKK10] K. Murota, Y. Kanno, M. Kojima, and S. Kojima. A numerical algorithm for block-diagonal decomposition of matrix ∗-algebras with application to semidefinite programming. *Japan J. Indust. Appl. Math.*, 27(1): 125–160, 2010. 4, 14

[MM10] T. Maehara and K. Murota. A numerical algorithm for block-diagonal decomposition of matrix ∗-algebras with general irreducible components. *Japan J. Indust. Appl. Math.*, 27(2): 263–293, 2010. 4, 14

[MPRW09] J. Malick, J. Povh, F. Rendl, and A. Wiegele. Regularization methods for semidefinite programming. *SIAM J. Optim.*, 20(1):336–356, 2009. 7

[NN94] Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*, volume 13 of *SIAM Studies in Applied Mathematics*. SIAM, Philadelphia, PA, 1994. 7

[Par03] P.A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Program.*, 96(2, Ser. B):293–320, 2003. 2, 4, 6

[PNA10] S. Pironio, M. Navascues, and A. Acin. Convergent relaxations of polynomial optimization problems with non-commuting variables. *SIAM J. Optim.* 20(5): 2157-2180, 2010. 1, 2, 4, 7

[PPSP05] S. Prajna, A. Papachristodoulou, P. Seiler, and P.A. Parrilo. SOSTOOLS and its control applications. In *Positive polynomials in control*, volume 312 of *Lecture Notes in Control and Inform. Sci.*, pages 273–292. Springer, Berlin, 2005. http://www.cds.caltech.edu/sostools/. 1

[PS03] P.A. Parrilo and B. Sturmfels. Minimizing polynomial functions. In *Algorithmic and quantitative real algebraic geometry (Piscataway, NJ, 2001)*, volume 60 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 83–99, Amer. Math. Soc., Providence, RI, 2003. 2, 4, 6

[PV09] K.F. Pál and T. Vértesi. Quantum bounds on Bell inequalities. *Phys. Rev. A (3)*, 79(2):022120, 12, 2009. 2

[Ram97] M.V. Ramana. An exact duality theory for semidefinite programming and its complexity implications. *Math. Program. Series B*, 77(2): 129–162, 1997. 7

[Sto01] J. Stochel. Solving the truncated moment problem solves the full moment problem. *Glasg. Math. J.*, 43(3):335–341, 2001. 9

[Stu99] J.F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.*, 11/12(1-4):625–653, 1999. http://sedumi.ie.lehigh.edu/. 7

[TTT99] K.-C. Toh, M.J. Todd, and R.-H. Tütüncü. SDPT3 a Matlab software package for semidenite programming, *Optim. Methods Softw.*, 11:545–581, 1999 http://www.math.nus.edu.sg/~mattohkc/sdpt3.html. 7

[WKKMS09] H. Waki, S. Kim, M. Kojima, M. Muramatsu, and H. Sugimoto. Algorithm 883: sparsePOP—a sparse semidefinite programming relaxation of polynomial optimization problems. *ACM Trans. Math. Software*, 35(2):Art. 15, 13, 2009. 1

[WSV00] H. Wolkowicz, R. Saigal, and L. Vandenberghe. *Handbook of Semidefinite Programming.* Kluwer, 2000. 7

[YFK03] M. Yamashita, K. Fujisawa, and M. Kojima. Implementation and evaluation of SDPA 6.0 (semidefinite programming algorithm 6.0). *Optim. Methods Softw.*, 18(4):491–505, 2003. http://sdpa.indsys.chuo-u.ac.jp/sdpa/. 7

Sabine Burgdorf, Universität Konstanz, Fachbereich Mathematik und Statistik, 78457 Konstanz, Germany, and Institut de Recherche Mathématiques de Rennes, Université de Rennes 1, Campus de Beaulieu, 35042 Rennes cedex, France

E-mail address: sabine.burgdorf@uni-konstanz.de

Kristijan Cafuta, Univerza v Ljubljani, Fakulteta za elektrotehniko, Laboratorij za uporabno matematiko, Tržaška 25, 1000 Ljubljana, Slovenia

E-mail address: kristijan.cafuta@fe.uni-lj.si

Igor Klep, Univerza v Mariboru, Fakulteta za naravoslovje in matematiko, Koroška 160, 2000 Maribor, and Univerza v Ljubljani, Fakulteta za matematiko in fiziko, Jadranska 19, 1111 Ljubljana, Slovenia

E-mail address: igor.klep@fmf.uni-lj.si

Janez Povh, Fakulteta za informacijske študije v Novem mestu, Novi trg 5, 8000 Novo mesto, Slovenia

E-mail address: janez.povh@fis.unm.si

NOT FOR PUBLICATION

Contents