MINIMIZER EXTRACTION IN POLYNOMIAL OPTIMIZATION IS ROBUST

IGOR KLEP[1], JANEZ POVH[2], AND JURIJ VOLČIČ[3]

Abstract. In this article we present a robustness analysis of the extraction of optimizers in polynomial optimization. Optimizers can be extracted by solving moment problems using flatness and the Gelfand-Naimark-Segal (GNS) construction. Here a modification of the GNS construction is presented that applies even to non-flat data, and then its sensitivity under perturbations is studied. The focus is on eigenvalue optimization for noncommutative polynomials, but it is also explained how the main results pertain to commutative and tracial optimization.

1. Introduction

Polynomial optimization (POP) studies optimization problems in which the objective and constraint functions are polynomials [Las09, BPT13]. It has a wide range of applications, e.g. to operations research, statistics, theoretical computer science and several branches of engineering and the physical sciences. The development of POP has been particularly fruitful since Putinar's Positivstellensatz [Put93] gave rise to the Lasserre relaxation scheme [Las01] which reformulates polynomial optimization problems as sequences of semidefinite programming (SDP) [WSV00, deK02] problems. Consequently, the area is nowadays intertwined with real algebraic geometry; see [Scw06, Lau09, Mar08, Par03, NDS06, HG05] for a small sample of the vast literature.

In parallel to these developments, the theory of noncommutative (nc) polynomial optimization (NCPOP) is growing rapidly [BKP16]. The fundamental problem that we consider in this paper is as follows. Given nc polynomials (see [BKP16] or Section 2 for definitions) $f, s_1, \dots, s_h$, compute

$$\lambda_{\min}(f) := \inf \left\{ v^{\intercal} f(A) v \colon s_i(A) \succeq 0 \text{ for all } i,\ \|v\| = 1 \right\}. \tag{1.1}$$

Hence $\lambda_{\min}(f)$ is the greatest lower bound on the eigenvalues of $f(A)$ taken over all tuples A of bounded self-adjoint operators on a separable infinite-dimensional Hilbert space which satisfy $s_i(A) \succeq 0$ for all i. Often it suffices to plug in tuples of matrices A only. Applications of NCPOPs can be found in control theory [dOHMP08] (cf. the textbook classics in [SIG97]), quantum theory [PNA10], and PDEs; [Cim10] uses NCPOPs to investigate PDEs and eigenvalues of polynomial partial differential operators. In [DLTW08] the authors investigate the quantum moment problem and entangled multi-prover games using NCPOPs.

A particularly interesting class of NCPOPs are tracial NCPOPs. Here one is interested in the smallest trace a nc polynomial attains, i.e.,

$$\mathrm{Tr}_{\min}(f) := \inf \left\{ \mathrm{Tr}\, f(A) \colon s_i(A) \succeq 0 \text{ for all } i \right\}. \tag{1.2}$$

A systematic study of these topics first arose out of an attempt to understand Connes' embedding conjecture in operator algebra [KS08], and later lead to the development of NCSOStools [CKP11, CKP12], an open source Matlab toolbox for handling NCPOPs. Tracial NCPOPs are intimately

Date: May 25, 2018.

2010 *Mathematics Subject Classification.* Primary 90C22, 14P10; Secondary 13J30, 47A57.

Key words and phrases. polynomial optimization, sum of squares, semidefinite programming, moment problem, Hankel matrix, flat extension, GNS construction, noncommutative polynomial, trace.

[1]Supported by the Marsden Fund Council of the Royal Society of New Zealand. Partially supported by the Slovenian Research Agency grants J1-8132, N1-0057 and P1-0222.

[2]Supported by the Slovenian Research Agency grants J1-8132 and N1-0057.

[3]Supported by the University of Auckland Doctoral Scholarship and the Deutsche Forschungsgemeinschaft (DFG) Grant No. SCHW 1723/1-1.

connected to quantum theory. Recently quantum analogues of the classical independence and chromatic graph parameters were studied using the cone of trace positive non-commutative polynomials [LP15]. Laurent with collaborators also presented how the entanglement dimension of a bipartite quantum correlations can be estimated using techniques from tracial polynomial optimization [GLL].

The optimization problems (1.1) and (1.2) are difficult, and are nowadays solved with a relaxation scheme based off noncommutative Positivstellensätze [HM04, KS08]; see [PNA10] for the relaxation scheme in the free NCPOP and [KP16] for its tracial analog. For instance, instead of (1.1) one solves

$$f^{(d)}_{\text{sohs}} := \sup \Big\{ \mu\colon f-\mu = \sigma + \sum_i \sigma_i(s_i) \Big\}, \tag{1.3}$$

where $\sigma=\sum_j h_j^\intercal h_j$ is a sum of hermitian squares of degree $\le 2d$, and $\sigma_i(s_i)=\sum_j h_{ij}^\intercal s_i h_{ij}$ are sums of weighted hermitian squares of degree $\le 2d$. The sequence $f^{(d)}_{\text{sohs}}$ is increasing and under mild conditions (archimedeanity of the constraint set) converges to $\lambda_{\min}(f)$ [PNA10, BKP16]. Further, the Gram matrix method allows us to rewrite (1.3) as an SDP making (1.3) a practical and effective approximation to (1.1).

The dual SDP problem to (1.3) can be presented as

$$\varphi^{(d)}_{\text{sohs}} = \inf \Big\{ \varphi(f)\colon \varphi(g^\intercal g)\ge 0,\ \varphi(h^\intercal s_i h)\ge 0,\ \varphi(1)=1 \Big\}, \tag{1.4}$$

where φ is a linear functional defined on nc polynomials of degree $\le 2d$. Under mild assumptions (e.g. the constraint set has nonempty interior), there is no duality gap [BKP16, Chapter 4] and thus $f^{(d)}_{\text{sohs}} = \varphi^{(d)}_{\text{sohs}}$. Besides computational advantages, the dual SDP makes it possible to establish tightness of a relaxation via flat extensions [CF96]. The first algorithm for extracting optimizers in the presence of a flat extension (of a linear functional φ or the associated Hankel matrix H_φ corresponding to the induced quadratic form) was described in [HL05] (cf. [Lau09]). Roughly speaking, if the optimal (Riesz) functional $\varphi^{(d+\delta)}$ corresponding to $\varphi^{(d+\delta)}_{\text{sohs}}$ is flat over $\varphi^{(d)}$ for some $\delta\ge 1$, meaning that the two matrices $H_{\varphi^{(d+\delta)}}$ and $H_{\varphi^{(d)}}$ are of the same rank, then there is a tuple of matrices A and a unit vector v with

$$\lambda_{\min}(f) = v^\intercal f(A) v = f^{(d)}_{\text{sohs}} = \varphi^{(d)}_{\text{sohs}} = \varphi^{(d+\delta)}(f) = \varphi^{(d)}(f). \tag{1.5}$$

To each linear functional φ on nc polynomials of degree $\le 2d$ as above one associates a symmetric Hankel matrix $H=H_\varphi$ carrying the same information as φ. The matrix H is indexed by words of degree $\le d$, and the (u,v) entry of H equals $\varphi(u^\intercal v)$. Hence extensions of linear functionals correspond to extensions of Hankel matrices.

One algorithm for extracting the tuple A and the vector v as in (1.5) is inspired by the Gelfand-Naimark-Segal (GNS) construction in functional analysis. It is particularly well suited for the robustness analysis undertaken in this article. A (truncated) GNS construction starts with a Hankel extension K of a positive semidefinite Hankel matrix H and produces a tuple of operators $\mathcal{X}^K$ on the finite-dimensional Hilbert space determined by H that serves as a candidate for an optimizer of NCPOP, see Section 2 for details. With straightforward modifications all of this also applies to commutative POP, and tracial NCPOP. Since the GNS construction in these two settings is also noncommutative in nature, the paper focuses mainly on NCPOP. Further, the noncommutative viewpoint makes it possible to give a mostly unified treatment of all three cases; the additional properties (commutativity or being tracial) are only used in the final part of the algorithm.

1.1. **Contributions.** The main goal of this paper is to present a robustness analysis of the extraction of optimizers via the so–called GNS construction in POP. This method can be modified to apply even for non-flat data, and we study how sensitive it is under perturbations. Our main focus is on the eigenvalue optimization in NCPOP, but we also address tracial NCPOP and commutative POP.

The main contributions of this paper are:

(i) In Section 3 we analyze robustness of optimizer extraction by GNS for global (i.e., constraint-free) NCPOP. The key result is Theorem 3.2, which for Hankel matrices K and K' extending a fixed Hankel matrix H gives explicit lower and upper bounds on the norm of $\mathcal{X}^K - \mathcal{X}^{K'}$ in terms of $\|K - K'\|$. This in turn allows us to estimate $f(\mathcal{X}^K) - f(\mathcal{X}^{K'})$ for a polynomial f; see Corollaries 3.5 and 3.6.

(ii) In Section 4 we apply the preceding results to NCPOP with constraints. If a Hankel extension K of H associated to a positive functional is not flat, meaning that the rank of K is larger than the rank of H, then the obtained matrix tuple $\mathcal{X}^K$ might not satisfy the given constraints. However, the error is small if K is close to being flat. We show in Theorem 4.1 that if K' is a flat Hankel extension of H, then we can quantify the violation of constraints in terms of $\|K - K'\|$.

(iii) In Section 4 we also show that to every K as above we can effectively assign a matrix $K^\flat$ (see Subsection 4.2 for the definition) such that K is flat if and only if $K = K^\flat$. If $\|K - K^\flat\|$ is small we say that K is *almost flat*. We establish in Theorem 4.6 the bounds on the violation of constraints in terms of $\|K - K^\flat\|$. We emphasize that $K^\flat$ is not necessarily Hankel, i.e. it does not need to correspond to a linear functional on polynomials.

(iv) In Section 5 we explain how our results pertain to the classical, commutative POP, and the tracial NCPOP.

(v) We provide extensive numerical examples that support the theory and illustrate the strength of our statements.

Our results imply that for the success of the Lasserre relaxation scheme for polynomial optimization one does not need to require flatness, which is evasive from a numerical point of view, but merely approximate flatness, which is much easier to establish when solving optimization problems numerically.

Acknowledgments. The authors thank the two anonymous referees for their insightful suggestions.

2. Preliminaries

This section presents background material on noncommutative Hankel matrices and the Gelfand-Naimark-Segal (GNS) construction used throughout the paper.

We would like to stress out that the GNS construction is neither the only nor the first method for extracting optimizers in POP. The original idea for applying results on flat Hankel extensions [CF96] in an extraction algorithm was first described by Henrion and Lasserre in [HL05, Section 2.2] using only the language of linear algebra; see also [Lau09, Section 6.7] for a slight variation. An application of the GNS construction in this context then appeared in [PNA10, Section 3.2] (but see also [HM04, MP05]) and its role was later highlighted in [BKP16, Sections 1.5 and 1.12]. The presentation of Hankel extensions and the GNS construction in this section serves as a preparation for the robustness analysis in the next section.

2.1. **Hankel matrices.** Let $\boldsymbol{x} = (x_1, \ldots, x_g)$ be a tuple of freely noncommuting variables and fix the graded lexicographic ordering of the free monoid $<\boldsymbol{x}>$ of words in $\boldsymbol{x}$. That is, given $w_1, w_2 \in <\boldsymbol{x}>$ we have $w_1 < w_2$ if and only if $|w_1| < |w_2|$ or $|w_1| = |w_2|$, $w_1 = ux_iv_1$, $w_2 = ux_jv_2$ and $i < j$. On the free algebra of noncommutative polynomials $\mathbb{R}<\boldsymbol{x}>$ we define the scalar product $\langle \cdot, \cdot \rangle_2$ by declaring that words form an orthonormal basis. To be more clear we identify polynomials $f = \sum_{w \in <\boldsymbol{x}>} \alpha_w w$ with vectors $\vec{f} = (\alpha_w)_{w \in <\boldsymbol{x}>}$ when considered with respect to this scalar product. Note that

$$\langle \overrightarrow{wf}, \overrightarrow{wf} \rangle_2 = \langle \vec{f}, \vec{f} \rangle_2$$

for $f \in \mathbb{R}<\boldsymbol{x}>$ and $w \in <\boldsymbol{x}>$. The corresponding norm on $\mathbb{R}<\boldsymbol{x}>$ is denoted $\|\cdot\|_2$. The induced operator norm of linear maps acting on subspaces of $\mathbb{R}<\boldsymbol{x}>$ is also denoted $\|\cdot\|_2$. To every

$f = \sum_w \alpha_w w \in \mathbb{R}<\boldsymbol{x}>$ we assign the univariate polynomial

$$\mathrm{err}_f(t) = \sum_{w \in <\boldsymbol{x}>} |\alpha_w|\,|w|\,t^{|w|-1} \in \mathbb{R}[t].$$

For example, if $f = 1 + x_1x_2 - x_2x_1 + x_2^3$, then $\mathrm{err}_f(t) = 4t + 3t^2$.

Next we endow $\mathbb{R}<\boldsymbol{x}>$ with the unique involution satisfying $x_j^\intercal = x_j$. In particular, if $w = x_{j_1}x_{j_2}\cdots x_{j_\ell}$, then $w^\intercal = x_{j_\ell}\cdots x_{j_2}x_{j_1}$. For $d \in \mathbb{N}$ let $<\boldsymbol{x}>_d \subset <\boldsymbol{x}>$ be the set of words of length at most d and $\mathbb{R}<\boldsymbol{x}>_d \subset \mathbb{R}<\boldsymbol{x}>$ the subspace spanned by $<\boldsymbol{x}>_d$. A **Hankel matrix of order** d is a symmetric matrix H whose rows and columns are indexed by words in $\boldsymbol{x}$ of length at most d in the chosen graded lexicographic ordering that satisfies

$$H_{u_1,v_1} = H_{u_2,v_2} \tag{2.1}$$

for all $u_i, v_i \in <\boldsymbol{x}>_d$ such that $u_1^\intercal v_1 = u_2^\intercal v_2$.

Notation 2.1. For a matrix H let $\sigma_{\min}(H)$ denotes its least nonzero singular value (in numerical experiments we take for $\sigma_{\min}(H)$ the least eigenvalue larger than 10^{-5}). In particular, if H is a positive semidefinite (psd) matrix, then $\sigma_{\min}(H)$ is its least positive eigenvalue.

Example 2.2. Let $g = 2$. Then

$$\begin{array}{c} \\ 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_2x_1 \\ x_2^2 \end{array}\begin{array}{c} \begin{array}{ccccccc} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2x_1 & x_2^2 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & \frac{1}{2} & -1 \\ 0 & 0 & 1 & 1 & 1 & -1 & 0 \\ 1 & 0 & 1 & 3 & 1 & -1 & 1 \\ 0 & 1 & 1 & 1 & 2 & -\frac{1}{2} & -1 \\ 0 & \frac{1}{2} & -1 & -1 & -\frac{1}{2} & \frac{3}{2} & -\frac{1}{2} \\ 1 & -1 & 0 & 1 & -1 & -\frac{1}{2} & 2 \end{pmatrix} \end{array}$$

is a psd Hankel matrix of order 2 with respect to the graded lexicographic monomial ordering on $<\boldsymbol{x}>_2$ induced by $x_1 < x_2$.

Let H be a psd Hankel matrix of order d and for $p \in \mathbb{R}<\boldsymbol{x}>_d$ let $[p] \in \mathbb{R}<\boldsymbol{x}>_d/\ker H$ denote its equivalence class.

Lemma 2.3. *On the vector space* $\mathcal{H} = \mathbb{R}<\boldsymbol{x}>_d/\ker H$ *there is a scalar product defined by* $\langle [p],[q]\rangle_H = \langle H\vec{p}, \vec{q}\rangle_2$. *If* $p,q,r \in \mathbb{R}<\boldsymbol{x}>_d$ *satisfy* $\deg p + \deg r, \deg q + \deg r \le d$, *then*

$$\langle [rp],[q]\rangle_H = \langle [p],[r^\intercal q]\rangle_H. \tag{2.2}$$

Furthermore, for

$$p = p_0 + p_1 \in \ker H \oplus (\ker H)^\perp = \mathbb{R}<\boldsymbol{x}>_d$$

we have $[p] = [p_1]$ *and*

$$\sqrt{\sigma_{\min}(H)}\|\vec{p_1}\|_2 \le \|[p_1]\|_H \le \sqrt{\|H\|_2}\|\vec{p_1}\|_2. \tag{2.3}$$

Proof. The bilinear map $\langle\cdot,\cdot\rangle_H$ is well-defined because H is symmetric, and it is a scalar product since H is psd. Equation (2.2) follows by the Hankel property of H and linearity. Finally, (2.3) is a direct consequence of the definition of $\langle\cdot,\cdot\rangle_H$. □

If

$$K = \begin{pmatrix} H & B \\ B^\intercal & C \end{pmatrix}, \tag{2.4}$$

then we say that K is an **extension of** H. Let H be a psd Hankel matrix of order d, and let K be a Hankel extension of H of order $d+\delta$ for some $\delta \ge 1$. For $1 \le j \le g$ let

$$K^{(j)} = \left(K_{u,x_jv}\right)_{u,v\in<\boldsymbol{x}>_d}$$

be a submatrix of K, the so-called localizing matrix associated to x_j.

Lemma 2.4. *The matrix $K^{(j)}$ is symmetric. If K is psd, then there exists a matrix W_j such that $K^{(j)} = HW_j$.*

Proof. For the first part observe that

$$K^{(j)}_{u,v} = K_{u,x_j v} = K_{x_j v,u} = K_{v,x_j u} = K^{(j)}_{v,u}$$

for $u, v \in <\boldsymbol{x}>_d$ since K is Hankel and symmetric.

The second part is equivalent to $\operatorname{im} K^{(j)} \subseteq \operatorname{im} H$, which is furthermore equivalent to $\ker H \subseteq \ker K^{(j)}$. Let $f \in \mathbb{R}<\boldsymbol{x}>_d$ be such that $\vec{f} \in \ker H$. Hence $\langle K\vec{f}, \vec{f}\rangle_2 = 0$ and so $K\vec{f} = 0$ since K is psd. Therefore $\left(\begin{smallmatrix} H \\ B^\intercal \end{smallmatrix}\right)\vec{f} = 0$, but since $K^{(j)}$ is a submatrix of $\left(\begin{smallmatrix} H \\ B^\intercal \end{smallmatrix}\right)$, we see that $\vec{f} \in \ker K^{(j)}$. □

2.2. **Gelfand-Naimark-Segal (GNS) construction.** Let H be a psd Hankel matrix of order d and let K be a psd Hankel extension of H of order $d + \delta$. As in the proof of Lemma 2.4 we see that $\ker H = \ker K \cap \mathbb{R}<\boldsymbol{x}>_d$, so we have the inclusion of Hilbert spaces

$$\mathcal{H} = \mathbb{R}<\boldsymbol{x}>_d / \ker H \subseteq \mathbb{R}<\boldsymbol{x}>_{d+\delta} / \ker K = \mathcal{K}.$$

If $\pi^K : \mathcal{K} \to \mathcal{H}$ is the orthogonal projection, then define linear operators

$$\mathcal{X}_j^K : \mathcal{H} \to \mathcal{H}, \qquad [p] \mapsto \pi^K([x_j p]).$$

Lemma 2.5. *Operators $\mathcal{X}_j^K$ are well-defined and symmetric with respect to $\langle\cdot,\cdot\rangle_H$.*

Proof. If $\vec{p} \in \ker H$, then

$$\langle \overrightarrow{x_j p}, H\vec{q}\rangle_2 = \langle K\overrightarrow{x_j p}, \vec{q}\rangle_2 = \langle H\vec{p}, \overrightarrow{x_j q}\rangle_2 = 0$$

for every $q \in \mathbb{R}<\boldsymbol{x}>_d$ because K is a Hankel matrix. Therefore $\mathcal{X}_j^K$ is well defined. Furthermore,

$$\langle \mathcal{X}_j^K[p], [q]\rangle_H = \langle K\overrightarrow{x_j p}, \vec{q}\rangle_2 = \langle K\vec{p}, \overrightarrow{x_j q}\rangle_2 = \langle \vec{p}, K\overrightarrow{x_j q}\rangle_2 = \langle [p], \mathcal{X}_j^K[q]\rangle_H$$

holds for every $p, q \in \mathbb{R}<\boldsymbol{x}>_d$ since K is Hankel, so $\mathcal{X}_j^K$ is symmetric. □

Let $\mathcal{X}^K = (\mathcal{X}_1^K, \dots, \mathcal{X}_g^K)$. We have

$$f(\mathcal{X}^K)[p] = [fp]$$

for all $f, p \in \mathbb{R}<\boldsymbol{x}>$ satisfying $\deg f + \deg p \leq d$ and consequently

$$\langle f(\mathcal{X}^K)[1], [1]\rangle_H = \langle H\vec{f}, \vec{1}\rangle_2$$

for all $f \in \mathbb{R}<\boldsymbol{x}>_d$.

2.2.1. *Explicit matrix computation.* Let

$$H = USU^\intercal \tag{2.5}$$

be a singular value decomposition of H, i.e., S is a positive definite (pd) diagonal matrix whose size is the rank of H ($\operatorname{rk} H$) and $U^\intercal U = I$. Then the equivalence classes of columns of $U\sqrt{S}^{-1}$ form an orthogonal basis $\mathcal{B}$ of the Hilbert space $\mathcal{H}$. Consequently, if $p \in \mathbb{R}<\boldsymbol{x}>_d$, then the expansion of $[p]$ with respect to $\mathcal{B}$ is given by $\sqrt{S}U^\intercal\vec{p}$ and hence

$$\|[p]\|_H = \|\sqrt{S}U^\intercal\vec{p}\|_2.$$

Remark 2.6. Concretely, for $w \in <\boldsymbol{x}>$ we have

$$\|[w]\|_H = \|\sqrt{S}U^\intercal\vec{w}\|_2 = \sqrt{\langle\sqrt{S}U^\intercal\vec{w}, \sqrt{S}U^\intercal\vec{w}\rangle_2} = \sqrt{H_{w,w}}.$$

Also, the vector $[1] \in \mathcal{H}$ is with respect to $\mathcal{B}$ given as the first column of the matrix $\sqrt{S}U^\intercal$.

Since the operator $\mathcal{X}_j^K$ is determined by

$$\langle \mathcal{X}_j^K[p],[q]\rangle_H = \langle K\overrightarrow{x_jp}, \vec{q}\rangle_2 = \langle K^{(j)}\vec{p}, \vec{q}\rangle_2$$

for $p,q \in \mathbb{R}\langle\boldsymbol{x}\rangle_d$, the matrix representing $\mathcal{X}_j^K$ with respect to $\mathcal{B}$ equals

$$\sqrt{S}^{-1}U^\intercal K^{(j)}U\sqrt{S}^{-1}. \tag{2.6}$$

Example 2.7. Let $g=2$. Then

$$K = \begin{pmatrix} 1.0000 & 0.5000 & 0.5001 & 1.0483 & -0.5483 & -0.5483 & 1.0484 \\ 0.5000 & 1.0483 & -0.5483 & 1.0627 & -0.0144 & -0.6090 & 0.0606 \\ 0.5001 & -0.5483 & 1.0484 & -0.0144 & -0.5340 & 0.0606 & 0.9878 \\ 1.0483 & 1.0627 & -0.0144 & 1.4622 & -0.3995 & -0.8006 & 0.7863 \\ -0.5483 & -0.0144 & -0.5340 & -0.3995 & 0.3852 & 0.1917 & -0.7256 \\ -0.5483 & -0.6090 & 0.0606 & -0.8006 & 0.1917 & 0.4411 & -0.3804 \\ 1.0484 & 0.0606 & 0.9878 & 0.7863 & -0.7256 & -0.3804 & 1.3682 \end{pmatrix}$$

is a psd Hankel extension of order 2 of the upper-left 3×3 submatrix H of K. Both matrices have rank 2 (only two eigenvalues of each matrix are larger than 10^{-8}).

We can decompose $H = USU^\intercal$ where

$$U = \begin{pmatrix} -0.8165 & -0.0009 \\ -0.4090 & 0.7067 \\ -0.4075 & -0.7075 \end{pmatrix}, \quad S = \begin{pmatrix} 1.5000 & 0.0000 \\ 0.0000 & 1.5966 \end{pmatrix}$$

and extract

$$K^{(1)} = \begin{pmatrix} 0.5000 & 1.0483 & -0.5483 \\ 1.0483 & 1.0627 & -0.0144 \\ -0.5483 & -0.0144 & -0.5340 \end{pmatrix}, \qquad K^{(2)} = \begin{pmatrix} 0.5001 & -0.5483 & 1.0484 \\ -0.5483 & -0.6090 & 0.0606 \\ 1.0484 & 0.0606 & 0.9878 \end{pmatrix}.$$

Note that $\ker H$ is spanned by $(1,-1,-1)^T$.

The orthonormal basis $\mathcal{B}$ of $\mathcal{H}$ is therefore given by equivalence classes of columns

$$V = \begin{pmatrix} -0.6667 & -0.0007 \\ -0.3339 & 0.5593 \\ -0.3327 & -0.5599 \end{pmatrix}.$$

Operators $\mathcal{X}_1^K, \mathcal{X}_2^K$ are in this basis represented by matrices

$$\hat{K}^{(1)} = \begin{pmatrix} 0.5019 & -0.8931 \\ -0.8931 & 0.1727 \end{pmatrix}, \qquad \hat{K}^{(2)} = \begin{pmatrix} 0.4981 & 0.8939 \\ 0.8939 & 0.0825 \end{pmatrix}.$$

2.2.2. *Flat extensions.* If K is an extension of H and $\operatorname{rk} K = \operatorname{rk} H$, then we say that K is a **flat** extension of H. Now let H be a psd Hankel matrix of order d and K a flat psd Hankel extension of H of order $d+\delta$. In this case $\mathcal{H} = \mathcal{K}$ and the Hankel property of K implies that

$$f(\mathcal{X}^K)[p] = [fp] \in \mathcal{K}$$

for every $f,p \in \mathbb{R}\langle\boldsymbol{x}\rangle$ satisfying $\deg f \le d$ and $\deg f + \deg p \le d+\delta$. Consequently

$$\langle f_1(\mathcal{X}^K)[p_1], f_2(\mathcal{X}^K)[p_2]\rangle_H = \langle K\overrightarrow{f_1p_1}, \overrightarrow{f_2p_2}\rangle_2 \tag{2.7}$$

for all $p_i \in \mathbb{R}\langle\boldsymbol{x}\rangle_d$ and $f_i \in \mathbb{R}\langle\boldsymbol{x}\rangle_\delta$.

2.2.3. *Putting it all together.* To motivate the concepts introduced in this section let us demonstrate how a relaxation hierarchy, flat extensions and the truncated GNS construction are used to extract global minimizer for a nc polynomial $f \in \mathbb{R}\langle\boldsymbol{x}\rangle_{2d}$. That is, if $\mathbb{S}_n$ denotes the space of symmetric $n\times n$ matrices, then we are interested in finding a tuple $X_0 \in \mathbb{S}_{n_0}^g$ and a unit vector $v_0 \in \mathbb{R}^{n_0}$ for some $n_0 \in \mathbb{N}$ such that

$$v_0^\intercal f(X_0)v_0 = f^* := \inf\left\{v^\intercal f(X)v \colon n\in\mathbb{N},\ X\in\mathbb{S}_n^g,\ \|v\|=1\right\}. \tag{2.8}$$

For the proofs of the following statements we refer to [BKP16, Section 4.2]. Consider

$$\varphi^{(d+1)} = \inf\left\{\varphi(f) \colon \varphi \text{ a linear functional on } \mathbb{R}{<}\boldsymbol{x}{>}_{2(d+1)},\ \varphi(g^\intercal g) \geq 0,\ \varphi(1) = 1\right\}.$$

A minimizer φ_0 for $\varphi^{(d+1)}$ can be found using SDP and can be given by a psd Hankel matrix K of order $d+1$, i.e.,

$$\varphi_0(u^* v) = \langle K\vec{u}, \vec{v}\rangle_2$$

for every $u, v \in \mathbb{R}{<}\boldsymbol{x}{>}_{d+1}$. Let H be the Hankel submatrix of order d of K (that is, H is the submatrix in the top-left corner of K of appropriate size). If K is a flat extension of H, then $f^* = \varphi^{(d+1)} = \varphi_0(f)$. Write $f = \sum_i p_i^* q_i$ for some $p_i, q_i \in \mathbb{R}{<}\boldsymbol{x}{>}_d$. Then

$$\varphi_0(f) = \sum_i \varphi_0(p_i^* q_i) = \sum_i \langle K\vec{p_i}, \vec{q_i}\rangle_2 = \sum_i \langle p_i(\mathcal{X}^K)[1], q_i(\mathcal{X}^K)[1]\rangle_H = \langle f(\mathcal{X}^K)[1], [1]\rangle_H$$

by (2.7), so f^* is attained at the tuple $\mathcal{X}^K$ of operators on $\mathcal{H}$ and the unit vector $[1] \in \mathcal{H}$.

3. Robustness of the GNS construction

This section contains our first main results. As explained in Subsection 2.2, to each noncommutative Hankel extension K of H we can associate a tuple of matrices $\mathcal{X}^K$ approximating freely noncommutative variables with respect to the scalar product induced by H. We explicitly quantify robustness of this process. That is, given Hankel matrices K, L both extending H, we give lower and upper bounds on $\|\mathcal{X}^K - \mathcal{X}^L\|_H$; see Theorem 3.2. (Here $\|\mathcal{X}^K - \mathcal{X}^L\|_H$ denotes the operator norm of $\mathcal{X}^K - \mathcal{X}^L$ on the Hilbert space $\mathcal{H}$ induced by H as in Subsection 2.2.) How this pertains to optimizer extraction in noncommutative polynomial optimization is explained in Corollaries 3.5 and 3.6.

Let H be a psd Hankel matrix of order d. Let $\mathrm{HExt}_{H,\delta} \subseteq \mathrm{End}_{\mathbb{R}}\, \mathbb{R}{<}\boldsymbol{x}{>}_{d+\delta}$ be the set of all psd Hankel extensions of H of order $d+\delta$ with the induced subspace topology. The GNS construction can be viewed as a map

$$\mathrm{HExt}_{H,\delta} \to (\mathrm{End}_{\mathbb{R}}\, \mathcal{H})^g, \qquad K \mapsto \mathcal{X}^K = \left(\mathcal{X}_1^K, \dots, \mathcal{X}_g^K\right).$$

By (2.6) this is a restriction of an affine linear map. In this section we quantify its boundedness and Lipschitz continuity, and test our estimates with several examples. Recall from Notation 2.1 that $\sigma_{\min}(H)$ denotes the least nonzero singular value of H.

Proposition 3.1. *If K is a psd Hankel extension of H, then*

$$\frac{\|K^{(j)}\|_2}{\|H\|_2} \leq \|\mathcal{X}_j^K\|_H \leq \frac{\|K^{(j)}\|_2}{\sigma_{\min}(H)}$$

for every $1 \leq j \leq g$.

Proof. By (2.6) we have

$$\|\mathcal{X}_j^K\|_H = \left\|\sqrt{S}^{-1} U^\intercal K^{(j)} U \sqrt{S}^{-1}\right\|_2.$$

Note that the columns of U span $\operatorname{im} H$, so $\|U^\intercal K^{(j)} U\|_2 = \|K^{(j)}\|_2$ by Lemma 2.4. The rest follows by

$$\frac{\|U^\intercal K^{(j)} U\|_2}{\|H\|_2} \leq \left\|\sqrt{S}^{-1} U^\intercal K^{(j)} U \sqrt{S}^{-1}\right\|_2 \leq \|S^{-1}\|_2 \|U^\intercal K^{(j)} U\|_2$$

and $\|S^{-1}\|_2 = \sigma_{\min}(H)^{-1}$. □

Theorem 3.2. *Let K and L be psd Hankel extensions of H. Then for every $1 \leq j \leq g$,*

$$\frac{\|K^{(j)} - L^{(j)}\|_2}{\|H\|_2} \leq \|\mathcal{X}_j^K - \mathcal{X}_j^L\|_H \leq \frac{\|K^{(j)} - L^{(j)}\|_2}{\sigma_{\min}(H)}.$$

Proof. Let $H = USU^\intercal$ be as in (2.5). By (2.6) we have

$$\|\mathcal{X}_j^K - \mathcal{X}_j^L\|_H = \|\sqrt{S}^{-1} U^\intercal (K^{(j)} - L^{(j)}) U \sqrt{S}^{-1}\|_2.$$

Since $\operatorname{im} K^{(j)}, \operatorname{im} L^{(j)} \subseteq \operatorname{im} H$, Lemma 2.4 implies

$$\|U^\intercal (K^{(j)} - L^{(j)}) U\|_2 = \|K^{(j)} - L^{(j)}\|_2$$

and the rest follows as in the proof of Proposition 3.1. □

Remark 3.3. All the inequalities in Proposition 3.1 and Theorem 3.2 are equalities if and only if $\sigma_{\min}(H) = \|H\|_2$, i.e., H is a scalar multiple of an orthogonal projection. If H is nonsingular, then $H = \alpha I$ for some $\alpha > 0$. If $d > 1$, then $0 = H_{x_1^2,1} = H_{x_1,x_1} = \alpha$, a contradiction. On the other hand, if $d = 1$, then the $(g+1) \times (g+1)$ matrix $H = \alpha I$ is indeed a pd Hankel matrix of order 1. If H is singular, then one can obtain arbitrary orders, for example by taking $H = (1)_{u,v}$.

Example 3.4. If H is nonsingular, then the ratio between lower and upper bound in Proposition 3.1 and Theorem 3.2 equals the condition number $\operatorname{cond}(H)$. Unfortunately it is known (see e.g. [Tyr94]) that for positive definite Hankel matrices, $\operatorname{cond}(H)$ grows exponentially with the order of H. However, while running some simulations for the case $g = 2$ we observed the following.

(1) We tested Proposition 3.1. For a random psd Hankel extension K of H we find $t \in [0, 1]$ such that

$$\|\mathcal{X}_1^K\|_H = (1-t)\frac{\|K^{(1)}\|_2}{\|H\|_2} + t\frac{\|K^{(1)}\|_2}{\sigma_{\min}(H)}.$$

If H is invertible, then t will likely be close to 0. On the other hand, if the rank of H is small, then t can get close to 1.

Concretely, we computed the value t (rounded to two decimal places) for 100000 random psd Hankel matrices H of order 3, whose most frequent rank was 3, and their psd Hankel extensions K of order 4. The distribution of the values t is presented in Figure 1.

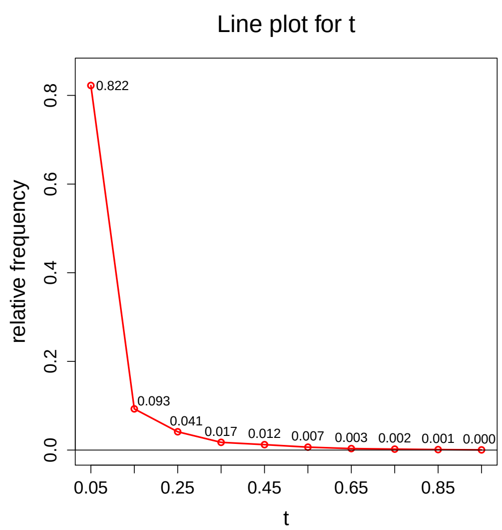


Figure 1. Distribution of values t for Proposition 3.1. We can see that out of 100000 random Hankel extensions more than 80 % have value t at most 0.05, which suggests that the inequality from Proposition 3.1 is left tight.

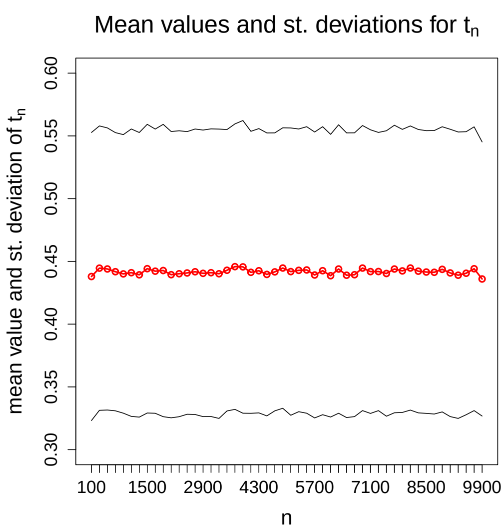


Figure 2. Mean values (red circles) and standard deviations (upper and lower black plots) for t_n from Example 3.4, part (2), i.e., each red circle represents the mean value of 2000 values of t_n (100 random extensions and 20 values of $n \in \{10, 20, \dots, 190, 200\} + 200k$, $k = 0, 1, 2, \dots, 49$).

(2) In a similar way we tested Theorem 3.2. After choosing a random psd Hankel extension K of H and a small $\varepsilon = 0.001$, we generated random Hankel extensions K_n ($n \in \mathbb{N}$) of H satisfying $\|K - K_n\| \leq \frac{\varepsilon}{n}$. Then we determined $t_n \in [0,1]$ such that

$$\|\mathcal{X}_1^K - \mathcal{X}_1^{K_n}\|_H = (1-t_n)\frac{\|K^{(1)} - K_n^{(1)}\|_2}{\|H\|_2} + t_n\frac{\|K^{(1)} - K_n^{(1)}\|_2}{\sigma_{\min}(H)}.$$

We repeat this for 100 random Hankel matrices H and for $n = 10, 20, ..., 10000$. We have observed that the minimum value of t_n was 0.12562 and the maximum value of t_n was 0.85242. We have computed for each $k = 0, 1, 2\ldots, 49$ the mean value and standard deviation of all values of t_n for $n \in \{10, 20, \ldots, 190, 200\} + 200k$, for all 100 random extensions. These 50 mean values and standard deviations are plotted in Figure 2.

In summary, the inequality of Proposition 3.1 is generically left tight (with sporadic values close to the upper bound), while the inequality of Theorem 3.2 is more balanced.

We next present the sensitivity of the values of a perturbed Riesz functional applied to a polynomial. This is of particular importance for polynomial optimization. In fact, this enables us to estimate how far from an optimum we are when applying the GNS construction to a Riesz functional that is only approximately optimal.

Corollary 3.5. *Let K and L be psd Hankel extensions of H and set $c = \frac{\max_j\{\|K^{(j)}\|_2, \|L^{(j)}\|_2\}}{\sigma_{\min}(H)}$. For every $f \in \mathbb{R}\langle\boldsymbol{x}\rangle$ we have*

$$\|f(\mathcal{X}^K) - f(\mathcal{X}^L)\|_H \leq \frac{\operatorname{err}_f(c)}{\sigma_{\min}(H)}\|K - L\|_2. \tag{3.1}$$

Proof. After reducing to the case $f = w \in \langle\boldsymbol{x}\rangle$ we prove (3.1) by induction on $|w|$ using Theorem 3.2 and Proposition 3.1. □

Corollary 3.6. *Let K and L be psd Hankel extensions of H and set $c = \frac{\max_j\{\|K^{(j)}\|_2, \|L^{(j)}\|_2\}}{\sigma_{\min}(H)}$. If $f \in \mathbb{R}\langle\boldsymbol{x}\rangle$ is written in the form*

$$f = f_0 + \sum_{|w|>0} p_w w q_w, \qquad \deg f_0 \leq 2d, \ \deg p_w = \deg q_w = d,$$

then

$$\big|\langle f(\mathcal{X}^K)[1],[1]\rangle_H - \langle f(\mathcal{X}^L)[1],[1]\rangle_H\big| \leq \frac{\sum_w |w|c^{|w|-1}\|[p_w]\|_H\|[q_w]\|_H}{\sigma_{\min}(H)}\|K - L\|_2. \tag{3.2}$$

Proof. If $f_0 = \sum_i h_i k_i$ and $\deg h_i, k_i \leq d$, then

$$\langle f_0(\mathcal{X}^K)[1],[1]\rangle_H = \sum_i \langle [k_i], [h_i^{\intercal}]\rangle = \langle f_0(\mathcal{X}^L)[1],[1]\rangle_H.$$

The rest now follows from the Cauchy-Schwarz inequality and Corollary 3.5. □

Example 3.7. In this example we illustrate the tightness of the inequality (3.2) of Corollary 3.6. We generated 100 random polynomials that were sums of hermitian squares of degree $2d_0$. For each polynomial we computed the psd Hankel matrix H of order $d_0 = 1$, which is an optimum of the dual problem (1.4) (or see ($\text{Eig}_{\text{DSDP}}^{d_0}$) in [BKP16, p. 65]). For each H we computed 10 psd Hankel extensions of order d_1, which was either $d_0 + 1 = 2$ or $d_0 + 2 = 3$, by solving a randomized SDP (inspired by Nie's method [Nie14]) ($\text{Constr-Eig}_{\text{RAND}}^{(d1)}$) from [BKP16, p. 72]; in this SDP we took the set of empty constraints, which means that we computed ($\text{Eig}_{\text{DSDP}}^{d_0}$) with a random positive definite objective function while forcing the submatrix corresponding to words of length $\leq d_0$ to be equal to H. Then we computed for each pair K, L of psd Hankel extensions and for each word $f = uwv \in \mathbb{R}\langle\boldsymbol{x}\rangle_{2d_1} \backslash \mathbb{R}\langle\boldsymbol{x}\rangle_{2d_0}$ (note that $|u| = |v| = d_0$) several

$d_0 = 1,\ d_1 = 2,\ n = 108{,}000$						
	LHS	RHS	errF	$\|K-L\|_2$	$\|\text{RHS}-\text{LHS}\|$	errF − Quot
median	2.5351e-11	9.6568e-10	5.9642e-02	1.6718e-08	7.0712e-10	9.953
Q1	9.0183e-13	8.9515e-11	1.2765e-02	5.6846e-09	6.2850e-11	2.898
Q3	1.8407e-09	1.8909e-08	2.3409e-01	8.5911e-08	1.0987e-08	15.460
average	2.0521e-05	4.0828e-04	4.9723e-01	9.1715e-05	3.8775e-04	10.420
Δ	5.0823e-02	3.3379e-01	2.3638e+01	1.4121e-02	3.3390e-01	23.560
st. dev.	5.1253e-04	7.7914e-03	1.7760e+00	8.1159e-04	7.5582e-03	7.97
$d_0 = 1,\ d_1 = 3,\ n = 540{,}000$						
	LHS	RHS	errF	$\|K-L\|_2$	$\|\text{RHS}-\text{LHS}\|$	errF − Quot
median	2.9215e-12	6.1935e-10	4.7106e-02	1.4594e-08	6.0173e-10	13.32
Q1	3.0391e-14	2.1174e-11	3.3607e-03	5.5781e-09	2.0627e-11	1.84
Q3	9.4635e-10	1.0773e-07	4.9995e-01	2.4627e-07	1.0162e-07	78.62
eaverage	7.9328e-04	3.9437e+01	3.0811e+01	3.9607e-02	3.9436e+01	215.30
Δ	9.7981e+00	6.6693e+04	6.7564e+03	9.8712e+00	6.6693e+04	6756.00
st. dev.	4.2421e-02	1.1710e+03	2.8765e+02	4.7148e-01	1.1710e+03	740.68

Table 1. Numerical results for LHS, RHS, errF, $\|K-L\|_2$, $|\text{RHS}-\text{LHS}|$ and errF − Quot, obtained by computing 100 psd Hankel matrices of order $d_0 = 1$ which were related to random polynomials of degree 2. For each Hankel matrix we computed 10 psd Hankel extensions of order 2 (the upper half of the table) and 10 psd Hankel extensions of order 3 (the lower part of the table). We evaluated (3.2) for each Hankel matrix, for each pair of extensions and for each monomial f of length 3,4 (the upper part of the table) and of lengths $3,\dots,6$ (the lower part of the table). For each of the quantities presented in this table we computed the median, the first and the third quartile, the average, the maximum difference in the data and the standard deviation. We can see in both parts of the table that the inequality (3.2) is in most cases very tight; the last two columns contain statistics about the difference $|\text{RHS}-\text{LHS}|$ (we take the absolute value since in some cases this difference is slightly negative for numerical reasons, for example the minimum of RHS − LHS is $-2.671\cdot 10^{-6}$) and errF − Quot. The third quartile of $|\text{RHS}-\text{LHS}|$ for the lower part of the table is $1.0162\cdot 10^{-7}$, which means that 75 % of these differences are below this value. The last column and the errF column can not be compared since the last column was computed over a much smaller set, as described in the main text.

parts of (3.2):

$$\text{LHS} = \left|\langle f(\mathcal{X}^K)[1],[1]\rangle_H - \langle f(\mathcal{X}^L)[1],[1]\rangle_H\right| \tag{3.3}$$

$$\text{RHS} = \frac{|w|c^{|w|-1}\sqrt{H_{u,u}}\sqrt{H_{v,v}}}{\sigma_{\min}(H)}\|K-L\|_2 \tag{3.4}$$

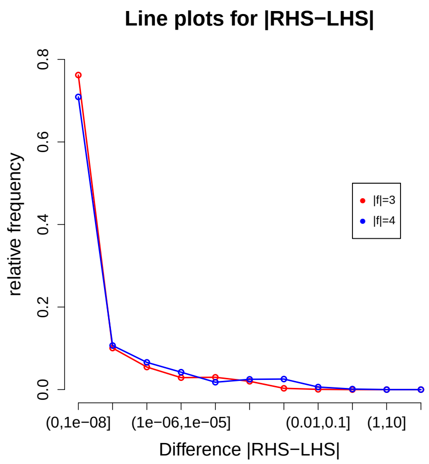
$$\text{errF} = \frac{|w|c^{|w|-1}\sqrt{H_{u,u}}\sqrt{H_{v,v}}}{\sigma_{\min}(H)} \tag{3.5}$$

$$\text{Quot} = \frac{\left|\langle f(\mathcal{X}^K)[1],[1]\rangle_H - \langle f(\mathcal{X}^L)[1],[1]\rangle_H\right|}{\|K-L\|_2} \tag{3.6}$$

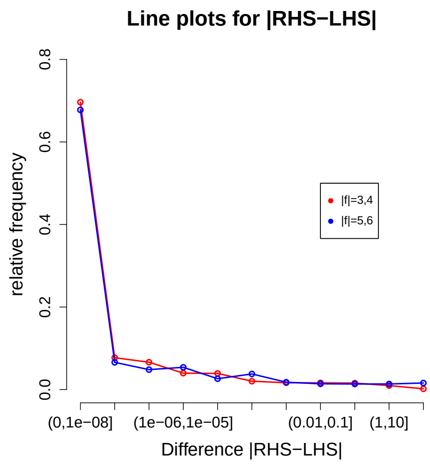
Here we used the fact that $\|w\|_H = \sqrt{H_{w,w}}$ for every $w \in \langle \boldsymbol{x}\rangle_{d_0}$. Note that RHS = errF · $\|K-L\|_2$ and Quot = LHS/$\|K-L\|_2 \leq$ errF. We point out that comparing Quot to errF needs special attention since there are many quadruples (f,H,K,L) such that RHS − LHS is very small (less than 10^{-10}), in some cases even slightly negative (e.g. around -10^{-15}) due to numerical errors. In these situations also the value $\|K-L\|_2$ is very small, even smaller than RHS − LHS. When we divide these two values the result might get big, even very negative. Therefore we evaluated Quot and compared it with errF only for quadruples (f,H,K,L) where RHS − LHS > 0 and $\|K-L\|_2 > 10^{-5}$, since the threshold for selecting $\sigma_{\min}(H)$ was 10^{-5}.

We report results in Table 1 and depict them in Figures 3–4.

We additionally evaluated the inequality (3.2) for all generated psd Hankel matrices of order $d_0 = 1$ and their extensions of order $d_1 = 2$, $d_1 = 3$, respectively, as described above, with monomials f of degree $2d_1+1, 2d_1+2$, to see if the distribution of the difference $|\text{RHS}-\text{LHS}|$



(a) Line plots for $|\mathrm{RHS} - \mathrm{LHS}|$ $(d_0 = 1, d_1 = 2)$.



(b) Line plots for $|\mathrm{RHS} - \mathrm{LHS}|$ $(d_0 = 1, d_1 = 3)$.

Figure 3. Line plots for $|\mathrm{RHS} - \mathrm{LHS}|$ demonstrate that the inequality (3.2) is in most cases very tight; in 75 % of random examples it is below $1.0162e - 07$. The left plot shows the differences $|\mathrm{RHS} - \mathrm{LHS}|$ separately for monomials f of length 3 and 4, while the right plot shows the differences for monomials f of length 3,4 and 5,6.

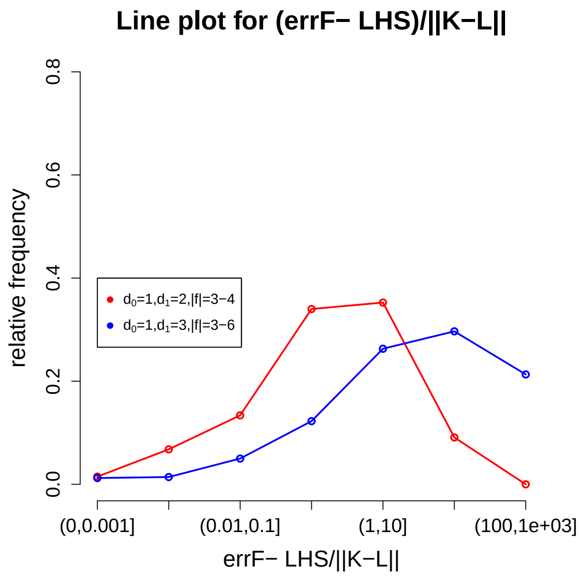
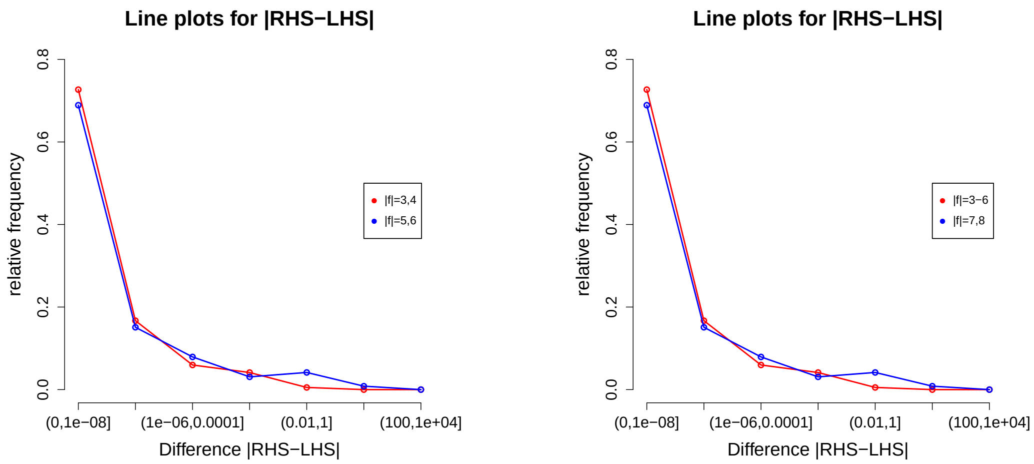


Figure 4. This figure depicts the differences $\mathrm{errF} - \mathrm{Quot}$ for $d_0 = 1$ and extensions of order $d_1 = 2$ with monomials f of degree $3, 4$ (red plot) and for $d_1 = 3$ with monomials f of degree $3, \ldots, 6$ (blue plot). We removed all values where RHS–LHS was negative or where $\|K - L\| < 10^{-5}$.

changes. In Figure 5 we show that the distribution of the differences does not change significantly, i.e., for a very large proportion of the cases the difference $|\mathrm{RHS} - \mathrm{LHS}|$ is very small and for a very small proportion it exceeds the value 100.

4. Robustness of the GNS construction with constraints

This section gives a robustness analysis of the GNS construction with constraints. Our main result, Theorem 4.1, quantifies how much constraints can be violated when performing a GNS construction on non-flat data. That is, given psd Hankel extensions K and L of H with L being flat, we bound the constraints violation when applying a GNS construction on K in terms of



(a) Line plots for $|\mathrm{RHS} - \mathrm{LHS}|$ $(d_0 = 1, d_1 = 2)$.

(b) Line plots for $|\mathrm{RHS} - \mathrm{LHS}|$ $(d_0 = 1, d_1 = 3)$.

Figure 5. Line plots for $|\mathrm{RHS} - \mathrm{LHS}|$ demonstrate that the inequality (3.2) remains very tight even if we consider f with degree larger than the order of the extension.

the norm of $K - L$. As a consequence we obtain a sufficient local condition for nonexistence of flat extensions in Corollary 4.2. In Subsection 4.2 we strengthen our robustness analysis of the GNS construction with constraints for almost flat matrices. Here a Hankel extension K of H is almost flat if K is close to $K^\flat$, the canonical flat (possibly non-Hankel) extension of H determined by K; see Subsection 4.2 for the definition of $K^\flat$. The advantage of this approach is that the error estimates are derived solely from K without assuming that there is an actual flat Hankel matrix near K. Finally, in Subsection 4.4 we interpret our results for NCPOP.

4.1. **Near flat extensions.** Let H be a psd Hankel matrix of order d, $s \in \mathbb{R}\langle \boldsymbol{x} \rangle_{2\delta}$ a symmetric nc polynomial and let K be a psd Hankel extension of H of order $d + \delta$. Write

$$s = \sum_{|u|,|v| \leq \delta} \alpha_{u,v} uv \tag{4.1}$$

and define the **localizing matrix**

$$K_s^{\Uparrow} = \left(\sum_{|u|,|v| \leq \delta} \alpha_{u,v} K_{u^\intercal u', vv'} \right)_{u',v'} \tag{4.2}$$

indexed by $u', v' \in \mathbb{R}\langle \boldsymbol{x} \rangle_d$. Observe that $K_s^{\Uparrow}$ depends only on s and not on the decomposition (4.1) by the Hankel property of K. Also, $K_s^{\Uparrow}$ is symmetric because s is symmetric. Note that $K_s^{\Uparrow}$ is psd if and only if

$$\sum_{|u|,|v| \leq \delta} \alpha_{u,v} \langle K \overrightarrow{vp}, \overrightarrow{u^\intercal p} \rangle_2 \geq 0 \tag{4.3}$$

holds for all $p \in \mathbb{R}\langle \boldsymbol{x} \rangle_d$. If K is a flat Hankel extension of H, then $K_s^{\Uparrow}$ is psd if and only if $s(\mathcal{X}^K)$ is psd; see e.g. [BKP16, Theorem 1.69].

For a symmetric matrix S let $\lambda_{\min}(S)$ be its smallest eigenvalue.

Theorem 4.1. *Let $s \in \mathbb{R}\langle \boldsymbol{x} \rangle_{2\delta}$ be a symmetric polynomial and let c_1 be the sum of absolute values of non-constant coefficients in s. Let K be a psd Hankel extension of H of order $d + \delta$ such that $K_s^{\Uparrow}$ is psd. Furthermore suppose that L is a flat psd Hankel extension of H of order*

$d+\delta$. *If* $c = \frac{\max_j\{\|K^{(j)}\|_2, \|L^{(j)}\|_2\}}{\sigma_{\min}(H)}$, *then*

$$\lambda_{\min}\left(s(\mathcal{X}^K)\right) \geq \frac{\lambda_{\min}\left(K_s^{\Uparrow}\right)}{\|H\|_2} - \|K-L\|_2 \frac{c_1 + \mathrm{err}_s(c)}{\sigma_{\min}(H)}.$$

Proof. Let $p \in (\ker H)^{\perp}$. By Corollary 3.5 we have

$$\left|\left\langle (s(\mathcal{X}^K) - s(\mathcal{X}^L))[p],[p]\right\rangle_H\right| \leq \frac{\mathrm{err}_s(c)}{\sigma_{\min}(H)}\|K-L\|_2\langle [p],[p]\rangle_H.$$

By the flatness of L, (4.3) and (2.3) we have

$$\begin{aligned}
\langle s(\mathcal{X}^L)[p],[p]\rangle_H &= \sum_{|u|,|v|\leq\delta} \alpha_{u,v}\langle v(\mathcal{X}^L)[p], u^{\intercal}(\mathcal{X}^L)[p]\rangle_H \\
&= \sum_{|u|,|v|\leq\delta} \alpha_{u,v}\langle L\overrightarrow{vp}, \overrightarrow{u^{\intercal}p}\rangle_2 \\
&= \langle K_s^{\Uparrow}\vec{p},\vec{p}\rangle_2 + \sum_{|u|,|v|\leq\delta,\ uv\neq 1} \alpha_{u,v}\langle (L-K)\overrightarrow{vp}, \overrightarrow{u^{\intercal}p}\rangle_2 \\
&\geq \lambda_{\min}\left(K_s^{\Uparrow}\right)\|\vec{p}\|_2^2 - \|K-L\|_2 c_1\|\vec{p}\|_2^2 \\
&\geq \left(\frac{\lambda_{\min}\left(K_s^{\Uparrow}\right)}{\|H\|_2} - \frac{c_1}{\sigma_{\min}(H)}\|K-L\|_2\right)\langle [p],[p]\rangle_H.
\end{aligned}$$

Hence

$$\begin{aligned}
\langle s(\mathcal{X}^K)[p],[p]\rangle_H &= \langle s(\mathcal{X}^L)[p],[p]\rangle_H + \langle (s(\mathcal{X}^K)-s(\mathcal{X}^L))[p],[p]\rangle_H \\
&\geq \left(\frac{\lambda_{\min}\left(K_s^{\Uparrow}\right)}{\|H\|_2} - \frac{c_1+\mathrm{err}_s(c)}{\sigma_{\min}(H)}\|K-L\|_2\right)\langle [p],[p]\rangle_H.
\end{aligned}$$ □

Corollary 4.2. *Let K be a psd Hankel extension of H of order $d+\delta$ and s a symmetric polynomial of degree at most 2δ. If $K_s^{\Uparrow}$ is psd and*

$$\frac{\lambda_{\min}\left(K_s^{\Uparrow}\right)}{\|H\|_2} > \lambda_{\min}\left(s(\mathcal{X}^K)\right),$$

then there exists a neighborhood of K not containing any flat psd Hankel extension of H.

Proof. Immediate consequence of Theorem 4.1. □

4.2. **Almost flat extensions.** Let H be a psd Hankel matrix of order d and let

$$K = \begin{pmatrix} H & B \\ B^{\intercal} & C\end{pmatrix}$$

be a Hankel extension of H of order $d+\delta$. As in Lemma 2.4 we see that $\ker H \subseteq \ker B^{\intercal}$, so there exists a matrix W such that $B = HW$. Let

$$K^{\flat} = \begin{pmatrix} H & B \\ B^{\intercal} & W^{\intercal}HW\end{pmatrix}.$$

Note that $K^{\flat}$ is psd and $\mathrm{rk}\, K^{\flat} = \mathrm{rk}\, H$ by

$$K^{\flat} = \begin{pmatrix} I & 0 \\ 0 & W^{\intercal}\end{pmatrix}\left(\begin{pmatrix} 1 & 1 \\ 1 & 1\end{pmatrix}\otimes H\right)\begin{pmatrix} I & 0 \\ 0 & W\end{pmatrix}.$$

In general, $K^{\flat}$ does not satisfy the Hankel property (2.1). However, in this subsection we show that $\|K-K^{\flat}\|_2$ can be used to estimate violation of constraints when performing a GNS construction with non-flat Hankel extensions.

Throughout the rest of the subsection we assume that H is invertible, i.e., $\mathcal{H} = \mathbb{R}\langle\boldsymbol{x}\rangle_d$. Then $\sigma_{\min}(H) = \lambda_{\min}(H) = \|H^{-1}\|_2^{-1}$. From (2.6) we see that the matrix representing $\mathcal{X}_j^K$ with respect to the standard word basis of $\mathcal{H}$ equals $H^{-1}K^{(j)}$.

Fix $j \in \{1,\dots,g\}$. For $p \in \mathbb{R}\langle\boldsymbol{x}\rangle_d$ let $p_\bullet \in \mathbb{R}\langle\boldsymbol{x}\rangle_d$ be such that $[p_\bullet] = \mathcal{X}_j^K[p]$; hence

$$\overrightarrow{p_\bullet} = H^{-1}K^{(j)}\overrightarrow{p}. \tag{4.4}$$

Lemma 4.3. *Let K be a psd Hankel extension of H of order $d+\delta$. Denote $c = \frac{\max_j\{\|K^{(j)}\|_2\}}{\sigma_{\min}(H)}$. If $p,q \in \mathbb{R}\langle\boldsymbol{x}\rangle_d$ and $w \in \langle\boldsymbol{x}\rangle_\delta$, then*

1. $\|\overrightarrow{p_\bullet}\|_2 \le \frac{\|K^{(j)}\|_2}{\sigma_{\min}(H)}\|\overrightarrow{p}\|_2$;
2. $K^\flat\overrightarrow{x_jp} = K\overrightarrow{p_\bullet}$ *and* $\|K(\overrightarrow{p_\bullet} - \overrightarrow{x_jp})\| \le \|K-K^\flat\|_2\|\overrightarrow{p}\|_2$;
3. $|\langle w(\mathcal{X}^K)[p],[q]\rangle_H - \langle K\overrightarrow{wp},\overrightarrow{q}\rangle_2| \le \|K-K^\flat\|_2\|\overrightarrow{p}\|_2\|\overrightarrow{q}\|_2\sum_{i=0}^{|w|-1}c^i$.

Proof. (1) is clear by (4.4).

(2) If $\deg p < d$, then $p_\bullet = x_jp$, so we can assume $\deg p = d$. If

$$K = \begin{pmatrix} H & B \\ B^\intercal & C\end{pmatrix}$$

with $B = HW$, then

$$\begin{aligned} H\overrightarrow{p_\bullet} &= H(H^{-1}K^{(j)}\overrightarrow{p}) = K^{(j)}\overrightarrow{p} = B\overrightarrow{x_jp},\\ B^\intercal\overrightarrow{p_\bullet} &= W^\intercal H\overrightarrow{p_\bullet} = W^\intercal B\overrightarrow{x_jp} = (W^\intercal HW)\overrightarrow{x_jp}\end{aligned}$$

and therefore

$$K^\flat\overrightarrow{x_jp} - K\overrightarrow{p_\bullet} = \begin{pmatrix} B\overrightarrow{x_jp} \\ W^\intercal HW\overrightarrow{x_jp}\end{pmatrix} - \begin{pmatrix} H\overrightarrow{p_\bullet} \\ B^\intercal\overrightarrow{p_\bullet}\end{pmatrix} = 0$$

and

$$K(\overrightarrow{x_jp} - \overrightarrow{p_\bullet}) = \begin{pmatrix} B\overrightarrow{x_jp} \\ C\overrightarrow{x_jp}\end{pmatrix} - \begin{pmatrix} H\overrightarrow{p_\bullet} \\ B^\intercal\overrightarrow{p_\bullet}\end{pmatrix} = \begin{pmatrix} 0 \\ C - W^\intercal HW\end{pmatrix}\overrightarrow{x_jp}.$$

(3) We prove this claim by induction on $|w|$, with the basis case $w=1$ being obvious. For $1 \le |w| < \delta$ let $r \in \mathbb{R}\langle\boldsymbol{x}\rangle_d$ be such that $[r] = w(\mathcal{X}^K)[p]$. Then

$$\begin{aligned}\langle \mathcal{X}_j^K w(\mathcal{X}^K)[p],[q]\rangle_H - \langle K\overrightarrow{x_jwp},\overrightarrow{q}\rangle_2 &= \langle K\overrightarrow{x_jr},\overrightarrow{q}\rangle_2 - \langle K\overrightarrow{x_jwp},\overrightarrow{q}\rangle_2\\ &= \langle \overrightarrow{r} - \overrightarrow{wp}, K\overrightarrow{x_jq}\rangle_2\\ &= \langle \overrightarrow{r} - \overrightarrow{wp}, K^\flat\overrightarrow{x_jq}\rangle_2 + \langle \overrightarrow{r} - \overrightarrow{wp}, (K-K^\flat)\overrightarrow{x_jq}\rangle_2.\end{aligned} \tag{4.5}$$

Firstly,

$$|\langle \overrightarrow{r} - \overrightarrow{wp}, (K-K^\flat)\overrightarrow{x_jq}\rangle_2| = |\langle \overrightarrow{wp}, (K-K^\flat)\overrightarrow{x_jq}\rangle_2| \le \|K-K^\flat\|_2\|\overrightarrow{p}\|_2\|\overrightarrow{q}\|_2 \tag{4.6}$$

by (2). Secondly,

$$\begin{aligned}\langle \overrightarrow{r} - \overrightarrow{wp}, K^\flat\overrightarrow{x_jq}\rangle_2 &= \langle \overrightarrow{r} - \overrightarrow{wp}, K\overrightarrow{q_\bullet}\rangle_2\\ &= \langle K\overrightarrow{r},\overrightarrow{q_\bullet}\rangle_2 - \langle K\overrightarrow{wp},\overrightarrow{q_\bullet}\rangle_2\\ &= \langle w(\mathcal{X}^K)[p],[q_\bullet]\rangle_H - \langle K\overrightarrow{wp},\overrightarrow{q_\bullet}\rangle_2\end{aligned}$$

and hence

$$\begin{aligned}|\langle \overrightarrow{r} - \overrightarrow{wp}, K^\flat\overrightarrow{x_jq}\rangle_2| &\le \|K-K^\flat\|_2\|\overrightarrow{p}\|_2\|\overrightarrow{q_\bullet}\|_2\sum_{i=0}^{|w|-1}c^i\\ &\le \|K-K^\flat\|_2\|\overrightarrow{p}\|_2\|\overrightarrow{q}\|_2\sum_{i=1}^{|w|}c^i\end{aligned} \tag{4.7}$$

by the induction hypothesis and (1). Therefore (3) follows by (4.5), (4.6) and (4.7). □

Lemma 4.4. *Let K be a psd Hankel extension of H of order $d+\delta$. Denote $c = \frac{\max_j\{\|K^{(j)}\|_2\}}{\sigma_{\min}(H)}$ and suppose $c \geq 1$. Let $w \in <\boldsymbol{x}>_\delta \setminus \{1\}$, $p, r \in \mathbb{R}<\boldsymbol{x}>_d$ be such that $[r] = w(\mathcal{X}^K)[p]$. Then*

$$\|K(\vec{r} - \overrightarrow{wp})\|_2 \leq (3 + (|w|-1)c^{|w|-1})\|K - K^\flat\|_2\|\vec{p}\|_2.$$

Proof. Since $K^\flat$ is flat we have

$$\max_{q\in\mathbb{R}<\boldsymbol{x}>_{d+\delta},\, \|\vec{q}\|_2=1} \langle K^\flat(\vec{r} - \overrightarrow{wp}),\, \vec{q}\rangle_2 = \max_{q\in\mathbb{R}<\boldsymbol{x}>_{d},\, \|\vec{q}\|_2=1} \langle K^\flat(\vec{r} - \overrightarrow{wp}),\, \vec{q}\rangle_2.$$

Hence there exists $q_0 \in \mathbb{R}<\boldsymbol{x}>_d$ such that $\|\vec{q_0}\|_2 = 1$ and

$$\langle K^\flat(\vec{r} - \overrightarrow{wp}), \vec{q_0}\rangle_2 = \|K^\flat(\vec{r} - \overrightarrow{wp})\|_2.$$

Therefore

$$\begin{aligned}
\|K(\vec{r} - \overrightarrow{wp})\|_2 &\leq \|(K - K^\flat)(\vec{r} - \overrightarrow{wp})\|_2 + \|K^\flat(\vec{r} - \overrightarrow{wp})\|_2 \\
&= \|(K - K^\flat)\overrightarrow{wp}\|_2 + \langle K^\flat(\vec{r} - \overrightarrow{wp}), \vec{q_0}\rangle_2 \\
&= \|(K - K^\flat)\overrightarrow{wp}\|_2 + \langle (K^\flat - K)(\vec{r} - \overrightarrow{wp}), \vec{q_0}\rangle_2 + \langle K(\vec{r} - \overrightarrow{wp}), \vec{q_0}\rangle_2 \\
&= \|(K - K^\flat)\overrightarrow{wp}\|_2 - \langle (K^\flat - K)\overrightarrow{wp}, \vec{q_0}\rangle_2 + \langle [r], [q_0]\rangle_H - \langle K\overrightarrow{wp}, \vec{q_0}\rangle_2 \\
&\leq 2\|K - K^\flat\|_2\|\vec{p}\|_2 + \|K - K^\flat\|_2\|\vec{p}\|_2 \sum_{i=0}^{|w|-1} c^i \\
&\leq \|K - K^\flat\|_2\|\vec{p}\|_2(3 + (|w|-1)c^{|w|-1})
\end{aligned}$$

by Lemma 4.3. ☐

Lemma 4.5. *Let K be a psd Hankel extension of H of order $d+\delta$. Denote $c = \frac{\max_j\{\|K^{(j)}\|_2\}}{\sigma_{\min}(H)}$ and suppose $c \geq 1$. If $w_1, w_2 \in <\boldsymbol{x}>_\delta$ and $p \in \mathbb{R}<\boldsymbol{x}>_d$, then*

$$|\langle w_1(\mathcal{X}^K)[p], w_2(\mathcal{X}^K)[p]\rangle_H - \langle K\overrightarrow{w_1p}, \overrightarrow{w_2p}\rangle_2| \leq M\|K - K^\flat\|_2\|\vec{p}\|_2^2,$$

where

$$M = 3(c^{|w_1|} + c^{|w_2|}) + (|w_1| + |w_2| - 2)c^{|w_1|+|w_2|-1}.$$

Proof. Let $r_i \in \mathbb{R}<\boldsymbol{x}>_d$ be such that $[r_i] = w_i(\mathcal{X}^K)[p]$. Then

$$\begin{aligned}
&|\langle K\vec{r_1}, \vec{r_2}\rangle_2 - \langle K\overrightarrow{w_1p}, \overrightarrow{w_2p}\rangle_2| \\
\leq &|\langle K\vec{r_1}, \vec{r_2} - \overrightarrow{w_2p}\rangle_2 + |\langle K(\vec{r_1} - \overrightarrow{w_1p}), \overrightarrow{w_2p}\rangle_2| \\
\leq &(3 + (|w_2|-1)c^{|w_2|-1})\|K - K^\flat\|_2\|\vec{r_1}\|_2\|\vec{p}\|_2 \\
&+ (3 + (|w_1|-1)c^{|w_1|-1})\|K - K^\flat\|_2\|\vec{r_2}\|_2\|\vec{p}\|_2 \\
\leq &((3 + (|w_2|-1)c^{|w_2|-1})c^{|w_1|} + (3 + (|w_1|-1)c^{|w_1|-1})c^{|w_2|})\|K - K^\flat\|_2\|\vec{p}\|_2^2.
\end{aligned}$$ ☐

Let $s \in \mathbb{R}<\boldsymbol{x}>_{2\delta}$. With respect to the decomposition

$$s = \ell + \sum_{1\leq|u|,|v|\leq\delta} \alpha_{u,v}uv,$$

where ℓ is an affine linear polynomial, let

$$M(c) = \sum_{1\leq|u|,|v|\leq\delta} |\alpha_{u,v}|(3(c^{|u|} + c^{|v|}) + (|u| + |v| - 2)c^{|u|+|v|-1}).$$

Theorem 4.6. *Let H be invertible and let K be a psd Hankel extension of H of order $d+\delta$ such that $K_s^\Uparrow$ is psd. Denote $c = \frac{\max_j\{\|K^{(j)}\|_2\}}{\sigma_{\min}(H)}$ and suppose $c \geq 1$. Then*

$$\lambda_{\min}\left(s(\mathcal{X}^K)\right) \geq \frac{\lambda_{\min}\left(K_s^\Uparrow\right)}{\|H\|_2} - \frac{M(c)}{\sigma_{\min}(H)}\|K - K^\flat\|_2. \tag{4.8}$$

Proof. Similarly as in the proof of Theorems 4.1 and 4.9 we see that Lemma 4.5 implies

$$\langle s(\mathcal{X}^K)[p],[p]\rangle_H \geq \langle K_s^{\Uparrow}\vec{p},\vec{p}\rangle_2 - M(c)\|K-K^\flat\|_2\langle\vec{p},\vec{p}\rangle_2. \qquad \square$$

Example 4.7 (TV screen)**.** Let us consider $d=1,\delta=2$ and $s=1-x_1^4-x_2^4$. Note that we have $M(c)=12c^2+4c^3$. Let

$$H=\begin{pmatrix}1.00 & 0.00 & 0.00\\ 0.00 & 0.75 & 0.20\\ 0.00 & 0.20 & 0.50\end{pmatrix}$$

be a pd Hankel matrix. We have generated 100 psd Hankel extensions K using the randomized Algorithm 1 from [KP16] (actually, we repeated Step 3 of this algorithm 100 times without checking the stopping criterion and each solution from Step 3 gave us one K). For each extension K we computed the left-hand side of (4.8), LHS $=\lambda_{\min}\left(s(\mathcal{X}^K)\right)$, the positive part of the right-hand side of (4.8), RHS1 $=\frac{\lambda_{\min}\left(K_s^{\Uparrow}\right)}{\|H\|_2}$, and the right-hand side of (4.8), RHS $=\frac{\lambda_{\min}\left(K_s^{\Uparrow}\right)}{\|H\|_2}-\frac{M(c)}{\sigma_{\min}(H)}\|K-K^\flat\|_2$. In Table 2 we report the median, the first and the third quartiles, average, $\Delta=\max-\min$ and the standard deviation for LHS, RHS1, RHS and for the difference LHS − RHS.

	LHS	RHS1	RHS	LHS − RHS
median	9.4616e-03	2.2707e-09	-2.1419e-06	2.9185e-03
Q1	1.9484e-09	2.3961e-10	-2.3735e+01	1.5954e-06
Q3	7.6957e-02	1.8933e-02	-5.4137e-07	2.3762e+01
average	3.8260e-02	2.1053e-02	-1.1973e+01	1.2011e+01
Δ	1.8648e-01	1.7902e-01	8.3478e+01	8.3390e+01
st. deviation	4.9192e-02	4.2062e-02	2.0090e+01	2.0107e+01

Table 2. Statistical parameters for (4.8): LHS $=\lambda_{\min}\left(s(\mathcal{X}^K)\right)$, RHS1 $=\frac{\lambda_{\min}(K_s^{\Uparrow})}{\|H\|_2}$, RHS $=\frac{\lambda_{\min}(K_s^{\Uparrow})}{\|H\|_2}-\frac{M(c)}{\sigma_{\min}(H)}\|K-K^\flat\|_2$ and the difference between the left hand side and the right hand side of (4.8): LHS − RHS. These numbers were computed on 100 random psd Hankel extensions K of order 3. We can see from the LHS column that the matrices $\mathcal{X}^K$ are almost always feasible, i.e., they satisfy $s(\mathcal{X}^K)\succeq 0$ in at least 75 % (actually in 84 %) of the cases. On the other hand, RHS column reveals that in at least 75 % (actually in 80 %) the right hand side of (4.8) is negative, but the column LHS − RHS additionally shows that inequality (4.8) is in our opinion still interesting, since in more than 50 % of the cases the difference LHS − RHS is smaller than $3\cdot 10^{-3}$.

We decided to plot the differences LHS − RHS on a logarithmic scale in Figure 6a. We provide two line plots: the red plot represents all 100 random extensions while the blue one corresponds only to those extensions that are very close to flatness, i.e., have exactly 3 eigenvalues larger 10^{-5}. There were 64 % of such instances.

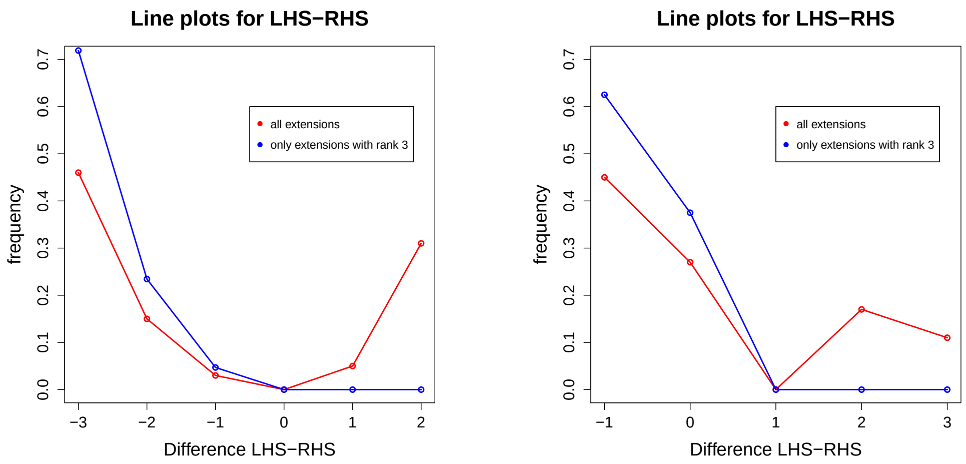
We can see that the inequality (4.8) is much tighter for the almost flat extensions. Indeed, some extensions have even 6 eigenvalues larger than 10^{-5} and therefore deviate a lot from flatness, leading to a large $\|K-K^\flat\|_2$ and therefore large difference LHS − RHS.

Example 4.8 (Bent TV screen)**.** Let $s=1-x_1^2-x_2^4$, $d=1$ and $\delta=2$. We have $M(c)=6c+6c^2+2c^3$. As in Example 4.7 we computed 100 psd Hankel extensions of the following pd Hankel matrix

$$H=\begin{pmatrix}1.00 & 0.00 & 0.00\\ 0.00 & 0.80 & 0.20\\ 0.00 & 0.20 & 0.30\end{pmatrix}.$$

Results, similar to those from Example 4.7 are reported in Table 3.

4.3. **Simple extensions.** Let H be a psd Hankel matrix of order d and K its psd Hankel extension of order $d+\delta$. As already stated, $K^\flat$ does not satisfy the Hankel property (2.1) in general. However, if $\delta=1$, then $K^\flat$ is Hankel since K is Hankel and the bottom-right block of



(a) Line plots for LHS − RHS for $s = 1 - x^4 - y^4$, in logarithmic scale.

(b) Line plots for LHS − RHS for $s = 1 - x^2 - y^4$, in logarithmic scale.

Figure 6. Line plots for differences LHS − RHS related to the inequality (4.8), for Examples 4.7 and 4.8, respectively, in logarithmic scale. The red curve corresponds to all 100 extensions while the blue curve represents these differences only for the extensions K which have rank equal to rank H, i.e., have 3 eigenvalues larger than 10^{-5}. We consider these extensions as almost flat. Both plots depicts that for the almost flat extensions the inequality (4.8) is very tight while the extensions which deviate significantly from the flatness (some of them have 6 eigenvalues larger than 10^{-5}) yield also large differences LHS − RHS.

	LHS	RHS1	RHS	LHS − RHS
median	9.5393e-02	7.1877e-10	-4.3748e-06	1.1656e-01
Q1	5.2930e-02	8.3273e-11	-2.7178e+01	5.9633e-02
Q3	1.5633e-01	1.2530e-03	-1.4011e-06	2.7206e+01
average	1.0665e-01	1.0567e-02	-2.4009e+01	2.4115e+01
Δ	2.4746e-01	1.0026e-01	1.9655e+02	1.9666e+02
st. deviation	6.2750e-02	2.2835e-02	4.6298e+01	4.6317e+01

Table 3. Statistical parameters for (4.8): LHS, RHS1, RHS and LHS − RHS, computed on 100 random psd Hankel extensions K of order 3. LHS column shows that the resulting tuples $\mathcal{X}^K$ are positive semidefinite in at least 75 % of cases (actually in all 100 cases), while RHS column is in more than 75 % of the cases negative. The inequality LHS $\geq$ RHS is rather loose, but in half of the extensions this difference is smaller than 0.1166. These small values are achieved only by the extensions that are very close to flatness, i.e., have only 3 eigenvalues larger than 10^{-5}.

$K^\flat$ corresponds to pairs of words of length $d+1$. Hence in this case $K^\flat$ is a flat Hankel extension of H of order $d+1$. Moreover, we have $\mathcal{X}_j^K = \mathcal{X}_j^{K^\flat}$. Indeed, for every $p, q \in \mathbb{R}\langle \boldsymbol{x} \rangle_d$ we have

$$\langle \mathcal{X}_j^K [p], [q] \rangle_H = \langle K \overrightarrow{x_j p}, \overrightarrow{q} \rangle_2 = \langle K^\flat \overrightarrow{x_j p}, \overrightarrow{q} \rangle_2 = \langle \mathcal{X}_j^{K^\flat} [p], [q] \rangle_H .$$

4.3.1. *Quadratic polynomial constraints.* Let $s \in \mathbb{R}\langle \boldsymbol{x} \rangle_2$ and write

$$s = \ell + \frac{1}{2} \sum_{j \leq j'} \alpha_{jj'} (x_j x_{j'} + x_{j'} x_j), \qquad \tilde{c}_1 = \sum_{j \leq j'} |\alpha_{jj'}|,$$

where ℓ is an affine linear polynomial.

Theorem 4.9. *Let K be a psd Hankel extension of H of order $d+1$ such that $K_s^{\Uparrow}$ is psd. Then*

$$\lambda_{\min} \left(s(\mathcal{X}^K) \right) \geq \frac{\lambda_{\min} \left(K_s^{\Uparrow} \right)}{\|H\|_2} - \|K - K^\flat\|_2 \frac{\tilde{c}_1}{\sigma_{\min}(H)}.$$

Proof. Let $p \in (\ker H)^\perp$. Then (4.3) and the flatness of $K^\flat$ imply

$$\begin{aligned}
\langle s(\mathcal{X}^K)[p],[p]\rangle_H &= \langle \ell(\mathcal{X}^K)[p],[p]\rangle_H + \sum_{j\le j'} \alpha_{jj'} \langle \mathcal{X}_j^K[p], \mathcal{X}_{j'}^K[p]\rangle_H \\
&= \langle \ell(\mathcal{X}^K)[p],[p]\rangle_H + \sum_{j\le j'} \alpha_{jj'} \langle \mathcal{X}_j^{K^\flat}[p], \mathcal{X}_{j'}^{K^\flat}[p]\rangle_H \\
&= \langle K\overrightarrow{\ell p}, \vec{p}\rangle_2 + \sum_{j\le j'} \alpha_{jj'} \langle K^\flat \overrightarrow{x_j p}, \overrightarrow{x_{j'} p}\rangle_2 \\
&= \langle K_s^{\Uparrow} \vec{p}, \vec{p}\rangle_2 + \sum_{j\le j'} \alpha_{jj'} \langle (K^\flat - K) \overrightarrow{x_j p}, \overrightarrow{x_{j'} p}\rangle_2 \\
&\ge \lambda_{\min}\left(K_s^{\Uparrow}\right) \|\vec{p}\|_2^2 - \sum_{j\le j'} |\alpha_{jj'}| \|K^\flat - K\|_2 \|\vec{p}\|_2^2 \\
&\ge \left(\frac{\lambda_{\min}\left(K_s^{\Uparrow}\right)}{\|H\|_2} - \frac{\tilde{c}_1}{\sigma_{\min}(H)} \|K - K^\flat\|_2 \right) \langle [p],[p]\rangle_H. \qquad \square
\end{aligned}$$

Proposition 4.10 (cf. [HKM12, Proposition 2.5])**.** *Let $s = \ell_0 - \sum_{k\ge1} \ell_k^\intercal \ell_k$, where ℓ_k are affine linear polynomials. If K is a psd Hankel extension of H of order $d+1$ such that $K_s^{\Uparrow}$ is psd, then $(K^\flat)_s^{\Uparrow}$ is psd.*

Proof. Since $K - K^\flat$ is psd, we have

$$\begin{aligned}
\langle (K^\flat)_s^{\Uparrow} \vec{p}, \vec{p}\rangle_2 &= \langle K^\flat \overrightarrow{\ell_0 p}, \vec{p}\rangle_2 - \sum_{k\ge1} \langle K^\flat \overrightarrow{\ell_k p}, \overrightarrow{\ell_k p}\rangle_2 \\
&= \langle K_s^{\Uparrow} \vec{p}, \vec{p}\rangle_2 + \sum_{k\ge1} \langle (K - K^\flat) \overrightarrow{\ell_k p}, \overrightarrow{\ell_k p}\rangle_2 \\
&\ge 0 + 0. \qquad \square
\end{aligned}$$

Remark 4.11. By [HM04', Theorem 3.1], polynomials s as in Proposition 4.10 are precisely concave polynomials. Two special cases of Proposition 4.10 (nc ball and nc polydisk) have already been established in [KP16, Theorem 3.1].

4.4. **Noncommutative polynomial optimization.** Let $\mathbb{S}_n$ denote the space of symmetric $n \times n$ matrices. Let $s_1, \dots, s_h \in \mathbb{R}\langle \boldsymbol{x} \rangle_{2\delta}$ be symmetric nc polynomials that generate an Archimedean quadratic module [BKP16] and consider the corresponding bounded free semi-algebraic set

$$\mathcal{S} = \bigcup_{n\in\mathbb{N}} \left\{ X \in \mathbb{S}_n^g \colon s_1(X) \succeq 0, \dots, s_h(X) \succeq 0 \right\}.$$

Let $f \in \mathbb{R}\langle \boldsymbol{x} \rangle$ be a symmetric polynomial and suppose we are interested in $X_0 \in \mathcal{S}$ such that

$$f(X_0) = f^* := \inf_{X\in\mathcal{S}} \lambda_{\min}(f(X)). \tag{4.9}$$

For every $d \ge \deg f$ let $K(d)$ be a psd Hankel matrix of order $d+\delta$ that is an optimal solution of the SDP relaxation (1.3) of (4.9) of order $d+\delta$; see e.g. [PNA10, Subsection 3.2] or [BKP16, Section 4.3] for terminology and further details. Write

$$f^{(d)} = \langle K(d) \vec{f}, \vec{1}\rangle_2.$$

Then there is $d_0 \in \mathbb{N}$, which can be deduced from s_i and f, such that the sequence $\{f^{(d)}\}_{d\ge d_0}$ is monotonically increasing [PNA10, Lemma 4]. Moreover, $\lim_d f^{(d)} = f^*$ [PNA10, Theorem 1].

Since $K(d)$ is the solution of the SDP relaxation of (4.9), $K(d)_{s_i}^{\Uparrow}$ is psd for every $1 \le i \le h$. Now view $K(d)$ as the Hankel extension of $H(d)$, which is its top-left Hankel submatrix of order d. Assume that $d \ge d_0$ is such that $K(d)_{s_i}^{\Uparrow}$ is pd for $1 \le i \le h$ and $K(d)$ is close to being flat in the sense that there is $\varepsilon \ll 1$ such that there exists a flat psd Hankel extension L of $H(d)$ of order $d+\delta$ with $\|K(d) - L\|_2 \le \varepsilon$, or that $H(d)$ is invertible and $\|K(d) - K(d)^\flat\|_2 \le \varepsilon$. If ε

is small enough, then $s_i(\mathcal{X}^{K(d)})$ is psd by Theorem 4.9 or Theorem 4.6, and hence $\mathcal{X}^{K(d)} \in \mathcal{S}$. Therefore

$$f^{(d)} = \langle K(d)\vec{f}, \vec{1}\rangle_2 = \left\langle f(\mathcal{X}^{K(d)})[1], [1]\right\rangle_{H(d)} \geq f^*$$

by the definition of f^*. On the other hand we have $f^{(d)} \leq f^*$, so $f^* = f^{(d)}$ and we conclude that f^* is attained at $\mathcal{X}^{K(d)}$ with eigenvector [1].

5. Commutative and tracial modifications of the theory

In this section we switch gears and explain how our results on Hankel matrices and polynomial optimization in the free noncommutative context pertain to the tracial noncommutative setting and the classical, commutative one.

5.1. **Tracial setting.** A Hankel matrix H is **tracial** if

$$H_{u_1,v_1} = H_{u_2,v_2} \quad \forall u_1^\intercal v_1 \overset{\text{cyc}}{\sim} u_2^\intercal v_2,$$

where $\overset{\text{cyc}}{\sim}$ denotes cyclic equivalence relation [KS08] on $<\boldsymbol{x}>$. Thus $w_1 \overset{\text{cyc}}{\sim} w_2$ if and only if $w_1 = uv$ and $w_2 = vu$ for some $u, v \in <\boldsymbol{x}>$. Equivalently, the word w_1 is a cyclic permutation of w_2.

Let H be a psd tracial Hankel matrix of order d and let K be a psd flat tracial Hankel extension of H of order $d+\delta$. Let $\mathcal{A} \subseteq \operatorname{End}_{\mathbb{R}} \mathcal{H}$ be the $\mathbb{R}$-subalgebra generated by $\mathcal{X}_j^K$. Since $\mathcal{X}_j^K$ are symmetric, $\mathcal{A}$ inherits the involution from $\operatorname{End}_{\mathbb{R}} \mathcal{H}$. Let $\delta' \in \mathbb{N}$ be such that $\mathcal{A}$ is spanned by $w(\mathcal{X}^K)$ for $w \in <\boldsymbol{x}>_{d+\delta'}$. If $\delta' > \delta$, then by [BK12, Theorem 3.18] there exists a unique psd flat tracial Hankel extension K' of K of order $d+\delta'$. Since $\mathcal{X}^{K'} = \mathcal{X}^K$ by flatness, we can without loss of generality assume $\delta' = \delta$. Because $\mathcal{H} = \mathcal{K}$, we can define a $\mathbb{R}$-linear functional

$$\varphi_K : \mathcal{A} \to \mathbb{R}, \qquad f(\mathcal{X}^K) \mapsto \langle f(\mathcal{X}^K)[1], [1]\rangle_H.$$

By flatness and the tracial property we have

$$\begin{aligned}
\langle (f_1 f_2)(\mathcal{X}^K)[1], [1]\rangle_H &= \langle f_2(\mathcal{X}^K)[1], f_1^\intercal(\mathcal{X}^K)[1]\rangle_H \\
&= \langle [f_2], [f_1^\intercal]\rangle_H \\
&= \langle [f_1], [f_2^\intercal]\rangle_H \\
&= \langle f_1(\mathcal{X}^K)[1], f_2^\intercal(\mathcal{X}^K)[1]\rangle_H \\
&= \langle (f_2 f_1)(\mathcal{X}^K)[1], [1]\rangle_H
\end{aligned}$$

for all $f_1, f_2 \in \mathbb{R}<\boldsymbol{x}>_{d+\delta}$. Consequently φ_K is a **tracial state** on $\mathcal{A}$ [BK12, Definition 3.9]. That is, φ_K is $\mathbb{R}$-linear functional with $\varphi_K(1) = 1$ and

$$\varphi_K(a^\intercal) = \varphi_K(a), \qquad \varphi_K(ab) = \varphi_K(ba)$$

for all $a, b \in \mathcal{A}$. If $\mathcal{A} = \operatorname{End}_{\mathbb{R}} \mathcal{H}$, then $\varphi_K = \operatorname{Tr}$ by [BK12, Lemma 3.11], where Tr is the normalized trace on $\operatorname{End}_{\mathbb{R}} \mathcal{H} \cong \operatorname{M}_{\dim \mathcal{H}}(\mathbb{R})$. In general, φ_K is a convex combination of trace evaluations [BK12, Theorem 3.14].

Proposition 5.1. *Let L be a psd flat tracial Hankel extension of H of order $d+\delta$ such that $\mathcal{X}_j^L$ generate $\operatorname{End}_{\mathbb{R}} \mathcal{H}$. If K is a psd Hankel extension of H of order $d+\delta$ such that $\mathcal{X}_j^K$ generate $\operatorname{End}_{\mathbb{R}} \mathcal{H}$, then*

$$\left|\langle K\vec{f}, \vec{1}\rangle_2 - \operatorname{Tr} f(\mathcal{X}^K)\right| \leq 2\frac{\operatorname{err}_f(c)}{\sigma_{\min}(H)}\|K - L\|_2$$

for all $f \in \mathbb{R}<\boldsymbol{x}>_{d+\delta}$.

Proof. Since L is flat we have $\langle f(\mathcal{X}^L)[1], [1]\rangle_H = \operatorname{Tr} f(\mathcal{X}^L)$ for all $f \in \mathbb{R}<\boldsymbol{x}>$. For every $n \times n$ matrix A we have $|\operatorname{Tr} A| \leq \|A\|_2$. Since

$$\left|\langle K\vec{f}, \vec{1}\rangle_2 - \operatorname{Tr} f(\mathcal{X}^K)\right| \leq \left|\langle f(\mathcal{X}^K)[1], [1]\rangle_H - \langle f(\mathcal{X}^L)[1], [1]\rangle_H\right| + \left|\operatorname{Tr} f(\mathcal{X}^K) - \operatorname{Tr} f(\mathcal{X}^L)\right|$$

for all $f \in \mathbb{R}<\boldsymbol{x}>_{d+\delta}$, the statement follows by Corollary 3.5. ∎

Remark 5.2. It is not hard to see that the set of all psd Hankel extensions of H, which yield operators that generate $\operatorname{End}_{\mathbb{R}} \mathcal{H}$, is Zariski open in $\operatorname{HExt}_{H,\delta}$ viewed as an affine space, therefore open and dense with respect to the Euclidean topology. Consequently, if K is a psd Hankel extension of H such that $\mathcal{X}_j^K$ generate $\operatorname{End}_{\mathbb{R}} \mathcal{H}$ and $\|K-L\|_2$ is small enough for some psd Hankel extension L of H, then $\mathcal{X}_j^L$ also generate $\operatorname{End}_{\mathbb{R}} \mathcal{H}$.

5.1.1. *Trace polynomial optimization.* Let us adopt the discussion of Subsection 4.4 to trace optimization [BKP16, Chapter 5]. For symmetric $s_1,\dots,s_h \in \mathbb{R}\langle\boldsymbol{x}\rangle_{2\delta}$ generating an Archimedean quadratic module let

$$\mathcal{S} = \bigcup_{n\in\mathbb{N}} \{X \in \mathbb{S}_n^g \colon s_1(X) \succeq 0,\dots,s_h(X) \succeq 0\}.$$

Given a symmetric polynomial $f \in \mathbb{R}\langle\boldsymbol{x}\rangle$ we are interested in $X_0 \in \mathcal{S}$ such that

$$\operatorname{Tr} f(X_0) = \operatorname{Tr}_{\min} f := \inf_{X\in\mathcal{S}} \operatorname{Tr} f(X). \tag{5.1}$$

For $d \geq \deg f$ let $K(d)$ be a psd tracial Hankel matrix of order $d+\delta$ that is an optimal solution of the SDP relaxation [KP16] of (5.1) of order $d+\delta$, and write

$$f^{(d)} = \langle K(d)\vec{f}, \vec{1}\rangle_2.$$

As in Subsection 4.4 we have $f^{(d)} \leq \operatorname{Tr}_{\min} f$ for $d \geq d_0$.

Assume that for some $d \geq d_0$ we have

1. $K(d)_{s_i}^{\Uparrow}$ is pd for all i;
2. $\mathcal{X}_j^{K(d)}$ generate $\operatorname{End}_{\mathbb{R}} \mathcal{H}(d)$;
3. $K(d)$ is close to being flat, i.e., there exists a flat psd tracial Hankel extension L of $H(d)$ of order $d+\delta$ with $\|K(d)-L\|_2 \ll 1$.

If $\|K(d)-L\|_2$ is small enough, then $s_i(\mathcal{X}^{K(d)})$ is psd by Theorem 4.6 and hence $\mathcal{X}^{K(d)} \in \mathcal{S}$. Thus

$$f^{(d)} \leq \operatorname{Tr}_{\min} f \leq \operatorname{Tr} f(\mathcal{X}^{K(d)}).$$

Moreover, the assumptions of Proposition 5.1 are satisfied by Remark 5.2 if $\|K(d)-L\|_2$ is small enough, so $\operatorname{Tr} f(\mathcal{X}^{K(d)}) - f^{(d)}$ grows (at most) as $\|K(d)-L\|_2$.

Note that this conclusion is weaker than the one of Subsection 4.4. While the Hankel property of K ensures $\langle K\vec{f}, \vec{1}\rangle_2 = \langle f(\mathcal{X}^K)[1],[1]\rangle_H$ in the freely noncommutative context, the tracial Hankel property of K and assuming $\mathcal{X}_j^K$ generate $\operatorname{End}_{\mathbb{R}} \mathcal{H}(d)$ do not suffice to conclude $\langle K\vec{f}, \vec{1}\rangle_2 = \operatorname{Tr} f(\mathcal{X}^K)$ in general.

5.2. **Commutative setting.** Now let $\boldsymbol{y} = (y_1,\dots,y_g)$ be a tuple of independent commuting variables and $[\boldsymbol{y}]$ the free commutative monoid generated by them. A **commutative Hankel matrix of order** d [Las01] is a symmetric matrix H indexed by elements in $[\boldsymbol{y}]$ of degree at most d that satisfies

$$H_{u_1,v_1} = H_{u_2,v_2} \quad \forall u_1v_1 = u_2v_2.$$

If H is psd, we obtain a scalar product on $\mathbb{R}[\boldsymbol{y}]_d/\ker H$. Analogously as in the noncommutative case let K be a psd commutative Hankel extension of H of order $d+\delta$ and consider the finite-dimensional Hilbert spaces

$$\mathcal{H} = \mathbb{R}[\boldsymbol{y}]_d/\ker H \subseteq \mathbb{R}[\boldsymbol{y}]_{d+\delta}/\ker K = \mathcal{K}.$$

We define operators

$$\mathcal{Y}_j^K \colon \mathcal{H} \to \mathcal{H}, \qquad [p] \mapsto [\pi^K(y_jp)],$$

where $\pi^K : \mathcal{K} \to \mathcal{H}$ is the orthogonal projection. Again, these are well-defined and symmetric, can be constructed without any flatness assumption and can be easily determined from the singular value decomposition of H and the submatrices of K analogously as in Subsection 2.2.1 (see Subsection 5.2.4 below for a worked example). However, the operators $\mathcal{Y}_j^K$ do not necessarily commute.

If K is a flat extension of H, then $\mathcal{H} = \mathcal{K}$ and the operators $\mathcal{Y}_j^K$ do commute: indeed, for every $p, q \in \mathbb{R}[\boldsymbol{y}]_d$ we have

$$\begin{aligned}\langle \mathcal{Y}_i^K \mathcal{Y}_j^K [p], [q] \rangle_H = \langle [y_j p], [y_i q] \rangle_H = \langle K \overrightarrow{y_i q}, \overrightarrow{y_j p} \rangle_2 \\ = \langle K \overrightarrow{y_j q}, \overrightarrow{y_i p} \rangle_2 = \langle [y_i p], [y_j q] \rangle_H = \langle \mathcal{Y}_j^K \mathcal{Y}_i^K [p], [q] \rangle_H.\end{aligned}$$

Since the self-adjoint operators $\mathcal{Y}_j^K$ on a finite-dimensional Hilbert space commute, they can be simultaneously diagonalized and thus lead to tuples of points in $\mathbb{R}^g$. In the non-flat case, where these operators do not commute, finding these points is trickier. We present a remedy in the near flat case in Subsection 5.2.3.

5.2.1. *Noncommutative lift.* Let $\pi : \mathbb{R}\langle \boldsymbol{x} \rangle \to \mathbb{R}[\boldsymbol{y}]$ be the canonical homomorphism defined by $\pi(x_j) = y_j$, the so-called commutative collapse. If H is a commutative Hankel matrix of order d, let H^{nc} be the (noncommutative) Hankel matrix of order d defined by

$$H^{\mathrm{nc}}_{u,v} = H_{\pi(u),\pi(v)}.$$

Observe that

$$\langle H^{\mathrm{nc}} \vec{p}, \vec{q} \rangle_2 = \langle H \overrightarrow{\pi(p)}, \overrightarrow{\pi(q)} \rangle_2 \tag{5.2}$$

for all $p, q \in \mathbb{R}\langle \boldsymbol{x} \rangle_d$, which implies the following.

Lemma 5.3. *If H is a commutative Hankel matrix, then the eigenvalues of H and H^{nc} coincide.*

In particular, H is psd if and only if H^{nc} is psd, and in this case π induces an isomorphism of Hilbert spaces

$$\tilde{\pi} : \mathbb{R}\langle \boldsymbol{x} \rangle_d / \ker H^{\mathrm{nc}} \to \mathbb{R}[\boldsymbol{y}]_d / \ker H. \tag{5.3}$$

If K is a commutative Hankel extension of H of order $d + \delta$, then K^{nc} is a Hankel extension of H^{nc} of order $d + \delta$. Moreover, using (5.2) it is easy to derive the following.

Proposition 5.4. *If the setup is as above, then $\tilde{\pi} \circ \mathcal{X}_j^{K^{\mathrm{nc}}} = \mathcal{Y}_j^K \circ \tilde{\pi}$ holds for $j = 1, \dots, g$.*

5.2.2. *Approximately flat GNS.* We can quantify how far the self-adjoint operators $\mathcal{Y}_j^K$ are from commuting by using a flat extension L of H.

Proposition 5.5. *Let K be a psd commutative Hankel extension of H. If L is a psd flat commutative Hankel extension of H, then*

$$\|\mathcal{Y}_i^K \mathcal{Y}_j^K - \mathcal{Y}_j^K \mathcal{Y}_i^K\|_H \le \frac{2(\|K\|_2 + \|L\|_2)}{\sigma_{\min}(H)^2} \|K - L\|_2.$$

Proof. Since L is flat, we have $\mathcal{Y}_i^L \mathcal{Y}_j^L = \mathcal{Y}_j^L \mathcal{Y}_i^L$. On the other hand, the bounds of Proposition 3.1 and Theorem 3.2 also hold for commutative Hankel extensions by Proposition 5.4. The rest now follows by

$$\begin{aligned}\|\mathcal{Y}_i^K \mathcal{Y}_j^K - \mathcal{Y}_j^K \mathcal{Y}_i^K\|_H &= \big\|(\mathcal{Y}_i^L + (\mathcal{Y}_i^K - \mathcal{Y}_i^L))\mathcal{Y}_j^K - \mathcal{Y}_j^K(\mathcal{Y}_i^L + (\mathcal{Y}_i^K - \mathcal{Y}_i^L))\big\|_H \\ &\le \|\mathcal{Y}_i^L \mathcal{Y}_j^K - \mathcal{Y}_j^K \mathcal{Y}_i^L\|_H + 2\|\mathcal{Y}_i^K - \mathcal{Y}_i^L\|_H \|\mathcal{Y}_j^K\|_H \\ &= \big\|\mathcal{Y}_i^L(\mathcal{Y}_j^L + (\mathcal{Y}_j^K - \mathcal{Y}_j^L)) - (\mathcal{Y}_j^L + (\mathcal{Y}_j^K - \mathcal{Y}_j^L))\mathcal{Y}_i^L\big\|_H \\ &\quad + 2\|\mathcal{Y}_i^K - \mathcal{Y}_i^L\|_H \|\mathcal{Y}_j^K\|_H \\ &\le 2\|\mathcal{Y}_j^K - \mathcal{Y}_j^L\|_H \|\mathcal{Y}_i^L\|_H + 2\|\mathcal{Y}_i^K - \mathcal{Y}_i^L\|_H \|\mathcal{Y}_j^K\|_H.\end{aligned}$$

□

5.2.3. *Near commuting matrices.* If $X_1, \dots, X_g$ are hermitian matrices such that the commutators $X_i X_j - X_j X_i$ are small (with respect to some norm), then by [LT70, PS79, Lin97, Gle] there exist commuting hermitian matrices $Y_1, \dots, Y_g$ such that $X_j - Y_j$ are small. Quantitative versions of these results are given in [Has09, Theorem 1] ($g = 2$) and [FK, Theorem 3] ($g \ge 3$). That is, their statements are of the following form. If X_j are hermitian and $\|X_i X_j - X_j X_i\| \le \delta$, then there exist commuting complex hermitian Y_j such that $\|X_j - Y_j\| < \varepsilon(\delta)$, where the function $\varepsilon(t)$ also depends on the norm $\|\cdot\|$ we are considering, and the dimension of the space if

$g \geq 3$. However, by carefully reading the proofs of [Has09, Theorem 1] and [FK, Theorem 3] one observes that their real versions also hold. Namely, if X_j are symmetric and $\|X_iX_j - X_jX_i\| \leq \delta$, then there exists commuting symmetric Y_j such that $\|X_j - Y_j\| < \varepsilon(\delta)$. Indeed, in the case of [FK, Theorem 3], their proof applies to real matrices without a change. On the other hand, the proof of [Has09, Theorem 1] proceeds in several steps involving constructions by integration and Lin's theorem [Lin97]. However, it is easy to check that the outputs of integrals appearing in the proof are real if the input information (matrices X_j) is real, and a real version of Lin's theorem holds by [LS16, Theorem 1].

If an extension K of a commutative Hankel matrix H is close to being flat, the norms of commutators of operators $\mathcal{Y}_j^K$ are small by Proposition 5.5. Therefore the previously mentioned results apply and there exist commuting symmetric matrices $Y_1, \dots, Y_g$ that are close to $\mathcal{Y}_1^K, \dots, \mathcal{Y}_g^K$. If $g = 2$, then there is a dimension-independent bound on the operator norm of $\mathcal{Y}_j^K - Y_j$ [Lin97] and we obtain the following corollary.

Corollary 5.6. *For $g = 2$ let K be a psd commutative Hankel extension of H and assume L is a psd flat commutative Hankel extension of H. If*

$$\Delta = \frac{2(\|K\|_2 + \|L\|_2)}{\|K\|_2^2}\|K - L\|_2,$$

then there exist commuting symmetric matrices Y_1, Y_2 on $\mathcal{H}$ such that

$$\|\mathcal{Y}_j^K - Y_j\|_H \leq \frac{\|K\|}{\sigma_{\min}(H)}\Delta^{1/5}\varepsilon(\Delta),$$

where the function $\varepsilon(t)$ grows slower than any power of t and is independent of $\dim \mathcal{H}$.

Proof. Using estimates $\|\mathcal{Y}_j^K\|_H \leq \frac{\|K\|_2}{\sigma_{\min}(H)}$ the statement follows by Proposition 5.5 and the real version of [Has09, Theorem 1]. □

If $g \geq 3$, the bounds cannot be chosen independently of dimension by [Voi83, Dav85]. However, by considering the normalized Hilbert-Schmidt norm we obtain the following.

Corollary 5.7. *For $g \geq 3$ let K be a psd commutative Hankel extension of H and assume L is a psd flat commutative Hankel extension of H. If*

$$\|K - L\|_2 \leq \frac{1}{16^{2\cdot 4^{g-2}}}\frac{\|K\|_2^2}{2(\|K\|_2 + \|L\|_2)}$$

and

$$\Delta = \frac{2(\|K\|_2 + \|L\|_2)}{\|K\|_2^2}\|K - L\|_2,$$

then there exist commuting symmetric matrices $Y_1, \dots, Y_g$ on $\mathcal{H}$ such that

$$\|\mathcal{Y}_j^K - Y_j\|_H \leq 5\sqrt{\dim \mathcal{H}}\frac{\|K\|_2}{\sigma_{\min}(H)}\Delta^{1/4^{g-1}}.$$

Proof. We apply Proposition 5.5 and the real version of [FK, Theorem 3] together with the estimates $\|\mathcal{Y}_j^K\|_H \leq \frac{\|K\|_2}{\sigma_{\min}(H)}$. The factor $\sqrt{\dim \mathcal{H}}$ appears because we replaced the Hilbert-Schmidt norm with the operator norm $\|\cdot\|_H$. □

5.2.4. *Solving commutative moment problems with nc techniques.* Let $g = 2$ and consider the following commutative Hankel matrix of order $d = 2$ that is a modification of [CF02, Example 1.13]:

$$H = \begin{pmatrix} 1 & 1 & 1 & 2 & 0 & 3 \\ 1 & 2 & 0 & 4 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 & 9 \\ 2 & 4 & 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 9 & 0 & 0 & 27+\varepsilon \end{pmatrix}.$$

For $\varepsilon > 0$, the matrix H is psd of rank 5 and does not admit a flat extension [CF02, Theorem 4.1]. We shall explain how the non-flat GNS construction together with the theory of near commuting matrices can be used to approximate the corresponding Riesz functional φ_H (and thus H itself) with a convex combination of Dirac measures. That is, we approximately solve the truncated moment problem (see e.g. [CF96]) associated to H.

Let

$$K = \begin{pmatrix} 1 & 1 & 1 & 2 & 0 & 3 & 4 & 0 & 0 & 9 \\ 1 & 2 & 0 & 4 & 0 & 0 & 9 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 & 9 & 0 & 0 & 0 & 27+\varepsilon \\ 2 & 4 & 0 & 9 & 0 & 0 & 18 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 9 & 0 & 0 & 27+\varepsilon & 0 & 0 & 0 & 81 \\ 4 & 9 & 0 & 18 & 0 & 0 & 42+\varepsilon & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 9 & 0 & 27+\varepsilon & 0 & 0 & 81 & 0 & 0 & 0 & 243+28\varepsilon+\varepsilon^2 \end{pmatrix}$$

be a psd Hankel extension of H. For $\varepsilon = 0.001$ we perform a GNS construction. Letting $USU^\intercal = H$ be a singular value decomposition, U is an isometry (i.e., $U^\intercal U = I$) and S is a 5×5 pd diagonal matrix. Then $\mathcal{Y}_j^K = \sqrt{S}^{-1} U^\intercal K^{(j)} U \sqrt{S}^{-1}$, where $K^{(j)} = (K_{u,y_j v})_{u,v \in [\boldsymbol{y}]_2}$, i.e.,

$$\mathcal{Y}_1^K = \begin{pmatrix} 0.00089 & -0.04199 & -0.00684 & 0.00878 & -0.00002 \\ -0.04199 & 2.06201 & 0.16149 & -0.55935 & 0.00252 \\ -0.00684 & 0.16149 & -0.35568 & 1.41864 & -0.00765 \\ 0.00878 & -0.55935 & 1.41864 & -1.70723 & 0.00503 \\ -0.00002 & 0.00252 & -0.00765 & 0.00503 & 0.00000 \end{pmatrix},$$

$$\mathcal{Y}_2^K = \begin{pmatrix} 2.99841 & 0.06282 & 0.02143 & 0.01261 & -0.03086 \\ 0.06282 & 0.00132 & 0.00045 & 0.00027 & -0.00116 \\ 0.02143 & 0.00045 & 0.00017 & 0.00011 & -0.01168 \\ 0.01261 & 0.00027 & 0.00011 & 0.00008 & -0.01735 \\ -0.03086 & -0.00116 & -0.01168 & -0.01735 & -5.99997 \end{pmatrix}.$$

The norm of the commutator $\mathcal{Y}_1^K \mathcal{Y}_2^K - \mathcal{Y}_2^K \mathcal{Y}_1^K$ is 0.0275054, so the two matrices are almost commuting. We diagonalize a random linear combination $\rho_1 \mathcal{Y}_1^K + \rho_2 \mathcal{Y}_2^K$ of these two matrices; $O^\intercal(\rho_1 \mathcal{Y}_1^K + \rho_2 \mathcal{Y}_2^K)O$ is diagonal for an orthogonal matrix O. Then we discard all off-diagonal entries of $O^\intercal \mathcal{Y}_j^K O$ and arrive at

$$Y_1 = \operatorname{diag}(-2.66908,\ 2.14507,\ 0.00004,\ 0.00000,\ 0.52396),$$
$$Y_2 = \operatorname{diag}(0.00001,\ -0.00009,\ -5.65386,\ 3.00003,\ 0.00008).$$

Letting v be the first column of $O^\intercal \sqrt{S} U^\intercal$,

$$v = (-0.07225,\ 0.64160,\ 0.00379,\ 0.57734,\ -0.49979)^\intercal,$$

we get

$$\varphi_H(f) \approx \sum_i v_i^2 f\big((Y_1)_{ii}, (Y_2)_{ii}\big).$$

The Hankel matrix associated with the right-hand side functional is

$$\begin{pmatrix} 1.00000 & 0.99996 & 0.99989 & 1.99990 & -0.00007 & 3.00047 \\ 0.99996 & 1.99990 & -0.00007 & 3.99973 & -0.00017 & 0.00000 \\ 0.99989 & -0.00007 & 3.00047 & -0.00017 & 0.00000 & 8.99754 \\ 1.99990 & 3.99973 & -0.00017 & 8.99931 & -0.00038 & 0.00000 \\ -0.00007 & -0.00017 & 0.00000 & -0.00038 & 0.00000 & 0.00000 \\ 3.00047 & 0.00000 & 8.99754 & 0.00000 & 0.00000 & 27.0154 \end{pmatrix}$$

and is close to H (the operator norm of the difference is 0.0148).

5.2.5. *Commutative polynomial optimization.* Lastly we discuss some consequences of the previous approximation results for commutative polynomial optimization. For the rest of the paper let H be a psd commutative Hankel matrix of order d. Let K a psd commutative Hankel extension of H of order $d+\delta$ and $s\in\mathbb{R}[\boldsymbol{y}]_{2\delta}$. Analogously as in Section 4 we define the localizing matrix of K with respect to s, denoted $K_s^{\Uparrow}$. Let $\hat{s}\in\mathbb{R}\langle\boldsymbol{x}\rangle_{2\delta}$ be an arbitrary symmetric nc polynomial such that $\pi(\hat{s})=s$. By Proposition 5.4 we have

$$\tilde{\pi}\circ\hat{s}(\mathcal{X}^{K^{\rm nc}})=\hat{s}(\mathcal{Y}^{K})\circ\tilde{\pi}. \tag{5.4}$$

It is also easy to verify that $K_s^{\Uparrow}$ is psd if and only if $(K^{\rm nc})_{\hat{s}}^{\Uparrow}$ is psd.

Corollary 5.8. *Let $s\in\mathbb{R}[\boldsymbol{y}]_{2\delta}$ and $e=4^{g-1}$. Let K be a psd commutative Hankel extension of H of order $d+\delta$ such that $K_s^{\Uparrow}$ is psd. Then there exists a constant $C>0$ depending on $s,\|K\|_2,\sigma_{\min}(H)$ (and on d if $g\ge 3$) such that the following holds. For every flat psd commutative Hankel extension L of H of order $d+\delta$ satisfying*

$$\|K-L\|_2\le\|K\|_2\left(\sqrt{1+2^{-(1+2e)}}-1\right) \tag{5.5}$$

if $g\ge 3$, there exists a point $\alpha\in\mathbb{R}^g$ such that

$$s(\alpha)\ge\frac{\lambda_{\min}\left(K_s^{\Uparrow}\right)}{\|H\|_2}-C\|K-L\|_2^{1/e}. \tag{5.6}$$

Proof. Since $\|L\|_2\le\|K\|_2+\|K-L\|_2$, (5.5) ensures that the assumptions of Corollary 5.7 are met if $g\ge 3$. Also, in Corollary 5.6 one can choose ε such that $t^{1/5}\varepsilon(t)$ grows slower than $t^{1/4}$ if $g=2$.

Now (5.4), Theorem 4.1 and Lemma 5.3 imply

$$\lambda_{\min}\left(\hat{s}(\mathcal{Y}^K)\right)=\lambda_{\min}\left(\hat{s}(\mathcal{X}^{K^{\rm nc}})\right)\ge\frac{\lambda_{\min}\left(K_s^{\Uparrow}\right)}{\|H\|_2}-C_1\|K-L\|_2, \tag{5.7}$$

where $C_1>0$ is a constant depending on $s,\|K\|_2,\sigma_{\min}(H)$.

Let $Y_1,\dots,Y_k$ be commuting symmetric matrices from Corollary 5.6 (if $g=2$) or Corollary 5.7 (if $g\ge 3$). Then

$$\|\hat{s}(Y)-\hat{s}(\mathcal{Y}^K)\|\le C_2\|K-L\|_2^{1/e} \tag{5.8}$$

for some constant C_2 depending on $s,\|K\|_2,\sigma_{\min}(H)$ (and $\dim\mathcal{H}$ if $g\ge 3$). Since Y_j commute, we have $s(Y)=\hat{s}(Y)$. Moreover, Y_j are jointly orthogonally diagonalizable, so from their eigenvalues we can make a tuple $\alpha\in\mathbb{R}^g$ such that $s(\alpha)$ equals the least eigenvalue of $s(Y)$. The conclusion now follows by (5.6) and (5.7). ☐

Now we apply the preceding results to the Lasserre relaxation scheme for commutative POP; see [Las09, Chapter 5] for a comprehensive explanation. Assume that polynomials $s_1,\dots,s_h\in\mathbb{R}[\boldsymbol{y}]_{2\delta}$ generate an Archimedean quadratic module and consider the corresponding semialgebraic set

$$\mathcal{S}=\{\alpha\in\mathbb{R}^g\colon s_1(\alpha)\ge 0,\dots,s_h(\alpha)\ge 0\}.$$

For $f\in\mathbb{R}[\boldsymbol{y}]_d$ we are interested in $\alpha_0\in\mathcal{S}$ such that

$$f(\alpha_0)=f^*:=\min_{\alpha\in\mathcal{S}}f(\alpha). \tag{5.9}$$

Following [Las01, Section 4] or [HL06, Subsection II.C], we can form SDP relaxations of (5.9) analogously as in Subsection 4.4. Let $K(d)$ be a psd commutative Hankel matrix of order $d+\delta$ that is an optimal solution of the relaxation of order $d+\delta$. Then $K(d)_{s_i}^{\Uparrow}$ is psd for every $1\le i\le h$. If

$$f^{(d)}=\langle K(d)\vec{f},\vec{1}\rangle_2,$$

then by [Las01, Theorem 4.2] or [HL06, Theorem 2.1] there exists $d_0\in\mathbb{N}$ such that the sequence $\{f^{(d)}\}_{d\ge d_0}$ is monotonically increasing and $\lim_d f^{(d)}=f^*$.

Assume that $d \geq d_0$ is such that $K(d)_{s_i}^{\Uparrow}$ is pd for $1 \leq i \leq h$ and $K(d)$ is close to being flat in the sense that there exists a flat psd commutative Hankel extension L of $H(d)$ of order $d+\delta$ with $\|K(d)-L\|_2 \ll 1$. If $\|K(d)-L\|_2$ is small enough, then by Corollary 5.8 there exists $\alpha \in \mathbb{R}^g$ such that $s_i(\alpha) \geq 0$ and hence $\alpha \in \mathcal{S}$. Therefore

$$f^{(d)} \leq f^* \leq f(\alpha).$$

By the proof of Corollary 5.8, $f(\alpha) - f^{(d)}$ grows (at most) as $\|K(d) - L\|_2^{1/e}$.

6. Conclusions

In this paper we presented a robustness analysis of the minimizer extraction via the Gelfand-Naimark-Segal (GNS) construction for polynomial optimization problems. We proved that in the case of constraint-free NCPOP we can bound the differences $\mathcal{X}^K - \mathcal{X}^{K'}$ and $f(\mathcal{X}^K) - f(\mathcal{X}^{K'})$ in terms of $\|K - K'\|$. In Section 4 we applied the preceding results to constrained NCPOPs and showed that the almost flat Hankel extensions of given matrix H associated to positive functionals are also almost feasible and we have quantified the deviation from feasibility in terms of deviation from flatness. We additionally explained how our results pertain to the classical, commutative POP, and to the tracial NCPOP. We also provided extensive numerical examples that support the theoretical results and show that the robustness analysis is often very tight.

When we are solving examples of POP and NCPOP using numerical methods we never end up with solutions of the dual (moment problems) that are precisely flat Hankel matrices but they are usually almost flat. For this situation the results presented in this paper imply that we can still use the GNS construction to extract solutions that are almost optimal and we can even estimate how far from the optimum they are.

References

[BPT13] G. Blekherman, P.A. Parrilo, R. R. Thomas: *Semidefinite optimization and convex algebraic geometry*, SIAM, 2013. 1

[BK12] S. Burgdorf, I. Klep: *The truncated tracial moment problem*, J. Operator Theory 68 (2012) 141–163. 19

[BKP16] S. Burgdorf, I. Klep, J. Povh: *Optimization of polynomials in non-commuting variables*, SpringerBriefs in Mathematics, Springer International Publishing, 2016. 1, 2, 3, 7, 9, 12, 18, 20

[CKP11] K. Cafuta, I. Klep, J. Povh: `NCSOStools`*: a computer algebra system for symbolic and numerical computation with non-commutative polynomials*, Optim. Methods. Softw. 26 (2011) 363–380. 1

[CKP12] K. Cafuta, I. Klep, J. Povh: *Constrained polynomial optimization problems with noncommuting variables*, SIAM J. Optim. 22 (2012) 363–383. 1

[Cim10] J. Cimprič: *A method for computing lowest eigenvalues of symmetric polynomial differential operators by semidefinite programming*, J. Math. Anal. Appl., 369 (2010) 443–452. 1

[CF96] R.E. Curto, L.A. Fialkow: *Solution of the truncated complex moment problem for flat data*, Mem. Amer. Math. Soc. 119 (1996), no. 568. 2, 3, 23

[CF02] R.E. Curto, L.A. Fialkow: *Solution of the singular quartic moment problem*, J. Operator Theory 48 (2002) 315–354. 22, 23

[Dav85] K. Davidson: *Almost commuting Hermitian matrices*, Math. Scand. 56 (1985) 222–240. 22

[deK02] E. de Klerk: *Aspects of semidefinite programming. Interior point algorithms and selected applications*, Applied Optimization, volume 65. Kluwer, 2002. 1

[dOHMP08] M.C. de Oliveira, J.W. Helton, S. McCullough, 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

[DLTW08] A.C. Doherty, Y.-C. Liang, B. Toner, S. Wehner: *The quantum moment problem and bounds on entangled multi-prover games*, in 23rd Annual IEEE Conference on Computational Complexity, pages 199–210, IEEE, 2008. 1

[FK] N. Filonov, I. Kachkovskiy: *A Hilbert-Schmidt analog of Huaxin Lin's Theorem*, preprint `arXiv:1008.4002v2`. 21, 22

[Gle] L. Glebsky: *Almost commuting matrices with respect to normalized Hilbert-Schmidt norm*, preprint `arXiv:1002.3082v1`. 21

[GLL] S. Gribling, D. de Laat, M. Laurent: *Bounds on entanglement dimensions and quantum graph parameters via noncommutative polynomial optimization*, preprint `https://arxiv.org/abs/1708.09696` 2

[Has09] M. B. Hastings: *Making Almost Commuting Matrices Commute*, Commun. Math. Phys. 291 (2009) 321–345. 21, 22

[HKM12] J. W. Helton, I. Klep, S. McCullough: *The convex Positivstellensatz in a free algebra*, Adv. Math. 231 (2012) 516–534. 18

[HM04] J.W. Helton, S. McCullough: *A Positivstellensatz for non-commutative polynomials*, Trans. Amer. Math. Soc. 356 (2004) 3721–3737. 2, 3

[HM04'] J.W. Helton, S. McCullough: *Convex noncommutative polynomials have degree two or less*, SIAM J. Matrix Anal. Appl. 25 (2004) 1124–1139. 18

[HG05] D. Henrion, A. Garulli (editors): *Positive polynomials in control*, Springer Science & Business Media, 2005. 1

[HL05] D. Henrion, J.-B. Lasserre: *Detecting global optimality and extracting solutions in GloptiPoly*, in: Positive polynomials in control 293–310, Lect. Notes Control Inf. Sci. 312, Springer, Berlin, 2005. 2, 3

[HL06] D. Henrion, J.-B. Lasserre: *Convergent relaxations of polynomial matrix inequalities and static output feedback*, IEEE Trans. Automat. Control 51 (2006) 192–202. 24

[KP16] I. Klep, J. Povh: *Constrained trace-optimization of polynomials in freely noncommuting variables*, J. Global Optim. 64 (2016) 325–348. 2, 16, 18, 20

[KS08] I. Klep, M. Schweighofer: *Connes' embedding conjecture and sums of Hermitian squares*, Adv. Math. 217 (2008) 1816-–1837. 1, 2, 19

[Las01] J.-B. Lasserre: *Global optimization with polynomials and the problem of moments*, SIAM J. Optim. 11 (2000/01) 796–817. 1, 20, 24

[Las09] J.-B. Lasserre: *Moments, positive polynomials and their applications*, Imperial College Press, London, 2009. 1, 24

[Lau09] M. Laurent: *Sums of squares, moment matrices and optimization over polynomials*, in: Emerging applications of algebraic geometry 157–270, IMA Vol. Math. Appl. 149, Springer, New York, 2009. 1, 2, 3

[LP15] M. Laurent, T. Piovesan: *Conic approach to quantum graph parameters using linear optimization over the completely positive semidefinite cone*, SIAM J. Optim. 25 (2015) 2461–2493. 2

[Lin97] H. Lin: *Almost commuting selfadjoint matrices and applications*, in Operator Algebras and Their Applications, Fields Inst. Commun. 13 (1997) 193–233. 21, 22

[LS16] T. A. Loring, A. P. W. Sørensen: *Almost commuting self-adjoint matrices: the real and self-dual cases*, Rev. Math. Phys. 28, 1650017 (2016). 22

[LT70] W. A. J. Luxembourg, R. F. Taylor: *Almost commuting matrices are near commuting matrices*, Indag. Math. 32 (1970) 96–98. 21

[Mar08] M. Marshall: *Positive polynomials and sums of squares*, American Mathematical Soc., 2008. 1

[MP05] S. McCullough, M. Putinar: *Noncommutative sums of squares*, Pacific J. Math. 218 (2005) 167–171. 3

[Nie14] J. Nie: *The $\mathcal{A}$-truncated K-moment problem*, Found. Comput. Math. 14 (2014) 1243–1276. 9

[NDS06] J. Nie, J. Demmel, B. Sturmfels: *Minimizing polynomials via sum of squares over the gradient ideal*, Math. Prog. 106 (2006) 587–606. 1

[Par03] P.A. Parrilo: *Semidefinite programming relaxations for semialgebraic problems*, Math. Prog. 96 (2003) 293–320. 1

[PS79] C. Pearcy, A. Shields: *Almost commuting matrices*, J. Funct. Anal. 33 (1979) 332–338. 21

[PNA10] S. Pironio, M. Navascués, A. Acín: *Convergent relaxations of polynomial optimization problems with noncommuting variables*, SIAM J. Optim. 20 (2010) 2157–2180. 1, 2, 3, 18

[Put93] M. Putinar: *Positive polynomials on compact semi-algebraic sets*, Indiana Univ. Math. J., 42 (1993) 969–984. 1

[Scw06] M. Schweighofer: *Global optimization of polynomials using gradient tentacles and sums of squares*, SIAM J. Optim. 17 (2006) 920–942. 1

[SIG97] R.E. Skelton, T. Iwasaki, D.E. Grigoriadis: *A unified algebraic approach to control design*, CRC Press, 1997. 1

[Tyr94] E. E. Tyrtyshnikov: *How bad are Hankel matrices?*, Numer. Math. 67 (1994) 261–269. 8

[Voi83] D. V. Voiculescu: *Asymptotically commuting finite rank unitary operators without commuting approximants*, Acta Sci. Math. 45 (1983) 429–431. 22

[WSV00] H. Wolkowicz, R. Saigal, L. Vandenberghe (editors): *Handbook of Semidefinite Programming*, Kluwer, 2000. 1

Igor Klep, Department of Mathematics, The University of Auckland
Email address: `igor.klep@auckland.ac.nz`

Janez Povh, Faculty Of Mechanical Engineering, University of Ljubljana
Email address: `janez.povh@fs.uni-lj.si`

Jurij Volčič, Department of Mathematics, Ben-Gurion University of the Negev, Israel
Email address: `volcic@post.bgu.ac.il`

NOT FOR PUBLICATION

Contents