RUTCOR
RESEARCH
REPORT

EXTENSION OF THE SEMIDEFINITE CHARACTERIZATION OF SUM OF SQUARES FUNCTIONAL SYSTEMS TO ALGEBRAIC STRUCTURES

Dávid Papp[a] Ricardo A. Collado[b]
Farid Alizadeh[c]

RRR 27–2009, DECEMBER 3, 2009

RUTCOR
Rutgers Center for
Operations Research
Rutgers University
640 Bartholomew Road
Piscataway, New Jersey
08854-8003
Telephone: 732-445-3804
Telefax: 732-445-5472
Email: rrr@rutcor.rutgers.edu
http://rutcor.rutgers.edu/∼rrr

[a]RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway, NJ 08854-8003; (dpapp@rutcor.rutgers.edu)
[b]RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway, NJ 08854-8003; (rsoto@rutcor.rutgers.edu)
[c]Rutgers Business School and RUTCOR, 640 Bartholomew Road, Piscataway, NJ 08854-8003; (allizadeh@rutcor.rutgers.edu)

Rutcor Research Report
RRR 27–2009, December 3, 2009

Extension of the semidefinite characterization of sum of squares functional systems to algebraic structures

Dávid Papp Ricardo A. Collado Farid Alizadeh

Abstract. We extend Nesterov's semidefinite programming (SDP) characterization of the cone of functions that can be expressed as sums of squares (SOS) of functions in finite dimensional linear functional spaces. Our extension is to algebraic systems that are endowed with a binary operation which map two elements of a finite dimensional vector space to another vector space; the binary operation must follow the distributive laws. We derive a number of previously known SOS characterizations as a special case of our framework. In addition to Nesterov's result [10] for finite dimensional linear functional spaces, we show that the cone of positive semidefinite univariate polynomials with symmetric matrices as coefficients, SOS polynomials with coefficients from Euclidean Jordan algebras (first studied by Kojima and Muramatsu [6]), and numerous other problems involving vector-valued functions not previously considered can be expressed in our framework. Some potential applications in geometric design problems with constraints on curvature of space curves, and in multivariate approximation theory problems with convexity as constraint are briefly discussed.

Acknowledgements: Research of DP and FA supported in part by National Science Foundation Grant number CMMI-0935305.

1 Introduction

Optimization models involving *sum of squares* (SOS) functional systems have raised significant interest, particularly in connection with *polynomial programming* (POP), that is, optimization models involving polynomial minimization, or equivalently, positive polynomial constraints. For example, POP was found to be applicable in combinatorial optimization problems [15]. A key underlying result in all POP models is that a univariate polynomial is nonnegative everywhere if and only if it is sum of perfect squares, and for multivariate polynomials, sums of squared polynomials form a proper subset of nonnegative polynomials. Nested hierarchies of SOS models, such as those proposed by Lasserre [8] and Parrilo [12] define nested sequences of SOS polynomial cones to approximate the cone of multivariate polynomials.

Another important family of applications of more general SOS functional systems (not only polynomials) is *shape constrained estimation*, that is, function estimation involving constraints, such as monotonicity or convexity, on the shape of the estimator [11]. Again, the key steps in expressing these estimation problems as tractable optimization problems are (i) the translation of the shape constraints to the nonnegativity of some linear transform of the estimator, and (ii) the replacement of the nonnegativity constraint by a constraint that the function belongs to some SOS functional system.

Nonnegativity is an intractable constraint in even the most fundamental functional systems. For example, recognizing nonnegative (multivariate) polynomials of degree four is known to be NP-hard, by a very simple reduction from the PARTITION problem. On the other hand, the constraint that a function belong to a specific SOS functional system can always be cast as a semidefinite programming constraint, using a result of Nesterov [10], which says that for every finite set of linearly independent, real valued functions $\mathcal{F} = \{f_1, \dots, f_m\}$, the set

$$\left\{ \sum_{i=1}^{k} f_i^2 \;\middle|\; f_i \in \operatorname{span}(\mathcal{F}) \right\}$$

is a linear image of the cone of $m \times m$ positive semidefinite matrices, and this mapping can be easily constructed from the set $\mathcal{F}$. This result makes it possible to cast optimization problems over SOS functional systems (including the common SOS restrictions of POP problems) as tractable semidefinite optimization problems.

Our paper is primarily motivated by shape constrained optimization problems. As the following examples show, not all natural shape constraints can be translated to nonnegativity of some linear transform of the shape constrained function.

1. (Convexity of a multivariate function.) A twice continuously differentiable real valued function f is convex over $S \subset \mathbb{R}^n$ if and only if its Hessian H is positive semidefinite over S, which we denote by $H(x) \succcurlyeq 0 \;\forall\, x \in S$. Magnani et. al [9] consider this problem in the case when f is a polynomial, and suggest the following solution: $H(x) \succcurlyeq 0$ for every $x \in S$ if and only if $y^\top H(x) y \geq 0$ for every $x \in S$ and $y \in \mathbb{R}^n$. This solution is not completely satisfying for two reasons. First, it calls for doubling the number

of variables in the problem, which is a serious complication, as the complexity of the resulting semidefinite program increases highly. Second, if f is not a polynomial, and H is not a polynomial matrix, then H is still a linear transform of f, however, the function $(x, y) \to y^\top H(x)y$ belongs to an entirely different functional system, in which it is generally difficult to establish a connection between nonnegativity and sum of squares.

2. (Functions with bounded curvature.) Consider a problem in which we are trying to find a twice differentiable curve given by its parametric representation $x(t) \in \mathbb{R}^n$, $t \in [0, T]$, under the constraint that the curvature of the curve must be bounded above by some constant $C \geq 0$. If t is the arc-length parameter of the curve, this constraint can be written as

$$\|x''(t)\|_2 \leq C \quad \forall\, t \in (0, T),$$

where x'' is the component-wise second derivative of the vector function x, and $\|\cdot\|_2$ denotes the Euclidean norm [14, Section 1-4]. Equivalently, the constraint can be written as

$$\left(\begin{smallmatrix} C \\ x''(t) \end{smallmatrix}\right) \in \mathcal{Q}_{n+1} \quad \forall\, t \in (0, T),$$

where $\mathcal{Q}_{n+1}$ is the $(n + 1)$-dimensional second order cone, or Lorentz cone [1].

Both examples suggest that we should consider constraints of the form $f(x) \in K \ \ \forall\, x \in S$, where f is a (perhaps multivariate) vector valued function, and K is some convex cone. As in the motivating one-dimensional case, this will generally be an intractable constraint, but we can try to find a tractable approximation for it, in the form of $f(x)$ being "sum of squares" with respect to some multiplication of vector-valued functions. This is particularly appealing when K is a symmetric cone, as symmetric cones are cones of squares with respect to a Euclidean (or, equivalently, formally real) Jordan algebra multiplication [2]. Kojima and Muramatsu [6] consider this problem in the special case when f is a vector valued polynomial whose coefficients multiplied according to a Euclidean Jordan algebra multiplication, and derive a semidefinite programming characterization for these sum of squares polynomials. As we show in Section 3, polynomials in this construction can be replaced by other functional systems, and the coefficients can be chosen from arbitrary (not Jordan) algebras. Multivariate SOS matrix polynomials are also considered in [7].

In this paper we consider the following, even more general problem. We take an arbitrary bilinear mapping $\diamond\colon A \times A \to B$, where A and B are finite dimensional real linear spaces, and show that the set of vectors that are sums of squares of vectors from A (with respect to the multiplication $\diamond$) is a linear image of positive semidefinite matrices. The linear transformation is explicitly constructed. This generalizes the above mentioned results from [10] and [6], and it is also applicable to the previous two shape constrained optimization problems.

2 Semidefinite Characterization of Sums of Squares

Consider two finite dimensional real linear spaces A and B, and a bilinear mapping $\diamond\colon A \times A \to B$. We define the cone of *sum of squares* vectors $\Sigma \subseteq B$ by

$$\Sigma \stackrel{\text{def}}{=} \left\{ \sum_{i=1}^{N} a_i \diamond a_i \;\middle|\; N \geq 1;\ a_1, \dots, a_N \in A \right\}.$$

Clearly, Σ is a convex cone, but it not necessarily proper, as it may not be either full-dimensional or pointed.

Let $E = \{e_1, \dots, e_m\}$ be a basis of A, and $V = \{v_1, \dots, v_n\}$ be a basis of a vector space $B' \supseteq B$. Furthermore, let the vector $\lambda_{ij} \in \mathbb{R}^n$ $(i, j = 1, \dots, m)$ be the coefficient vector of $e_i \diamond e_j \in B$ in basis V:

$$e_i \diamond e_j = \sum_{\ell} \lambda_{ij\ell} v_\ell. \tag{1}$$

Finally, we define the linear operator $\Lambda\colon \mathbb{R}^n \to \mathbb{R}^{m \times m}$, coordinate-wise, by the formula

$$(\Lambda(w))_{ij} \stackrel{\text{def}}{=} \langle w, \lambda_{ij} \rangle \quad \forall\, w \in \mathbb{R}^n \qquad (i, j = 1, \dots, m);$$

its adjoint operator is denoted by Λ^*. If $\diamond$ is commutative, then Λ attains only symmetric values, and it is more natural to define Λ as an $\mathbb{R}^n \to \mathbb{S}_m$ operator, where $\mathbb{S}_m$ is the space of $m \times m$ real symmetric matrices.

Our main theorem is the characterization of the sum of squares cone Σ as a linear image of the cone of positive semidefinite matrices.

Theorem 1. *Let $u = (u_1, \dots, u_n)^\top \in \mathbb{R}^n$ be arbitrary. Then $\sum_{\ell=1}^n u_\ell v_\ell \in \Sigma$ if and only if there exists a real symmetric positive semidefinite matrix $Y \in \mathbb{R}^{m \times m}$ satisfying $u = \Lambda^*(Y)$.*

Proof. Let us assume first that $\sum_{\ell=1}^n u_\ell v_\ell \in \Sigma$, that is,

$$\sum_{\ell=1}^{n} u_\ell v_\ell = \sum_{k=1}^{N} a_k \diamond a_k$$

for some $a_k \in A$. Each a_k can be written in the basis E as $a_k = \sum_{j=1}^m y_j^{(k)} e_j$, with $y_j^{(k)} \in \mathbb{R}$.

By choosing $Y = \sum_{k=1}^{N} y^{(k)} {y^{(k)}}^\top$, which is clearly positive semidefinite, we obtain

$$\begin{aligned}
\sum_{\ell=1}^{n} u_\ell v_\ell = \sum_{k=1}^{N} a_k \diamond a_k &= \sum_{k=1}^{N} \left(\left(\sum_{i=1}^{m} y_i^{(k)} e_i \right) \diamond \left(\sum_{j=1}^{m} y_j^{(k)} e_j \right) \right) \\
&= \sum_{k=1}^{N} \sum_{i=1}^{m} \sum_{j=1}^{m} y_i^{(k)} y_j^{(k)} \left(e_i \diamond e_j \right) \\
&= \sum_{i=1}^{m} \sum_{j=1}^{m} Y_{ij} \left(e_i \diamond e_j \right) \\
&= \sum_{i=1}^{m} \sum_{j=1}^{m} Y_{ij} \sum_{\ell=1}^{n} \lambda_{ij\ell} v_\ell,
\end{aligned}$$

where the last equation comes from Equation (1). Using the symbol 1_ℓ for the ℓ-th unit vector $(0, \ldots, 0, 1, 0 \ldots, 0) \in \mathbb{R}^n$, the last expression can be simplified further:

$$\sum_{i=1}^{m} \sum_{j=1}^{m} Y_{ij} \sum_{\ell=1}^{n} \lambda_{ij\ell} v_\ell = \sum_{\ell=1}^{n} \langle Y, \Lambda(1_\ell) \rangle v_\ell = \sum_{\ell=1}^{n} \langle \Lambda^*(Y), 1_\ell \rangle v_\ell.$$

Since $\{v_\ell\}$ is a basis, comparing the coefficients in $\sum_\ell u_\ell v_\ell$ to those in the last expression yields $u = \Lambda^*(Y)$, completing the proof in the "if" direction.

To prove the converse claim, reverse the steps of the above proof: if $u = \Lambda^*(Y)$, and Y is positive semidefinite, then obtain vectors $y^{(k)}$ satisfying $Y = \sum_k y^{(k)} {y^{(k)}}^\top$ (for example, from the spectral decomposition of Y), and use the above identities to deduce

$$\sum_{\ell=1}^{n} u_\ell v_\ell = \sum_{k} \left(\left(\sum_{i=1}^{m} y_i^{(k)} e_i \right) \diamond \left(\sum_{j=1}^{m} y_j^{(k)} e_j \right) \right) \in \Sigma,$$

as claimed. □

After fixing the basis V for B', the cone Σ is naturally identified with the cone $\{u \mid \sum_{\ell=1}^{n} u_\ell v_\ell \in \Sigma\}$. It shall raise no ambiguities to denote the latter cone by Σ, too.

In optimization applications it is necessary to characterize the dual cone of Σ, denoted by Σ^*. Using our main theorem, and the well-known fact that the cone of positive semidefinite matrices is self-dual (in the space of symmetric matrices), Σ^* is easily characterized, especially when $\diamond$ is commutative.

Theorem 2. *Using the notation above,*

$$\Sigma^* = \left\{ v \;\middle|\; \exists\, S \succcurlyeq 0, A = -A^\top \colon \Lambda(v) = S + A \right\}.$$

In particular, if $\diamond$ is commutative, then

$$\Sigma^* = \{ v \mid \Lambda(v) \succcurlyeq 0 \}.$$

Proof. By definition, a vector $v \in \mathbb{R}^n$ is in Σ^* if and only if $\langle u, v\rangle \geq 0$ for every $u \in \Sigma$. By Theorem 1, $u \in \Sigma$ if and only if $u = \Lambda^*(Y)$ for some $Y \succcurlyeq 0$, consequently $v \in \Sigma^*$ if and only if

$$\langle u, v\rangle = \langle \Lambda^*(Y), v\rangle = \langle Y, \Lambda(v)\rangle \geq 0 \tag{2}$$

for every $Y \succcurlyeq 0$. The dual cone of positive semidefinite matrices (embedded in $\mathbb{R}^{n\times n}$ as opposed to $\mathbb{S}_n$) is the cone of matrices that can be written in the form of $S + A$, where S is positive semidefinite and A is skew-symmetric. This proves our first claim.

If $\diamond$ is commutative, then $\Lambda(v)$ is symmetric for every v. Hence, in the decomposition $\Lambda(v) = S + A$ we must have $A = 0$. This proves the second part of the theorem. □

2.1 Examples

First, we consider two very simple examples.

Example 1. Let $A = B = B' = \mathbb{C}$, the algebra of complex numbers, viewed as a two-dimensional space over $\mathbb{R}$, with $\diamond$ being the usual multiplication: $\left(\begin{smallmatrix} a_1 \\ b_1 \end{smallmatrix}\right) \diamond \left(\begin{smallmatrix} a_2 \\ b_2 \end{smallmatrix}\right) = \left(\begin{smallmatrix} a_1a_2 - b_1b_2 \\ a_1b_2 + a_2b_1 \end{smallmatrix}\right)$. Using the standard basis as V, we obtain that $\Lambda(w) = \left(\begin{smallmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{smallmatrix}\right)$, and that

$$\Sigma = \left\{ \left(\begin{smallmatrix} Y_{11} - Y_{22} \\ 2Y_{12} \end{smallmatrix}\right) \;\middle|\; \left(\begin{smallmatrix} Y_{11} & Y_{12} \\ Y_{12} & Y_{22} \end{smallmatrix}\right) \succcurlyeq 0 \right\} = \mathbb{R}^2$$

is the entire space. We conclude that every complex number is sum of squares, in concordance with the fact that every complex number is, in fact, a square. Similar situation holds for quaternions and octonions.

Example 2. Consider any finite dimensional anticommutative algebra $(\mathbf{A}, \diamond)$ over $\mathbb{R}$, that is, assume $x \diamond y = -y \diamond x$ for every $x, y \in \mathbf{A}$. Then $e_i \diamond e_j = -e_j \diamond e_i$; thus $\Lambda(v)$ is skew-symmetric for every v. Comparing this to Theorem 2 we conclude that $\Sigma^* = \mathbf{A}$, therefore $\Sigma = \{0\}$, as expected, since zero is the only vector that can be obtained as sum of squares.

Alternatively, since $\Lambda(v)$ is skew-symmetric, $\Lambda^*(Y) = -\Lambda^*(Y^\top)$. Therefore, $\Lambda^*(Y) = 0$ for every symmetric Y, and in particular for every positive semidefinite Y. Now we conclude, using Theorem 1 directly, that $\Sigma = \{0\}$.

To avoid trivial examples such as the first one, it is useful to consider when Σ is a pointed cone. A convex cone K is *pointed* if it does not contain a line, or equivalently, if $0 \neq x \in K$ implies $-x \notin K$. As the following lemma shows, a condition sufficient to obtain a pointed Σ is that the multiplication $\diamond$ be formally real: $\diamond$ is said to be *formally real* if for for every $a_1, \ldots, a_k \in A$, $\sum_{i=1}^k (a_i \diamond a_i) = 0$ implies that each $a_i = 0$.

Lemma 3. *If $\diamond$ is formally real, then Σ is pointed.*

Proof. Suppose that for some nonzero vector x, both x and $-x$ are in K. But then $0 = x + (-x)$ is sum of squares with respect to $\diamond$: $0 = \underbrace{\sum_i a_i \diamond a_i}_{x} + \underbrace{\sum_i b_i \diamond b_i}_{-x}$, implying that each a_i and b_i are zero. Consequently $x = 0$. □

Another special case of Theorem 1 is Nesterov's well-known characterization of sum of squares functional systems [10].

Example 3. Let $f_1, \ldots, f_n$ be arbitrary functions mapping a set Δ to $\mathbb{R}$, and $A = \mathrm{span}(\{f_1, \ldots, f_n\})$. Then $\Sigma \subseteq \mathrm{span}(\{f_1^2, f_1 f_2, \ldots, f_n^2\})$. The semidefinite characterization of Σ obtained from Theorem 1 is identical to the one in [10].

It is well-known [4, 13] that a univariate polynomial of degree $2n$ is nonnegative over the real line if and only if it is the sum of squares of polynomials of degree n. This yields a semidefinite characterization of nonnegative univariate polynomials that is also well-known. For completeness we repeat this important special case of Example 3, and generalize it below, in Corollary 8.

Corollary 4. *A polynomial $p(t) = \sum_{i=0}^{2n} p_i t^i$ is nonnegative for every $t \in \mathbb{R}$ if and only if there exists a positive semidefinite matrix $Y \in \mathbb{R}^{(n+1)\times(n+1)}$ such that $p_\ell = \sum_{i+j=l} y_{ij}$, where y_{ij} is the (i,j)-th entry of Y.*

Proof. Immediate from the previous paragraph and Theorem 1, using the standard monomial bases $1, t, \ldots, t^n$ as the basis of A and $1, \ldots, t^{2n}$ as the basis of $B = B'$. Following the general construction of the Λ operator we have that for every $w = (w_0, \ldots, w_{2n})$,

$$\Lambda(w) = \begin{pmatrix} w_0 & w_1 & \cdots & w_n \\ w_1 & \iddots & & w_{n+1} \\ \vdots & & \ddots & \vdots \\ w_n & w_{n+1} & \cdots & w_{2n} \end{pmatrix}.$$

Now the claim follows from Theorem 1. □

2.2 Simplified characterizations

If A' is a proper subspace of A, then the corresponding sum of squares cone Σ' is a subset of Σ, but it need not be a proper subset. The simplest example where $\Sigma' = \Sigma$ is obtained when we define the product of any two vectors in A to be the zero vector. A non-trivial example is given by the Cracovian algebra [5], which will be instrumental in the characterization of positive semidefinite matrix polynomials. (See Corollary 8.)

Example 4. (Cracovian algebra.) Let $A = B = B'$ be the space of $k \times k$ real matrices $\mathbb{R}^{k\times k}$ equipped with the product $\diamond$ defined as $U \diamond V = UV^\top$. This non-commutative, non-associative, but formally real multiplication is also known as the *Cracovian* multiplication. A basis for $A = \mathbb{R}^{k\times k}$ is the set of unit matrices E_{ij}. (E_{ij} is the zero-one matrix with a single 1 in the position (i,j).) Consider the subspace A' spanned by the matrices E_{i1}. The sum of squares cones corresponding to A and A' are identical. (This is the same as saying that any positive semidefinite matrix $VV^\top$ can be written as a sum of rank one positive definite matrices $vv^\top$.)

The order of the matrix Y in Theorem 1 is the dimension of the space A. Thus, using the space A' instead of A may considerably simplify the semidefinite characterization of Σ. In optimization models, this reduces the size of the semidefinite constraint $Y \succcurlyeq 0$.

2.3 Weighted sums of squares cones

The above theory extends to the semidefinite representability of *weighted sum of squares* vectors. (We call a set *semidefinite representable* if it is the linear image or preimage of the cone of real symmetric positive semidefinite matrices of a fixed order.) To simplify notation, let us assume that $A = B = B'$. Given a set of *weights* $w_1, \dots, w_r \in A$, we define the set of *weighted sum of squares* (or WSOS) vectors Σ^w as

$$\Sigma^w \stackrel{\text{def}}{=} \left\{ \sum_{k=1}^{r} \left(w_k \diamond \sum_{i=1}^{N_k} (a_i \diamond a_i) \right) \,\middle|\, N_k \geq 1;\ a_1, \dots, a_N \in A \right\}.$$

Our treatment of weighted sum of squares systems differs from the one in [10], as we cannot assume to have access to “square roots” of the weights.

The intuition behind the semidefinite representability of Σ^w is the following. The cone Σ^w corresponding to r weights $w_1, \dots, w_r$ can be written as a Minkowski sum of r cones, each corresponding to a single weight:

$$\Sigma^w = \Sigma^1 + \cdots + \Sigma^r; \quad \Sigma^k = \left\{ w_k \diamond \sum_{i=1}^{N_k} (a_i \diamond a_i) \,\middle|\, N_k \geq 1;\ a_1, \dots, a_N \in A \right\}\ (k = 1, \dots, r),$$

where $+$ denotes the Minkowski sum. Now each Σ^i has a semidefinite representation analogous to that of Σ, in fact, each Σ^i is a linear transform of Σ. Finally, the Minkowski sum of semidefinite representable sets is semidefinite representable by definition. Theorem 5 below gives the precise formulation of this result.

As before, let $E = \{e_1, \dots, e_m\}$ be a basis of A, but now let the vector $\lambda_{ij}^k \in \mathbb{R}^m$ $(i, j = 1, \dots, m,\ k = 1, \dots, r)$ be the coefficient vector of $w_k \diamond (e_i \diamond e_j) \in A$ in basis E:

$$w_k \diamond (e_i \diamond e_j) = \sum_{\ell} \lambda_{ij\ell}^k e_\ell. \tag{3}$$

Finally, for each $k = 1, \dots, r$ we define the linear operator $\Lambda^k \colon \mathbb{R}^m \to \mathbb{R}^{m \times m}$ coordinate-wise, by the formula

$$(\Lambda^k(w))_{ij} \stackrel{\text{def}}{=} \langle w, \lambda_{ij}^k \rangle \quad \forall\, w \in \mathbb{R}^m \qquad (i, j = 1, \dots, m);$$

its adjoint operator is denoted by $(\Lambda^k)^*$. If $\diamond$ is commutative, then Λ attains only symmetric values, and it is more natural to define Λ as an $\mathbb{R}^n \to \mathbb{S}_m$ operator, where $\mathbb{S}_m$ is the space of $m \times m$ real symmetric matrices.

With this notation, the above argument leads to the following semidefinite characterization of WSOS vectors.

Theorem 5. *Let $u = (u_1, \dots, u_m)^\top \in \mathbb{R}^m$ be arbitrary. Then $\sum_{\ell=1}^m u_\ell e_\ell \in \Sigma^w$ if and only if there exists real symmetric positive semidefinite matrices $Y^{(1)}, \dots, Y^{(r)} \in \mathbb{R}^{m\times m}$ satisfying $u = \sum_{k=1}^r (\Lambda^k)^*(Y^{(k)})$.*

3 Application to Vector-valued Functions

While many of the examples in the previous sections can be regarded as basic, they serve as building blocks of some truly non-trivial results, such as the semidefinite characterization of (weighted) sums of squares of *vector-valued functions*, and positive semidefinite matrix polynomials. All the results of this section could be obtained by the direct application of Theorem 1; however, the following lemma considerably simplifies their presentation.

Lemma 6. *Let A_i and B_i, $i = 1, \dots, k$ be finite dimensional real vector spaces, and $\diamond_i \colon A_i \times A_i \to B_i$ be bilinear mappings. Define $A = A_1 \times \cdots \times A_k$, $B = B_1 \times \cdots \times B_k$, and $\diamond \colon A \times A \to B$ by the identity*

$$(x_1, \dots, x_k) \diamond (y_1, \dots, y_k) = (x_1 \diamond_1 y_1, \dots, x_k \diamond_k y_k).$$

Then the Λ operator corresponding to $\diamond$ is $\Lambda_1 \otimes \cdots \otimes \Lambda_k$, where Λ_i is the Λ operator corresponding to $\diamond_i$, and $\otimes$ denotes the Kronecker product.

Proof. Apply the construction preceding Theorem 1 to the direct product of the bases of A_i, which is a basis of A. □

Our next example relies on a consequence of Youla's spectral factorization theorem [16]. For completeness we recall the specific corollary we use.

Proposition 7 ([16, Theorem 2 and Corollary 2])**.** *Let $P(t)$ be an $m \times m$ real symmetric polynomial matrix, and let r be the largest integer such that $P(t)$ has at least at least one minor of order r that does not vanish identically. Then there exists an $m \times r$ polynomial matrix $Q(t)$ satisfying the identity $P(t) = Q(t)Q(t)^\top$.*

Example 5. By Proposition 7, a univariate real matrix polynomial $P(t)$ of degree $2n$ is positive semidefinite for every t if and only if it is the sum of squares of degree n matrix polynomials, where squaring is with respect to the Cracovian multiplication. This yields a semidefinite characterization analogous to the characterization of complex semidefinite matrix polynomials in [3]:

Corollary 8. *The $k \times k$ matrix polynomial $P(t) = \sum_{i=0}^{2n} P_i t^i$ is positive semidefinite for every $t \in \mathbb{R}$ if and only if there exists a positive semidefinite block matrix $Y \in \mathbb{R}^{(n+1)k\times(n+1)k}$ consisting of blocks Y_{ij}, $i, j = 0, \dots, n$ or order k such that $P_\ell = \sum_{i+j=\ell} Y_{ij}$ for each $\ell = 0, \dots, 2n$.*

Proof. By Proposition 7, $P(t)$ is nonnegative for every $t \in \mathbb{R}$ if and only if it is sum of squares with respect to the multiplication $R(t) \diamond S(t) = R(t)S(t)^\top$. We can characterize separately the sum of squares cone of the coefficient matrices (with respect to the Cracovian multiplication) and the sum of squares cone of ordinary polynomials (with respect to the ordinary polynomial multiplication), and then use Lemma 6 to obtain the characterization of sum of squares matrix polynomials.

In fact, we have already determined the Λ operators corresponding to the Cracovian algebra in Example 4 and to the nonnegative polynomials in Corollary 4. Now, the theorem follows from Lemma 6 and Theorem 1. □

Similarly, Lemma 6 can be used to characterize the sums of squares of *arbitrary* vector-valued functions, not only of semidefinite valued matrix polynomials. More precisely, let $(\mathbb{A}, \circ)$ be a (not necessarily commutative or associative) finite dimensional algebra, $f_1, \dots, f_n$ be given real valued functions, and let A be the space

$$\left\{ \sum_{i=0}^{n} v_i f_i \mid v_i \in \mathbb{A} \right\}.$$

Then the cone $\Sigma = A^2$ is semidefinite representable via Theorem 1. This generalizes Lemma 2 of [6], where a similar characterization is obtained in the special case when $f_i = t^i$, and $(\mathbb{A}, \circ)$ is a Euclidean (or formally real) Jordan algebra.

4 A Coordinate-free Approach

The semidefinite representations of Σ and Σ^* given in Theorems 1 and 2 use an explicitly constructed operator Λ, which in turn depends on the selected bases of the linear spaces A and B'. To better understand the structure of the cones Σ and Σ^*, or to prove theorems about them, a coordinate-free approach might be more fruitful. In this section we outline such an approach for vector valued functions whose coefficients are vectors from a (not necessarily associative) algebra. We also hope that this direction may lead to efficient algorithms for optimization problems involving sum of squares cones, without the (sometimes inefficient) translation of these constraints to semidefinite programming constraints.

As before, we consider a bilinear mapping $\diamond \colon A \times A \to B$, where A and B are finite dimensional real linear spaces. A particularly interesting special case is $\diamond$ is the multiplication of an algebra $(\mathbb{A}, \diamond)$, that is, when $\mathbb{A} = A = B$; in the remainder of this section we only consider this special case. We define an $\mathbb{A}$-*vector* of dimension n and an $\mathbb{A}$-*matrix* of dimension $m \times n$ as an element of $\mathbb{A}^n$ and $\mathbb{A}^{m \times n}$, respectively. $\mathbb{A}$-matrices and $\mathbb{A}$-vectors will be denoted by boldface letters. Matrix and vector multiplication is extended in the straightforward way: if $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times \ell$, then $\mathbf{A} \diamond \mathbf{B}$ is $m \times \ell$:

$$(\mathbf{A} \diamond \mathbf{B})_{ij} \stackrel{\text{def}}{=} \sum_k \mathbf{A}_{ik} \diamond \mathbf{B}_{kj}.$$

This definition also extends to the case when $\mathbf{A}$ or $\mathbf{B}$ (or both) are are replaced by real matrices of the same dimension.

We also extend the notation of an inner product $\langle\cdot,\cdot\rangle$ to $\mathbb{A}$-vectors and matrices, with the usual meaning, except that multiplication of $\mathbb{A}$-vectors and $\mathbb{A}$-matrices are with respect to $\diamond$. For example

$$\langle \mathbf{a}, \mathbf{b} \rangle = \sum_i (\mathbf{a}_i \diamond \mathbf{b}_i)$$

We say that an $\mathbb{A}$-matrix $\mathbf{A}$ is $\mathbb{A}$-*positive semidefinite* or $\mathbb{A}$-*psd*, denoted by $\mathbf{A} \succcurlyeq_{\mathbb{A}} 0$, if there are $\mathbb{A}$-vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ such that

$$\mathbf{A} = \sum_{i=1}^{k} \mathbf{a}_i \diamond \mathbf{a}_i^\top.$$

Turning to $\mathbb{A}$-vector valued functions, let $f_1, \ldots, f_m$ be linearly independent real valued functions (similarly to Example 3), and consider the set of vector valued functions

$$\mathcal{F} = \left\{ \sum_{i=1}^{m} a_i f_i \;\middle|\; a_i \in \mathbb{A} \right\}.$$

The set of sum of squares functions $\Sigma_{\mathbb{A}}$ is then defined as

$$\Sigma_{\mathbb{A}} = \left\{ \sum_{i=1}^{N} g_i^2 \;\middle|\; N \geq 1 \;\; g_i \in \mathcal{F} \right\}.$$

If $V = \{v_1, \ldots, v_n\}$ denotes a basis of $\operatorname{span}\{f_1^2, f_1 f_2, \ldots, f_1 f_m, f_2^2, \ldots, f_m^2\}$, then elements of $\Sigma_{\mathbb{A}}$ can be expressed as $\sum_{i=1}^{n} a_i v_i$ for some $a_i \in \mathbb{A}$, and the question of our interest is for which $\mathbb{A}$-vectors $\mathbf{a} = (a_1, \ldots, a_n)^\top$ does the function $\sum_{i=1}^{n} a_i v_i$ belong to $\Sigma_{\mathbb{A}}$.

Theorems 1 and 2 give a semidefinite characterization of Σ, but this characterization depends on the selected basis of A. As we shall see, it is also possible to characterize Σ in terms of the above defined $\mathbb{A}$-psd $\mathbb{A}$-matrices in a coordinate-free manner. To state our main theorem, we need some more notation.

For every $i, j = 1, \ldots, m$ let the vector $\lambda_{ij} \in \mathbb{R}^n$ be the coefficient vector of $f_i f_j$ in basis V:

$$f_i f_j = \sum_{\ell} \lambda_{ij\ell} v_\ell, \tag{4}$$

and we define the linear operator $\Lambda \colon \mathbb{R}^n \to \mathbb{R}^{m \times m}$, coordinate-wise, by the formula

$$(\Lambda(w))_{ij} \stackrel{\text{def}}{=} \langle w, \lambda_{ij} \rangle \quad \forall\, w \in \mathbb{R}^n \qquad (i, j = 1, \ldots, m).$$

In other words, Λ is the unique operator satisfying $uu^\top = \Lambda(v)$. Finally, we define another linear operator, $\Lambda^* \colon \mathbb{A}^{n \times n} \to \mathbb{A}^n$ such that for every $\mathbb{A}$-matrix Y,

$$\Lambda^*(Y) = \left(\langle Y, \lambda^{(1)} \rangle, \ldots, \langle Y, \lambda^{(n)} \rangle \right)^\top,$$

where $\lambda^{(k)} \in \mathbb{R}^{m\times m}$ is the matrix made up of the kth entries of the vectors λ_{ij}. While Λ^* is not the adjoint of Λ, nor is $\langle\cdot,\cdot\rangle$ an inner product, this notation is motivated by the following identity, which resembles to the definition of the adjoint.

Lemma 9. *For every $Y \in \mathbb{A}^{m\times m}$ and $w \in \mathbb{R}^n$, $\langle Y, \Lambda(w)\rangle = \langle \Lambda^*(Y), w\rangle$.*

Proof.

$$\langle Y, \Lambda(w)\rangle = \sum_{i,j} Y_{ij}\langle w, \lambda_{ij}\rangle = \sum_{i,j} Y_{ij} \sum_k w_k \lambda_{ij}^{(k)} = \sum_k w_k \sum_{i,j} Y_{ij}\lambda_{ij}^{(k)} = \langle \Lambda^*(Y), w\rangle$$

□

Denoting the vector $(f_1,\dots,f_m)^\top$ by f and the vector $(u_1,\dots,u_n)^\top$ by u, the $\mathbb{A}$-vector valued functional version of Theorem 1 can be formulated as follows.

Theorem 10. *An $\mathbb{A}$-vector valued function $\mathbf{p} = \langle \mathbf{a}, u\rangle$ is in $\Sigma_{\mathbb{A}}$ if and only if there exists an $\mathbb{A}$-psd matrix $\mathbf{Y}$ satisfying*

$$\mathbf{p} = \langle \Lambda^*(\mathbf{Y}), v\rangle$$

Proof. The proof is analogous to that of Theorem 1. If $\mathbf{p} = \sum_i \mathbf{p}_i^2$ with $\mathbf{p}_i = \langle \mathbf{a}_i, u\rangle$, then

$$\begin{aligned}
\mathbf{p} &= \sum_i (\langle \mathbf{a}_i, u\rangle)^2 \\
&= \sum_i \langle \mathbf{a}_i \diamond \mathbf{a}_i^\top, uu^\top\rangle \\
&= \left\langle \sum_i \mathbf{a}_i \diamond \mathbf{a}_i^\top, uu^\top \right\rangle \\
&= \left\langle \sum_i \mathbf{a}_i \diamond \mathbf{a}_i^\top, \Lambda(v) \right\rangle \\
&= \left\langle \Lambda^*\left(\sum_i (\mathbf{a}_i \diamond \mathbf{a}_i^\top)\right), v \right\rangle,
\end{aligned}$$

using, in the last step, Lemma 9. Setting $\mathbf{Y} = \sum_i (\mathbf{a}_i \diamond \mathbf{a}_i^\top)$ we see that $\mathbf{Y} \succcurlyeq_{\mathbb{A}} 0$. Conversely, if $\mathbf{p} = \langle \Lambda^*(\mathbf{Y}), v\rangle$, then by reversing the above argument we obtain that $\mathbf{p} = \sum_i \mathbf{p}_i^2$ for some $\mathbf{p}_i = \langle \mathbf{a}_i, u\rangle$. □

It follows from the definition that $\Sigma_{\mathbb{A}}$ is a convex cone, but it is not necessarily a proper (closed, full-dimensional, pointed) cone. As the following lemma shows, to make $\Sigma_{\mathbb{A}}$ pointed, it is sufficient to choose an $\mathbb{A}$ that is formally real: An algebra $(\mathbb{A}, \diamond)$ is said to be *formally real* if for for every $a_1,\dots,a_k \in \mathbb{A}$, $\sum_{i=1}^k (a_i \diamond a_i) = 0$ implies that each $a_i = 0$.

Lemma 11. *If $(\mathbb{A}, \diamond)$ is formally real, then the cone of $\mathbb{A}$-psd matrices is pointed.*

Proof. Suppose that $\mathbf{B} \succcurlyeq_{\mathbb{A}} 0$. For every $\mathbb{A}$-vector $\mathbf{b}$, the diagonal entries of $\mathbf{b} \diamond \mathbf{b}^\top$ are squares, consequently the diagonal entries of $\mathbf{B} = \sum_k (\mathbf{b}_k \diamond \mathbf{b}_k^\top)$ are sums of squares. If $(\mathbb{A}, \diamond)$ is formally real, then the (i,i)-th entry of $\mathbf{B}$ can be zero only if the ith entry of every $\mathbf{b}_k$ is zero. Hence, the diagonal of $\mathbf{B}$ can only be zero if every $\mathbf{b}_k$ is zero. This implies that the zero matrix cannot be expressed as the sum of two nonzero $\mathbb{A}$-psd matrix, and the cone of $\mathbb{A}$-psd matrices is indeed pointed. □

5 Conclusion

A couple of questions related to $\mathbb{A}$-positive semidefiniteness remain open. From the algorithmic viewpoint, it would be interesting to see if numerically efficient self-concordant barrier functions can be designed for $\mathbb{A}$-psd cones, at least under some assumptions on the underlying algebra. Theorem 1 can be used to give a characterization of $\mathbb{A}$-psd cones via real positive semidefinite matrices, but this is not necessarily the most efficient way of solving optimization problems involving $\mathbb{A}$-psd cones. Exploring the connections between the properties of the algebra $(\mathbb{A}, \diamond)$ and the properties of $\mathbb{A}$-psd cones might also be an interesting direction for future research. Finally, it would be interesting to see how the semidefinite programming models suggested by Theorems 1 and 2 perform in practice, when used in shape constrained optimization problems like the ones mentioned in Section 1.

References

[1] F. Alizadeh and D. Goldfarb, *Second-order cone programming*, Mathematical Programming Series B **95** (2003), 3–51.

[2] J. Faraut and A. Korányi, *Analysis on symmetric cones*, Oxford University Press, Oxford, UK, 1994.

[3] Y. Genin, Y. Hachez, Yu. Nesterov, and P. Van Dooren, *Optimization problems over positive pseudo-polynomial matrices*, SIAM Journal on Matrix Analysis and Applications **25** (2003), no. 1, 57–79.

[4] S. Karlin and W. J. Studden, *Tchebycheff Systems, with Applications in Analysis and Statistics*, Wiley Interscience Publishers, 1966.

[5] Jerzy Kociński, *Cracovian algebra*, Nova Science, 2004.

[6] M. Kojima and M. Muramatsu, *An extension of sums of squares relaxations to polynomial optimization problems over symmetric cones*, Mathematical Programming, Ser. A **110** (2007), no. 2, 315–336.

[7] Masakazu Kojima, *Sums of squares relaxations of polynomial semidefinite programs*, Tech. Report B-397, Dept. of Math and Comp. Sci, Tokyo Institute of Technology, November 2003.

[8] Jean B. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM Journal on Optimization **11** (2001), no. 3, 796–817.

[9] Alessandro Magnani, Sanjay Lall, and Stephen Boyd, *Tractable fitting with convex polynomials via sum-of-squares*, 44th IEEE Conference on Decision and Control and 2005 European Control Conference. CDC–ECC '05. (Seville, Spain), December 2005, pp. 1672–1677.

[10] Yu. Nesterov, *Squared functional systems and optimization problems*, High Performance Optimization (H. Frenk, K. Roos, T. Terlaky, and S. Zhang, eds.), Appl. Optim., Kluwer Acad. Publ., Dordrecht, 2000, pp. 405–440.

[11] Dávid Papp and Farid Alizadeh, *Linear and second order cone programming approaches to statistical estimation problems*, Tech. Report RRR 13–2008, Rutgers Center for Operations Research (RUTCOR), Rutgers University, September 2008.

[12] Pablo A. Parrilo, *Semidefinite programming relaxations for semialgebraic problems*, Mathematical Programming **96** (2003), no. 2, 293–320.

[13] George Pólya and Gabor Szegö, *Problems and theorems in analysis*, vol. 2, Springer, 1976.

[14] Dirk J. Struik, *Lectures on classical differential geometry*, Dover, New York, 1988.

[15] H. van Maaren, L. van Norden, and M.J.H. Heule, *Sums of squares based approximation algorithms for MAX-SAT*, Discrete Applied Mathematics **156** (2008), no. 10, 1754–1779.

[16] D. C. Youla, *On the factorization of rational matrices*, IRE Transactions on Information Theory (1961), 172–189.