Convex Noncommutative Polynomials Have Degree Two or Less

J. William Helton[*] Scott McCullough [†]

Abstract

A polynomial p (with real coefficients) in noncommutative variables is matrix convex provided

$$p(tX + (1-t)Y) \leq tp(X) + (1-t)p(Y)$$

for all $0 \leq t \leq 1$ and for all tuples $X = (X_1, \ldots, X_g)$ and $Y = (Y_1, \ldots, Y_g)$ of symmetric matrices on a common finite dimensional vector space of a sufficiently large dimension (depending upon p). The main result of this paper is that every matrix convex polynomial has degree two or less. More generally, the polynomial p has degree at most two if convexity holds only for all matrices X and Y in an "open set". An analogous result for nonsymmetric variables is also obtained.

Matrix convexity is an important consideration in engineering system theory. This motivated our work and our results suggest that matrix convexity in conjunction with a type of "system scalability" produces surprisingly heavy constraints.

1 Introduction

Let $x = \{x_1, \ldots, x_g\}$ denote noncommuting indeterminates and let $\mathcal{N}(x)$ denote the set of polynomials in the indeterminates x. For example,

$$p = x_1 x_2^3 + x_2^3 x_1 + x_3 x_1 x_2 + x_2 x_1 x_3$$

is a symmetric polynomial in $\mathcal{N}(x)$.

A symmetric polynomial p is **matrix convex** if for each positive integer n, each pair of tuples $X = (X_1, \ldots, X_g)$ and $Y = (Y_1, \ldots, Y_g)$ of symmetric $n \times n$ matrices, and each $0 \leq t \leq 1$,

$$p(tX + (1-t)Y) \leq tp(X) + (1-t)p(Y), \tag{1.1}$$

where for an $n \times n$ matrix A, the notation $A \geq 0$ means A is positive semi-definite; i.e., A is symmetric and $\langle Ax, x \rangle \geq 0$ for all vectors x. Even in one-variable, convexity in the noncommutative setting differs from convexity in the commuting case because here Y need not commute with X. For example, to see $p = x^4$ is not matrix convex, let

$$X = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} \text{ and } Y = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

and compute,

$$\frac{1}{2}X^4 + \frac{1}{2}Y^4 - (\frac{1}{2}X + \frac{1}{2}Y)^4 = \begin{pmatrix} 164 & 120 \\ 120 & 84 \end{pmatrix}$$

[*]Partially supported by the the NSF, DARPA and Ford Motor Co.

[†]partially supported by NSF grant DMS-0140112

which is not positive semi-definite. On the other hand, to verify that x^2 is a matrix convex polynomial, observe that

$$\begin{aligned} tX^2 + (1-t)Y^2 \; - \; & (tX + (1-t)Y)^2 \\ = \; & t(1-t)(X^2 - XY - YX + Y^2) = t(1-t)(X-Y)^2 \geq 0. \end{aligned}$$

Our main theorem, Theorem 3.1, says (in several contexts) that

> *any noncommutative polynomial which is "matrix convex" on an "open set" has degree two or less.*

Historical background for this result appears in Section 8.2. The paper begins with the formal setup and definitions including that of "open set", see Section 2. After stating Theorem 3.1 we prove the theorem for symmetric variables X in two special cases, first when the polynomial is matrix convex everywhere and second when the polynomial is "matrix convex on the polydisc," since these are both important special cases and their proofs illustrate the general approach. The everywhere positive case is taken up in section 4. Section 5 contains a key lemma and the proof of the main result in the case that the polynomial is "matrix convex on the polydisc." The proof of the general case for both symmetric and nonsymmetric variables is presented in section 6. As an aside we mention, in Section 6.3, alternative proofs which yield partial results. A refinement of the main result which connects the work with linear matrix inequalities (LMIs) is discussed in section 7. The paper concludes with a section, Section 8, which indicates engineering motivation.

Here is the idea of the proof. A noncommutative polynomial p has a calculus second directional derivative q which is also a polynomial with degree the same as that of p, unless p has degree less than or equal one in which case $q = 0$. Our working definition of matrix convex, as discussed in Section 2.4, corresponds to the second directional derivative q of p being a "matrix positive" polynomial. Earlier results [M01] [H02], and [MPprept] say that "matrix positive" noncommutative polynomials are all sums of squares [1]. We compute that, if the degree of p exceeds two, then q has terms which preclude it from being a sum of squares. This settles the matrix convex everywhere case.

Convexity on an "open set" corresponds to positivity of the second derivative q on that set, but now q is not likely to be a sum of squares. In this case, we apply a type of noncommutative Positivstellensatz from [CHSYprept]. In the symmetric variable and "positive on the polydisc" case, the Positivstellensatz of [HMPprept] suffices.

2 Definitions

We shall now give formal definitions at appropriate levels of generality.

2.1 Noncommutative Polynomials

Of interest are two classes of noncommutative variables $x = \{x_1, \ldots, x_g\}$. In the first the x_j are symmetric and in the second they are free of relations. (So far in the Introduction we discussed only the symmetric situation.) In both cases, the definition of a convex polynomial requires g new noncommutative variables $\{h_1, \ldots, h_g\}$ either symmetric or free in correspondence with the nature of x. Now we give more details.

[1] This is in contrast with situation in the commutative case emanating from Hilbert's work and his 17^{th} problem, see [R00] for results and a survey and [PV00] for a general closely related Positivstellensatz.

Let $\mathcal{F}(x)$ denote the free semi-group on the noncommutative generators $x = \{x_1, \dots, x_g\}$. In common language, $\mathcal{F}(x)$ is the semi-group of words in $x_1, \cdots, x_g$. Note that the empty word $\emptyset$ is the identity in $\mathcal{F}(x)$.

Let $\mathcal{N}(x)$ denote the polynomials, over the field of real numbers $\mathbb{R}$, in the noncommuting generators $x = \{x_1, \cdots, x_g\}$. Thus $\mathcal{N}(x)$ is the free $\mathbb{R}$-algebra on x. As a vector space, $\mathcal{N}(x)$ consists of real linear combinations of words w from $\mathcal{F}(x)$. Concretely, a $p \in \mathcal{N}(x)$ is an expression of the form

$$p = \sum_{w \in \mathcal{F}(x)} p_w w, \tag{2.1}$$

where the sum is finite and each $p_w \in \mathbb{R}$. The algebra $\mathcal{N}(x)$ has a natural involution ^{T}, which behaves in the following way. Given a word $w = x_{j_1} x_{j_2} \cdots x_{j_n}$ from $\mathcal{F}(x)$ viewed as an element of $\mathcal{N}(x)$, the involution applied to w is

$$w^T = x_{j_n} \cdots x_{j_2} x_{j_1}.$$

In general, given p as in (2.1), $p^T = \sum p_w w^T$. A polynomial p **in** $\mathcal{N}(x)$ **is symmetric** provided $p^T = p$.

Define $\mathcal{F}(x)[h]$ and $\mathcal{N}(x)[h]$ by analogy with $\mathcal{F}(x)$ and $\mathcal{N}(x)$ as the free semi-group and free $\mathbb{R}$-algebra in the $2g$ variables $\{x, h\} = \{x_1, \dots, x_g, h_1, \dots, h_g\}$ respectively. While $\mathcal{F}(x)[h]$ and $\mathcal{N}(x)[h]$ are the same as $\mathcal{F}(x)$ and $\mathcal{N}(x)$ with g replaced by $2g$, in the sequel the variables x and h will play a somewhat different role. Often we will write $\mathcal{N}$ (resp. $\mathcal{F}$) instead of either $\mathcal{N}(x)$ or $\mathcal{N}(x)[h]$ (resp. $\mathcal{F}(x)$ or $\mathcal{F}(x)[h]$).

Let $\mathcal{F}_*(x)$ and $\mathcal{N}_*(x)$ denote the free semi-group and free $\mathbb{R}$-algebra on the $2g$ variables $\{x, x^T\} = \{x_1, \dots, x_g, x_1^T, \dots, x_g^T\}$. The involution in this setting is determined by $x_j \mapsto x_j^T$ and $x_j^T \mapsto x_j$, so that if w is a word in $\{x, x^T\}$, say $w = z_1 \cdots z_n$, then

$$w^T = z_n^T \cdots z_1^T.$$

Here $z_j \in \{x, x^T\}$. The involution extends from $\mathcal{F}_*(x)$ to $\mathcal{N}_*(x)$ in the canonical way.

Finally, the notations $\mathcal{F}_*(x)[h]$ and $\mathcal{N}_*(x)[h]$ will denote the free semi-group and free $\mathbb{R}$-algebra on the $4g$ generators

$$\{x_1, \dots, x_g, x_1^T, \dots, x_g^T, h_1, \dots, h_g, h_1^T, \dots, h_g^T\}$$

with involution defined by analogy with $\mathcal{F}_*(x)$ and $\mathcal{N}_*(x)$. Often we will write $\mathcal{N}_*$ (resp. $\mathcal{F}_*$) instead of either $\mathcal{N}_*(x)$ or $\mathcal{N}_*(x)[h]$ (resp. $\mathcal{F}_*(x)$ or $\mathcal{F}_*(x)[h]$).

2.2 Matrix Noncommutative Polynomials

Given a finite index set $\mathcal{J}$ and a set $\mathcal{S}$, let $M_{\mathcal{J}}(\mathcal{S})$ denote the matrices with entries from $\mathcal{S}$ indexed by $\mathcal{J}$. Thus, an $M \in M_{\mathcal{J}}(\mathcal{S})$ has the form $M = (M_{j,\ell})_{j,\ell \in \mathcal{J}}$ for some $M_{j,\ell} \in \mathcal{S}$. In the case $\mathcal{J} = \{1, \dots, n\}$, the set $M_{\mathcal{J}}(\mathcal{S})$ is simply $M_n(\mathcal{S})$, the $n \times n$ matrices with entries from $\mathcal{S}$. Similarly, view $\mathcal{S}^{\mathcal{J}}$ as (column) vectors indexed by $\mathcal{J}$. For instance, when $\mathcal{J} = \{1, \dots, n\}$, we find $\mathcal{S}^{\mathcal{J}} = \mathcal{S}^n$ is the set of n-vectors with entries from $\mathcal{S}$.

If $\mathcal{J}$ is a finite subset of $\mathcal{F}$ and $\mathcal{S} = \mathcal{N}$, then $M_{\mathcal{J}}(\mathcal{N})$ is an algebra with involution

$$M^T = (M_{v,w})^T_{v,w \in \mathcal{J}} = (M^T_{w,v})_{v,w \in \mathcal{J}}.$$

Further, given $V \in \mathcal{N}^{\mathcal{J}}$ and $M \in M_{\mathcal{J}}(\mathcal{N})$, the definition

$$V^T M V = \sum_{u,w \in \mathcal{J}} V_u^T M_{u,w} V_w$$

is unavoidable. Elements of $M_{\mathcal{J}}(\mathcal{N})$ are naturally identified with noncommutative matrix-valued polynomials by writing $p \in M_{\mathcal{J}}(\mathcal{N})$ as

$$p = \sum_{w \in \mathcal{F}} p_w w \tag{2.2}$$

just as in (2.1), but now where $p_w \in M_{\mathcal{J}}(\mathbb{R})$. With this notation, the involution is given by

$$p^T = \sum_{w \in \mathcal{F}} p_w^T w^T.$$

A matrix-valued noncommutative polynomial of degree one is a linear pencil. Explicitly, in the $\mathcal{N}(x)$ case, a linear pencil Λ has the form

$$\Lambda = \Lambda_0 + \sum_1^g \Lambda_j x_j,$$

where $\Lambda_j \in M_n(\mathbb{R})$ for some n (or more generally, the Λ_j are operators on a Hilbert space).

2.3 Substituting Matrices for Indeterminates

Often we shall be interested in evaluating a polynomial p in $\mathcal{N}(x)$ at a tuple of bounded symmetric operators $X = (X_1, \dots, X_g)$ on a common real Hilbert space $\mathcal{H}$. Define $X^{\emptyset} = I$, the identity operator on $\mathcal{H}$; given a word $w \in \mathcal{F}(x)$ different from the empty word, $w = x_{j_1} x_{j_2} \cdots x_{j_n}$, let

$$X^w = X_{j_1} X_{j_2} \cdots X_{j_n};$$

and given p as in (2.1), define $p(X) = \sum p_w X^w$. Note that the involution on $\mathcal{N}$ is compatible with the transpose operation on operators on real Hilbert space,

$$p(X)^T = p^T(X),$$

where $p(X)^T$ denotes the transpose of the operator $p(X)$ (with respect to the native inner product). Often the Hilbert space is $\mathbb{R}^n$ and so the operators X_j are real symmetric $n \times n$ matrices and $p(X)^T$ is just the usual transpose of the $n \times n$ matrix $p(X)$.

Let $\mathcal{B}(\mathcal{H})$ denote the bounded linear operators on $\mathcal{H}$. A fixed tuple $X = (X_1, \dots, X_g)$ of symmetric elements of $\mathcal{B}(\mathcal{H})$ determines an algebra homomorphism $\mathcal{N}(x) \longrightarrow \mathcal{B}(\mathcal{H})$ which preserves ^{T} by evaluation, $p \mapsto p(X)$. This evaluation mapping extends to matrix polynomials $M_{\mathcal{J}}(\mathcal{N}(x)) \longrightarrow M_{\mathcal{J}}(\mathcal{B}(\mathcal{H}))$ by defining, for a p in $M_{\mathcal{J}}(\mathcal{N}(x))$ with entries $p_{j,\ell}$, the matrix $p(X)$ as the matrix with entries $p_{j,\ell}(X)$. In other words, we apply the evaluation map entrywise. Note that $M_{\mathcal{J}}(\mathcal{B}(\mathcal{H}))$ is naturally identified with $\mathcal{B}(\oplus_{\mathcal{J}} \mathcal{H})$ and that, in the notation of (2.2),

$$p(X) = \sum p_w \otimes X^w,$$

where the coefficients are matrices. If $X = (X_1, \dots, X_g)$ and $H = (H_1, \dots, H_g)$ are tuples of symmetric operators on $\mathcal{H}$, then the evaluation homomorphism defined by

$$p(x, h) = p(x)[h] \mapsto p(X, H) = p(X)[H]$$

acts as a mapping $\mathcal{N}(x)[h] \longrightarrow \mathcal{B}(\mathcal{H})$.

In the $\mathcal{N}_*(x)$ case evaluation is allowed at arbitrary tuples $X = (X_1, \dots, X_g)$ of operators on a common real Hilbert space $\mathcal{H}$ where now X_j^T is substituted for x_j^T. Evaluation of $p \in M_{\mathcal{J}}(\mathcal{N}_*(x))$ or $p \in M_{\mathcal{J}}(\mathcal{N}_*(x)[h])$ at tuples X or (X, H) is defined also.

Lemma 2.1 *Given d, there exists a real Hilbert space $\mathcal{K}$ of dimension $\sum_0^{2d} g^j$ and a tuple $Y = (Y_1, \ldots, Y_g)$ of symmetric operators on $\mathcal{K}$ such that if $p \in \mathcal{N}(x)$ has degree at most d and if $p(Y) = 0$, then $p = 0$.*

Similarly, there exists a Hilbert space $\mathcal{K}$ of dimension $\sum_0^{2d}(2g)^j$ and a tuple of operators $Y = (Y_1, \ldots, Y_g)$ on $\mathcal{K}$ such that if $p \in \mathcal{N}_*(x)$ has degree at most d and if $p(Y) = 0$, then $p = 0$.

We will have use of the following variant of Lemma 2.1, which uses only that for each p there is a Y (perhaps depending upon p) in Lemma 2.1. Let $\mathcal{B}^{\mathrm{sym}}(\mathcal{H})^g$ denote g-tuples $X = (X_1, \ldots, X_g)$ of symmetric operators on $\mathcal{H}$. Let $\mathcal{B}(\mathcal{H})^g$ denote all g-tuples of operators on $\mathcal{H}$. In the case $\mathcal{H} = \mathbb{R}^n$, we write $(M_n^{\mathrm{sym}})^g$ and M_n^g in place of $\mathcal{B}^{\mathrm{sym}}(\mathbb{R}^n)^g$ and $\mathcal{B}(\mathbb{R}^n)^g$.

Lemma 2.2 *Given d, there exists a Hilbert space $\mathcal{K}$ of dimension $\sum_0^{2d} g^j$ such that if G is an open subset of $\mathcal{B}^{sym}(\mathcal{K})^g$, if $p \in \mathcal{N}(x)$ has degree at most d, and if $p(X) = 0$ for all $X \in G$, then $p = 0$.*

Similarly, there exists a Hilbert space $\mathcal{K}$ of dimension $\sum_0^{2d}(2g)^j$ such that if G is an open subset of $\mathcal{B}(\mathcal{K})^g$, if $p \in \mathcal{N}_*(x)$ has degree at most d, and if $p(X) = 0$ for all $X \in G$, then $p = 0$.

Proof. Choose a $Z \in G$ and let $h, k \in \mathcal{K}$ be given. Define, the old fashion polynomial on $t \in \mathbb{R}$,

$$s(t) = \langle p((1-t)Z + tY)h, k\rangle,$$

where Y is the tuple from Lemma 2.1 and $(1-t)Z + tY$ is the tuple

$$((1-t)Z_1 + tY_1, \ldots, (1-t)Z_g + tY_g).$$

Since G is open and $p(X) = 0$ for $X \in G$, $s(t) = 0$ for small t. Since s is a polynomial, $s = 0$ and hence, substituting $t = 1$ gives $\langle p(Y)h, k\rangle = 0$. Thus, $p(Y) = 0$. ∎

2.4 Matrix Convexity and Positivity

A polynomial $q \in \mathcal{N}(x)$ is matrix positive if $q(X) \geq 0$ for all tuples $X = (X_1, \ldots, X_g)$ of symmetric operators on finite dimensional Hilbert space. Matrix positive for q in either $\mathcal{N}(x)[h]$, $\mathcal{N}_*(x)$, or $\mathcal{N}_*(x)[h]$ is defined in a similar fashion.

Matrix positive polynomials are sums of squares.

Theorem 2.3 *Given d, there exists a Hilbert space $\mathcal{K}$ of dimension $N(d) = \sum_0^d g^j$ such that if $q \in \mathcal{N}(x)$, the degree of q is at most d, and $q(X) \geq 0$ for all tuples $X = (X_1, \ldots, X_g)$ on $\mathcal{K}$, then there exists $r_j \in \mathcal{N}(x)$, $1 \leq j \leq N(d)$, such that $q = \sum r_j^T r_j$.*

Similarly, there exists a Hilbert space $\mathcal{K}$ of dimension $N(d) = \sum_0^d (2g)^j$ such that if $q \in \mathcal{N}_*(x)$, q has degree at most d, and $q(X) \geq 0$ for all tuples $X = (X_1, \ldots, X_g)$ on $\mathcal{K}$, then there exists $r_j \in \mathcal{N}_*(x)$, $1 \leq j \leq N(d)$, such that $q = \sum r_j^T r_j$.

Versions of this sum of squares (SoS) result can be found in [H02],[M01], and [MPprept].

2.4.1 Matrix Convexity

Matrix convexity can be formulated in terms of the second derivative and positivity, just as in the case of a real variable. Given a polynomial $p \in \mathcal{N}(x)$,

$$r(x)[h] := p(x+h) - p(x)$$

is a polynomial in $\mathcal{N}(x)[h]$. Define the Hessian q of p to be the part of $r(x)[h]$ which is homogeneous of degree two in h. Alternatively, the Hessian is the second directional derivative of p,

$$q(x)[h] := \frac{d^2 p(x+th)}{dt^2}|_{t=0}.$$

For example, $p = x_1^2 x_2$ has Hessian

$$q(x)[h] = h_1^2 x_2 + h_1 x_1 h_2 + x_1 h_1 h_2.$$

If $q \neq 0$, that is if p has degree ≥ 2, then the degree of q equals the degree of p.

Theorem 2.4 *([HMer98]) A polynomial $p \in \mathcal{N}$ is matrix convex if and only if its Hessian $q(x)[h]$ is matrix positive.*

A polynomial $p \in \mathcal{N}_*(x)$ is matrix convex if (1.1) holds for all tuples X and Y whether symmetric or not. The Hessian of p is again the homogeneous of degree two in h part of $p(x+h)-p(x)$. For instance, the Hessian of $p(x) = xx^Tx$, is $xh^Th + hx^Th + hh^Tx$. Theorem 2.4 is true with $\mathcal{N}$ replaced by $\mathcal{N}_*$.

2.4.2 Positivity Domains

Let $M_\infty(\mathcal{N})$ and $M_\infty(\mathcal{N}_*)$ denote the unions $\cup_{n=1}^\infty M_n(\mathcal{N}(x))$ and $\cup_{n=1}^\infty M_n(\mathcal{N}_*(x))$ respectively. Fix a subset $\mathcal{P}$ of $M_\infty(\mathcal{N})$ or $M_\infty(\mathcal{N}_*)$. The case that $\mathcal{P}$ consists of symmetric polynomials is of primary interest, but we will have occasion to consider more general collections. Given a real Hilbert space $\mathcal{H}$, let $\mathcal{D}_\mathcal{P}(\mathcal{H})$ denote the tuples $X = (X_1, \ldots, X_g)$ such that each X_j is an operator on $\mathcal{H}$ and $p(X) \geq 0$ for each $p \in \mathcal{P}$. In the $\mathcal{N}$ case each X_j is, of course, assumed symmetric.

The **positivity domain of** $\mathcal{P}$, denoted $\mathcal{D}_\mathcal{P}$, is the collection of tuples X such that $X \in \mathcal{D}_\mathcal{P}(\mathcal{H})$ for some $\mathcal{H}$. The fact that $\mathcal{D}_\mathcal{P}$ is not actually a set presents no logical difficulties and typically it may be assumed that the Hilbert spaces are separable and even finite dimensional.

2.4.3 Matrix Convexity on a Positivity Domain

Given a collection $\mathcal{P} \subset M_\infty(\mathcal{N}(x))$ with corresponding positivity domain $\mathcal{D}_\mathcal{P}$, a polynomial $q \in \mathcal{N}(x)[h]$ is **matrix positive on** $\mathcal{D}_\mathcal{P}$ if $q(X)[H]$ is positive semi-definite for all tuples $X = (X_1, \ldots, X_g)$ and $H = (H_1, \ldots, H_g)$ of symmetric operators on a common Hilbert space such that $X \in \mathcal{D}_\mathcal{P}$. The polynomial $p \in \mathcal{N}(x)$ is **matrix convex** on $\mathcal{D}_\mathcal{P}$ provided its Hessian is matrix positive on $\mathcal{D}_\mathcal{P}$. When $\mathcal{D}_\mathcal{P}$ is all matrices, for example if $\mathcal{P}$ consists of the polynomial 1, then matrix convexity on $\mathcal{D}_\mathcal{P}$ is the same as matrix convexity.

Matrix convex on a positivity domain is defined in the $\mathcal{N}_*$ case in the expected manner.

2.4.4 The Openness Condition

Definition 2.5 (Openness property) *The positivity domain $\mathcal{D}_{\mathcal{P}}$ has the* **openness property** *provided that there is an integer n_0 with the property that when $n > n_0$, the set of matrices $\mathcal{D}_{\mathcal{P}} \cap M_n$ is equal to the closure of the interior of $\mathcal{D}_{\mathcal{P}} \cap M_n$. Often we say such a $\mathcal{D}_{\mathcal{P}}$ is an* **open positivity domain**.

3 The Main Theorem

Theorem 3.1 *If a noncommutative symmetric polynomial p is matrix convex on some positivity domain which satisfies the openness condition, then p has degree 2 or less. Here either $p \in \mathcal{N}(x)$ or $p \in \mathcal{N}_*(x)$ with matrix convex interpreted accordingly.*

4 Proof of Theorem 3.1 for Everywhere Convex Polynomials

We first treat the special case of Theorem 3.1 in which $p \in \mathcal{N}$ is matrix positive everywhere, since it is easy and serves as a guide to part of the proof for Theorem 3.1.

Proposition 4.1 *If a noncommutative symmetric polynomial p in symmetric variables is matrix convex everywhere, then p has degree 2 or less. That is, if $p \in \mathcal{N}(x)$ is matrix convex (everywhere), then the degree of p is at most two.*

Given $p \in \mathcal{N}$,

$$p = \sum_w p_w w,$$

we say p **contains** the word u or u **appears** in p if $p_u \neq 0$.

Proof. Let $q(x)[h]$ denote the second directional derivative of p in direction h. It is a symmetric polynomial which is homogeneous of degree two in h. By Theorem 2.4 the polynomial p is matrix convex if and only if q is matrix positive. Thus, by Theorem 2.3, q is a sum of squares so that there exists an m and polynomials r_j in x and h such that q has the form

$$q = \sum_{j=1}^{m} r_j^T r_j. \tag{4.1}$$

Write each r_j as

$$r_j = \sum_{w \in \mathcal{F}(x)[h]} r_j(w) w,$$

where all but finitely many of the coefficients $r_j(w) \in \mathbb{R}$ are 0.

We begin our analysis of the r_j by showing that each r_j has degree in h no greater than 1. For a polynomial $r \in \mathcal{N}(x)[h]$, let $\deg_h(r)$ denote the degree of r in h and $\deg_x(r)$ denote the degree of r in x. Let

$$d_h = \max\{\deg_h(r_j) : j\},$$

let

$$d_x = \max\{\deg_x(w) : \text{ there exists } j \text{ so that } r_j \text{ contains } w \text{ and } \deg_h(w) = d_h\}$$

and let

$$\mathcal{S}_{d_x,d_h} := \{w : r_j \text{ contains } w \text{ for some } j,\ \deg_h(w) = d_h,\ \text{and } \deg_x(w) = d_x\}.$$

The portion of q homogeneous of degree $2d_h$ in h and $2d_x$ in x is

$$Q = \sum_{\{j=1,\ldots,m,\ v,w\in\mathcal{S}_{d_x,d_h}\}} r_j(v)r_j(w)v^T w.$$

Since, for $v_j, w_j \in \mathcal{S}_{d_x,d_h}$, the equality $v_1^T w_1 = v_2^T w_2$ can occur if and only if $v_1 = v_2$ and $w_1 = w_2$, we see that $Q \neq 0$ and thus $\deg_h(q) = 2d_h$. Since q has degree two in h, we obtain $2d_h = 2$, so $d_h = 1$.

Now we turn to bounding the total degree of q. The asymptotics of a matrix positive q dictate that it have even degree. Accordingly, denote the degree of q by $2N$. Recall $2N$ is also the degree of p, since we may assume degree $p \geq 3$, or the Corollary is proved. Thus the polynomial p contains a term of the form

$$t := x_{\ell_1}x_{\ell_2}x_{\ell_3}\cdots x_{\ell_{2N}}. \tag{4.2}$$

The second derivative of t in the direction h contains a term of the form

$$\mu := h_{\ell_1}h_{\ell_2}x_{\ell_3}\cdots x_{\ell_{2N}}$$

and consequently $q(x)[h]$ contains the term μ. Thus, at least one of the products $r_{j_0}^T\ r_{j_0}$ must contain μ. Use now the finding in the previous paragraph that r_{j_0} has degree at most one in h to conclude that r_{j_0} must contain the term $h_{\ell_2}x_{\ell_3}\cdots x_{\ell_{2N}}$ and therefore the polynomial r_{j_0} has (total) degree at least $2N-1$.

Next observe canceling the terms of largest (total) degree in $\sum r_j^T r_j$ is impossible, so each r_j is a polynomial of degree half of the degree of q or less. That is $\deg(r_j) \leq N$ for each j, including r_{j_0}. It follows that $N \leq 1$. ■

5 Gram Representations

In this section we lay ground work for proving Theorem 3.1 and prove a special case which illustrates the general idea.

5.1 A Gram Representation for a Polynomial

The analog of the sum of squares representation (4.1) used in the proof of Corollary 4.1 required for the proof in the general case is a Gram representation for a polynomial $q(x)[h] = q(x,h)$ which is homogeneous of degree two in h and matrix positive on a positivity domain. We discuss the case of symmetric variables. The case of non-symmetric variables is similar, but notationally more complicated.

Since q is homogeneous of degree two in h, it may be written as

$$q(x,h) = V(x)[h]^T M(x)V(x)[h] = V^T MV \tag{5.1}$$

where the *border* vector $V(x)[h]$ is linear in h and has the form

$$V(x)[h] := \begin{pmatrix} V^1(x)[h_1] \\ \vdots \\ V^k(x)[h_k] \end{pmatrix} \text{ where } V^j(x)[h_j] = \begin{pmatrix} h_j m_1^j(x) \\ h_j m_2^j(x) \\ \vdots \\ h_j m_{\ell_j}^j(x) \end{pmatrix} \tag{5.2}$$

the m_r^j are monomials in x, and the matrix M is symmetric and its entries are noncommutative polynomials in x. The following lemma says we may (and we will) take V to have the property for each fixed j all of the $m_r^j(x)$ are distinct monomials.

Lemma 5.1 *There is a V^TMV representation (5.1) for $q(x)[h]$ in which for each fixed j all of the $m_i^j(x)$ are distinct monomials. Here distinct precludes one monomial being a scalar multiple of another.*

Proof. One can represent $q(x)[h]$ as in (5.1) with m_i^j being monomials. Clearly the only issue is whether two of these monomials are collinear. The proof that collinearity in $V(x)[h]$ is removable can be done with induction where the key induction step goes as follows. Suppose we have a q with the representation

$$q(x)[h] = \begin{pmatrix} m \\ \alpha m \\ n \end{pmatrix}^T \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} \begin{pmatrix} m \\ \alpha m \\ n \end{pmatrix}$$

with α a real number and m and n noncollinear monomials. Write q as

$$\begin{aligned} q(x)[h] \;&=\; m^T(p_{11} + \alpha^2 p_{22} + \alpha p_{21} + \alpha p_{12})m \\ &+\; m^T(p_{13} + \alpha p_{23})n + n^T(p_{31} + \alpha p_{32})m + n^T p_{33} n \end{aligned}$$

which leads to the representation

$$q(x)[h] = \begin{pmatrix} m \\ \\ n \end{pmatrix}^T \begin{pmatrix} p_{11} + \alpha^2 p_{22} + \alpha p_{21} + \alpha p_{12} & p_{13} + \alpha p_{23} \\ \\ p_{31} + \alpha p_{32} & p_{33} \end{pmatrix} \begin{pmatrix} m \\ \\ n \end{pmatrix}$$

which has linearly independent borders. ∎

Here distinct precludes one monomial being a scalar multiple of another. It will be convenient at times to index the polynomial entries of the matrix M by the monomials $\mathcal{J}$ in $V(x)[h]$ as in subsection 2.2. In this way M has entries $M_{h_j m_\ell^j, h_{j'} m_{\ell'}^{j'}}$.

For convenience, arrange

$$m_1^j < m_2^j < \cdots < m_{\ell_j}^j$$

in, say, graded lexicographic order. (That is, low degree is less than high degree and after that dictionary order breaks ties). Also we assume that each monomial is essential to representing q, that is, no proper subset of $\{m_1^j, \ldots, m_{\ell_j}^j\}$ produces such a representation of q. In particular, no row (or column) of M is identically zero. Such a "Gram" representation always exists, which along with a surprising positivity property, is proved in [CHSYprept]. This will be recalled formally later, see Theorem 6.1 (Theorem 8.3 of [CHSYprept]) stated near the end of the proof. See also [HMPprept] for a result which is more general in certain directions.

In the next subsection we prove a property of M special to the fact that it represents q, the Hessian of a polynomial.

5.2 The Degree of q vs Positivity of its Representer

The following key lemma is presented for symmetric as well as nonsymmetric variables, since this does not complicate notation.

Lemma 5.2 *Let p be a symmetric polynomial in either $\mathcal{N}(x)$ or $\mathcal{N}_*(x)$. Suppose the Hessian $q(x,h)$ of p (which is in either in $\mathcal{N}(x)[h]$ or $\mathcal{N}_*(x)[h]$ depending upon p and is homogeneous of degree two in h) is represented by V^TMV as in (5.1). If the degree of q in x and h together exceeds two, then there is an integer n_0 such that $M(X)$ is not positive semidefinite on any open set of tuples X of matrices of dimension n greater than or equal to n_0. In fact, if d is the degree of M in x and h jointly, then n_0 can be chosen equal to either $\sum_0^d g^j$ or $\sum_0^d (2g)^j$ in the $\mathcal{N}$ and $\mathcal{N}_*$ cases respectively.*

Proof. First we treat p with general non-symmetric x and h.

Let N denote the degree[2] of p, then p must contain a term of one of the following forms

$$t := x_i^T x_j m \quad \text{or} \quad t := x_i x_j m$$

or

$$t := x_i x_j^T m \quad \text{or} \quad t := x_i^T x_j^T m,$$

where m is a monomial of degree $N-2$ in x, x^T. We work through in detail what happens when a t of the form $t = x_i^T x_j m$ appears in p, since the other cases go similarly. In the second directional derivative $q(x,h)$ of p, a term of the form

$$\mu := h_i^T h_j m \tag{5.3}$$

appears. Thus, in the j part, V^j, of the border vector V, the monomial $h_j m$ appears. The monomial $m_{\ell_j}^j$ has largest degree (in x as it is a monomial in x only) of those monomials in V^j and thus the degree of $m_{\ell_j}^j$ is at least $N-2$.

Suppose $M_{h_j m_{\ell_j}^j, h_j m_{\ell_j}^j} \neq 0$. In this case

$$q_{\ell_j} := m_{\ell_j}^{j\,T} h_j^T M_{h_j m_{\ell_j}^j, h_j m_{\ell_j}^j} h_j m_{\ell_j}^j$$

is a nonzero polynomial which is part of q and which can not be canceled by any other part of q by the nature of the $q = V^TMV$ representation and the fact that $m_{\ell_j}^j$ is largest in the monomial ordering. (The key property here is the distinctness of the terms in V which prevents the $h_j m_\ell^j$ from repeating). It follows that the degree of q is at least twice the degree of $h_j m_{\ell_j}^j$ and hence the degree of q is at least $2(N-1)$. On the other hand, the degree of q is N. Hence, $2(N-1) \leq \deg(q) \leq N$ and it follows that $N \leq 2$. Thus p and q have degree no greater than 2.

From the preceding paragraph, if q has degree exceeding two, then

$$M_{h_j m_{\ell_j}^j, h_j m_{\ell_j}^j} = 0.$$

Fix $n \geq \sum_0^d (2g)^j$ and let $\mathcal{O} = \{X \in (M_n)^g : M(X) \geq 0\}$. For each $X \in \mathcal{O}$ and monomial w appearing in V, we see that the entries $M_{h_j m_{\ell_j}^j, w}(X)$ of the matrix $M(X)$ are zero. This is because $M(X)$ is positive semidefinite and the diagonal entry $M_{h_j m_{\ell_j}^j, h_j m_{\ell_j}^j}(X)$ is zero. If $\mathcal{O}$ contains an open set, then, from Lemma 2.2, each $M_{h_j m_{\ell_j}^j, w} = 0$ which contradicts our standing assumption that $h_j m_{\ell_j}^j$ is actually needed to represent q. Thus, $\mathcal{O}$ contains no open set and this is the conclusion of the lemma. Thus we have proved the lemma for the $\mathcal{N}_*$

[2] Convexity is assumed on a region only, so we can not use asymptotic arguments to conclude immediately that N is even.

case when p contains $t = x_i^T x_j m$. If p contains $t := x_i x_j m$ or $t := x_i x_j^T m$ or $t := x_i^T x_j^T m$, times an irrelevant scalar multiple, the proof proceeds exactly as before with $\mu := h_i h_j m$ respectively $\mu := h_i h_j^T m$ or respectively $\mu := h_i^T h_j^T m$ replacing $\mu = h_i^T h_j m$. The lemma is proved for the $\mathcal{N}_*$ case.

The proof for the case with symmetric variables x, h, is a minor variation of the proof we just gave. ∎

5.3 Proof of a Special Case

Theorem 3.1 for polynomials in $\mathcal{N}$ and special $\mathcal{D}_{\mathcal{P}}$ follows from Lemma 5.2 and either the main result of [HMPprept] or specialization of Theorem 8.3 of [CHSYprept] about rational functions to polynomials. The main value of presenting this case is that the proof is short, yet informative.

Theorem 5.3 *If $p \in \mathcal{N}(x)$ is matrix convex on the collection $\mathcal{D}$ of all tuples $X = (X_1, \ldots, X_g)$ of symmetric operators acting on a common Hilbert space with each X_j a contraction, that is $\|X_j\| \leq 1$, then p has degree at most two. We emphasize the conclusion holds whenever p is matrix convex on a positivity domain $\mathcal{D} = \mathcal{D}_{\mathcal{P}}$ which contains all tuples of symmetric contractions.*

Proof. The hypothesis on p implies that its Hessian q satisfies $q(X)[H] \geq 0$ for all tuples $X = (X_1, \ldots, X_g)$ of symmetric contractions and all tuples $H = (H_1, \ldots, H_g)$ of symmetric operators (all on the same Hilbert space). As a special case of the main result of [HMPprept], it follows that q has a representation $q = V^T M V$ as in (5.1) with $M(X) \geq 0$ for all tuples $X = (X_1, \ldots, X_g)$ of symmetric contractions. Lemma 5.2 implies q has degree at most two. Since $\deg(p) = \deg(q)$, we conclude that the degree of p is at most two. ∎

Note that this is Theorem 3.1 for polynomials in $\mathcal{N}$ except here we have a special type of set, a polydisk, which satisfies the openness condition. It is tempting to conclude that Theorem 3.1 follows immediately from this by scaling and translating the unit polydisk. However, in our noncommutative setting, translation is only permissible by a multiple of the identity.

The restriction to $\mathcal{D}_{\mathcal{P}}$ consisting of contractions is occasioned by use of [HMPprept]. However, as we soon see, the substitution of a key result from [CHSYprept] permits the extension of the result to any positivity domain which satisfies the openness condition.

6 Proof of Theorem 3.1

Our proof of Theorem 3.1 for matrix convex polynomials in either $\mathcal{N}$ or $\mathcal{N}_*$ and general positivity domains $\mathcal{D}_{\mathcal{P}}$ requires Theorem 8.3 of [CHSYprept] which analyzes, very generally, positivity of the M in $V^T M V$ representations.

6.1 Background

Theorem 8.3 in [CHSYprept] actually was stated at a sufficient level of generality for the case at hand. The statement requires considerable notation which explains why we did not do this earlier. The first subsection follows the layout of [CHSYprept] and describes the general structure. The statement of Theorem 8.3 in [CHSYprept] is in the second subsection.

6.1.1 $V(x)[h]$ for the General Case

In a slight change of notation, we now consider quadratic functions in the tuple of variables h, some of which are constrained to be symmetric and some not.

Define h as

$$h := \{h_{-N}, \dots, h_{-1}, h_1, \dots, h_N, h_{N+1}, \dots, h_r, h_{r+1}, \dots, h_k\}, \tag{6.1}$$

where $\{h_j\}_{j=r+1}^{k}$ are constrained to be symmetric and $h_j = h_{-j}^T$, for $j = 1, \dots, N$. That is, separate h into three different parts as follows: the first part[3] $\{h_j\}_{j=-N}^{N}$ has the pairwise restriction that $h_{-j} = h_j^T$, for $j = 1, \dots, N$, the second part $\{h_j\}_{j=N+1}^{r}$ has no restriction, the third part $\{h_j\}_{j=r+1}^{N}$ has each h_j constrained to be symmetric. Let $\mathcal{I}$ denote the integers between $-N$ and k except for 0. Thus, $\mathcal{I}$ is the index set for the h_j which are the entries of h.

Any noncommutative symmetric quadratic $q(x)[h]$ can be put in the form

$$V(x)[h]^T M_q V(x)[h],$$

where M_q is a rational function in x which can be taken to be a polynomial in x in the case that q is a polynomial, and where the border $V(x)[h]$ has the form

$$V(x)[h] := \begin{pmatrix} V^{mix}(x)[h] \\ V^{pure}(x)[h] \\ V^{sym}(x)[h] \end{pmatrix}, \tag{6.2}$$

with $V^{mix}(x)[h]$, $V^{pure}(x)[h]$, and $V^{sym}(x)[h]$ defined as follows:

$$V^{mix}(x)[h] = \begin{pmatrix} h_{-N} m_1^{-N}(x) \\ \vdots \\ h_{-N} m_{\ell_{-N}}^{-N}(x) \\ \vdots \\ h_{-1} m_1^{-1}(x) \\ \vdots \\ h_{-1} m_{\ell_{-1}}^{-1}(x) \\ h_1 m_1^1(x) \\ \vdots \\ h_1 m_{\ell_1}^1(x) \\ \vdots \\ h_N m_1^N(x) \\ \vdots \\ h_N m_{\ell_N}^{h}(x) \end{pmatrix} \qquad V^{pure}(x)[h] = \begin{pmatrix} h_{N+1} m_1^{N+1}(x) \\ \vdots \\ h_{N+1} m_{\ell_{N+1}}^{N+1}(x) \\ \vdots \\ h_r m_1^r(x) \\ \vdots \\ h_r m_{\ell_r}^r(x) \end{pmatrix}$$

$$V^{sym}(x)[h] = \begin{pmatrix} h_{r+1} m_1^{r+1}(x) \\ \vdots \\ h_{r+1} m_{\ell_{r+1}}^{r+1}(x) \\ \vdots \\ h_k m_1^k(x) \\ \vdots \\ h_k m_{\ell_k}^k(x) \end{pmatrix}.$$

In order to illustrate the above definitions, we give a simple example of a quadratic function and its border vector representation. Let the quadratic function $q(x)[h]$ be given

[3] The integer 0 is not included in the index set $j = -N, \dots, N$ of the first part, but for simplicity of notation we do not make this explicit, since it is clear from context.

by $q(x)[h] = h_1^T\rho_1(x)h_1 + h_1\rho_2(x)h_1^T + h_2\rho_3(x)h_2^T + h_3^T\rho_4(x)h_3 + h_4\rho_5(x)h_4$, where h_1, h_2, and h_3 are not symmetric and $h_4 = h_4^T$. The ρ_j are rational functions in x. For this quadratic $q(x)[h]$, the border vector has the following structure:

$$V(x)[h] = \begin{pmatrix} \left.\begin{matrix} h_1 \\ h_1^T \end{matrix}\right\} \text{Mixed} \\ \\ \left.\begin{matrix} h_2^T \\ h_3 \end{matrix}\right\} \text{Pure} \\ \\ \left.\begin{matrix} h_4 \end{matrix}\right\} \text{Symmetric} \end{pmatrix}$$

Note that this representation of $q(x)[h]$ might require simple relabeling of variables. For example, if $q(x)[\{h,k\}] = h^TA(x)h + kB(x)k^T$, then $h_1 = h$, $h_2 = k^T$ and

$$V(x)[h] = V^{pure}(x)[h] = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}. \tag{6.3}$$

Allowing simple relabeling of variables increases the scope of such representations to include all cases.

6.1.2 Positive Quadratic Functions: Theorem 8.3 of [CHSYprept]

The main result Theorem 8.3 of [CHSYprept] for a noncommutative rational function $q(x)[h]$ which is quadratic in h when specialized to polynomials gives the following theorem.

Theorem 6.1 *(Theorem 8.3 of [CHSYprept])*
Assumptions*:*

Consider a noncommutative polynomial $q(x)[h]$ which is a quadratic in the variables h and a set of polynomials $\mathcal{P}$ and its positivity domain $\mathcal{D}_{\mathcal{P}}$. Write $q(x)[h]$ in the form $q(x)[h] = V(x)[h]^TM(x)V(x)[h]$. Suppose that the following two conditions hold:

i. the positivity domain $\mathcal{D}_{\mathcal{P}}$ satisfies the openness property for some big enough n_0;

ii. the border vector $V(x)[h]$ of the quadratic function $q(x)[h]$ has for each fixed j distinct monomials m_i^j, $i = 1, 2, \cdots \ell_j$.

Conclusion*: The following statements are equivalent:*

a. $q(X)[H]$ is a positive semidefinite matrix for each pair of tuples of matrices X and H for which $X \in \mathcal{D}_{\mathcal{P}}$;

b. $M(X) \geq 0$ for all X in $\mathcal{D}_{\mathcal{P}}$.

6.2 Proof for the General Case

.

Now we finish the proof of Theorem 3.1. Choose a representation V^TMV for q, the Hessian of p, where M is a matrix with entries which are polynomial in x. We wish to apply Theorem 6.1 so must check its hypotheses (i) and (ii). Hypothesis (i) follows immediately from the fact that Theorem 3.1 requires p to be matrix convex (hence q to be matrix positive) on an open positivity domain. Hypothesis (ii) follows immediately from the fact that a representing M exists and from Lemma 5.1 which says that such a representation $V^T MV$ can always be replaced by one with distinct monomials in the border V. Theorem 6.1 implies that $M(X) \geq 0$ for all tuples X, either symmetric or general as the case may be. An application of Lemma 5.2 just as in the proof of Theorem 5.3 completes the proof. ■

6.3 Alternate Proofs

We make a few remarks about the possibility of alternate proofs.

First directly proving Theorem 3.1 for $f(s) = s^n$ where s is a single variable ($g = 1$) is easy and well known [A79] [RS79]. More generally, suppose that $g = 1$ and that p has degree n and is matrix convex everywhere. Then $\lim_{t\to\infty} \frac{1}{t^n} p(ts) = s^n$ is matrix convex. Thus $n = 0, 1$, or 2. Note matrix convexity on an open set is not strong enough to accommodate this asymptotic argument, but, although we do not include it, it is possible to give elementary proofs for various open sets.

Next consider a polynomial p in $g > 1$ variables which is matrix convex everywhere. Make a linear change of (collapsing of) variables $Ly = x$, where L is any $g \times 1$ matrix with real entries. Then $k(y) := p(Ly)$ is a matrix convex polynomial in one variable and so has degree less than are equal to 2. However, the fact that each such k has degree at most two does not necessarily imply that p has degree two. For example, if p has the property that whenever all variables x_i and x_j commute then $p = 0$, then $k = 0$, since $(Ly)_i, (Ly)_j$ commute. Thus any polynomial which has the form

$$\sum_j l_j c_j r_j \tag{6.4}$$

where c_j is the commutator of two polynomials has the "$k = 0$" property. Conversely, if p has the $k = 0$ property, then p has a representation as in (6.4). Thus there are many polynomials which the one variable result says nothing about.

7 Representing Quadratic Polynomials as LMI

The following corollary of Theorem 3.1 gives a little more detail.

Corollary 7.1 *A matrix convex noncommutative symmetric polynomial p as in Theorem 3.1 can be written as*

$$p(x) = c_0 + \Lambda_0(x) + \sum_{j=1}^{N} \Lambda_j(x)^T \Lambda_j(x)$$

where $\Lambda_0, \cdots, \Lambda_N$ are linear in x and c_0 is a constant.

Proof. Convexity and Theorem 3.1 tell us that p has degree two or less. Set $\phi(x) := p(x) - c_0 - \Lambda_0(x)$, where $c_0 + \Lambda_0(x)$ is the affine linear part of p. The polynomial ϕ is a homogeneous quadratic by construction. Thus the Hessian of ϕ in direction h, which is of course homogeneous quadratic, equals $\phi(h)$. Matrix convexity says that this Hessian is matrix positive, so ϕ is matrix positive. Every matrix positive noncommutative polynomial is is a sum of squares, see [H02] [M01] [MPprept]. Thus ϕ is a sum of squares,

$$\phi = \sum_{j=1}^{N} \Lambda_j(x)^T \Lambda_j(x).$$

Each of the Λ_j have degree at most one in x, as ϕ has degree two in x and since it is impossible to cancel highest degree terms in this sum of squares representation for ϕ. ∎

Remark: If q is concave, so that $p = -q$ is convex, and is represented as in Corollary 7.1, then the linear pencil

$$L(x) := \begin{pmatrix} c_0 + \Lambda_0(x) & \Lambda(x)^T \\ \Lambda(x) & -I \end{pmatrix} \tag{7.1}$$

has the same negativity domain as q, where

$$\Lambda(x) := \begin{pmatrix} \Lambda_1(x) \\ \Lambda_2(x) \\ \vdots \\ \Lambda_N(x) \end{pmatrix}.$$

This is because q is a Schur complement of $-L$ and

$$p(x) = c_0 + \Lambda_0(x) + \sum_{j=1}^{N} \Lambda_j(x)^T \Lambda_j(x) \leq 0$$

implies $c_0 + \Lambda_0(x) \leq 0$.

Those familiar with linear matrix inequalities (LMIs) see immediately that $L(x) \leq 0$ is an LMI. Thus Corollary 7.1 associates any matrix convex polynomial with an LMI. This and a variety of examples suggest to the authors that problems which correspond to concave or convex rational functions can be "converted" to equivalent LMI problems. Our speculation is bound up with the issue of convex positivity domains $\mathcal{D}_{\mathcal{P}}$, an issue not addressed in this paper (since our focus has been on noncommutative polynomials). To prove something along the lines we suggest will require vast machinery beyond that constructed here.

8 History and Engineering Motivation

We begin with motivation for our convexity results and then turn to history.

8.1 Engineering Motivation

Motivation for this paper comes from engineering system theory. One of the main practical advances in the 1990's was methodology for converting many many linear systems problems directly to matrix inequalities. See for example, [SIG97] and [GN99] which give collections of fairly recent results along these line.

These methods are well behaved numerically (up to modest size matrices) provided the inequalities are convex in some sense. Further system problems where the statement of the problem does not explicitly mention system size (as is true with most classical textbook problems of control theory), typically convert to matrix inequalities where the variables are matrices. The key point is that statements which are made for these matrices must hold for matrices of any size. That is, all of the formulae in these problems scale automatically with system size (the system dimension is not explicitly mentioned). We informally call these **dimensionless** or **scalable problems**, see [H02m]. Dimensionless problems typically produce collections of noncommutative rational functions.

Thus a key issue is to analyze matrix convexity of collections of noncommutative rational functions. While this article treats only the special case of a single polynomial the result is so strong that one suspects that even at great levels of generality noncommutative convex situations are rare and very rigid.

The author's impression (vastly incomplete, since there are thousands of engineering matrix inequality papers) of the systems literature is that whenever a dimensionless problem converts to a "convex problem", possibly by change of variables, it converts to an LMI. This is how convexity is acquired and proved in practice. The (vague) speculation in the remark in Section 7, that any matrix convex problem is "associated" with some LMI, implies that matrix convexity is not fundamentally less restrictive than are LMIs for dimensionless problems.

8.2 History

Matrix convex functions have been studied since the 1930's as in the very early papers by [K36] [BS55] and followed closely after the ground breaking work of Löwner [L34]. The focus of work until the 1990's, when engineering became an influence, was on functions of one (matrix) variable. Functions such as logs and fractional powers were studied and the closest result to the one for polynomials in this paper is

Theorem 8.1 *The function $f(X) = X^r$ on positive definite symmetric matrices X is matrix convex if $1 \leq r \leq 2$ or $-1 \leq r \leq 0$ and matrix concave if $0 \leq r \leq 1$.*

Theorem 8.1 is due to Ando [A79]. Conversely Shorrock and Rizvi [RS79] show that for other values of r, the function f is neither convex or concave. We have not seen the early derivative consequence of this that a monic polynomial in one variable is matrix convex if and only if its degree is less than or equal to two.

More recent advances on matrix convexity are summarized in [LM00] which proves at considerable generality matrix convexity of Schur compliments. Also the special type of matrix convex structure, Linear Matrix Inequalities, recently popular with engineers, was discussed above.

References

[A79] Ando, T., *Concavity of certain maps on positive definite matrices and applications to Hadamard products*, Linear Algebra Appl. **26**(1979), 203-241.

[BS55] Bendat, J. and Sherman, S., *Monotone and convex operator functions*, Trans. Amer. Math. Soc., **79**(1955), 58-71.

[CHSYprept] Camino J. F., Helton J. W., Skelton R.E., and Ye J., Matrix inequalities: A Symbolic Procedure to Determine Convexity Automatically, preprint 1 -49. to appear July 2003 in Integral Equations and Operator Theory.

[GN99] El-Ghaoui L. and Niculescu S.. *Advances in Linear Matrix Inequality Methods in Control*, SIAM, 1999.

[H02] Helton J. W., Positive non commutative polynomials are sums of squares, Annals of Math vol 56 num 2, Sept 2002 pp. 675-694.

[H02m] Helton J. W., Manipulating Matrix Inequalities Automatically, volume of Plenary Talks at the MTNS 2002, Springer Verlag.

[HMer98] Helton J. W., and Merino 0., "Sufficient Conditions for Optimization of Matrix Functions", pp. 1-5, Conf on Decision and Control 98.

[HMPprept] Helton J.W., McCullough S., and Putinar M., *Matrix representations for positive non-commutative polynomials*, preprint (http://www.math.ufl.edu/∼sam/xandh.dvi(ps)(pdf)).

[K36] Kraus, F., *Über Konvexe Matrixfunktionen*, Math. Z., **41**(1936), 18-42.

[L34] Löwner, C. *Über Monotone Matrixfunktionen*, Math. Z., **38**(1934), 177-216.

[LM00] Li, Chi-Kwong and Mathias, Roy, *Extremal characterizations of the Schur complement and resulting inequalities*, SIAM Rev., **42**(2000), no. 2, 233-246.

[M01] McCullough S., Factorization of operator-valued polynomials in several non-commuting variables, Linear Algebra Appl. **326** (2001), no. 1-3, 193–203.

[MPprept] McCullough S., and Putinar M., Non-commutative Sums of Squares, preprint (http://www.math.ufl.edu/∼sam/sos.dvi(ps)(pdf)).

[PV99] Putinar, Mihai; Vasilescu, Florian-Horia Solving moment problems by dimensional extension, Annals of Math. (2) **149** (1999), no. 3, 1087–1107.

[R00] Reznick B. *Some Concrete Aspects of Hilbert's 17th Problem*, Real Algebraic Geometry and Ordered Structures, C. Delzell and J. Madden, editors, Cont. Math 253, pp.251-272, 2000.

[RS79] Rizvi, M. Haseeb and Shorrock, R. W., *A note on matrix-convexity*, Canada. H. Statist., **7**(1979), no. 1, 39-41.

[SIG97] R. E. Skelton, T. Iwasaki, and K. M. Grigoriadis. *A Unified Algebraic Approach to Linear Control Design.* Taylor & Francis, 1997.

Department of Mathematics
University of California
San Diego, CA 92093, USA
E-mail: helton@osiris.ucsd.edu
WWW URL: http://www.ucsd.edu

Department of Mathematics
University of Florida
Gainesville, Florida 32611-8105
E-mail: sam@mail.math.ufl.edu

NOT FOR PUBLICATION

Contents