NCSOSTOOLS: A COMPUTER ALGEBRA SYSTEM FOR SYMBOLIC AND NUMERICAL COMPUTATION WITH NONCOMMUTATIVE POLYNOMIALS

KRISTIJAN CAFUTA, IGOR KLEP[1], AND JANEZ POVH[2]

ABSTRACT. `NCSOStools` is a Matlab toolbox for

- symbolic computation with polynomials in noncommuting variables;
- constructing and solving sum of hermitian squares (with commutators) programs for polynomials in noncommuting variables.

It can be used in combination with semidefinite programming software, such as SeDuMi, SDPA or SDPT3 to solve these constructed programs.

This paper provides an overview of the theoretical underpinning of these sum of hermitian squares (with commutators) programs, and provides a gentle introduction to the primary features of `NCSOStools`.

1. INTRODUCTION

Starting with Helton's seminal paper [Hel02], *free semialgebraic geometry* is being established. Among the things that make this area exciting are its many facets of applications. A nice survey on applications to control theory, systems engineering and optimization is given in [dOHMP08], while applications to mathematical physics and operator algebras have been given by the second author [KS08a, KS08b].

Unlike classical semialgebraic (or real algebraic) geometry where real polynomial rings in *commuting* variables are the objects of study, free semialgebraic geometry deals with real polynomials in *noncommuting* (NC) variables and their finite-dimensional representations. Of interest are various notions of *positivity* induced by these. For instance, positivity via positive semidefiniteness or the positivity of the trace. Both of these can be reformulated and studied using sums of hermitian squares (with commutators) and semidefinite programming.

We developed `NCSOStools` as a consequence of this recent interest in noncommutative positivity and sums of (hermitian) squares (SOHS). `NCSOStools` is an open source Matlab toolbox for solving SOHS problems using semidefinite programming. As a side product our toolbox implements symbolic computation with noncommuting variables in Matlab.

Date: May 30, 2010.

2000 *Mathematics Subject Classification.* Primary 90C22, 13J30; Secondary 11E25, 08B20, 90C90.

Key words and phrases. noncommutative polynomial, sum of hermitian squares, commutator, semidefinite programming, Matlab toolbox.

[1]Partially supported by the Slovenian Research Agency (project no. Z1-9570-0101-06).

[2]Supported by the Slovenian Research Agency (project no. 1000-08-210518).

There is a small overlap in features with Helton's `NCAlgebra` package for Mathematica [HdOMS]. However, `NCSOStools` performs only basic manipulations with noncommuting variables, while `NCAlgebra` is a fully-fledged add-on for symbolic computation with polynomials, matrices and rational functions in noncommuting variables. However our primary interest is with the different notions of positivity and sum of hermitian squares (with commutators) problems, where semidefinite programming plays an important role, and we feel that for constructing and solving these problems Matlab is the optimal framework.

Readers interested in solving sums of squares problems for commuting polynomials are referred to one of the many great existing packages, such as SOSTOOLS [PPSP05], SparsePOP [WKK$^+$09], GloptiPoly [HLL09], or YALMIP [Löf04].

This paper is organized as follows. The first section fixes notation and introduces terminology. Then in Section 2 we introduce the central objects, *sums of hermitian squares* and use these to study positive semidefinite NC polynomials. The natural correspondence between sums of hermitian squares and semidefinite programming is also explained in some detail. The main theoretical contribution here is an algorithm to extract an eigenvalue minimizer of an NC polynomial. Section 3 is brief, works on the symbolic level and introduces commutators and cyclic equivalence. These notions are used in Section 4 to study trace-positive NC polynomials using sums of hermitian squares and commutators. Such representations can again be found using semidefinite programming. Section 5 touches upon two notions of convexity. The last section contains an expanded example demonstrating some of the features of our toolbox. For a list of all available commands and more detailed documentation we refer the reader to our website:

`http://ncsostools.fis.unm.si/documentation`

1.1. **Notation.** We write $\mathbb{N} := \{1, 2, \dots\}$, $\mathbb{R}$ for the sets of natural and real numbers. 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 algebra $\mathbb{R}\langle \underline{X} \rangle$ of polynomials in n noncommuting variables $\underline{X} = (X_1, \dots, X_n)$ with coefficients from $\mathbb{R}$. 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*. 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 $\deg f$. We shall also consider the degree of f in X_i, $\deg_i f$. Similarly, the length of the shortest word appearing in $f \in \mathbb{R}\langle \underline{X} \rangle$ is called the *min-degree* of f and denoted by $\operatorname{mindeg} f$. Likewise, $\operatorname{mindeg}_i f$ is introduced. If the variable X_i does not occur in some monomial in f, then $\operatorname{mindeg}_i f = 0$. For instance, if $f = X_1^3 - 3X_3X_2X_1 + 2X_4X_1^2X_4$, then

$$\deg f = 4, \quad \deg_1 f = 3, \quad \deg_2 f = \deg_3 f = 1, \quad \deg_4 f = 2,$$

$$\operatorname{mindeg} f = 3, \quad \operatorname{mindeg}_1 f = 1, \quad \operatorname{mindeg}_2 f = \operatorname{mindeg}_3 f = \operatorname{mindeg}_4 f = 0.$$

An element of the form aw where $0 \neq a \in \mathbb{R}$ and $w \in \langle \underline{X} \rangle$ is called a *monomial* and a its *coefficient*. Hence words are monomials whose coefficient is 1.

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.

$$(X_1^2 - X_2X_3X_1)^* = X_1^2 - X_1X_3X_2.$$

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,

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

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 shall also use V^t to denote the row vector with components v_i.

2. Positive semidefinite NC polynomials

A symmetric matrix $A \in \mathbb{R}^{s\times s}$ is positive semidefinite if and only if it is of the form B^tB for some $B \in \mathbb{R}^{s\times s}$. In this section we introduce the notion of *sum of hermitian squares* (SOHS) and explain its relation with semidefinite programming.

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. A polynomial $f \in \mathbb{R}\langle \underline{X} \rangle$ is SOHS if it belongs to Σ^2. Clearly, $\Sigma^2 \subsetneq \operatorname{Sym}\mathbb{R}\langle \underline{X} \rangle$. For example,

$$X_1X_2 + 2X_2X_1 \notin \operatorname{Sym}\mathbb{R}\langle \underline{X} \rangle, \quad X_1^2X_2X_1^2 \in \operatorname{Sym}\mathbb{R}\langle \underline{X} \rangle \setminus \Sigma^2,$$

$$2 + X_1X_2 + X_2X_1 + X_1X_2^2X_1 = 1 + (1 + X_2X_1)^*(1 + X_2X_1) \in \Sigma^2.$$

If $f \in \mathbb{R}\langle \underline{X} \rangle$ is SOHS and we substitute symmetric matrices $A_1, \dots, A_n$ of the same size for the variables $\underline{X}$, then the resulting matrix $f(A_1, \dots, A_n)$ is positive semidefinite. Helton [Hel02] and McCullough [McC01] proved (a slight variant of) the converse of the above observation: if $f \in \mathbb{R}\langle \underline{X} \rangle$ and $f(A_1, \dots, A_n) \succeq 0$ for *all* symmetric matrices A_i of the same size, then f is SOHS. For a beautiful exposition, we refer the reader to [MP05].

The following proposition (cf. [Hel02, §2.2] or [MP05, Theorem 2.1]) is the noncommutative version of the classical result due to Choi, Lam and Reznick ([CLR95, §2]; see also [Par03, PW98]). The easy proof is included for the sake of completeness.

Proposition 2.1. *Suppose $f \in \operatorname{Sym}\mathbb{R}\langle \underline{X} \rangle$ is of degree $\leq 2d$. Then $f \in \Sigma^2$ if and only if there exists a positive semidefinite matrix G satisfying*

$$f = W_d^* G W_d = \sum_{i,j} G_{i,j}(W_d)_i^*(W_d)_j, \tag{1}$$

where W_d is a vector consisting of all words in $\langle \underline{X} \rangle$ of degree $\leq d$.

Conversely, given such a positive semidefinite matrix G with rank r, one can construct NC polynomials $g_1, \dots, g_r \in \mathbb{R}\langle \underline{X} \rangle$ of degree $\leq d$ such that

$$f = \sum_{i=1}^{r} g_i^* g_i. \tag{2}$$

The matrix G is called *a Gram matrix* for f.

Proof. If $f = \sum_i g_i^* g_i \in \Sigma^2$, then $\deg g_i \leq d$ for all i as the highest degree terms cannot cancel. Indeed, otherwise by extracting all the appropriate highest degree terms h_i with degree $> d$ from the g_i we would obtain $h_i \in \mathbb{R}\langle \underline{X} \rangle \setminus \{0\}$ satisfying

$$\sum_i h_i^* h_i = 0. \tag{3}$$

By substituting symmetric matrices for variables in (3), we see that each h_i vanishes for all these substitutions. But then the nonexistence of (dimension-free) polynomial identities for tuples of symmetric matrices (cf. [Row80, §2.5, §1.4]) implies $h_j = 0$ for all j. Contradiction.

Hence we can write $g_i = G_i^t W_d$, where G_i^t is the (row) vector consisting of the coefficients of g_i. Then $g_i^* g_i = W_d^* G_i G_i^t W_d$ and setting $G := \sum_i G_i G_i^t$, (1) clearly holds.

Conversely, given a positive semidefinite $G \in \mathbb{R}^{N \times N}$ of rank r satisfying (1), write $G = \sum_{i=1}^r G_i G_i^t$ for $G_i \in \mathbb{R}^{N \times 1}$. Defining $g_i := G_i^t W_d$ yields (2). ■

Example 2.2. In this example we consider NC polynomials in 2 variables which we denote by X, Y. Let

$$f = 1 - 2X + 2X^2 + Y^2 - 2X^2Y - 2YX^2 + 2YXY + 2YX^2Y.$$

Let V be the subvector $\begin{bmatrix} 1 & X & Y & XY \end{bmatrix}^t$ of W_2. Then the Gram matrix for f with respect to V is given by

$$G(a) := \begin{bmatrix} 1 & -1 & 0 & a \\ -1 & 2 & -a & -2 \\ 0 & -a & 1 & 1 \\ a & -2 & 1 & 2 \end{bmatrix}.$$

(That is, $f = V^* G(a) V$.) This matrix is positive semidefinite if and only if $a = 1$ as follows easily from the characteristic polynomial of $G(a)$. Moreover, $G(1) = C^t C$ for

$$C = \begin{bmatrix} 1 & -1 & 0 & 1 \\ 0 & 1 & -1 & -1 \end{bmatrix}.$$

From

$$CV = \begin{bmatrix} 1 - X + XY & X - Y - XY \end{bmatrix}^t$$

it follows that

$$f = (1 - X + XY)^*(1 - X + XY) + (X - Y - XY)^*(X - Y - XY) \in \Sigma^2.$$

The problem whether a given polynomial is SOHS is therefore a special instance of a semidefinite feasibility problem. This is explained in detail in the following two subsections.

2.1. **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, $A_1, \ldots, A_m$ of the same size over $\mathbb{R}$ and a vector $b \in \mathbb{R}^m$, we formulate a *semidefinite program in standard primal form* as follows:

$$\text{(PSDP)} \qquad \begin{array}{rrcl} \inf & \langle C, G \rangle & & \\ \text{s. t.} & \langle A_i, G \rangle & = & b_i, \quad i = 1, \ldots, m \\ & G & \succeq & 0. \end{array}$$

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

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

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

The importance of semidefinite programming was spurred by the development of efficient methods which can find an ε-optimal solution in a polynomial time in s, m and $\log \varepsilon$, where s is the order of matrix variables G and Z and m is the number of linear constraints. There exist several open source packages which find such 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 [PRW06, MPRW09]).

Our standard reference for SDP is [Tod01].

2.2. **Sums of hermitian squares and SDP.** In this subsection we present a *conceptual algorithm* based on SDP for checking whether a given $f \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle$ is SOHS. Following Proposition 2.1 we must determine whether there exists a positive semidefinite matrix G such that $f = W_d^* G W_d$, where W_d is the vector of all words of degree $\leq d$. This is a semidefinite feasibility problem in the matrix variable G, where the constraints $\langle A_i, G\rangle = b_i$ are implied by the fact that for each product of monomials $w \in \{p^*q \mid p, q \in W_d\}$ the following must be true:

$$\sum_{\substack{p,q \in W_d \\ p^*q = w}} G_{p,q} = a_w, \tag{4}$$

where a_w is the coefficient of w in f ($a_w = 0$ if the monomial w does not appear in f).

Any input polynomial f is symmetric, so $a_w = a_{w^*}$ for all w, and equations (4) can be rewritten as

$$\sum_{\substack{u,v \in W_d \\ u^*v = w}} G_{u,v} + \sum_{\substack{u,v \in W_d \\ u^*v = w^*}} G_{u,v} = a_w + a_{w^*} \quad \forall w \in \{p^*q \mid p, q \in W_d\}, \tag{5}$$

or equivalently,

$$\langle A_w, G\rangle = a_w + a_{w^*} \quad \forall w \in \{p^*q \mid p, q \in W_d\}, \tag{6}$$

where A_w is the symmetric matrix defined by

$$(A_w)_{u,v} = \begin{cases} 2; & \text{if } u^*v \in \{w, w^*\},\ w^* = w, \\ 1; & \text{if } u^*v \in \{w, w^*\},\ w^* \neq w, \\ 0; & \text{otherwise.} \end{cases}$$

Note: $A_w = A_{w^*}$ for all w.

As we are interested in an arbitrary positive semidefinite $G = [G_{u,v}]_{u,v \in W}$ satisfying the constraints (6), we can choose the objective function freely. However, in practice one prefers solutions of small rank leading to shorter SOHS decompositions. Hence we minimize the trace, a commonly used heuristic for matrix rank minimization [RFP]. Therefore our SDP in the primal form is as follows:

$$(\mathrm{SOHS_{SDP}}) \qquad \begin{array}{rrcl} \inf & \langle I, G\rangle & & \\ \text{s.t.} & \langle A_w, G\rangle & = & a_w + a_{w^*} \quad \forall w \in \{p^*q \mid p, q \in W_d\} \\ & G & \succeq & 0. \end{array}$$

(Here and in the sequel, I denotes the identity matrix of appropriate size.) To reduce the size of this SDP (i.e., to make W_d smaller), we may employ the following simple observation:

Proposition 2.3. *Let $f \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle$, let $m_i := \frac{\operatorname{mindeg}_i f}{2}$, $M_i := \frac{\deg_i f}{2}$, $m := \frac{\operatorname{mindeg} f}{2}$, $M := \frac{\deg f}{2}$. Set*

$$V := \{w \in \langle \underline{X}\rangle \mid m_i \leq \deg_i w \leq M_i \text{ for all } i,\ m \leq \deg w \leq M\}.$$

Then $f \in \Sigma^2$ if and only if there exists a positive semidefinite matrix G satisfying $f = V^*GV$.

Proof. This follows from the fact that the highest or lowest degree terms in a SOHS decomposition cannot cancel. ■

Example 2.4 (Example 2.2 revisited)**.** Let us return to

$$f = 1 - 2X + 2X^2 + Y^2 - 2X^2Y - 2YX^2 + 2YXY + 2YX^2Y.$$

We shall describe in some detail ($\mathrm{SOHS_{SDP}}$) for f. From Proposition 2.3, we obtain

$$V = \begin{bmatrix} 1 & X & Y & XY & YX \end{bmatrix}^t.$$

Thus G is a symmetric 5×5 matrix and there will be 17 matrices A_w, as $|\{u^*v \mid u, v \in V\}| = 17$. In fact, there are only 13 different matrices A_w as $A_w = A_{w^*}$. Here is a sample:

$$A_{YX} = A_{XY} = \begin{bmatrix} 0&0&0&1&1\\ 0&0&1&0&0\\ 0&1&0&0&0\\ 1&0&0&0&0\\ 1&0&0&0&0 \end{bmatrix}, \quad A_{XY^2X} = \begin{bmatrix} 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&2 \end{bmatrix}.$$

These two give rise to the following linear constraints in ($\mathrm{SOHS_{SDP}}$):

$$\begin{aligned} G_{1,XY} + G_{X,Y} + G_{XY,1} + G_{1,YX} + G_{Y,X} + G_{YX,1} &= \langle A_{XY}, G\rangle \\ &= a_{XY} + a_{YX} = 0, \\ 2G_{YX,YX} = \langle A_{XY^2X}, G\rangle &= 2a_{XY^2X} = 0, \end{aligned}$$

where we have used a_w to denote the coefficients of f and the entries of V enumerate the columns, while the entries of V^* enumerate the rows of G. Observe that the second constraint tells us that the (YX, YX) entry of G is zero. As we are looking for a positive semidefinite G, the corresponding row and column of G can be assumed to be identically zero. That is, the last entry of V is redundant (cf. Example 2.2).

A further reduction in the vector of words needed is presented in [KP10] (the so-called Newton chip method) and its implementation in `NCSOStools` is `NCsos`.

2.3. **Eigenvalue optimization of NC polynomials and flat extensions.** One of the features of our freely available Matlab software package `NCSOStools` [CKP] is `NCmin` which uses sum of hermitian squares and semidefinite programming to compute a global (eigenvalue) minimum of a symmetric NC polynomial f. This is discussed in detail in [KP10, §5]. Here we present the theoretical underpinning of an algorithm to extract the minimizers of f, implemented in `NCopt`.

The main ingredients are the noncommutative moment problem and its solution due to McCullough [McC01], and the Curto-Fialkow theory [CF96] of how flatness governs the truncated moment problem. Our results are influenced by the method of Henrion and Lasserre [HL05] for the commutative case, which has been implemented in GloptiPoly [HLL09]. For an investigation of the non-global case in the free noncommutative setting see [PNA].

2.3.1. *Eigenvalue optimization is an SDP.* Let $f \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}$. We are interested in the smallest eigenvalue $f^\star \in \mathbb{R}$ of the polynomial f. That is,

(7)
$$f^\star = \inf\big\{\langle f(A)v, v\rangle \mid A \text{ an } n\text{-tuple of symmetric matrices, } v \text{ a unit vector}\big\}.$$

Hence $f^\star$ is the greatest lower bound on the eigenvalues $f(A)$ can attain for n-tuples of symmetric matrices A, i.e., $(f - f^\star)(A) \succeq 0$ for all n-tuples of symmetric matrices A, and $f^\star$ is the largest real number with this property. Given that a polynomial is positive semidefinite if and only if it is a sum of hermitian squares (the Helton-McCullough SOHS theorem), we can compute $f^\star$ conveniently with SDP. Let

$$(\mathrm{SDP}_{\mathrm{eig-min}}) \qquad \begin{array}{rcl} f^{\mathrm{sohs}} & = & \sup\ \lambda \\ \text{s. t.} & & f - \lambda \ \in\ \Sigma^2. \end{array}$$

Then $f^{\mathrm{sohs}} = f^\star$.

In general $(\mathrm{SDP}_{\mathrm{eig-min}})$ does not satisfy the Slater condition. That is, there does not always exist a *strictly feasible* solution. Nevertheless $(\mathrm{SDP}_{\mathrm{eig-min}})$ satisfies strong duality [KP10, Theorem 5.1], i.e., its optimal value f^{sohs} coincides with the optimal value L_{sohs} of the dual SDP:

$$(\mathrm{DSDP}_{\mathrm{eig-min}})_d \qquad \begin{array}{rcl} L_{\mathrm{sohs}} & = & \inf L(f) \\ \text{s. t.} & & L : \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{\leq 2d} \to \mathbb{R} \quad \text{is linear} \\ & & L(1) = 1 \\ & & L(p^*p) \geq 0 \quad \text{for all } p \in \mathbb{R}\langle \underline{X}\rangle_{\leq d}. \end{array}$$

2.3.2. *Extract the optimizers.* In this section we investigate the attainability of $f^\star$ and explain how to extract the minimizers A, v for f if the lower bound $f^\star$ is attained. That is, A is an n-tuple of symmetric matrices and v is a unit eigenvector for $f(A)$ satisfying

$$f^\star = \langle f(A)v, v\rangle. \tag{8}$$

Of course, in general f will not be bounded from below. Another problem is that even if f is bounded, the infimum $f^\star$ need not be attained.

Example 2.5. Let $f = Y^2 + (XY-1)^*(XY-1)$. Clearly, $f^{\mathrm{sohs}} \geq 0$. However, $f\big(1/\varepsilon, \varepsilon\big) = \varepsilon^2$, so $f^{\mathrm{sohs}} = 0$ and hence $L_{\mathrm{sohs}} = 0$. On the other hand, $f^\star$ from (7) and the dual optimum L_{sohs} are not attained.

Let us first consider $f^\star$. Suppose (A, B) is a pair of matrices yielding a singular $f(A, B)$ and let v be a nullvector. Then

$$B^2 v = 0 \quad \text{and} \quad (AB - I)^*(AB - I)v = 0.$$

From the former we obtain $Bv = 0$, whence

$$v = Iv = (AB - I)v = 0,$$

a contradiction.

We now turn to the nonexistence of a dual optimizer. Suppose otherwise and let $L : \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{\leq 4} \to \mathbb{R}$ be a minimizer with $L(1) = 1$. We extend L to $\mathbb{R}\langle \underline{X}\rangle_{\leq 4}$ by symmetrization. That is,

$$L(p) := \frac{1}{2} L(p + p^*).$$

We note L induces a semi-scalar product (i.e., a positive semidefinite bilinear form) $(p, q) \mapsto L(p^*q)$ on $\mathbb{R}\langle \underline{X}\rangle_{\leq 2}$ due to the positivity property. Since $L(f) = 0$, we have

$$L(Y^2) = 0 \quad \text{and} \quad L\big((XY - 1)^*(XY - 1)\big) = 0.$$

Hence by the Cauchy-Schwarz inequality, $L(XY) = L(YX) = 0$. Thus

$$0 = L\big((XY - 1)^*(XY - 1)\big) = L\big((XY)^*(XY)\big) + L(1) \geq L(1) \geq 0,$$

implying $L(1) = 0$, a contradiction.

Hence despite the strong duality holding for $(\mathrm{SDP}_{\mathrm{eig-min}})$, the eigenvalue infimum $f^\star$ and the dual optimum L_{sohs} need not be attained, so some caution is necessary. In the sequel our main interest lies in the case where $f^\star$ is attained. We shall see later below (see Corollary 2.10) that this happens if and only it the infimum $L_{\mathrm{sohs}} = f^{\mathrm{sohs}} = f^\star$ for $(\mathrm{DSDP}_{\mathrm{eig-min}})_{d+1}$ is attained.

Definition 2.6. To each linear functional $L : \mathbb{R}\langle \underline{X}\rangle_{\leq 2d} \to \mathbb{R}$ we associate a matrix M_d (called an *NC Hankel matrix*) indexed by words $u, v \in \langle \underline{X}\rangle$ of length $\leq d$, with

$$(M_d)_{u,v} = L(u^*v). \tag{9}$$

If L is *positive*, i.e., $L(p^*p) \geq 0$ for all $p \in \mathbb{R}\langle \underline{X}\rangle_{\leq d}$, then M_d is positive semidefinite. We say that L is *unital* if $L(1) = 1$.

Note that a matrix M indexed by words of length $\leq d$ satisfying the *NC Hankel condition* $M_{u_1,v_1} = M_{u_2,v_2}$ if $u_1^*v_1 = u_2^*v_2$, yields a linear functional L on $\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}$ as in (9). If M is positive semidefinite, then L is positive.

Definition 2.7. Let $A \in \mathbb{R}^{s\times s}$ be a symmetric matrix. A (symmetric) extension of A is a symmetric matrix $\tilde{A} \in \mathbb{R}^{(s+\ell)\times(s+\ell)}$ 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 $C = Z^tAZ$ for some matrix Z.

Proposition 2.8. *Let $f \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}$ be bounded from below. If the infimum L_{sohs} for $(\mathrm{DSDP}_{\mathrm{eig-min}})_{d+1}$ is attained, then it is attained at a linear map L that is flat over its own restriction to $\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}$.*

Proof. For this proof it is beneficial to work with NC Hankel matrices. Let L be a minimizer for $(\mathrm{DSDP}_{\mathrm{eig-min}})_{d+1}$. To it we associate M_{d+1} and its restriction M_d. Then

$$M_{d+1} = \begin{bmatrix} M_d & B \\ B^t & C \end{bmatrix}$$

for some B, C. Since M_{d+1} and M_d are positive semidefinite, $B = M_d Z$ and $C \succeq Z^t M_d Z$ for some Z (this is easy to verify using Schur complements; or see [CF96]) Now form a "new" M_{d+1}:

$$\tilde{M}_{d+1} = \begin{bmatrix} M_d & B \\ B^t & Z^t M_d Z \end{bmatrix} = \begin{bmatrix} I & Z \end{bmatrix}^t M_d \begin{bmatrix} I & Z \end{bmatrix}.$$

This matrix is obviously flat over M_d, positive semidefinite, and satisfies the NC Hankel condition (it is inherited from M_{d+1} since for all quadruples u, v, z, w of words of degree $d+1$ we have $u^*v = z^*w \iff u = z$ and $z = w$). So it yields a positive linear map $\tilde{L}$ on $\mathbb{R}\langle \underline{X}\rangle_{\leq 2d+2}$ flat over $\tilde{L}|_{\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}} = L|_{\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}}$. Moreover, $\tilde{L}(f) = L(f) = L_{\mathrm{sohs}}$. ∎

The following is a solution to a free noncommutative moment problem in the truncated case. It resembles the classical results of Curto and Fialkow [CF96] in the commutative case. For the free noncommutative moment problem see [McC01] or also [PNA]. A similar statement (with a positive definiteness assumption) is given in [MP05].

Theorem 2.9. *Suppose $L : \mathbb{R}\langle \underline{X}\rangle_{\leq 2d+2} \to \mathbb{R}$ is positive and flat over $L|_{\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}}$. Then there is an n-tuple A of symmetric matrices of size $s \leq \dim \mathbb{R}\langle \underline{X}\rangle_{\leq d}$ and a vector v such that*

$$L(p^*q) = \langle p(A)v, q(A)v\rangle \tag{10}$$

for all $p, q \in \mathbb{R}\langle \underline{X}\rangle$ with $\deg p + \deg q \leq 2d$.

Proof. For this we use the Gelfand-Naimark-Segal (GNS) construction. To L we associate two positive semidefinite matrices, M_{d+1} and its restriction M_d. Since M_{d+1} is flat over M_d, there exist s linear independent columns of M_d labeled by words $w \in \langle \underline{X}\rangle$ with $\deg w \leq d$ which form a basis $\mathcal{B}$ of $E = \operatorname{range} M_{d+1}$. Now L (or M_{d+1}) induces a positive definite bilinear form (i.e., a scalar product) $\langle \_, \_\rangle_E$ on E.

Let A_i be the left multiplication with X_i on E, i.e., if $\overline{w}$ denotes the column of M_{d+1} labeled by $w \in \langle \underline{X}\rangle_{\leq d+1}$, then $A_i\colon \overline{u} \mapsto \overline{X_i u}$ for $u \in \langle \underline{X}\rangle_{\leq d}$. The operator A_i is well defined and symmetric:

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

Let $v := \overline{1}$, and $A = (A_1, \ldots, A_n)$. Note it suffices to prove (10) for words $u, w \in \langle \underline{X}\rangle$ with $\deg u + \deg w \leq 2d$. Since the A_i are symmetric, there is no harm in assuming $\deg u, \deg w \leq d$. Now compute

$$L(u^*w) = \langle \overline{u}, \overline{w}\rangle_E = \langle u(A)\overline{1}, w(A)\overline{1}\rangle_E = \langle u(A)v, w(A)v\rangle_E. \qquad ∎$$

Corollary 2.10. *Let $f \in \mathbb{R}\langle \underline{X}\rangle_{\leq 2d}$. Then $f^\star$ is attained if and only if there is a feasible point L for $(\mathrm{DSDP}_{\mathrm{eig-min}})_{d+1}$ satisfying $L(f) = f^\star$.*

Proof. (⇒) If (8) holds for some A, v, then $L(p) := \langle p(A)v, v\rangle$ is the desired feasible point. (⇐) By Proposition 2.8, we may assume L is flat over $L|_{\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}}$. Now Theorem 2.9 applies and yields A, v. By definition, $\|v\| = \sqrt{\langle v, v\rangle} = \sqrt{L(1)} = 1$. Hence $f(A)$ has (unit) eigenvector v with eigenvalue $f^\star$. ■

2.3.3. *Implementing the extraction of optimizers.* Let $f \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}$.

Step 1: Solve $(\mathrm{DSDP}_{\mathrm{eig-min}})_{d+1}$. If the problem is unbounded or the optimum is not attained, stop. Otherwise let L denote an optimizer.

Step 2: To L we associate the positive semidefinite matrix $M_{d+1} = \begin{bmatrix} M_d & B \\ B^t & C \end{bmatrix}$. Modify M_{d+1}: $\tilde{M}_{d+1} = \begin{bmatrix} M_d & B \\ B^t & Z^t M_d Z \end{bmatrix}$, where Z satisfies $M_d Z = B$. This matrix yields a positive linear map $\tilde{L}$ on $\mathbb{R}\langle \underline{X}\rangle_{\leq 2d+2}$ which is flat over $\tilde{L}|_{\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}} = L|_{\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}}$. In particular, $\tilde{L}(f) = L(f) = f^\star$.

Step 3: As in the proof of Theorem 2.9, use the GNS construction on $\tilde{L}$ to compute symmetric matrices A_i and a vector v with $\tilde{L}(f) = f^\star = \langle f(A)v, v\rangle$.

In Step 3, to construct symmetric matrix representations $A_i \in \mathbb{R}^{s\times s}$ of the multiplication operators we calculate their image according to a chosen basis $\mathcal{B}$ for $E = \operatorname{range}\tilde{M}_{d+1}$. To be more specific, $A_i\overline{u}_1$ for $u_1 \in \langle \underline{X}\rangle_{\leq d}$ 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 $\begin{bmatrix}\lambda_1 & \dots & \lambda_s\end{bmatrix}^t$ will be the first column of A_i. The vector v is the eigenvector of $f(A)$ corresponding to the smallest eigenvalue.

Warning 2.11. Running the above algorithm raises several challenges in practice. Since the primal problem $(\mathrm{SDP}_{\mathrm{eig-min}})$ often has no strictly feasible point we have no guarantee that the optimal value L_{sohs} of $(\mathrm{DSDP}_{\mathrm{eig-min}})_{d+1}$ is attained. We do not know how to test for attainability efficiently, since all state-of-the-art SDP solvers return only an ε-optimal solution (a point which is feasible and gives optimal value up to some rounding error).

Detecting unboundedness of $(\mathrm{DSDP}_{\mathrm{eig-min}})_{d+1}$ seems easier. First of all, the SDP solver is likely to detect it directly. Otherwise numerical problems will be mentioned, and we then solve the (usually much smaller) primal problem $(\mathrm{SDP}_{\mathrm{eig-min}})$ to detect its infeasibility, which is equivalent to the unboundedness of $(\mathrm{DSDP}_{\mathrm{eig-min}})_{d+1}$.

In summary, the performance of our algorithm to extract the optimizers depends heavily on the quality of the underlying SDP solver.

Remark 2.12. We finish this section by emphasizing that the extraction of eigenvalue optimizers (theoretically) *always* works if the optimum for $(\mathrm{DSDP}_{\mathrm{eig-min}})_{d+1}$ is attained. This is in sharp contrast with the commutative case; cf. [Las09].

We implemented the procedure explained in Steps 1–3 under `NCSOStools` as `NCopt`, while solving $(\mathrm{SDP}_{\mathrm{eig-min}})$ can be done by calling `NCmin` - see also Subsection 6.2 for a demonstration.

3. Commutators and zero trace NC polynomials

It is well-known and easy to see that trace zero matrices are sums of commutators. Less obvious is the fact (not needed in this paper) that trace zero matrices are commutators. In this section we present the corresponding theory for NC polynomials and describe how it is implemented in `NCSOStools`. Most of the results are taken from [KS08a].

Definition 3.1. An element of the form $[p, q] := pq - qp$ for $p, q \in \mathbb{R}\langle \underline{X} \rangle$ is called a *commutator*. Two 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:

$$f - g = \sum_{i=1}^{k} (p_i q_i - q_i p_i) \text{ for some } k \in \mathbb{N} \cup \{0\} \text{ and } p_i, q_i \in \mathbb{R}\langle \underline{X} \rangle.$$

Example 3.2. $2X^3Y + 3XYX^2 \overset{\text{cyc}}{\sim} X^2YX + 4YX^3$ as

$$2X^3Y + 3XYX^2 - (X^2YX + 4YX^3) = [2X, X^2Y] + [X, XY^2] + [4X, YX^2].$$

The following remark shows that cyclic equivalence can easily be tested.

Remark 3.3.
(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) Two 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 $v \in \langle \underline{X} \rangle$,

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

Given $f \overset{\text{cyc}}{\sim} g$ and an n-tuple of symmetric matrices A of the same size, $\operatorname{tr} f(A) = \operatorname{tr} g(A)$. The converse is given by the following tracial Nullstellensatz:

Theorem 3.4 (Klep-Schweighofer [KS08a])**.** *Let $d \in \mathbb{N}$ and $f \in \operatorname{Sym} \mathbb{R}\langle \underline{X} \rangle$ be of degree $\leq d$ satisfying*

$$\operatorname{tr}(f(A_1, \ldots, A_n)) = 0 \tag{11}$$

for all symmetric $A_1, \ldots, A_n \in \mathbb{R}^{d \times d}$. Then $f \overset{\text{cyc}}{\sim} 0$.

The cyclic equivalence test has been implemented under `NCSOStools` - see `NCisCycEq`.

4. Trace positive NC polynomials

A notion of positivity of NC polynomials weaker than that via positive semidefiniteness considered in Section 2, is given by the trace: $f \in \mathbb{R}\langle \underline{X} \rangle$ is called *trace-positive* if $\operatorname{tr} f(A) \geq 0$ for all tuples of symmetric matrices A of the same size. Clearly, every $f \in \Sigma^2$ is trace-positive and the same is true for every NC polynomial cyclically equivalent to SOHS. However, unlike in the positive semidefinite case, the converse fails. That is, there are trace-positive polynomials which are not cyclically equivalent to SOHS, see [KS08a, Example 4.4] or

[KS08b, Example 3.5]. Nevertheless, the obvious certificate for trace-positivity has been shown to be useful in applications to e.g. operator algebras [KS08a] and mathematical physics [KS08a], so deserves a further study here.

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 SOHS. By definition the elements in Θ^2 are exactly the polynomials which are sums of hermitian squares and commutators.

Testing whether a given $f \in \mathbb{R}\langle \underline{X} \rangle$ is an element of Θ^2 can be done again using SDP (the so-called *Gram matrix method*) as observed in [KS08b, §3]. A slightly improved algorithm reducing the size of the SDP needed is given by the following theorem.

Theorem 4.1. *Let $f \in \mathbb{R}\langle \underline{X} \rangle$, let $m_i := \frac{\operatorname{mindeg}_i f}{2}$, $M_i := \frac{\deg_i f}{2}$, $m := \frac{\operatorname{mindeg} f}{2}$, $M := \frac{\deg f}{2}$. Set*

$$V := \{w \in \langle \underline{X} \rangle \mid m_i \leq \deg_i w \leq M_i \text{ for all } i,\, m \leq \deg w \leq M\}.$$

Then $f \in \Theta^2$ if and only if there exists a positive semidefinite matrix G satisfying $f \overset{\text{cyc}}{\sim} V^*GV$.

Proof. Suppose $f \overset{\text{cyc}}{\sim} V^*GV$ for some positive semidefinite G. Then $G = \sum G_i G_i^t$ for some vectors G_i and $V^*GV = \sum g_i^* g_i$, where $g_i = G_i^t V$. Thus $f \in \Theta^2$.

Conversely, suppose $f \overset{\text{cyc}}{\sim} \sum g_i^* g_i$. We claim that each g_i is in the linear span of V. Assume otherwise, say one of the g_i contains a word w with $\deg_j w < m_j$. Let h_i denote the sum of all monomials of g_i whose corresponding words have $\deg_j$ less than m_j. Let $r_i = g_i - h_i$. Then

$$f \overset{\text{cyc}}{\sim} \sum g_i^* g_i = \sum (h_i + r_i)^* (h_i + r_i) = \sum h_i^* h_i + \sum h_i^* r_i + \sum r_i^* h_i + \sum r_i^* r_i. \tag{12}$$

Since each monomial w in $h_i^* r_i$, $r_i^* h_i$ and $r_i^* r_i$ has $\deg_i w \geq 2m_i$, none of these can be cyclically equivalent to a monomial in $h_i^* h_i$. Thus

$$0 \overset{\text{cyc}}{\sim} \sum h_i^* h_i, \quad f \overset{\text{cyc}}{\sim} \sum h_i^* r_i + \sum r_i^* h_i + \sum r_i^* r_i.$$

However, this implies $h_i = 0$ for all i (see [KS08b, Lemma 3.2]; or also the proof of Proposition 2.1), contradicting the choice of w.

The remaining cases (i.e., a word w in one of the g_i with $\deg_i w > M_i$ or $\deg w < m$ or $\deg w > M$) can be dealt with similarly, so we omit the details. ∎

Testing whether a NC polynomial is a sum of hermitian squares and commutators (i.e., an element of Θ^2) has been implemented under `NCSOStools` as `NCcycSos`.

4.1. Trace-optimization of NC polynomials.

In this subsection we present a "practical" application of SOHS decompositions modulo cyclic equivalence, namely approximating global minima of NC polynomials.

The minimum of an NC polynomial with respect to positive semidefiniteness has been discussed above, and here we focus on the *trace minimum* of an NC polynomial f, that is, the largest number $f^\star$ making $f - f^\star$ trace-positive. Equivalently,

$$f^\dagger = \inf\{\operatorname{tr} f(A) \mid A \text{ an } n\text{-tuple of symmetric matrices of the same size}\}.$$

(A word of caution: tr denotes the *normalized* trace, i.e., $\operatorname{tr} I = 1$.) This number is hard to compute but a good approximation can be given using sums of hermitian squares and commutators. For this we define

$$(\mathrm{SDP}_{\mathrm{tr-min}}) \qquad \begin{array}{rcl} f^{\mathrm{c-sohs}} & = & \sup\ \lambda \\ \text{s.t.} & & f - \lambda \ \in\ \Theta^2. \end{array}$$

We denote the problem above by $\mathrm{SDP}_{\mathrm{tr-min}}$, since it is an instance of semidefinite programming. Suppose $f \in \operatorname{Sym} \mathbb{R}\langle \underline{X}\rangle$. Let W be a vector consisting of all monomials from $\langle \underline{X}\rangle$ with degree $\le \frac{1}{2}\deg f$ and degree in X_i at most $\frac{1}{2}\deg_i f$. Assume the first entry of W is 1. Then $(\mathrm{SDP}_{\mathrm{tr-min}})$ rewrites into

$$\begin{array}{lrcl} \sup & f_0 - \langle E_{11}, G\rangle & & \\ \text{s.t.} & f - f_0 & \overset{\mathrm{cyc}}{\sim} & W^*(G - G_{11}E_{11})W \\ & G & \succeq & 0. \end{array}$$

(Here f_0 is the constant term of f and E_{11} is the matrix with all entries 0 except for the $(1,1)$ entry which is 1.) The condition $f - f_0 \overset{\mathrm{cyc}}{\sim} W^*(G - G_{11}E_{11})W$ translates into linear constraints on the entries of G by Remark 3.3.

Proposition 4.2. $f^\dagger \ge f^{\mathrm{c-sohs}}$. *The inequality might be strict.*

Proof. If $\lambda \in \mathbb{R}$ is such that $f - \lambda \in \Theta^2$, then $f - \lambda \overset{\mathrm{cyc}}{\sim} g \in \Sigma^2$ for some $g \in \mathbb{R}\langle \underline{X}\rangle$. Thus $\operatorname{tr} f(A) = \operatorname{tr}(g(A)) + \lambda \ge \lambda$.

The second statement follows from the fact that trace-positive NC polynomials need not be sums of hermitian squares and commutators. For an explicit example we refer the reader to [KS08a, Example 4.4] or [KS08b, Example 3.5]. ∎

In general $(\mathrm{SDP}_{\mathrm{tr-min}})$ does not satisfy the Slater condition. Nevertheless:

Theorem 4.3. $(\mathrm{SDP}_{\mathrm{tr-min}})$ *satisfies strong duality.*

Proof. Suppose $f \in \mathbb{R}\langle \underline{X}\rangle$ is of degree $\le 2d$ and its trace is bounded from below. Let $\Theta^2_{\le 2d}$ denote the cone of all sums of hermitian squares and commutators of degree $\le 2d$, i.e.,

$$\Theta^2_{\le 2d} = \{f \in \mathbb{R}\langle \underline{X}\rangle \mid \deg f \le 2d,\ f \overset{\mathrm{cyc}}{\sim} \sum_{i=1}^{t} g_i^* g_i\,, t \in \mathbb{N},\ g_i \in \mathbb{R}\langle \underline{X}\rangle \text{ of degree} \le d\}.$$

Then $(\mathrm{SDP}_{\mathrm{tr-min}})$ can be rewritten as:

$$(\text{Primal}) \qquad \begin{array}{ll} \sup & \varepsilon \\ \text{s.t.} & f - \varepsilon \in \Theta^2_{\le 2d}. \end{array}$$

The dual cone of $\Theta^2_{\le 2d}$ is the set of all linear maps $\operatorname{Sym} \mathbb{R}\langle \underline{X}\rangle_{\le 2d} \to \mathbb{R}$ which are nonnegative on $\Theta^2_{\le 2d}$; note that these automatically vanish on all commutators. (We use $\operatorname{Sym} \mathbb{R}\langle \underline{X}\rangle_{\le 2d}$ to denote the set of all (symmetric) NC polynomials of degree $\le 2d$.)

Claim: The cone $\Theta^2_{\leq 2d}$ is closed in $\operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}$.
Proof: This is a straightforward modification of [MP05, Proposition 3.4], so we omit it.

Let us now return to the SDP. The dual problem to (Primal) is given by:

$$\text{(Dual)}\qquad \begin{array}{ll} \inf & L(f) \\ \text{s.t.} & L : \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{\leq 2d} \to \mathbb{R} \quad \text{is linear} \\ & L(1) = 1 \\ & L(p^*p) \geq 0 \quad \text{for all } p \in \mathbb{R}\langle \underline{X}\rangle_{\leq d} \\ & L(pq - qp) = 0 \quad \text{for all } p, q \in \mathbb{R}\langle \underline{X}\rangle_{\leq d}. \end{array}$$

Let $f^{\text{c-sohs}}$ and $L_{\text{c-sohs}}$ denote the optimal value of (Primal) and (Dual), respectively. We claim that $f^{\text{c-sohs}} = L_{\text{c-sohs}}$. Clearly, $f^{\text{c-sohs}} \leq L_{\text{c-sohs}}$. To prove the converse note that $L(f - L_{\text{c-sohs}}) \geq 0$ for all L in the dual cone of $\Theta^2_{\leq 2d}$. This means that $f - L_{\text{c-sohs}}$ belongs to the closure of $\Theta^2_{\leq 2d}$, so by the Claim, $f - L_{\text{c-sohs}} \in \Theta^2_{\leq 2d}$. Hence also $f^{\text{c-sohs}} \geq L_{\text{c-sohs}}$.

Now suppose $f \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{\leq 2d}$ is not bounded from below. Then for every $\lambda \in \mathbb{R}$ there exists a tuple of symmetric matrices A such that $\operatorname{tr}(f - \lambda)(A) = \operatorname{tr} f(A) - \lambda < 0$. Define

$$L : \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{\leq 2d} \to \mathbb{R}, \quad g \mapsto \operatorname{tr} g(A).$$

Then $L(f) < \lambda$. As λ was arbitrary, this shows that (Dual) is unbounded, hence strong duality holds in this case as well. ■

Trace-optimization of NC polynomials is implemented in `NCSOStools`, where the optimal solution of (Primal) is computed by calling the routine `NCcycMin`.

5. Convex and trace convex NC polynomials

5.1. **Convex NC polynomials.**

Motivated by consideration in engineering system theory (cf. [dOHMP08] for a modern treatment), Helton and McCullough [HM04] studied convex NC polynomials. An NC polynomial $p \in \mathbb{R}\langle \underline{X}\rangle$ is *convex* if it satisfies

$$p(tA + (1-t)B) \preceq tp(A) + (1-t)p(B)$$

for all $0 \leq t \leq 1$ and for all tuples A, B of symmetric matrices of the same size.

Convexity can be rephrased using second directional derivatives and sums of hermitian squares. Given $p \in \mathbb{R}\langle \underline{X}\rangle$, consider

$$r(\underline{X}, \underline{H}) = p(\underline{X} + \underline{H}) - p(\underline{X}) \in \mathbb{R}\langle \underline{X}, \underline{H}\rangle.$$

Then the *second directional derivative* $p''(\underline{X}, \underline{H}) \in \mathbb{R}\langle \underline{X}, \underline{H}\rangle$ is defined to be twice the part of $r(\underline{X}, \underline{H})$ which is homogeneous of degree two in $\underline{H}$. Alternatively,

$$p''(\underline{X}, \underline{H}) = \frac{d^2 p(\underline{X} + t\underline{H})}{dt^2}\Big|_{t=0}.$$

For example, if $p(\underline{X}) = X_1X_2X_1$, then $p''(\underline{X}, \underline{H}) = 2(X_1H_2H_1 + H_1X_2H_1 + H_1H_2X_1)$.

By [HM04, Theorem 2.4], $p \in \mathbb{R}\langle \underline{X}\rangle$ is convex if and only if p'' is a sum of hermitian squares in $\mathbb{R}\langle \underline{X}, \underline{H}\rangle$. Thus this is easily tested using `NCSOStools` and has been implemented under `NCisConvex0`.

However, the convexity test can be simplified and greatly improved using [HM04, Theorem 3.1]: every convex NC polynomial p is of degree ≤ 2. Hence only variables $\underline{H}$ will appear in p''. The Gram matrix G for p'' is therefore a *unique scalar matrix*, so testing for convexity of p is simply checking whether G is positive semidefinite. See `NCisConvex`.

For a different and more general type of convexity test we refer the reader to [CHSY03].

5.2. **Trace convex NC polynomials.** An NC polynomial $p \in \mathbb{R}\langle\underline{X}\rangle$ is *trace convex* if it satisfies

$$\operatorname{tr} p(tA + (1-t)B) \leq t \operatorname{tr} p(A) + (1-t) \operatorname{tr} p(B)$$

for all $0 \leq t \leq 1$ and for all tuples A, B of symmetric matrices of the same size.

As with convexity, tracial convexity can be rephrased using second directional derivatives: p is trace convex if and only if p'' is trace-positive. However, this is hard to check, so we have instead implemented the test for the stronger condition $p'' \in \Theta^2$, see `NCisCycConvex`. We remark that for NC polynomials in *one* variable, p is trace convex if and only if $p'' \in \Theta^2$, a result due to Chris Nelson et al. at UCSD (in preparation). Equivalently: p is convex as a polynomial of one *commuting* variable.

We do not know whether $p \in \operatorname{Sym} \mathbb{R}\langle\underline{X}\rangle$ is trace convex if and only if $p'' \in \Theta^2$ in general.

6. Two examples

In this section we give two sample applications of our toolbox. One concerns the Bessis-Moussa-Villani (BMV) conjecture from quantum statistical mechanics, while the other presents the eigenvalue minimization of polynomials based on the algorithm presented in Section 2.3.

6.1. **The BMV conjecture.** In an attempt to simplify the calculation of partition functions of quantum mechanical systems Bessis, Moussa and Villani [BMV75] conjectured in 1975 that for any two symmetric matrices A, B, where B is positive semidefinite, the function $t \mapsto \operatorname{tr}(e^{A-tB})$ is the Laplace transform of a positive Borel measure with real support. This would permit the calculation of explicit upper and lower bounds of energy levels in multiple particle systems. In their 2004 paper [LS04], Lieb and Seiringer have given the following purely algebraic reformulation:

Conjecture 6.1. *For all positive semidefinite matrices A and B and all $m \in \mathbb{N}$, the polynomial $p(t) := \operatorname{tr}((A + tB)^m) \in \mathbb{R}[t]$ has only nonnegative coefficients.*

The coefficient of t^k in $p(t)$ for a given m is the trace of $S_{m,k}(A, B)$, where $S_{m,k}(A, B)$ is the sum of all words of length m in the letters A and B in which B appears exactly k times. For example $S_{4,2}(A, B) = A^2B^2 + ABAB + AB^2A + BABA + B^2A^2 + BA^2B$. $S_{m,k}(X, Y)$ is thus an NC polynomial; it is the sum of all words in two variables X, Y of degree m with degree k in Y.

In the last few years there has been much activity around the question for which pairs (m, k) does $S_{m,k}(X^2, Y^2) \in \Theta^2$ or $S_{m,k}(X, Y) \in \Theta^2$ hold? An affirmative answer (for all m, k) to the former would suffice for the BMV conjecture to hold; this question has been resolved completely (see e.g. [KS08b, CDTA, Bur]), however only finitely many nontrivial $S_{m,k}(X^2, Y^2)$ admit a Θ^2-certificate. Here we give a quick proof of the main result of [KS08b] establishing $S_{14,6}(X^2, Y^2) \in \Theta^2$. Together with [Hil07] this proves the BMV conjecture for $m \leq 13$.

Example 6.2. Consider the polynomial $f = S_{14,6}(X^2, Y^2)$. To prove that $f \in \Theta^2$ with the aid of `NCSOStools`, proceed as follows:

(1) Define two noncommuting variables:

```
>> NCvars x y;
```

(2) Our polynomial f is constructed using `BMVq(14,6)`.

```
>> f=BMVq(14,6);
```

For a numerical test whether $f \in \Theta^2$, we first construct a small monomial vector V [KS08b, Proposition 3.3] to be used in the Gram matrix method.

```
>> [v1,v2,v3]=BMVsets(14,6); V=[v1;v2;v3];
>> params.obj = 0; params.V=V;
>> [IsCycEq,X,V,sohs,g,SDP_data] = NCcycSos(f, params);
```

This yields a *floating point* positive definite 70×70 Gram matrix X. The rest of the output: `IsCycEq` $= 1$ since f is (numerically) in Θ^2; `sohs` is a vector of polynomials g_i with $f \overset{\text{cyc}}{\sim} \sum_i g_i^* g_i =$ `g`; `SDP_data` is the SDP data run for testing whether $f \in \Theta^2$.

To obtain an *exact* Θ^2-certificate, we can round and project the obtained solution X (cf. [PP08] for details).

6.2. **Eigenvalue minimization.** In this section we present a toy example of eigenvalue optimization as presented in Section 2.3.

Example 6.3. Assuming we have already introduced NC variables x, y, let us define

```
>> f = (1-y+x*y+y*x)'*(1-y+x*y+y*x) + (-2+y^2)^2 + (-x+x^2)^2;
```

As is usual in Matlab, the prime ' denotes an involution, in our case acting on NC polynomials. By definition, f is a sum of hermitian squares. We shall compute the eigenvalue minimum $f^\star$ of f and determine the minimizers A, B, v satisfying $\langle f(A,B)v, v\rangle = f^\star$. Here A, B are symmetric matrices, and v is a unit eigenvector of $f(A,B)$, corresponding to $\lambda_{\min}(f(A,B))$. Running

```
>> NCmin(f)
```

yields an eigenvalue minimum $f^\star = 0.0000$. We next run the algorithm presented in Subsection 2.3.3 to extract optimizers:

```
>> [X,fX,eig_val,eig_vec]=NCopt(f)
```

The output: `X` is a 2×16 matrix, whose rows represent symmetric matrices A, B; `fX` is the 4×4 matrix $f(A,B)$; `eig_val` are the eigenvalues of `fX`, and `eig_vec` are the corresponding unit eigenvectors. In our example,

$$A = \begin{bmatrix} 0.2117 & 0.2911 & 0.2907 & 0.0671 \\ 0.2911 & 0.8360 & -0.0328 & -0.2921 \\ 0.2907 & -0.0328 & 0.7916 & 0.3179 \\ 0.0671 & -0.2921 & 0.3179 & -0.2987 \end{bmatrix}$$

$$B = \begin{bmatrix} 1.1832 & -0.7438 & 0.2035 & -0.0920 \\ -0.7438 & -1.1270 & 0.0527 & -0.5790 \\ 0.2035 & 0.0527 & -1.2469 & -1.1049 \\ -0.0920 & -0.5790 & -1.1049 & 0.9849 \end{bmatrix}$$

$$f(A,B) = \begin{bmatrix} 0.0164 & 0.0120 & 0.1019 & -0.1058 \\ 0.0120 & 0.3019 & 0.4787 & 0.0576 \\ 0.1019 & 0.4787 & 1.2163 & -0.2926 \\ -0.1058 & 0.0576 & -0.2926 & 1.9334 \end{bmatrix}$$

and the smallest eigenvalue $f^\star$ of $f(A,B)$ is (only 4 decimal digits displayed) 0.0000, with the corresponding unit eigenvector

$$v = \begin{bmatrix} 0.9663 & 0.2038 & -0.1555 & 0.0233 \end{bmatrix}^t .$$

This was computed on a Mac using SDPA. We note the minimum of f on $\mathbb{R}^2$ (i.e., the minimum of f considered as a polynomial in *commuting* variables) can be computed exactly using Mathematica. It is approximately 0.0146.

Conclusions

In this paper we present `NCSOStools`: a computer algebra system for working with noncommutative polynomials with a special focus on methods determining whether a given NC polynomial is a sum of hermitian squares (SOHS) or is cyclically equivalent to SOHS (i.e., is a sum of hermitian squares and commutators). `NCSOStools` is an open source Matlab toolbox freely available from our web site:

`http://ncsostools.fis.unm.si/`

The package contains several extensions, like computing SOHS lower bounds, extracting minimizers, and checking for convexity or trace convexity of given NC polynomials. Moreover, functions have been implemented to handle cyclic equivalence. Most of the methods rely on semidefinite programming therefore the user should provide an SDP solver. Currently SeDuMi [Stu99], SDPA [YFK03] and SDPT3 [TTT] are supported, while other solvers might be added in the future.

`NCSOStools` can handle NC polynomials of medium size, while larger problems may run into trouble for two reasons: the underlying SDP is too big for the state-of-the-art SDP solvers or the (combinatorial) process of constructing the SDP is too exhaustive. The ongoing research will mainly concern the second issue (with improvements for sparse NC polynomials or NC polynomials with symmetries). Also methods to produce exact rational solutions from numerical solutions given by SDP solvers are being implemented, in the spirit of [PP08].

Acknowledgments. The authors thank both anonymous referees for helpful suggestions. Sabine Burgdorf provided valuable feedback on the software.

References

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

[Bur] S. Burgdorf. Sums of hermitian squares as an approach to the BMV conjecture. Linear Multilinear Algebra, to appear, `http://arxiv.org/abs/0802.1153`.

[CDTA] B. Collins, K.J. Dykema, and F. Torres-Ayala. Sum-of-squares results for polynomials related to the Bessis-Moussa–Villani conjecture. Preprint, `http://arxiv.org/abs/0905.0420`.

[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.

[CHSY03] J.F. Camino, J.W. Helton, R.E. Skelton, and Jieping Ye. Matrix inequalities: a symbolic procedure to determine convexity automatically. *Integral Equations Operator Theory*, 46(4):399–454, 2003.

[CKP] 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`.

[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.

[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.

[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/`.

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

[Hil07] C.J. Hillar. Advances on the Bessis-Moussa-Villani trace conjecture. *Linear Algebra Appl.*, 426(1):130–142, 2007.

[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.

[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/`.

[HM04] J.W. Helton and S. McCullough. Convex noncommutative polynomials have degree two or less. *SIAM J. Matrix Anal. Appl.*, 25(4):1124–1139, 2004.

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

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

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

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

[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`.

[LS04] E.H. Lieb and R. Seiringer. Equivalent forms of the Bessis-Moussa-Villani conjecture. *J. Stat. Phys.*, 115(1-2):185–190, 2004.

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

[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`.

[MP05] S. McCullough and M. Putinar. Noncommutative sums of squares. *Pacific J. Math.*, 218(1):167–171, 2005.

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

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

[PNA] S. Pironio, M. Navascues, and A. Acin. Convergent relaxations of polynomial optimization problems with non-commuting variables. To appear in SIAM J. Optim. `http://arxiv.org/abs/0903.4368`.

[PP08] H. Peyrl and P.A. Parrilo. Computing sum of squares decompositions with rational coefficients. *Theoret. Comput. Sci.*, 409(2):269–281, 2008.

[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/`.

[PRW06] J. Povh, F. Rendl, and A. Wiegele. A boundary point method to solve semidefinite programs. *Computing*, 78:277–286, 2006.

[PW98] V. Powers and T. Wörmann. An algorithm for sums of squares of real polynomials. *J. Pure Appl. Algebra*, 127(1):99–104, 1998.

[RFP] B. Recht, M. Fazel, and P.A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. To appear in SIAM Rev. `http://arxiv.org/abs/0706.4138`.

[Row80] L.H. Rowen. *Polynomial identities in ring theory*, volume 84 of *Pure and Applied Mathematics*. Academic Press Inc., New York, 1980.

[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/`.

[Tod01] M.J. Todd. Semidefinite optimization. *Acta Numerica*, 10:515–560, 2001.

[TTT] K.-C. Toh, M.J. Todd, and R.-H. Tütüncü. SDPT3 version 4.0 (beta) – a MATLAB software for semidefinite-quadratic-linear programming. `http://www.math.nus.edu.sg/~mattohkc/sdpt3.html`.

[WKK+09] 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.

[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/`.

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, Institute of mathematics, physics and mechanics Ljubljana, Jadranska 19, 1111 Ljubljana, Slovenia
E-mail address: `janez.povh@fis.unm.si`

NOT FOR PUBLICATION

Contents