"Positive" Non-Commutative Polynomials are Sums of Squares

J. William Helton *
Mathematics Department
University of California at San Diego 92093

July 21, 2002

Abstract

Hilbert's 17th problem concerns expressing polynomials on R^n as a sum of squares. It is well known that many positive polynomials are not sums of squares; see [R00] [deA preprt] for excellent surveys. In this paper we consider symmetric non-commutative polynomials and call one "matrix positive", if whenever matrices of any size are substituted for the variables in the polynomial the matrix value which the polynomial takes is positive semidefinite. The result in this paper is:

A polynomial is matrix positive if and only if it is a sum of squares.

1 Introduction

We consider polynomials, that is, weighted sums of words on $2n$ generators which are closed under an involution, denoted by ^{T}, somewhat loosely called **transpose**. We denote the generators by $X_1, \cdots, X_n, X_1^T, \cdots, X_n^T$ and abbreviate them with the notation $X = \{X_1, X_2, \ldots, X_n\}$ and $X^T = \{X_1^T, X_2^T, \ldots, X_n^T\}$.

We call a symmetric polynomial, $\mathcal{Q}$, in X and X^T **matrix-positive** provided that when we substitute into $\mathcal{Q}$ any real matrices $\mathcal{X}_1, \cdots, \mathcal{X}_n$ of any dimension $r \times r$ for $X_1, \cdots, X_n$, and their transposes, $\mathcal{X}_1^\tau, \cdots, \mathcal{X}_n^\tau$ for $X_1^T, \cdots, X_n^T$, the resulting matrix $\mathcal{Q}(\mathcal{X}_1, \cdots, \mathcal{X}_n, \mathcal{X}_1^\tau, \cdots, \mathcal{X}_n^\tau)$ is positive semi-definite. Consider the following example in two indeterminants

$$\mathcal{Q}(X) = X_1^2 + (X_1^2)^T + X_2^T X_2. \tag{1.1}$$

If $\mathcal{X}_1$ and $\mathcal{X}_2$ are one dimensional, then $\mathcal{Q}(\mathcal{X}) = \mathcal{X}_1^2 + (\mathcal{X}_1^2)^T + \mathcal{X}_2^T \mathcal{X}_2$ is a sum of squares of numbers and so is positive semi-definite. However, if we substitute $\mathcal{X}_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ for X_1 and $\mathcal{X}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ for X_2, then we get

$$\mathcal{Q}(\mathcal{X}) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.2}$$

*Partly supported by the NSF, the ONR, DARPA and the Ford Motor Co.

which is not positive semi-definite. Thus $\mathcal{Q}(\mathcal{X})$ is not matrix positive.

We say a polynomial $\mathcal{Q}$ is a **Sum of Squares, SoS,** provided

$$\mathcal{Q}(X) = \sum_{i=1}^{k} h_i(X)^T h_i(X) \tag{1.3}$$

where each h_i is a polynomial in X and X^T.

Theorem 1.1 *Suppose $\mathcal{Q}$ is a non-commutative symmetric polynomial. If $\mathcal{Q}$ is a SoS, then $\mathcal{Q}$ is matrix-positive. If $\mathcal{Q}$ is matrix-positive, then $\mathcal{Q}$ is a SoS.*

The remainder of this paper is devoted to the proof, with the exception being a brief section at the end motivating the study of matrix positivity and describing implications of our Theorem. The proof is not entirely self contained in that it requires Corollary 3.3 from [CHSY preprt], so the serious reader might want to obtain that paper. The case where all operators are complex unitary was proved in [M].

This paper owes a serious debt to Jeff Ovall for numerous suggestions and a careful reading. Thanks are also due to Daniel Curtis who conducted some valuable computer experiments.

2 Representing Symmetric Polynomials

In this section we give a standard "Gram" representation for a polynomial. Also we characterize the non-uniqueness in the representation.

2.1 The Represention

Lemma 2.1 *If $\mathcal{Q}(X)$ is a symmetric polynomial, then there exists a symmetric matrix $M_{\mathcal{Q}}$ with real entries, not dependent on X, and a vector $V(X)$ of monomials in X such that*

$$\mathcal{Q}(X) = V(X)^T M_{\mathcal{Q}} V(X). \tag{2.1}$$

Furthermore, the vector $V(X)$ can always be chosen to be $V^d(X)$ where d is the least integer bigger than $\frac{1}{2}$ (degree of $\mathcal{Q}$).

Here we let V^d denote the column vector of all monic monomials of degree less than or equal to d in X_j and X_j^T for $j = 1, \ldots, n$, listed in graded lexicographic order. The length of V^d is $\nu(d) := 1 + (2n) + (2n)^2 + \cdots + (2n)^d$, since that is the number of monomials in X, X^T of length $= d$. For example, if $X = \{X_1, X_2\}$, then $V^2(X)$ is the column vector with entries

$$\{I, X_1, X_2, X_1^T, X_2^T, X_1^2, X_1X_2, X_1X_1^T, X_1X_2^T, X_2X_1, \ldots, (X_2^T)^2\}.$$

We think of $V(X)$ as a vector of monomials which often will be denoted

$$V(X) = \begin{pmatrix} V(X)_0 \\ \vdots \\ V(X)_{p-1} \end{pmatrix}. \tag{2.2}$$

Examples of the representation (2.2) are

$$\mathcal{Q} = X_1X_1^T + X_1X_2 + X_2^TX_1^T + 2 + 2X_2^TX_1^TX_1X_2 = \begin{pmatrix} I \\ X_1^T \\ X_1X_2 \end{pmatrix}^T \begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} I \\ X_1^T \\ X_1X_2 \end{pmatrix}.$$

Here

$$V(X) = \begin{pmatrix} I \\ X_1^T \\ X_1X_2 \end{pmatrix} \quad \text{and } M_{\mathcal{Q}} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix}. \tag{2.3}$$

Another example is

$$\mathcal{Q} = 2X_2^TX_1^TX_2^TX_2X_1X_2 = \begin{pmatrix} X_2X_1X_2 \end{pmatrix}^T \begin{pmatrix} 2 \end{pmatrix} \begin{pmatrix} X_2X_1X_2 \end{pmatrix}.$$

Proof of Lemma 2.1 If m is a monic monomial in X and X^T of degree less than or equal to $2d$, then we can write m as a product of two monomials

$$m = m_L m_R$$

each of degree $\leq d$. Thus $m + m^T$ can be written $m + m^T = V^d(X)^T E_{ij} V^d(X)$ where E_{ij} is a self adjoint matrix whose entries are all 0 except the ij^{th} and ji^{th} entry corresponding to

$$[V^d(X)^T]_i = m_L \quad \text{and} \quad [V^d(X)]_j = m_R$$

are equal to 1.

Suppose that $i \neq j$. Then

$$\begin{aligned} V^d(X)E_{ij}V^d(X) &= \begin{pmatrix} V^d(X)_i \\ V^d(X)_j \end{pmatrix}^T \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} V^d(X)_i \\ V^d(X)_j \end{pmatrix} \\ &= V^d(X)_j^T V^d(X)_i + V^d(X)_i^T V^d(X)_j \\ &= m_R^T\, m_L^T + m_L\, m_R \\ &= m^T + m\,. \end{aligned}$$

We would like to use representing matrices having only zeroes on the diagonal. Any monomial m of degree strictly less than $2d$ and greater than 0 has such a representation. Also even if m has degree $2d$ and is not symmetric, it has such a representation, which is unique. Thus for monomials m with $1 \leq$ degree $m \leq 2d-1$ or which are not symmetric, we define M^m to be the set of matrices E_{ij} having only zeroes on the diagonal and representing $m + m^T$ via

$$m + m^T = V^d(X)^T E_{ij} V^d(X).$$

When degree $m = 2d$ and $m = m^T$, the monic monomial m has a *unique representation* $m = m_L m_R$ with deg $m_L = d$ and deg $m_R = d$. In this case $m_L^T = m_R$. Thus the representing set M^m for m consists of one matrix which is all 0's except for one diagonal entry which is 1. When degree m is 0 this is also true.

To represent a symmetric $\mathcal{Q}$ write it as a weighted sum of symmetrized monic monomials, $w^\ell(m^\ell + (m^\ell)^T)$. Thus we get a representation for $\mathcal{Q}$ with

$$M_{\mathcal{Q}} := \sum_{\ell=1}^{L} w^\ell M_\ell \tag{2.4}$$

where we choose one matrix M_ℓ from each set M^{m^ℓ}. •

The representation

$$\mathcal{Q}(X) = V^d(X)^T M_{\mathcal{Q}} V^d(X)$$

for a fixed $\mathcal{Q}$ can be done with many symmetric matrices $M_{\mathcal{Q}}$. We characterize this non-uniqueness in §2.2.

2.2 The Non-uniqueness in the Represention

Define $SR^{p\times p}$ to be the symmetric $p \times p$ matrices and abbreviate positive semidefinite by PSD. A non-commutative polynomial $\mathcal{Q}$ with a representation

$$\mathcal{Q}(X) = V(X)^T M_{\mathcal{Q}}^0 V(X) \tag{2.5}$$

may also be represented by different M's and the same V. These representations of $\mathcal{Q}$ have M of the form

$$M = M_{\mathcal{Q}}^0 + B \tag{2.6}$$

where $B \varepsilon \mathcal{B}_V$ with $\mathcal{B}_V \subset SR^{p\times p}$ defined by

$$\mathcal{B}_V := \{B : V(X)^T B V(X) = 0 \qquad \text{and} \qquad B = B^T\}. \tag{2.7}$$

Here $M_{\mathcal{Q}}^0 \varepsilon SR^{p\times p}$. *If one matrix M of the form (2.6) is PSD, then the Cholesky decomposition of M implies $\mathcal{Q}$ is a SoS.* This is discussed more fully in Proposition 6.1. It is well known structure which was exploited quite successfully by [PW98] for computational purposes in attempting to express commutative polynomials as SoS.

$\mathcal{B}_V$ can be neatly characterized as the span of a certain set of "**fundamental matrices for V**" each having 3 or 4 nonzero entries. This is needed later. The fundamental matrices sit in correspondence with those monomials which are entries of $V(X)$ satisfying

$$V(X)_i^T V(X)_j = V(X)_k^T V(X)_\ell \tag{2.8}$$

and we define a **fundamental matrix** corresponding to the pair of pairs of integers $\{(i,j),(k,\ell)\}$ satisfying (2.8) to be any matrix F satisfying

$$\begin{aligned} F_{ij} &= -F_{k\ell} \\ F_{ji} &= -F_{\ell k} \\ F_{st} &= 0 \text{ otherwise.} \end{aligned}$$

We denote the set of such fundamental matrices by $\mathcal{F}(\{(i,j),(k,\ell)\})$. Clearly any F in $\mathcal{F}\{(i,j),(k,\ell)\}$ is a symmetric matrix. Note that if $(i,j) = (k,\ell)$, then $\mathcal{F}\{(i,j),(i,j)\}$ contains only the zero matrix.

Lemma 2.2 *The fundamental matrices are contained in $\mathcal{B}_V$ and $\mathcal{B}_V$ is the span of them.*

Proof:

$$V(X)^T B V(X) = \sum_{u,s=0}^{p-1} B_{us} V(X)_u^T V(X)_s \tag{2.9}$$

is a sum of monomials, so this is identically 0 if and only if the coefficient of each monomial is 0. Consider one monomial $\mu(X)$ which can be written in r different ways

$$\mu(X) = V(X)_{i_\ell}^T V(X)_{j_\ell} \qquad\qquad \ell = 1, \ldots, r. \tag{2.10}$$

The coefficients of $\mu(X)$ in (2.9) which make $\mu(X)$ disappear are

$$\phi_{\mu(X)} := \{(B_{i_1 j_1}, \ldots, B_{i_r j_r}) : \sum_{\ell=1}^{r} B_{i_\ell j_\ell} = 0\}.$$

This is a vector space spanned by the subset consisting of all vectors in $\phi_{\mu(X)}$ having exactly two nonzero entries, say $\{i_{\ell_u}, j_{\ell_u}\}$ and $\{i_{\ell_s}, j_{\ell_s}\}$. Symmetry of B forces the coefficient of $\mu(X)$ to vanish if and only if the coefficient of $\mu(X)^T$ vanishes. More specifically, each such vector corresponds to fundamental matrices $\mathcal{F}(\{(i_{\ell_u}, j_{\ell_u}), (i_{\ell_s}, j_{\ell_s})\})$ for V. Here (i_{ℓ_u}, j_{ℓ_u}) is not allowed to equal (i_{ℓ_s}, j_{ℓ_s}).

Clearly, an F in $\mathcal{F}(\{(i_{\ell_u}, j_{\ell_u}), (i_{\ell_s}, j_{\ell_s})\})$ is in $\mathcal{B}_V$, because

$$\begin{aligned} V(X)^T F V(X) &= \sum_{u,s=1}^{p-1} F_{us} V(X)_u^T V(X)_s \\ &= F_{i_{\ell_u} j_{\ell_u}} V(X)_{i_{\ell_u}}^T V(X)_{j_{\ell_u}} + F_{i_{\ell_s} j_{\ell_s}} V(X)_{i_{\ell_s}}^T V(X)_{j_{\ell_s}} \\ &\quad + F_{j_{\ell_u} i_{\ell_u}} V(X)_{j_{\ell_u}}^T V(X)_{i_{\ell_u}} + F_{j_{\ell_s} i_{\ell_s}} V(X)_{j_{\ell_s}}^T V(X)_{i_{\ell_s}} \\ &= [F_{i_{\ell_u} j_{\ell_u}} + F_{i_{\ell_s} j_{\ell_s}}]\mu(X) + [F_{j_{\ell_u} i_{\ell_u}} + F_{j_{\ell_s} i_{\ell_s}}]\mu(X) = 0. \end{aligned} \tag{2.11}$$

Let P_μ denote the **set of all pairs of pairs of integers** associated with the monomial $\mu(X)$ by (2.10). We do not allow repetition inside a pair of pairs, that is, the two pairs cannot be the same. The span of the fundamental matrices arising from $\phi_{\mu(X)}$ as we sweep through the monomials $\mu(X)$ equals $\mathcal{B}_V$. That is,

$$\mathcal{B}_V = \text{span } \{\mathcal{F}(\{(i,j),(k,\ell)\}) : \quad \{(i,j),(k,\ell)\} \in P_\mu \text{ for some monomial } \mu\}.$$

•

3 The Range of a Representation of a List of Monomials

We shall be substituting matrices $\mathcal{X}_1, \ldots, \mathcal{X}_n \in R^{r\times r}$ for the indeterminants $X_1, \ldots, X_n$ in V^d to obtain a vector denoted $V^d(\mathcal{X})$ whose entries are $r \times r$ matrices. Fix $v_0 \in R^r$ and define

$$\mathcal{R}_d^{v_0,r} := \{V^d(\mathcal{X})v_0 : \text{all } \mathcal{X} \in R^{r\times r}\}.$$

Our goal in this section will be to give a simple characterization of $\mathcal{R}^{v_0,r}$ whenever r is sufficiently large.

As an example consider the range not for $V^d(X)$ but for the shorter list of monomials $V(X)$ given by (2.3) when $r = 2$ and $v_0 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$. This range is

$$\mathcal{R} = \left\{ \begin{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} \\ \begin{pmatrix} a & c \\ b & a \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} \end{pmatrix} : a,b,c,d,e,f,g,h \in R \right\}$$

$$= \left\{ \begin{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} \\ \begin{pmatrix} x \\ y \end{pmatrix} \\ \begin{pmatrix} z \\ w \end{pmatrix} \end{pmatrix} : x,y,z,w \in R \right\}.$$

To see the nature of our characterization, think of $v \in \mathcal{R}_d^{v_0,r}$ as $\nu(d)$ vectors $v_0, \ldots, v_{\nu(d)-1}$ in R^r. They correspond to monomials in some matrix tuple $\mathcal{X}$ applied to v_0, that is, to $V(\mathcal{X})_0 v_0, \ldots, V(\mathcal{X})_{\nu(d)-1} v_0$ and as a consequence certain dot products of them are equal, for example,

$$\mathcal{X}_1\mathcal{X}_2 v_0 \cdot \mathcal{X}_3\mathcal{X}_1 v_0 = \mathcal{X}_2 v \cdot \mathcal{X}_1^T\mathcal{X}_3\mathcal{X}_1 v_0$$

which in notation we soon formally introduce says that the component vectors of v satisfy

$$v_{\{1,2\}} \cdot v_{\{3,1\}} = v_{\{2\}} \cdot v_{\{-1,3,1\}}. \tag{3.1}$$

There are many similar relationships and the main result of this section is that this set of relationships precisely characterizes the closure of $\mathcal{R}_d^{v_0,r}$. To give details we need considerable notation.

3.1 Notation

A monomial $X_{j_1}X_{j_2}\ldots X_{j_\ell}$ is determined by a tuple $j := \{j_1, j_2, \ldots, j_\ell\}$ of integers $-k \le j_i \le k$, $j_i \neq 0$ with $1 \le i \le \ell$; we abbreviate the corresponding monomial by X^j. Here X_{-j_i} means $X_{j_i}^T$. Note, $X_{-(-j_i)} = X_{j_i}^{T^T} = X_{j_i}$. Denote all such tuples of ℓ integers by U^ℓ. The only idiosyncrasy with this rather intuitive notation is that we do not allow any j_i to be 0. For example, we exclude tuples like

$$j = \{2, 0, -8, 4\}$$

which naturally would correspond to the monomial $X_2 I X_8^T X_4 = X_2 X_8^T X_4 = X^{\{2,-8,4\}}$. Excluding 0 as an index avoids parameterizing the same monomial in two different ways. Indeed tuples in $U^1 \cup U^2 \cup \ldots U^\ell$ sit in one to one correspondence with monomials in $X_1, \ldots, X_n, X_1^T, \ldots X_n^T$ with degree bigger than 0 and no bigger than ℓ. To include the monomial I use the notation $X^0 = I$ and U^0 denotes 0. We denote by $\mathcal{U}^\ell$ all tuples

$$\mathcal{U}^\ell := U^0 \cup U^1 \cup \cdots \cup U^\ell$$

of length less than or equal to ℓ; together with U^0 which is 0.

If matrices $\mathcal{X}$ are substituted for X, then X^j becomes $\mathcal{X}^j$ and we denote the vector $\mathcal{X}^j v_0$ by

$$v_j := \mathcal{X}^j v_0. \tag{3.2}$$

This notation is illustrated by (3.1). Also it is easy to confuse with notation used in §2.1 and §5 where v_j has a subscript which is an integer rather than a tuple of integers.

Define **concatenation** of two sequences of integers j and w, denoted $j * w$, by

$$j * w := \{j_1, \ldots, j_\ell, w_1, \ldots, w_r\}$$

where $j := \{j_1, \ldots, j_\ell\}$ and $w := \{w_1, \ldots, w_r\}$. Define the **transpose** of the sequence j by $\widetilde{j} := \{j_\ell, \ldots, j_1\}$, and define $-j$ to be $\{-j_1, \ldots, -j_\ell\}$. Thus

$$(X^j)^T = (X_{j_1} X_{j_2} \ldots X_{j_\ell})^T = X_{-j_\ell} \ldots X_{-j_1} = X^{-\tilde{j}}.$$

The **key dot product relation** on the components of $v = V^d(\mathcal{X})v_0$ is

$$v_{j*w} \cdot v_s = v_w \cdot v_{(-\widetilde{j})*s} \tag{3.3}$$

which is true because

$$\begin{aligned} X^{j*w} v_0 \cdot X^s v_0 &= X^w v_0 \cdot [X^j]^T X^s v_0 \\ &= X^w v_0 \cdot X^{-j} X^s v_0 \end{aligned}$$

for any sequences $j \varepsilon\, \mathcal{U}^d, w \varepsilon\, \mathcal{U}^d, s \varepsilon\, \mathcal{U}^d$ for which $j * w \varepsilon\, \mathcal{U}^d$ and $(-\widetilde{j}) * s \varepsilon\, \mathcal{U}^d$.

3.2 Result on the Range

Proposition 3.1 *Fix $v_0 \in R^r$. Suppose we have a set $b\mathcal{S}^d$ of vectors v_j in R^r indexed by tuples*

$$j \in\ U^0 \cup\ U^1 \cup \cdots \cup\ U^d = \mathcal{U}^d$$

which satisfy

1. *the dot product relations (3.3)*
2. *the vectors in $b\mathcal{S}^{d-1} := \{v_j \text{ in } b\mathcal{S}^d : j \in \mathcal{U}^{d-1}\}$ are linearly independent.*

Then there are matrices $\mathcal{X}_1, \ldots, \mathcal{X}_n \in R^{r\times r}$, *such that*

$$v_j = \mathcal{X}^j v_0 = \mathcal{X}_{j_1}\mathcal{X}_{j_2}\ldots\mathcal{X}_{j_\ell} v_0$$

for all $j \in U^\ell$ *for* $1 \le \ell \le d$.

Proof: For each $1 \le t \le n$, define a $r \times r$ matrix $\mathcal{X}_t$ on $b\mathcal{S}^{d-1}$ and $\widehat{\mathcal{X}}_{-t}$ on $b\mathcal{S}^{d-1}$ by

$$\mathcal{X}_t v_j := v_{\{t\}*j} \qquad \text{and} \qquad \widehat{\mathcal{X}}_{-t} v_j := v_{(-\{t\})*j} \quad \text{for all } v_j \varepsilon b\mathcal{S}^{d-1}. \tag{3.4}$$

Thus $\mathcal{X}_t$ and $\widehat{\mathcal{X}}_{-t}$ are defined on the span $\mathcal{S}^{d-1}$ of $b\mathcal{S}^{d-1}$ and not defined on its orthogonal complement, $\mathcal{S}^{(d-1)^\perp}$ in R^r. We can think of fully defining $\mathcal{X}_t$ and $\widehat{\mathcal{X}}_{-t}$ on R^r in terms of matrices

$$\mathcal{X}_t = \begin{pmatrix} \mathcal{X}_{11} & \mathcal{X}_{12} \\ \mathcal{X}_{21} & \mathcal{X}_{22} \end{pmatrix} \qquad \widehat{\mathcal{X}}_{-t} = \begin{pmatrix} \widehat{\mathcal{X}}_{11} & \widehat{\mathcal{X}}_{12} \\ \widehat{\mathcal{X}}_{21} & \widehat{\mathcal{X}}_{22} \end{pmatrix}$$

partitioned with respect to the subspace $\mathcal{S}^{d-1}$ and $\mathcal{S}^{(d-1)\perp}$, where $\mathcal{X}_{11}, \mathcal{X}_{21}, \widehat{\mathcal{X}}_{11}\widehat{\mathcal{X}}_{21}$ are known and

$$\mathcal{X}_{12}, \mathcal{X}_{22}, \widehat{\mathcal{X}}_{12}, \widehat{\mathcal{X}}_{22} \tag{3.5}$$

are free to be determined.

We wish to choose the free blocks (3.5) to make $\mathcal{X}_t^T = \widehat{\mathcal{X}}_{-t}$, and now we show how this is done. Since both $\mathcal{X}_{11}$ and $\widehat{\mathcal{X}}_{11}$ are defined we must, verify the compatibility condition $\mathcal{X}_{11}^T = \widehat{\mathcal{X}}_{11}$. This follows from the "dot product" relations (3.3), namely, for any v_j and v_s in $b\mathcal{S}^{d-1}$, a basis for $\mathcal{S}^{d-1}$, we have

$$\begin{aligned} v_j \cdot \mathcal{X}_{11}^T v_s = \mathcal{X}_{11} v_j \cdot v_s &= \mathcal{X}_t v_j \cdot v_s = v_{\{t\}*j} \cdot v_s = v_j \cdot v_{(-\{t\})*s} \\ &= v_j \cdot \widehat{\mathcal{X}}_{-t} v_s. \end{aligned}$$

Thus $\mathcal{X}_{11}^T = [\widehat{\mathcal{X}}_{-t}]_{11} = \widehat{\mathcal{X}}_{11}$. From here on it is just a matter of picking undetermined blocks in the obvious way. Set

$$\mathcal{X}_{12} := \widehat{\mathcal{X}}_{21}^T \quad \widehat{\mathcal{X}}_{12} := \mathcal{X}_{21}^T$$

thereby determining them completely. The choice of $\mathcal{X}_{22}$ and $\widehat{\mathcal{X}}_{22}$ is completely open subject to the constraint that

$$\mathcal{X}_{22}^T = \widehat{\mathcal{X}}_{22},$$

so we make any such choice. This constructs $\mathcal{X}_t$ and $\widehat{\mathcal{X}}_{-t} = \mathcal{X}_t^T$.

Now we wish to show that with this choice of $\mathcal{X}$ we have

$$v_j = \mathcal{X}^j v_0 \quad \text{for } j \in \mathcal{U}^\delta \tag{3.6}$$

for $\delta = d$. The proof proceeds by induction on the degree δ of the monomial. For $\delta = 0$ we have the first component of v is indeed v_0, since $\mathcal{X}^0 = I$; that is, our notation is consistent. Suppose (3.6) holds for $\delta < d_1$ and consider $s = \{t\} * j$ where $t \neq 0, -n \le t \le n$ and length $j < \delta$. Then

$$\mathcal{X}^s v_0 = \mathcal{X}_t \mathcal{X}^j v_0 = \mathcal{X}_t v_j.$$

This together with (3.4) says $\mathcal{X}^s v_0 = v_{\{t\}*j} = v_s$ as is required by (3.6) for length $s = d_1$. •

Theorem 3.2 *For sufficiently large r, the subset*

$$\mathcal{R}_d^r := \text{closure}\{\mathcal{R}_d^{v_0,r} : v_0 \in R^r\}$$

of $R^{r\nu(d)}$, equals the set

$$\widehat{\mathcal{R}}_d^r := \left\{ \begin{pmatrix} v_0 \\ v_{\{1\}} \\ \vdots \\ v_{\{-n,\ldots,-n\}} \end{pmatrix} \varepsilon R^{r\nu(d)} : \begin{array}{l} \textit{the entries } v_j \textit{ in } R^r \textit{ satisfy the} \\ \textit{dot product constraint (3.3)} \end{array} \right\}.$$

Proof (the beginning): Pick any v in $\widehat{\mathcal{R}}_d^r$. Let v_0 denote the first entry of v. If v_0 is not 0 and if the entries v_j for $j \in \mathcal{U}^{d-1}$ are linearly independent, then by Proposition 3.1 we have $v \in \mathcal{R}_d^r$. If v_0 is 0 or linear dependence occurs, then perturb the v_j slightly while maintaining the dot product conditions (3.3) to obtain a linearly independent set. The next section proves that such perturbations are possible. •

3.2.1 Perturbations Overcome Linear Dependence

Lemma 3.3 *Suppose that v_j for $j \in \mathcal{U}^d$ are vectors in R^r which satisfy the dot product relations (3.3).*

(a) If ϕ_j, for $j \in \mathcal{U}^d$ in R^r are linearly independent, satisfy the relations (3.3), and if v_i is orthogonal to ϕ_j for each $i, j \in \mathcal{U}^d$, then the vectors

$$\tilde{v}_j = v_j + \phi_j \quad \textit{for} \quad j \in \mathcal{U}^d$$

are linearly independent and satisfy the relations (3.3).

(b) If there exist linearly independent ψ_j in R^q for $j \in \mathcal{U}^d$ satisfying the relations (3.3), and if $r \geq \nu(d)+q$, then there exist arbitrarily small perturbations ϕ_j of v_j in R^r, that is, $\tilde{v}_j = v_j + \phi_j$ which are linearly independent and satisfy (3.3).

Proof: That (3.3) holds for the $\tilde{v}_j$ follows from orthogonality in (a) which implies

$$\tilde{v}_i \cdot \tilde{v}_j = v_i \cdot v_j + \phi_i \cdot \phi_j$$

and from the fact that both the set of v_j and the set of ϕ_j satisfy (3.3). To see the linear independence of the $\tilde{v}_j$'s consider a linear combination

$$0 = \sum_{j \in \mathcal{U}^d} \alpha_j \tilde{v}_j = \sum_{j \in \mathcal{U}^d} \alpha_j v_j + \sum_{j \in \mathcal{U}^d} \alpha_j \phi_j$$

which, since all v_i are orthogonal to all, ϕ_j implies

$$\sum_{j \in \mathcal{U}^d} \alpha_j \phi_j = 0.$$

By linear independence of the ϕ_j, the α_j are all 0. This proves part (a).

To prove part (b) note that the orthogonal complement ρ of the span of v_j for $j \in \mathcal{U}^d$ has dimension greater than or equal to q. Thus we can embed R^q, which contains ϕ_j for $j \in \mathcal{U}^d$, isometrically into ρ, and we denote the image of these vectors under embedding by $\tilde{\phi}_j$ for $j \in \mathcal{U}^d$. Pick any ε and set

$$\psi_j := v_j + \varepsilon\tilde{\phi}_j.$$

The vectors v_j and $\varepsilon\tilde{\phi}_j$ satisfy the hypothesis of part (a) Lemma 3.3. The conclusion of part (a) yields the conclusion we are trying to prove for part (b). •

The next step is to prove

Lemma 3.4 . *There exists a set $\{\phi_j$ for $j \in \mathcal{U}^d\}$ of linearly independent vectors in R^q for some (large) q which satisfies (3.3).*

Proof: Our approach to constructing the ϕ_j's is to show that a vector consisting of linearly independent ϕ_j's exists in $\mathcal{R}_d^{v_0,q}$. Since the entries of any vector in $\mathcal{R}_d^{v_0,q}$ satisfy the dot product relations (3.3) this accomplishes our goal. Fix v_0. Now we assume any vector in $\mathcal{R}_d^{v_0,q}$ has linearly dependent entries. Equivalently, for any $q \times q$ matrices $\mathcal{X} = \mathcal{X}_1, \ldots, \mathcal{X}_n$, we have real numbers $\lambda_j(\mathcal{X}, v_0)$ for $j \in \mathcal{U}^d$ satisfying

$$\sum_{j \in \mathcal{U}^d} \lambda_j(\mathcal{X}, v_0) V^d(\mathcal{X})_j = 0 \tag{3.7}$$

In [CHSY preprt] Corollary 3.3 says[1] that if (3.7) holds for all large enough $\mathcal{X}$ and v_0, then there exist real numbers α_j for $j \in \mathcal{U}^d$ such that the non-commutative polynomial

$$\sum_{j \in \mathcal{U}^d} \alpha_j V^d(X)_j \tag{3.8}$$

in indeterminates X is identically 0. The point here is that the α_j do not depend on X. The $V^d(X)_j$ for $j \in \mathcal{U}^d$ are distinct monomials, hence linearly independent. This contradicts (3.8).

Proof of Theorem 3.2 (the finale): One finishes the proof of Theorem 3.2 simply by putting together Lemma 3.3 and Lemma 3.4. •

4 Making Partially Defined Matrices Positive

4.1 Matrix Positivising Problems

The SoS problem when reformulated as in equation (2.6), (3.7) and the subsequent comment leads us to a special case of the following problem.

Matrix positivising problem, $M^0, \mathcal{B}$: *Given a subspace $\mathcal{B} \subset SR^{p\times p}$ and given $M^0 \varepsilon SR^{p\times p}$. Is there a $B\varepsilon\mathcal{B}$ such that*

$$M = M^0 + B$$

is positive semidefinite?

[1] The proof in [CHSY preprt] goes through quite a few steps, each reducing the dependence of the λ_j on $\mathcal{X}$ and v_0.

4.2 The Dual Problem

Now we describe the dual problem which is in fact equivalent to this problem. Our interest stems from the fact that the matrix positivity condition at the center of our interest is equivalent to the positivity condition at the core of the dual problem. A good reference on positivity in matrix situations is [P86].

Suppose $\mathcal{B}$ is a subspace of the symmetric matrices $SR^{p\times p}$. We shall use the inner product $< \cdot, \cdot >$ on $SR^{p\times p}$, defined by

$$< A, B > = \text{tr } AB. \tag{4.1}$$

Let $\mathcal{B}^{\perp}$ denote the orthogonal complement

$$\mathcal{B}^{\perp} := \{A \in SR^{p\times p} : \text{tr } AB = 0 \text{ all } B\varepsilon\mathcal{B}\}$$

of $\mathcal{B}$ in $SR^{p\times p}$. Clearly $\mathcal{B}^{\perp}$ is a linear subspace.

If the answer to the matrix positivising problem is yes, that is, $M = M^0 + B^0 \geq 0$ exists, the linear functional ℓ_{M^0} defined on $\mathcal{B}^{\perp}$ by

$$\ell_{M^0}(A) = \text{tr } M^0 A = \text{tr}(M^0 + B)A$$

(for any $B\varepsilon\mathcal{B}$) is nonnegative for each $A\varepsilon\mathcal{B}^{\perp +}$, the set of positive semidefinite matrices which lie in $\mathcal{B}^{\perp}$. This is because

$$\ell_{M^0}(A) = \text{tr}(M^0 + B^0)A \geq 0$$

since A is positive semidefinite. The converse is also true, which gives:

Lemma 4.1 *Suppose that $\mathcal{B}^{\perp +}$ contains an invertible matrix. The matrix positivising problem $M^0, \mathcal{B}$ has a positive solution if and only if the linear functional ℓ_{M^0} is nonnegative on $\mathcal{B}^{\perp +}$.*

Proof: One side of the result is proved, so we turn to proving the side which starts with the linear functional ℓ_{M^0}. We invoke a variation on the Hahn-Banach Theorem, often called the Krein Extension Theorem, see Ch. 10.4 Exercise 22 [R64], which says when positive linear functionals on a subspace extend to positive linear functionals on the whole space. In our situation we are assuming that the linear functional ℓ_{M_0} takes positive values on $\mathcal{B}^{\perp}$ intersect the cone of PSD matrices. The PSD matrices satisfy $P \in$ PSD and $-P \in$PSD implies $P = 0$, one hypothesis of the Krein extension Theorem. The other key hypothesis is that if W is in $SR^{p\times p}$, then there is a $\mathcal{A} \in \mathcal{B}^{\perp +}$ which dominates W, that is $\mathcal{A} - W \geq 0$. The existence of an invertible matrix $\tilde{\mathcal{A}}$ in $\mathcal{B}^{\perp +}$ implies this, since any W is dominated by some scalar multiple of $\tilde{\mathcal{A}}$. Thus we have verified the hypothesis guaranteeing that the linear functional ℓ_{M^0} extends to a linear functional $\tilde{\ell}$ which takes a nonnegative value on any nonnegative matrix.

We can represent $\tilde{\ell}$ as

$$\tilde{\ell}(A) = \text{tr } MA \quad \text{for all} \quad A \in R^{p\times p}$$

using a matrix $M \in R^{p\times p}$. The positivity of $\tilde{\ell}$ implies that M is a PSD matrix. Also for any $A \in \mathcal{B}^{\perp}$,

$$0 = \tilde{\ell}(A) - \ell_{M^0}(A) = \text{tr}([M - M^0]A).$$

Thus $B := M - M^0 \in \mathcal{B}^{\perp\perp} = \mathcal{B}$. Therefore we have produced a PSD matrix

$$M = M^0 + B$$

with $B \in \mathcal{B}$ as required.

This was stated (with slightly less generality) in Theorem 2.1 [AHMR88]. •

Ultimately proving Theorem 1.1 depends on Lemma 4.1 applied to $M^0_Q, \mathcal{B}_{V^d}$ in §2. Thus we need to compute $(\mathcal{B}_{V^d})^{\perp+}$ and the next section does something a little more general than that.

5 Ranges and the Dual Problem

In §2.2 we saw that the set $\mathcal{B}_V$, characterizing non-uniqueness of the Gram representation, has the form

$$\mathcal{B}^{\mathcal{P}} := \operatorname*{span}_{\substack{\{(i,j),(k,\ell)\} \in P \\ P \in \mathcal{P}}} \mathcal{F}\left(\{(i,j),(k,\ell)\}\right)$$

where $\mathcal{P}$ is a collection of sets P of pair of pairs of finite integers between 0 and $p-1$; no two pairs in a element of P can be the same. Here we recall the notation

$$\mathcal{F}(\{(i,j),(k,\ell)\}) = \{B : B_{ij} = -B_{k\ell}, B_{ji} = -B_{\ell k}, \text{ otherwise } B_{us} = 0\}.$$

While in §2.2 each P was associated to a monomial, in this section we study an arbitrary collection $\mathcal{P}$ of sets P of pairs of pairs subject to the condition that any two sets P^1, P^2 in $\mathcal{P}$ are disjoint. It is this that we formally call a **pair of pair collection**. To $\mathcal{P}$ we associate $\mathcal{T}(\mathcal{P})$ the set of all pairs of pairs in $\mathcal{P}$, that is,

$$\mathcal{T}(\mathcal{P}) := \{\{(i,j),(k,\ell)\} \in P : P \in \mathcal{P}\}.$$

It is useful because

$$\mathcal{B}^{\mathcal{P}} = \operatorname*{span}_{\{(i,j),(k,\ell)\} \in \mathcal{T}(\mathcal{P})} \mathcal{F}(\{(i,j),(k,\ell)\}).$$

$\mathcal{B}^{\mathcal{P}}$ is a subset of the symmetric $p \times p$ matrices, and in this section we characterize its orthogonal complement $(\mathcal{B}^{\mathcal{P}})^{\perp}$ inside $SR^{p\times p}$ with respect to the *trace* dot product $\langle\ ,\ \rangle$. This characterization of the orthogonal complement will be done in terms of sets of vectors in R^{pr} of the form

$$\mathcal{R}^{P,r} := \left\{ \begin{pmatrix} v_0 \\ v_1 \\ \vdots \\ v_{p-1} \end{pmatrix} \varepsilon R^{pr} : \text{each } \ v_j \in R^r \text{ and } v_i \cdot v_j = v_k \cdot v_\ell \text{ for } \{(i,j),(k,\ell)\} \in P \right\}$$

where P denotes a set of pairs of pairs of integers between 0 and $p-1$. The motivation comes from recalling that such sets occurred in our key Theorem 3.2 which characterized the range of $V^d(\mathcal{X})$ applied to $v_0 \in R^r$. In fact what was found there is that $\mathcal{R}^r_d$ is

$$\mathcal{R}^r_d = \bigcap \{\mathcal{R}^{P_\mu,r} : \mu \text{ a monomial in } X, X^T \text{ of degree} \le d\}$$

though in Theorem 3.2 the notation was different. Here P_μ stands for the particular set of pairs of pairs defined in §2.1. This section lays out a strong link between the range of a list of monomials V characterized in §3 and the $\perp$ of the null space of representations based on V.

We shall consider a matrix A_v in $SR^{p\times p}$ associated to v in $\mathcal{R}^{P,r}$ by specifying that its $k\ell^{th}$ entry is

$$[A_v]_{k\ell} = v_k \cdot v_\ell. \tag{5.1}$$

Note that A_v is a PSD matrix. The main result of this section is

Lemma 5.1 *Suppose $\mathcal{P}$ is a pair of pair collection. Then*

$$\mathcal{B}^{\mathcal{P}} = \left[\bigcap_{P\in\mathcal{P}} \{A_v : v \in \mathcal{R}^{P,r}\}\right]^{\perp} \subset SR^{p\times p}$$

If $r \geq p$, then

$$(\mathcal{B}^{\mathcal{P}})^{\perp+} = \bigcap_{P\in\mathcal{P}} \{A_v : v \in \mathcal{R}^{P,r}\} \subset SR^{p\times p}. \tag{5.2}$$

Moreover if $r \geq p$, the set $(\mathcal{B}^{\mathcal{P}})^{\perp+}$ contains no invertible matrix if and only if there is a nonzero vector $(x_0, x_1, \ldots, x_{p-1}) \in R^p$ satisfying

$$\sum_{j=0}^{p-1} v_j x_j = 0 \tag{5.3}$$

for all $v \in \bigcap_{P\in\mathcal{P}} \mathcal{R}^{P,r}$.

5.1 Proof of Lemma 5.1

Some notation helps. Define

$$\begin{aligned} \mathcal{A}_P &:= \{A_v : v \in \mathcal{R}^{P,r}\} \subset SR^{p\times p} \\ \mathcal{A}^{\mathcal{P}} &:= \bigcap_{P\in\mathcal{P}} \mathcal{A}_P. \end{aligned} \tag{5.4}$$

In this notation the lemma says

$$(\mathcal{A}^{\mathcal{P}})^{\perp} = \mathcal{B}^{\mathcal{P}} \tag{5.5}$$

$$\mathcal{A}^{\mathcal{P}} = (\mathcal{B}^{\mathcal{P}})^{\perp} \cap (\text{the PSD matrices}) =: (\mathcal{B}^{\mathcal{P}})^{\perp+}.$$

Thus $\mathcal{A}^{\mathcal{P}}$, for $r \geq p$, is the intersection of the PSD matrices with a subspace of matrices.

We shall use

$$(\mathcal{A}^{\mathcal{P}})^{\perp} = \text{span}\left\{\mathbf{a}(\{(i,j),(k,\ell)\})^{\perp} : \quad \{(i,j),(k,\ell)\} \in \mathcal{T}(\mathcal{P})\right\} \tag{5.6}$$

where **a** is defined by

$$\mathbf{a}(\{(i,j),(k,\ell)\}) := \{A_v : v_i \cdot v_j = v_k \cdot v_\ell \quad v \in \mathbb{R}^{pr}\}.$$

The first step is to compute $\mathbf{a}(\{(i,j),(k,\ell)\})^\perp \subset SR^{p\times p}$, and with this in mind note that the symmetric matrix B is in it if and only if

$$\text{tr } BA_v = 0 \quad \text{for all } v \in \mathbb{R}^{pr} \text{ with } v_i \cdot v_j = v_k \cdot v_\ell.$$

That is,

$$\begin{aligned} 0 = \text{tr}BA_v &= \sum_{s,t=0}^{p-1} B_{st}[A_v]_{st} = \sum_{s,t=0}^{p-1} B_{st} v_s \cdot v_t \\ &= 2[B_{ij} + B_{k\ell}]v_i \cdot v_j + \sum_{\substack{s,t=0 \\ (s,t)\neq(i,j)or(j,i)or \\ (k,\ell)\text{or}(\ell,k)}}^{p-1} B_{st} v_s \cdot v_t. \end{aligned} \tag{5.7}$$

Since $r \geq p$ we can choose a rich enough family of v satisfying $v_i \cdot v_j = v_k \cdot v_\ell$ to conclude that

$$\begin{aligned} &B_{ij} + B_{k\ell} = 0 \\ &B_{st} = 0 \text{ if } (s,t) \neq (i,j) \text{ or } (j,i) \text{ or } (k,\ell) \text{ or } (\ell,k). \end{aligned}$$

Clearly these conditions are sufficient as well as necessary. Thus we have proved

$$\mathbf{a}\{(i,j),(k,\ell)\}^\perp = \mathcal{F}\{(i,j),(k,\ell)\}.$$

Which because of (5.6) implies

$$(\mathcal{A}^{\mathcal{P}})^\perp = \mathcal{B}^{\mathcal{P}}$$

as required.

Next we compute $(\mathcal{B}^{\mathcal{P}})^{\perp+}$. We need a lemma

Lemma 5.2 *If A is a PSD matrix in $SR^{p\times p}$ with the rank q, then it can be written as $A_v \in SR^{p\times p}$ for some $v = (v_0, v_1, \ldots, v_{p-1})^T$ with $v_\ell \varepsilon R^q$ for $\ell = 0, \ldots, p-1$. We call this the q-***vector representation*** *of A.*

Proof. Start with A. By the Cholesky decomposition , LDL^T, any PSD matrix A in $SR^{p\times p}$ has a representation

$$A = \sum_{j=1}^{q} w^j \ {w^j}^T \tag{5.8}$$

where $w^j \varepsilon R^p$ is a column vector with entries $w_0^j, w_1^j, \ldots, w_{p-1}^j \in R$ and with and $q \leq p$. Note q can be taken to be the rank of A. Take v to be the vector

$$v^T := (w_0^1, \ldots, w_0^q, w_1^1, \ldots, w_1^q, \ldots, w_{p-1}^1, \ldots, w_{p-1}^q)$$

which we use to define subvectors v_j by

$$v^T =: (v_0, v_2, \ldots, v_{p-1}),$$

that is, $v_k := (w_k^1, \ldots, w_k^q)$.

Now we check that A is A_v by showing that the $k\ell^{th}$ entry of A is $[A]_{k\ell} = v_k \cdot v_\ell$. To prove this write (5.8) in long form as

$$A = \begin{pmatrix} w_0^1 \\ w_1^1 \\ \vdots \\ w_p^1 \end{pmatrix} (w_0^1, \ldots, w_p^1) + \cdots + \begin{pmatrix} w_0^q \\ w_1^q \\ \vdots \\ w_p^q \end{pmatrix} (w_0^q, \ldots, w_p^q).$$

to see that

$$[A]_{k\ell} = w_k^1 w_\ell^1 + w_k^2 w_\ell^2 + \cdots + w_k^q w_\ell^q = v_k \cdot v_\ell. \quad \bullet$$

To complete the proof of Lemma 5.1 consider $A \in (\mathcal{B}^{\mathcal{P}})^{\perp +} \subset RF^{p \times p}$ and note that A has a q-vector representation as $A = A_v$ for some v in R^{pq}. By the definition of $\mathcal{B}^{\mathcal{P}}$ equation (5.7) implies the components $v_j \in R^q$ of v satisfy $v_i \cdot v_j = v_k \cdot v_\ell$ for all $\{(i,j),(k,\ell)\} \in \mathcal{T}(\mathcal{P})$. Thus, for $q \leq r$, we have $A_v \in \mathcal{A}_{\mathcal{P}}$. Since the rank of A is less than or equal to $p \leq r$, we see that A does have a r-representation, so we have proved the second formula in the lemma.

It remains only to prove (5.3). Begin with the observation that $(\mathcal{B}^{\mathcal{P}})^{\perp +}$ is a cone, so if A^1 and A^2 belong to it, then so does $A^1 + A^2$. Since PSD implies

$$\text{Null}(A^1 + A^2) = \text{Null}(A^1) \cap \text{Null}(A^2),$$

we have by adding more $A^j \subseteq (\mathcal{B}^{\mathcal{P}})^{\perp +}$, that if $(\mathcal{B}^{\mathcal{P}})^{\perp +}$ does not contain an invertible matrix, then there is a non trivial space $\mathcal{N} \subset R^p$ satisfying

$$\mathcal{N} \subset \bigcap_{A \in (\mathcal{B}^{\mathcal{P}})^{\perp +}} \text{Null}(A).$$

Pick $x \in R^p$ satisfying $x \in \text{Null}(A_v)$ and observe

$$0 = A_v x \cdot x = \sum_{i=0}^{p-1} \left[\sum_{j=0}^{p-1} v^i \cdot v^j x_j \right] x_i$$

$$0 = \left\| \sum_{j=0}^{p-1} v^j x_j \right\|^2$$

Thus (5.3) is true. $\bullet$

Note that if r is too small, then we have not proved (and in fact it is not always true) that $\{A_v : v \varepsilon \mathcal{R}^{\mathcal{P},r}\}$ is a subspace intersect PSD.

6 Proof of the Main Theorem

This brief section stitches together the earlier ones to prove Theorem 1.1.

We begin with a non-commutative symmetric polynomial $\mathcal{Q}$. Assume that $\mathcal{Q}$ is matrix positive. This forces its degree to be an even number which we denote $2d$. We begin our proof by representing $\mathcal{Q}$ as in §2 with a matrix we denote $M_{\mathcal{Q}}$ in $R^{p\times p}$ where $p = \nu(d)$ and with V^d, the vector which lists all monomials of degree $\leq d$.

Now we will show that

$$\ell_{M_Q}(A) := \text{tr } M_Q A \geq 0 \quad \text{for all} \quad A \in \mathcal{A}_{V^d,q} \tag{6.1}$$

where

$$\mathcal{A}_{V^d,q} := \{A_{V^d(\mathcal{X})v_0} : \mathcal{X} \text{ is a tuple of } q\times q \text{ matrices and } v_0 \in R^q\}.$$

Indeed, fix $B \in R^{p\times p}$. For $\mathcal{X}$ with entries $\mathcal{X}_t \in R^{q\times q}$ for $|t| = 1, \cdots, n$ and $v_0 \varepsilon R^q$, we have $V^d(\mathcal{X})v_0 \varepsilon R^{qp}$. Then

$$\begin{aligned}\ell_B(A) = \text{tr } BA_{V^d(\mathcal{X})v_0} &= \sum_{i,j=0}^{p-1} B_{ij} v_i \cdot v_j \\ &= v_0^T V^d(\mathcal{X})^T B V^d(\mathcal{X}) v_0.\end{aligned}$$

Consequently,

$$\ell_{M_{\mathcal{Q}}}(A_{V^d(\mathcal{X})v_0}) = \text{tr } M_{\mathcal{Q}} A_{V^d(\mathcal{X})v_0} = v_0^T V^d(\mathcal{X})^T M_{\mathcal{Q}} V^d(\mathcal{X}) v_0 = v_0^T\, \mathcal{Q}(\mathcal{X}) v_0 \geq 0,$$

since Q is matrix positive.

Define $\mathcal{B}_{V^d,q}$ to be

$$\mathcal{B}_{V^d,q} := \{B \in SR^{p\times p} : V^d(\mathcal{X})^T B V(\mathcal{X}) = 0 \text{ for all } \mathcal{X} \in R^{q\times q}\}.$$

The first part of Lemma 5.1 says that

$$\mathcal{A}_{V^d,q} \subset (\mathcal{B}_{V^d,q})^{\perp+} \subset SR^{p\times p}$$

for q sufficiently large. Clearly, for large enough q, we have

$$\mathcal{B}_{V^d,q} = \mathcal{B}_{V^d}. \tag{6.2}$$

Thus

$$\mathcal{A}_{V^d,q} \subset (\mathcal{B}_{V^d})^{\perp+} \subset SR^{p\times p}. \tag{6.3}$$

Assertions (6.1) and (6.3) together with Lemma 4.1 imply

Proposition 6.1 *Matrix positivity of $\mathcal{Q}$ implies $\mathcal{Q}$ is a sum of squares if for some q the closure of $\mathcal{A}_{V^d,q}$ is all of $(\mathcal{B}_{V^d})^{\perp+} \subset SR^{p\times p}$ and if $(\mathcal{B}_{V^d})^{\perp+}$ contains an invertible matrix.*

Proof: We saw immediately above that, $\ell_{M_0} \geq 0$ on $\mathcal{A}_{V^d,q}$, so if the closure of $\mathcal{A}_{V^d,q}$ is all of $(\mathcal{B}_{V^d})^{\perp+}$, then this implies that $\ell_{M_0} \geq 0$ on $(\mathcal{B}_{V^d})^{\perp+}$. Lemma 4.1 implies there is a PSD matrix M of the form (2.6), that is,

$$\mathcal{Q}(X) = V^d(X)^T M V^d(X).$$

The Cholesky decomposition of M, namely $M = L^T D L$ with D diagonal and D nonnegative gives a SoS decomposition of $\mathcal{Q}$. •

To finish proving Theorem 1.1 we apply §5 and §3. Theorem 3.2 says that if q is sufficiently large then the closure of $\mathcal{A}_{V^d,q}$ has the form $\mathcal{A}^{\mathcal{P}}$ for a certain pair of pair collection $\mathcal{P}$. $\mathcal{P}$ is given explicitly by the dot product relations (3.3) for V^d. Thus the second part of Lemma 5.1 and (6.2) combine with (6.3) to say that closure $\mathcal{A}_{V^d,q} = (\mathcal{B}_{V^d})^{\perp+}$.

Moreover, to see that $(\mathcal{B}_{V^d})^{\perp+}$ contains an invertible matrix we apply (5.3). Again we use that for large enough q, the set equality $\mathcal{B}_{V^d} = \mathcal{B}_{V^d,q} = \mathcal{B}^{\mathcal{P}}$ holds for the pair of pair collection $\mathcal{P}$. The only way $(\mathcal{B}_{V^d})^{\perp+}$ could fail to contain an invertible matrix is if every set of vectors $\{v_0, v_1, \ldots, v_{\nu(d)-1}\}$ in R^q which satisfies

$$v_i \cdot v_j = v_k \cdot v_\ell \text{ for all } \{(i,j),(k,\ell)\} \in \mathcal{P} \tag{6.4}$$

is linearly dependent. But Lemma 3.4 guarantees for large enough q that there does exist a linearly independent set of vectors satisfying (6.4).

All hypotheses of Proposition 6.1 have been validated, thus we may conclude that Q is a SoS. •

7 Motivation

This work was motivated by the question, what " non-commutative inequality" features should be put into a computer algebra package. The development of our package NCAlgebra, which gives Mathematica capability in a general non-commutative algebra, is motivated largely by linear systems engineering problems. Within the last 10 years matrix inequalities have come to dominate this subject; so finding helpful computer algebra techniques while unexplored is important. The viewpoint we take is that of an engineering researcher who tries to find formulas solving systems problems with the goal of having his formulas included in a workstation (eg. Matlab) tool box. Such formulas typically involve non-commuting variables and the user of the tool box (of which there are often thousands) substitutes matrices for the non-commutative variables, quite possibly without knowing that is what he is doing. The matrices vary tremendously in size depending on which physical system is being designed. The point is that formulas in the tool box have non-commuting variables and any property required of them must hold when matrices of any size are substituted in. This is what motivated our definition of matrix positive polynomial.

Once convinced of the need to study matrix positive polynomials there is the disappointment of observing that checking the condition even approximately by brute force faces the curse of dimensionality and so is impractical. For example, to test a polynomial in two variables X, Y with 5×5 matrices, where we use a 40 point grid on each matrix entry yields$(40)^{2\times 25} \approx 10^{80}$

different matrices on which we need to evaluate the polynomial. This would take on the scale of 10^{60} years.

On the other hand computing if a commutative polynomial with a modest number of terms is a SoS can be done very quickly by a mixture of algebraic techniques due to Reznik, Powers-Wormann, see[PW98], and semidefinite programming, see Parrilo [P00]. Their techniques apply directly to non-commutative polynomials, so the main theorem of this paper tells us that matrix positivity is a property which can be checked in practice.

References

[AHMR88] J. Agler, J. W. Helton, S. McCullough and L. Rodman, *Positive semidefinite matrices with a given sparsity pattern*, J. Linear Algebra Appl., 107 (1988), p. 101–149.

[CHSY00] J. Camino, J.W. Helton, R.E. Skelton, and Jeiping Ye, *A Symbolic Algorithm For Determining Convexity of A Matrix Function: How To Get Schur Complements Out Of Your Life*, Conf. on Decision and Control, Sydney 2000.

[CHSYpreprt] J. Camino, J.W. Helton, R.E. Skelton, and Jeiping Ye, www.math.ucsd.edu/∼helton.

[DApreprt] J. D. D'Angelo. *Proper holomorphic mappings, positivity conditions, and isometric embedding.* preprint.

[Mpreprt] S. McCullough. *Factorization of operator valued polynomials in several non-commuting variables.* preprint, January 2000.

[P00] P. Parrilo *Structured Semidefinite Programs and Semialgebaric Geometry Methods in Robustness and Optimization.* Cal. Inst. of Tech. Phd. Thesis, May 2000.

[P86] V. Paulsen, *Completely bounded maps and dilations*, Pitman Research Notes in Math., 1986.

[PW98] V. Powers and T. Wormann, *An algorithm for sums of squares of real polynomials*, Journal of Pure and Applied Algebra 127, 1998, p. 99-104.

[R00] B. Reznick, *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.

[R64] H. Royden, *Real Analysis*, Third printing McMillan, 1964.

Contents