ON A NEW CLASS OF MATRIX SUPPORT FUNCTIONALS WITH APPLICATIONS

JAMES V. BURKE AND TIM HOHEISEL

ABSTRACT. A new class of matrix support functionals is presented which establish a connection between optimal value functions for quadratic optimization problems, the matrix-fractional function, the pseudo matrix-fractional function, and the nuclear norm. The support function is based on the graph of the product of a matrix with its transpose. Closed form expressions for the support functional and its subdifferential are derived. In particular, the support functional is shown to be continuously differentiable on the interior of its domain, and a formula for the derivative is given when it exists.

1 Introduction

One of the first topics studied in an elementary course on optimization is that of optimizing a quadratic function over an affine set:

$$\min_{u\in\mathbb{R}^n} \frac{1}{2}u^TVu - x^Tu \quad \text{s.t.} \quad Au = b, \tag{1}$$

where $(x, V, A, b) \in \mathbb{R}^n \times \mathbb{S}^n \times \mathbb{R}^{p\times n} \times \mathbb{R}^p$, with $\mathbb{S}^n$ the linear space of real symmetric $n \times n$ matrices. A complete characterization of the set of solutions and the optimal value are easily obtained (see Theorem 3.2 below) without the aid of calculus. Less well studied are the properties of the optimal value function

$$v(x, V) := \inf_{u\in\mathbb{R}^n} \left\{ \frac{1}{2}u^TVu - x^Tu \mid Au = b \right\}, \tag{2}$$

for given $(A, b) \in \mathbb{R}^{p\times n} \times \mathbb{R}^p$ with $b \in \operatorname{rge} A$. The variational properties of v, and its extensions, are easily derived as a consequence of the more general results of this paper. In particular, we show that v is a concave function, and, more specifically, it is the negative of a convex support functional on $\mathbb{R}^n \times \mathbb{S}^n$. In addition, we show $-v$ equals the (pseudo) matrix-fractional function when $(A, b) = (0, 0)$ [2, 5, 7, 13, 14] (see Section 5.2).

The central object of study is the support functional $\sigma_{\mathcal{D}(A,B)} : \mathbb{E} \to \bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ given by

$$\sigma_{\mathcal{D}(A,B)}(X, V) := \sup \{\langle (X, V), (U, W)\rangle \mid (U, W) \in \mathcal{D}(A, B)\}, \tag{3}$$

where $\mathbb{E} := \mathbb{R}^{n\times m} \times \mathbb{S}^n$, $\operatorname{rge} B \subset \operatorname{rge} A$, and

$$\mathcal{D}(A, B) := \left\{ \left(Y, -\frac{1}{2}YY^T\right) \in \mathbb{E} \;\middle|\; Y \in \mathbb{R}^{n\times m} : \; AY = B \right\}. \tag{4}$$

2010 *Mathematics Subject Classification.* 46N10, 49J52, 49J53, 90C20, 90C46.

Key words and phrases. Support function, quadratic programming, value function, matrix-fractional function, nuclear norm, subdifferential, normal cone.

This research was partially supported by the DFG (Deutsche Forschungsgemeinschaft) under grant HO 4739/2-1.

Here, we use the Frobenius inner product for matrix spaces (see Section 2.2). The set $\mathcal{D}(A,B)$ is the graph of the mapping $Y \mapsto -\frac{1}{2}YY^T$ over the affine manifold $\{Y \mid AY = B\}$. This mapping is a common transformation in numerous applications to statistics and *semidefinite programming* (e.g., see [1, 2]). In Section 4, we obtain a closed form expression for $\sigma_{\mathcal{D}(A,B)}$ in terms of X, V, A, and B, a description of its domain and its interior, a characterization of the closed convex hull of the set $\mathcal{D}(A,B)$, and a characterization of the convex subdifferential $\partial\sigma_{\mathcal{D}(A,B)}$. In particular, we show that $\sigma_{\mathcal{D}(A,B)}$ is differentiable on the interior of its domain and provide a formula for the derivative. By taking $m = 1$, we obtain the representation

$$v = -\sigma_{\mathcal{D}(A,b)},$$

where v is the optimal value function defined in (2) (see Theorem 5.1). In addition, we find that $\sigma_{\mathcal{D}(0,0)}$ is precisely the *pseudo matrix-fractional function* studied in [2, 5, 7, 13, 14]. The pseudo matrix-fractional function coincides with the matrix-fractional function when V is positive definite (see Theorem 5.3).

Our derivation of the subdifferential $\partial\sigma_{\mathcal{D}(A,B)}$ makes use of techniques from nonconvex, nonsmooth variational analysis. For this reason, our study begins in Section 2 by recalling the necessary background material from convex and nonsmooth variational analysis [3, 16] as well as matrix analysis [12]. In Section 3, we review the elementary results concerning the problem (1) reformulating them in a manner more suitable for our study. These results are the key to our analysis. The core results of the paper are presented in Section 4 where we develop the properties of the support functional $\sigma_{\mathcal{D}(A,B)}$. In Section 5 we present three applications of the results of Section 4. The first of these is the representation of the optimal value function v described above. The second is the application to the matrix-fractional function and its generalizations. The final application establishes an interesting relationship between $\sigma_{\mathcal{D}(A,B)}$ and the nuclear norm. In particular, we recover the relationship discussed in [13] which is used to obtain smooth approximations to the nuclear norm in optimization.

Notation: The *support* of a vector $d \in \mathbb{R}^p$ is the index set

$$\operatorname{supp} d := \{i \in \{1, \dots, p\} \mid d_i \neq 0\}.$$

The *unit simplex* in $\mathbb{R}^p$ is given by

$$\Delta_p := \left\{\lambda \in \mathbb{R}^p \mid \lambda_i \geq 0 \ (i = 1, \dots, p), \quad \sum_{i=1}^p \lambda_i = 1\right\},$$

while the *unit sphere* in $\mathbb{R}^p$ is the set

$$\mathcal{S}_{p-1} := \{x \in \mathbb{R}^p \mid \|x\|_2 = 1\}.$$

For a set $B \subset \mathbb{R}^p$ its *orthogonal complement* is defined by

$$B^\perp := \left\{x \in \mathbb{R}^p \;\middle|\; x^T b = 0 \quad \forall b \in B\right\}.$$

Its *linear hull* or *span* is the set

$$\operatorname{span} B := \left\{x \in \mathbb{R}^p \;\middle|\; \exists n \in \mathbb{N}, b^i \in B, \alpha_i \in \mathbb{R} \ (i = 1, \dots, n) : \ x = \sum_{i=1}^n \alpha_i b^i\right\}.$$

For two sets $A, B \subset \mathbb{R}^n$ its *Minkowski sum* is the set

$$A + B := \{a + b \mid a \in A, \ b \in B\}.$$

Moreover, if $A = \{a\}$ is a singleton, we loosely write $a + B := \{a\} + B$.

2 Preliminaries

2.1 Tools for variational analysis in Euclidean space

Let $(\mathcal{E}, \langle\cdot,\cdot\rangle)$ be a Euclidean space, i.e., a finite dimensional real inner product space. We denote the norm on $\mathcal{E}$ induced by the inner product $\langle\cdot,\cdot\rangle$ by $\|\cdot\|$, i.e., $\|x\| := \sqrt{\langle x, x\rangle}$ for all $x \in \mathcal{E}$. In particular, we equip $\mathcal{E} = \mathbb{R}^n$ with the standard scalar product giving $\|\cdot\| = \|\cdot\|_2$.

For an extended real-valued function $f : \mathcal{E} \to \overline{\mathbb{R}}$ its *epigraph* is the set

$$\operatorname{epi} f := \{(x, \alpha) \in \mathcal{E} \times \mathbb{R} \mid f(x) \leq \alpha\},$$

and its *domain* is the set

$$\operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\}.$$

The notion of the epigraph allows for very handy definitions for a number of properties of extended real-valued functions. A function $f : \mathcal{E} \to \overline{\mathbb{R}}$ is called *lower semicontinuous (lsc)* (or *closed*) if $\operatorname{epi} f$ is a closed set, and *upper semicontinuous (usc)* if $-f$ is lsc. The lower semicontinuous hull of f, denoted $\operatorname{cl} f$, is the function whose epigraph is given by $\operatorname{cl}(\operatorname{epi} f)$. f is said to be *convex* if $\operatorname{epi} f$ is a convex set, and concave if $-f$ is convex. Morever, f is said to be *proper* if $\operatorname{dom} f \neq \emptyset$ and $f > -\infty$.

The *(convex) indicator function* of a set $C \subset \mathcal{E}$, $\delta_C : \mathcal{E} \to \mathbb{R} \cup \{+\infty\}$, is given by

$$\delta_C(x) := \begin{cases} 0 & \text{if} \quad x \in C, \\ +\infty & \text{if} \quad x \notin C. \end{cases}$$

The indicator function δ_C is convex if and only if C is convex, and δ_C is lsc if and only if C is closed. The set C is a cone if $\lambda C \subset C$ for all $\lambda \geq 0$. It is a convex cone if, in addition, $C + C \subset C$.

The *(convex) subdifferential* of a function $f : \mathcal{E} \to \overline{\mathbb{R}}$ at a point $\bar{x} \in \operatorname{dom} f$ is the set

$$\partial f(\bar{x}) := \{v \in \mathcal{E} \mid f(x) \geq f(\bar{x}) + \langle v, x - \bar{x}\rangle \quad \forall x \in \mathcal{E}\}$$

with $\operatorname{dom} \partial f := \{\bar{x} \mid \partial f(\bar{x}) \neq \emptyset\}$. The *(convex) conjugate* $f^* : \mathcal{E} \to \overline{\mathbb{R}}$ of f is defined by

$$f^*(y) := \sup_{x \in \operatorname{dom} f} \{\langle x, y\rangle - f(x)\}.$$

We call a function $f : \mathcal{E} \to \overline{\mathbb{R}}$ *positively homogeneous* if

$$0 \in \operatorname{dom} f \quad \text{and} \quad f(\lambda v) = \lambda f(v) \quad \forall v \in \mathcal{E} \text{ and } \lambda > 0.$$

If, in addition,

$$f(v + w) \leq f(v) + f(w) \quad \forall v, w \in \mathcal{E},$$

we say that h is sublinear. Note that f is positively homogeneous if and only if $\operatorname{epi} f$ is a cone, in which case either $f(0) = 0$ or $f(0) = -\infty$. Moreover, h is sublinear if and only if it is convex and positively homogeneous, which holds if and only if $\operatorname{epi} f$ is a convex cone [16, Ex. 3.19].

For $C \subset \mathcal{E}$, its *convex hull* is defined to be the set

$$\operatorname{conv} C := \bigcap \{D \subset \mathcal{E} \mid C \subset D,\ D \text{ convex}\}.$$

By *Carathéodory's Theorem*, e.g., see [3, Sec. 2.3, Ex. 5], we can represent $\mathrm{conv}\, C$ as

$$\mathrm{conv}\, C = \left\{ x \in \mathcal{E} \;\middle|\; \exists \lambda \in \Delta_{\kappa+1}, c_1, \dots, c_{\kappa+1} \in C:\ x = \sum_{i=1}^{\kappa+1} \lambda_i c_i \right\},$$

where $\kappa := \dim \mathcal{E}$. Furthermore, the *closed convex hull* of C is the closure of the convex hull of C, i.e.,

$$\overline{\mathrm{conv}}\, C := \mathrm{cl}\,(\mathrm{conv}\, C).$$

The function

$$\sigma_C(v) := \delta_C^*(v) = \sup_{w \in C} \langle v, w \rangle$$

is called the *support functional* for C, which is always sublinear (hence convex), lsc and proper when C is nonempty, cf., e.g., [3, Sec. 4.2, Ex. 9], and the bi-conjugacy theorem [3, Th. 4.2.1] gives

$$\sigma_C^* = \delta_{\overline{\mathrm{conv}}\, C}. \tag{5}$$

Given a set $C \subset \mathcal{E}$ the *regular normal cone* of C at $\bar{x} \in C$ is the set

$$\hat{N}_C(\bar{x}) := \left\{ d \in \mathcal{E} \;\middle|\; \liminf_{x \to_C \bar{x}} \frac{\langle d, x - \bar{x} \rangle}{\|x - \bar{x}\|} \le 0 \right\},$$

and the *limiting normal cone* of C at $\bar{x} \in C$ is the set

$$N_C(\bar{x}) := \left\{ v \in \mathcal{E} \;\middle|\; \exists \{x^k \in C\} \to \bar{x},\ v^k \in \hat{N}_C(x^k):\ v^k \to v \right\},$$

i.e., the limiting normal cone is the *outer limit* of the regular normal cone in the sense of *Painlevé-Kuratowski*, cf. [16, Def. 5.4].

We call a closed set $C \subset \mathcal{E}$ *(Clarke) regular at a point* $\bar{x} \in C$ if $N_C(\bar{x}) = \hat{N}_C(\bar{x})$. If this holds true for every point $\bar{x} \in C$, we simply say C is *(Clarke) regular*. In particular, every closed convex set $D \subset \mathcal{E}$ is regular and

$$N_D(\bar{x}) = \{ v \in \mathcal{E} \mid \langle v, x - \bar{x} \rangle \le 0 \quad \forall x \in D \} = \partial \delta_D(\bar{x}) \quad \forall \bar{x} \in D.$$

2.2 Tools from matrix analysis

A general reference for the material in this section is [12]. For a matrix $A \in \mathbb{R}^{p \times m}$ we denote the ith column vector $(i = 1, \dots, m)$ by $a^i \in \mathbb{R}^p$, i.e., $A = [a^1, \dots, a^m]$. The *range* of A is the set $\mathrm{rge}\, A := \{ Ax \in \mathbb{R}^p \mid x \in \mathbb{R}^m \} = \mathrm{span}\, \{a^1, \dots, a^m\}$, and the *kernel* or *null space* is $\ker A := \{ x \in \mathbb{R}^m \mid Ax = 0 \}$. The following basic subspace relationships are elementary and used freely throughout:

$$\ker A = \left(\mathrm{rge}\, A^T\right)^\perp \quad \text{and} \quad \mathrm{rge}\, A = (\ker A^T)^\perp.$$

The trace of a square matrix $M \in \mathbb{R}^{n \times n}$ is the sum of all diagonal elements, and we write $\mathrm{tr}\,(M) = \sum_{i=1}^n m_{ii}$. If $\lambda_1(M), \dots, \lambda_n(M)$ are the eigenvalues of M, counting multiplicities, then $\mathrm{tr}\,(M) = \sum_{i=1}^n \lambda_i(M)$.

The singular value decomposition (SVD) of $A \in \mathbb{R}^{p \times m}$ is a factorization of the form $A = U \hat{\Sigma} Q^T$, where $U \in \mathbb{R}^{p \times p}$ and $Q \in \mathbb{R}^{m \times m}$ are orthogonal and the only nonzero entries of the matrix $\hat{\Sigma} \in \mathbb{R}^{p \times m}$ are $\hat{\Sigma}_{ii} > 0$ for $i = 1, 2, \dots, k := \mathrm{rank}\, A$. These nonzero entries, $\sigma_1(A) \ge \sigma_2(A) \ge \dots \ge \sigma_k(A)$, are called the singular values of A, and they equal the square roots of the nonzero eigenvalues of $A^T A$ (or, equivalently, AA^T). The *reduced* SVD is the factorization $A = U_1 \Sigma Q_1$, where $\Sigma \in \mathbb{R}^{k \times k}$ is the nonsingular square diagonal matrix of singular values, $U_1 \in \mathbb{R}^{p \times k}$ consists of the first k columns of U, and $Q_1 \in \mathbb{R}^{k \times m}$ consists of the first k columns of Q. The remaining columns of both U and Q can be given any desired but commensurate left-right ordering. The singular values can be used to define a family of norms on $\mathbb{R}^{p\times m}$ called the *Schatten norms*. Of particular interest to us is the *nuclear norm* (or Schatten 1-norm) defined by $\|A\|_n = \operatorname{tr}\Sigma$.

For $A \in \mathbb{R}^{p\times n}$, the matrix S is called the *Moore-Penrose pseudoinverse* of A if and only if

$$SA = (SA)^T,\ AS = (AS^T),\ ASA = A, \text{ and } SAS = S. \tag{6}$$

We denote the Moore-Penrose pseudoinverse by $A^\dagger$ (cf. [11, p. 139] or [15, Ch. 2]). It is straightforward to show that $A^\dagger = Q_1\Sigma^{-1}U_1^T$, where $A = U_1\Sigma Q_1^T$ is the reduced SVD for A. In particular, $A^\dagger A$ and $AA^\dagger$ are the orthogonal projections onto $\operatorname{rge} A^T$ and $\operatorname{rge} A$, respectively. Consequently, given $b \in \operatorname{rge} A$, we have

$$\{x \in \mathbb{R}^n \mid Ax = b\} = A^\dagger b + \ker A,$$

which is of particular importance in our study.

For two matrices $A \in \mathbb{R}^{r\times s}, B \in \mathbb{R}^{p\times q}$, we define their *Kronecker product* by

$$A \otimes B := \begin{pmatrix} a_{11}B & \cdots & a_{1s}B \\ \vdots & \vdots & \vdots \\ a_{r1}B & \cdots & a_{rs}B \end{pmatrix}.$$

The space of real symmetric matrices of dimension $p \times p$ is denoted by $\mathbb{S}^p$. The subset of symmetric positive semidefinite matrices is denoted by $\mathbb{S}^p_+$, and the subset of symmetric positive definite matrices is $\mathbb{S}^p_{++}$. In addition, we use the following notation:

$$S \succeq T \quad :\Longleftrightarrow \quad S - T \in \mathbb{S}^p_+,$$

and

$$S \succ T \quad :\Longleftrightarrow \quad S - T \in \mathbb{S}^p_{++}.$$

Whenever we consider a matrix (sub-)space of the form $\mathcal{E} \subset \mathbb{R}^{r\times s}$ we equip it with the inner product $\langle\cdot,\cdot\rangle : \mathcal{E} \to \mathbb{R}$ defined by

$$\langle A, B\rangle := \operatorname{tr}(A^T B),$$

which makes $\mathcal{E}$ a Euclidean space. The induced norm is called the Frobenius norm. This procedure is also adopted for matrix product-spaces. In particular, the space $\mathbb{E} := \mathbb{R}^{n\times m} \times \mathbb{S}^n$ is equipped with the inner product $\langle\cdot,\cdot\rangle : \mathbb{E} \to \mathbb{R}$, defined by

$$\langle (X, V), (Y, W)\rangle := \operatorname{tr}(X^T Y) + \operatorname{tr}(VW)$$

yielding a Euclidean space (of dimension $\kappa := mn + \frac{n(n+1)}{2}$) with induced norm $\|\cdot\| : \mathbb{E} \to \mathbb{R}$ given by

$$\|(X, V)\| = \sqrt{\operatorname{tr}(X^T X) + \operatorname{tr}(V^2)} = \sqrt{\|X\|_F^2 + \|V\|_F^2},$$

where $\|\cdot\|_F$ denotes the *Frobenius norm* on the respective space.

3 Optimization of quadratic functions

We now recall the elementary properties of the optimization problem (1) and its associated optimal value function (2).

Lemma 3.1 (Solvabilty of equality constrained QPs)**.** *For* $(x, V, A, b) \in \mathbb{R}^n \times \mathbb{S}^n \times \mathbb{R}^{p\times n} \times \mathbb{R}^p$ *consider the quadratic optimization problem* (1)*. Then* (1) *has a solution if and only if the following three conditions hold:*

i) $b \in \operatorname{rge} A$ *(i.e.* (1) *is feasible).*
ii) $x \in \operatorname{rge} \begin{bmatrix} V & A^T \end{bmatrix}$.
iii) $u^T V u \geq 0 \quad \forall u \in \ker A$.

If, however, $b \in \operatorname{rge} A$*, but ii) or iii) are violated, we have* $v(x, V, A, b) = -\infty$*.*

Proof. This is a standard result, see, for example, [4, Sec. 4.3]. □

In the sequel, we use the following notation:

$$V \succeq_{\ker A} 0 \quad :\Longleftrightarrow \quad u^T V u \geq 0 \quad \forall u \in \ker A$$

and

$$V \succ_{\ker A} 0 \quad :\Longleftrightarrow \quad u^T V u > 0 \quad \forall u \in \ker A \setminus \{0\}.$$

A complete description of the optimal value function v defined in (2) is given in our next result. This is the foundation on which our discussion of the support functional $\sigma_{\mathcal{D}(A,B)}$ is based.

Theorem 3.2. *At* (x, V)*, the optimal value-function* v *in* (2) *is given by*

$$\begin{cases} -\frac{1}{2}\begin{pmatrix} x \\ b \end{pmatrix}^T \begin{pmatrix} V & A^T \\ A & 0 \end{pmatrix}^\dagger \begin{pmatrix} x \\ b \end{pmatrix} & \text{if } \begin{pmatrix} x \\ b \end{pmatrix} \in \operatorname{rge} \begin{pmatrix} V & A^T \\ A & 0 \end{pmatrix},\ V \succeq_{\ker A} 0, \\ -\infty & \text{if } b \in \operatorname{rge} A \ \wedge\ (x \notin \operatorname{rge}[V\ A^T] \ \vee\ V \not\succeq_{\ker A} 0), \\ +\infty & \text{if } b \notin \operatorname{rge} A. \end{cases}$$

Proof. First, if $b \notin \operatorname{rge} A$, then (1) is infeasible, hence, by convention $v(V, x, A, b) = +\infty$. Second, if $b \in \operatorname{rge} A$, Lemma 3.1 tells us that if $x \notin \operatorname{rge} \begin{bmatrix} V\ A^T \end{bmatrix}$ or V is not positive semidefinite on $\ker A$, we have $v(V, x, A, b) = -\infty$. Hence, we need only show the expression for v when $\begin{pmatrix} x \\ b \end{pmatrix} \in \operatorname{rge} \begin{pmatrix} V & A^T \\ A & 0 \end{pmatrix}$ (i.e. $b \in \operatorname{rge} A$ and $x \in \operatorname{rge} \begin{bmatrix} V\ A^T \end{bmatrix}$) and V is positive semidefinite on $\ker A$. Again, Lemma 3.1 tells us that, in this case, a solution to (1) exists. The first-order necessary optimality conditions at

$$\bar{u} \in \operatorname*{argmin}_{u \in \mathbb{R}^n} \left\{ \frac{1}{2} u^T V u - x^T u \quad \text{s.t.} \quad Au = b \right\} \neq \emptyset$$

are that there exists $\bar{y} \in \mathbb{R}^p$ for which

$$\begin{pmatrix} V & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \bar{u} \\ \bar{y} \end{pmatrix} = \begin{pmatrix} x \\ b \end{pmatrix},$$

or equivalently,

$$\begin{pmatrix} \bar{u} \\ \bar{y} \end{pmatrix} \in \begin{pmatrix} V & A^T \\ A & 0 \end{pmatrix}^\dagger \begin{pmatrix} x \\ b \end{pmatrix} + \ker \begin{pmatrix} V & A^T \\ A & 0 \end{pmatrix}.$$

Plugging such a pair $\binom{\bar u}{\bar y}$ into the objective function yields

$$\begin{aligned}\frac{1}{2}\bar u^T V\bar u - x^T\bar u &= \frac{1}{2}\bar u^T\underbrace{(V\bar u - x)}_{=-A^T\bar y} - \frac{1}{2}x^T\bar u\\ &= -\frac{1}{2}\underbrace{\bar u^T A^T}_{=b^T}\bar y - \frac{1}{2}x^T\bar u\\ &= -\frac{1}{2}\binom{x}{b}^T\binom{\bar u}{\bar y}\\ &= -\frac{1}{2}\binom{x}{b}^T\begin{pmatrix} V & A^T\\ A & 0\end{pmatrix}^\dagger\binom{x}{b}^T,\end{aligned}$$

where the last equation is due to the fact that

$$\binom{x}{b}\in \operatorname{rge}\begin{pmatrix} V & A^T\\ A & 0\end{pmatrix} = \left(\ker\begin{pmatrix} V & A^T\\ A & 0\end{pmatrix}\right)^\perp.$$

Since all such points yield the same optimal value, this concludes the proof. □

The matrix

$$M(V) := \begin{pmatrix} V & A^T\\ A & 0\end{pmatrix} \tag{7}$$

in the foregoing theorem plays a pivotal role in our analysis. The next result sheds some light on properties of $M(V)$ which is often referred to as a *bordered matrix* in the literature [6, 8, 15].

Proposition 3.3 (Finsler's Lemma and bordered matrices)**.** *Let $A\in\mathbb{R}^{p\times n}$ and $V\in\mathbb{S}^n$. Then*

$$V\succ_{\ker A} 0 \quad\iff\quad \exists\varepsilon>0:\ V+\varepsilon A^TA\succ 0.$$

Moreover, the bordered matrix $M(V)$ is invertible if and only if $\operatorname{rank} A = p$ *and $V\succ_{\ker A} 0$ in which case its inverse $M(V)^{-1}$ is given by*

$$\begin{pmatrix} P(P^TVP)^{-1}P^T & \left(I-P(P^TVP)^{-1}P^TV\right)A^\dagger\\ (A^\dagger)^T\left(I-VP(P^TVP)^{-1}P^T\right) & (A^\dagger)^T\left(VP(P^TVP)^{-1}P^TV-V\right)A^\dagger\end{pmatrix},$$

where $P\in\mathbb{R}^{n\times(n-p)}$ is any matrix whose columns form an orthonormal basis of $\ker A$*.*

Proof. For the first statement see, e.g., [6, Th. 2] and for the characterization of the invertibility of $M(V)$ see, e.g., [6, Th. 7]. The inversion formula can be found in [8, Th. 1] and is easily verified by direct matrix multiplication. □

Remark 3.4. *A general formula for the pseudoinverse $M(V)^\dagger$ can be found in* [15, p. 58].

The first statement in Proposition 3.3 is referred to in the literature as *Finsler's Lemma* and originally goes back to [9].

Corollary 3.5. *Let $A\in\mathbb{R}^{p\times n}$ and $V\in\mathbb{S}^n$. Then the following hold:*

a) $V\succ_{\ker A} 0 \quad\Longrightarrow\quad \operatorname{rge}[V\ A^T]=\mathbb{R}^n$.
b) $V\succ_{\ker A} 0 \quad\iff\quad [V\succeq_{\ker A} 0]\ \wedge\ [\exists U\subset\mathbb{R}^p:\ \ker M(V)=\{0\}^n\times U]$, *in which case* $U=\ker A^T$.

Proof. a) Let $x \in \mathbb{R}^n$. Due to Finsler's Lemma, there exists $\varepsilon > 0$ and $r \in \mathbb{R}^n$ such that $(V + \varepsilon A^T A)r = x$. Putting $s := \varepsilon Ar$ gives $[V \ A^T]\binom{r}{s} = Vr + \varepsilon A^T Ar = x$, which proves a).
b) '$\Rightarrow$:' Let $\binom{x}{y} \in \ker M(V)$, i.e.

$$Vx + A^T y = 0 \quad \text{and} \quad Ax = 0.$$

Multiplying the first condition by x^T yields $x^T V x = 0$. By assumption, as $x \in \ker A$, we get $x = 0$. Hence, $y \in \ker A^T$, and thus $\ker M(V) = \{0\}^n \times \ker A^T$. The condition $V \succeq_{\ker A} 0$ holds trivially.
'$\Leftarrow$:' Let $x \in \ker A \setminus \{0\}$. By assumption $\ker M(V) = \{0\}^n \times U$, and so $\binom{x}{y} \notin \ker M(V)$ for all $y \in \mathbb{R}^p$. Thus, since $Ax = 0$, we must have $Vx + A^T y \neq 0$ for all $y \in \mathbb{R}^p$ or, equivalently, $Vx \notin \operatorname{rge} A^T = (\ker A)^\perp$. Hence, there exists $\hat{x} \in \ker A$ such that $0 < \hat{x}^T V x$. We now compute that for all $\varepsilon > 0$, we have

$$0 \leq (\varepsilon\hat{x} - x)^T V (\varepsilon\hat{x} - x) = \varepsilon^2 \hat{x}^T V \hat{x} - 2\varepsilon \hat{x}^T V x + x^T V x,$$

as $\varepsilon\hat{x} - x \in \ker A$ and $V \succeq_{\ker A} 0$ by assumption. Rearranging the terms and dividing by $\varepsilon > 0$ yields

$$0 < 2\hat{x}^T V x - \varepsilon \hat{x}^T V \hat{x} \leq \frac{x^T V x}{\varepsilon}$$

for all ε sufficiently small. Consequently, $x^T V x > 0$. □

4 The class of matrix support functionals $\sigma_{\mathcal{D}(A,B)}$

We are finally positioned to develop the properties of the support functional $\sigma_{\mathcal{D}(A,B)}$ defined in (3), where $A \in \mathbb{R}^{p\times n}$ and $B \in \mathbb{R}^{p\times m}$ are such that $\operatorname{rge} B \subset \operatorname{rge} A$. We begin by establishing a closed form expression for $\sigma_{\mathcal{D}(A,B)}$. The key fact used in this derivation is Theorem 3.2.

Theorem 4.1. *Let $A \in \mathbb{R}^{p\times n}$ and $B \in \mathbb{R}^{p\times m}$ such that $\operatorname{rge} B \subset \operatorname{rge} A$ and let $\mathcal{D}(A,B)$ be given by (4). Then*

$$\sigma_{\mathcal{D}(A,B)}(X,V) = \begin{cases} \frac{1}{2}\operatorname{tr}\left(\binom{X}{B}^T M(V)^\dagger \binom{X}{B}\right) & \text{if } \operatorname{rge}\binom{X}{B} \subset \operatorname{rge} M(V),\ V \succeq_{\ker A} 0, \\ +\infty & \text{else.} \end{cases}$$

In particular,

$$\operatorname{dom} \sigma_{D(A,B)} = \operatorname{dom} \partial\sigma_{D(A,B)} = \left\{ (X,V) \in \mathbb{E} \;\middle|\; \operatorname{rge}\begin{pmatrix} X \\ B \end{pmatrix} \subset \operatorname{rge} M(V),\ V \succeq_{\ker A} 0 \right\},$$

with this set being closed. Moreover, we have

$$\operatorname{int}(\operatorname{dom} \sigma_{D(A,B)}) = \{(X,V) \in \mathbb{E} \mid V \succ_{\ker A} 0\}.$$

Proof. By direct computation,

$$\begin{aligned}
\sigma_{\mathcal{D}(A,B)}(X,V) &= \sup_{(U,W)\in\mathcal{D}(A,B)} \langle (X,V),(U,W)\rangle \\
&= \sup_{U:\ AU=B} \left\{ \mathrm{tr}\,(X^TU) - \frac{1}{2}\mathrm{tr}\,(UU^TV)\right\} \\
&= \sup_{U:\ AU=B} \left\{ \mathrm{tr}\,(X^TU) - \frac{1}{2}\mathrm{tr}\,\left(\sum_{i=1}^m u^i(u^i)^TV\right)\right\} \\
&= \sup_{U:\ AU=B} \left\{ \sum_{i=1}^m (x^i)^Tu^i - \frac{1}{2}(u^i)^TVu^i\right\} \\
&= -\sum_{i=1}^m \left\{ \inf_{u:\ Au=b^i} \frac{1}{2}u^TVu - (x^i)^Tu\right\} \qquad (8)\\
&= \sum_{i=1}^m \begin{cases} \frac{1}{2}\binom{x^i}{b^i}^T M(V)^\dagger \binom{x^i}{b^i} & \text{if } \binom{x^i}{b^i}\in \mathrm{rge}\, M(V),\ V\succeq_{\ker A} 0,\\ +\infty & \text{else}\end{cases} \\
&= \begin{cases} \frac{1}{2}\mathrm{tr}\left(\begin{pmatrix} X\\ B\end{pmatrix}^T M(V)^\dagger \begin{pmatrix} X\\ B\end{pmatrix}\right) & \text{if } \mathrm{rge}\begin{pmatrix} X\\ B\end{pmatrix}\subset \mathrm{rge}\, M(V),\ V\succeq_{\ker A} 0,\\ +\infty & \text{else.}\end{cases}
\end{aligned}$$

Here, the sixth equation exploits Theorem 3.2. This establishes the representation for $\sigma_{\mathcal{D}(A,B)}$ as well as for its domain. In addition, since

$$\partial\sigma_{\mathcal{D}(A,B)}(X,V) = \arg\max\,\{\langle (X,V),\,(U,W)\rangle \mid (U,W)\in\mathcal{D}(A,B)\},$$

Theorem 3.2 also yields the equivalence $\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)} = \mathrm{dom}\,\partial\sigma_{\mathcal{D}(A,B)}$.

In order to see that $\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}$ is closed, let $\{(X_k,V_k)\in \mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}\}\to (V,X)$. In particular, $\mathrm{rge}\,X_k\subset \mathrm{rge}\,[V_k\ \ A^T]$ and $V_k\succeq_{\ker A} 0$. These properties are preserved by passing to the limit, hence $(X,V)\in\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}$, i.e., $\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}$ is closed.

It remains to prove the expression for $\mathrm{int}\,(\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)})$. First we show $O := \{(X,V)\in\mathbb{E}\mid V\succ_{\ker A} 0\}$ is open. For this purpose, let $(X,V)\in O$. Suppose $q:=\mathrm{rank}\,A$ and let $P\in\mathbb{R}^{n\times(n-q)}$ be such that its columns form a basis of $\ker A$ so that P^TVP is positive definite. Using the fact that $\mathbb{S}^n_{++} = \mathrm{int}\,\mathbb{S}^n_+$ (e.g., see [3, Ch. 1, Ex. 1]), there exists an $\varepsilon>0$ such that

$$\|W-V\|\le\varepsilon \quad\Rightarrow\quad P^TWP\succ 0 \quad \forall W\in\mathbb{S}^n.$$

Hence, $B_\varepsilon(X,V)\subset O$ so that O is open.

Next, we show that $O\subset \mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}$. Again let $(X,V)\in O$, so that $V\succ_{\ker A} 0$. By Corollary 3.5, $\mathrm{rge}\,[V\ \ A^T]=\mathbb{R}^n$ and so $\mathrm{rge}\,X\subset\mathrm{rge}\,[V\ \ A^T]$. Hence $(X,V)\in \mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}$ yielding $O\subset \mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}$. All in all, since O is open and $O\subset \mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}$, we have $O\subset \mathrm{int}\,\big(\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}\big)$.

We now show the reverse inclusion. Let $(X,V)\in\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}\setminus O$. Hence V is positive semidefinite, but not positive definite on $\ker A$. Therefore, there exists $x\in\ker A\setminus\{0\}$ such that $x^TVx=0$. Define

$$V_k := V - \frac{1}{k}I\in\mathbb{S}^n \quad (k\in\mathbb{N}).$$

Then, for all $k \in \mathbb{N}$, $x^T V_k x = -\frac{\|x\|^2}{k} < 0$, so that $(X, V_k) \notin \operatorname{dom} \sigma_{\mathcal{D}(A,B)}$ for all $k \in \mathbb{N}$ while $(X, V_k) \to (X, V)$. Hence $(X, V) \in \operatorname{bd}\left(\operatorname{dom} \sigma_{\mathcal{D}(A,B)}\right)$, that is, $(X, V) \notin \operatorname{int}(\operatorname{dom} \sigma_{\mathcal{D}(A,B)})$. Consequently, $O \supset \operatorname{int}\left(\operatorname{dom} \sigma_{\mathcal{D}(A,B)}\right)$. □

Our next goal is to compute the subdifferential of $\sigma_{D(A,B)}$. For this purpose, we make use of an alternative representation of the closed convex hull of $\mathcal{D}(A, B)$ which we now develop. Set

$$\mathcal{F}(A,B) := \left\{ (d,Z) \in \mathbb{R}^{\kappa+1} \times \mathbb{R}^{n \times m(\kappa+1)} \;\middle|\; \begin{array}{l} d \geq 0,\ \|d\| = 1, \\ AZ_i = d_i B \ (i = 1, \dots, \kappa+1) \end{array} \right\}, \quad (9)$$

where, for $(d, Z) \in \mathcal{F}(A, B)$, we interprete Z as a block matrix of the form

$$Z = [Z_1, \dots, Z_{\kappa+1}], \quad Z_i \in \mathbb{R}^{n \times m} \ (i = 1, \dots, \kappa+1).$$

Lemma 4.2. *For the set $\mathcal{D}(A, B)$ defined in (4), we have*

$$\operatorname{conv}\mathcal{D}(A,B) = \left\{ \left(Z(d \otimes I_m), -\frac{1}{2} Z Z^T \right) \,\middle|\, (d,Z) \in \mathcal{F}(A,B) \colon Z_i = 0 \ (i \notin \operatorname{supp} d) \right\}. \quad (10)$$

Proof. Recall that $\dim \mathbb{E} = \kappa$, hence, by Carathéodory's Theorem (see Section 2.1), every element of $\operatorname{conv} \mathcal{D}(A, B)$ can be represented as a convex combination of no more than $\kappa + 1$ elements of $\mathcal{D}(A, B)$.

In order to see the $\subset$ −inclusion, let $(Y, -\frac{1}{2} W) \in \operatorname{conv} \mathcal{D}(A, B)$. Then there exist $Y_i \in \mathbb{R}^{n \times m}$ with $AY_i = B$ $(i = 1, \dots, \kappa + 1)$ and $\lambda \in \Delta_{\kappa+1}$ such that

$$Y = \sum_{i=1}^{\kappa+1} \lambda_i Y_i \quad \text{and} \quad W = \sum_{i=1}^{\kappa+1} \lambda_i Y_i Y_i^T.$$

For $i = 1, \dots, \kappa+1$ define $d_i := \sqrt{\lambda_i}$ and $Z_i := \sqrt{\lambda_i} Y_i$. With $Z = [Z_1, Z_2, \dots, Z_{\kappa+1}]$ and $d = (d_1, \dots, d_{\kappa+1})^T$, we have

$$\|d\|^2 = \sum_{i=1}^{\kappa+1} \lambda_i = 1 \quad \text{and} \quad AZ_i = d_i B \quad (i = 1, \dots, \kappa+1).$$

That is,

$$(d, Z) \in \mathcal{F}(A, B) \quad \text{and} \quad Z_i = 0 \quad (i \notin \operatorname{supp} d).$$

Moreover, it follows that

$$Z(d \otimes I_m) = \sum_{i=1}^{\kappa+1} d_i Z_i = \sum_{i=1}^{\kappa+1} \lambda_i Y_i = Y$$

and

$$Z Z^T = \sum_{i=1}^{\kappa+1} Z_i Z_i^T = \sum_{i=1}^{\kappa+1} \lambda_i Y_i Y_i^T = W.$$

Hence, $\operatorname{conv} \mathcal{D}(A, B)$ is contained in the set on the right-hand side of (10).
To show the reverse inclusion, suppose that $(Z(d \otimes I_m), -\frac{1}{2} Z Z^T)$ is an element of the set on the right-hand side of (10), and define

$$\lambda_i := d_i^2 \quad \text{and} \quad Y_i := \begin{cases} \frac{1}{d_i} Z_i & \text{if } i \in \operatorname{supp} d, \\ A^\dagger B & \text{else} \end{cases} \qquad (i = 1, \dots, \kappa+1).$$

Then, in particular, $\lambda \in \Delta_{\kappa+1}$ and $AY_i = B$ for all $i = 1, \dots, \kappa+1$. Finally, we have

$$\sum_{i=1}^{\kappa+1} \lambda_i Y_i = \sum_{i\in \operatorname{supp} d} d_i Z_i = \sum_{i=1}^{\kappa+1} d_i Z_i = Z(d\otimes I_m)$$

and

$$\sum_{i=1}^{\kappa+1} \lambda_i Y_i Y_i^T = \sum_{i\in \operatorname{supp} d} Z_i Z_i^T = \sum_{i=1}^{\kappa+1} Z_i Z_i^T,$$

where we exploit the fact that $Z_i = 0$ if $i \notin \operatorname{supp} d$. Hence

$$\left(Z(d\otimes I_m), -\frac{1}{2}ZZ^T\right) \in \operatorname{conv} \mathcal{D}(A,B)$$

which concludes the proof. □

Next, we compute the closed convex hull of $\mathcal{D}(A,B)$.

Proposition 4.3. *For the set $\mathcal{D}(A,B)$ defined in* (4)*, we have*

$$\overline{\operatorname{conv}}\, \mathcal{D}(A,B) = \left\{ (Z(d\otimes I_m), -\frac{1}{2}ZZ^T) \mid (d,Z) \in \mathcal{F}(A,B) \right\}. \tag{11}$$

Proof. The $\subset$ −inclusion follows as soon as we establish that the set on the right-hand side is closed (as it obviously contains $\operatorname{conv} \mathcal{D}(A,B)$): For this purpose, suppose $(W, -\frac{1}{2}H) \in \mathbb{E}$ is such that there is a sequence $\{(Z_k(d^k \otimes I_m), -\frac{1}{2}Z_k Z_k^T)\} \to (W, -\frac{1}{2}H)$ with $(d^k, Z_k) \in \mathcal{F}(A,B)$ for all $k \in \mathbb{N}$. Let $U_k \Sigma_k Q_k^T = Z_k$ be the singular value decomposition of Z_k with $U_k^T U_k = I_n$ and $Q_k^T Q_k = I_{m(\kappa+1)}$ for all $k \in \mathbb{N}$. For each $k \in \mathbb{N}$, define $\hat{\Sigma}_k \in \mathbb{R}^{n\times n}$ to be the non-negative diagonal matrix comprised of the first n columns of Σ_k. Note that the diagonal of $\hat{\Sigma}_k$ contains all of the nonzero singular values of Z_k. Since $U_k \hat{\Sigma}_k^2 U_k^T = Z_k Z_k^T \to H$ with U_k orthogonal, it must be the case that $\{\Sigma_k\}$ is bounded. Hence, with no loss in generality, there exist $(U, Q, \Sigma, d) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{m(\kappa+1)\times m(\kappa+1)} \times \mathbb{R}_+^{n\times m} \times \mathbb{R}_+^{(\kappa+1)}$ with U and Q orthogonal, Σ diagonal, and $\|d\|_2 = 1$ such that $(U_k, Q_k, \Sigma_k, d^k) \to (U, Q, \Sigma, d)$.

In particular, it holds that $Z_k \to Z := U\Sigma Q^T$, $ZZ^T = H$, $W = Z(d\otimes I_m)$ and, clearly, $AZ_i = d_i B$ for all $i = 1, \dots, \kappa+1$. This shows that $(W, -\frac{1}{2}H)$ is an element of the set on the right-hand side of (11) and so this set is closed.

To prove the reverse inclusion, let $(d,Z) \in \mathcal{F}(A,B)$. We are done once we find a sequence $\{(d^k, Z_k) \in \mathcal{F}(A,B)\} \to (d,Z)$ such that $Z_i^k = 0$ if $d_i^k = 0$. To this end, define the index set

$$Q := \{i \in \{1, \dots, \kappa+1\} \mid d_i = 0,\ Z_i \neq 0\},$$

and put $q := |Q|$. If $q = 0$ (i.e. $Q = \emptyset$), the definitions $d^k := d$ and $Z^k := Z$ for all $k \in \mathbb{N}$ will suffice.

Otherwise, choose $i_0 \in \operatorname{supp} d$ (which is nonempty, as $\|d\| = 1$). Then define

$$d_i^k := \begin{cases} 0 & \text{if} \quad d_i = 0, Z_i = 0, \\ \sqrt{\frac{2k-1}{q}\frac{d_{i_0}}{k}} & \text{if} \quad i \in Q, \\ \frac{k-1}{k} d_{i_0} & \text{if} \quad i = i_0, \\ d_i & \text{if} \quad i \in \operatorname{supp} d \setminus \{i_0\} \end{cases}$$

and

$$Z_i^k := \begin{cases} Z_i & \text{if} \quad d_i = 0, Z_i = 0, \\ d_i^k A^\dagger B + Z_i & \text{if} \quad i \in Q, \\ \frac{d_{i_0}^k}{d_{i_0}} Z_{i_0} & \text{if} \quad i = i_0, \\ Z_i & \text{if} \quad i \in \operatorname{supp} d \setminus \{i_0\}. \end{cases}$$

Then $d^k \to d$ and $Z^k \to Z$. Also, we have $AZ_i^k = d_i^k B$ for all $i = 1, \dots, \kappa+1$, and $Z_i^k = Z_i = 0$ if $d_i^k = 0$. In addition,

$$\begin{aligned} \|d^k\|^2 &= \sum_{i \in Q} \frac{2k-1}{qk^2} d_{i_0}^2 + \frac{(k-1)^2}{k^2} d_{i_0}^2 + \sum_{i \in \operatorname{supp} d \setminus \{i_0\}} d_i^2 \\ &= \frac{2k-1}{k^2} d_{i_0}^2 + \frac{(k-1)^2}{k^2} d_{i_0}^2 + \sum_{i \in \operatorname{supp} d \setminus \{i_0\}} d_i^2 \\ &= d_{i_0}^2 + \sum_{i \in \operatorname{supp} d \setminus \{i_0\}} d_i^2 \\ &= \sum_{i \in \operatorname{supp} d} d_i^2 \\ &= \|d\|^2 \\ &= 1. \end{aligned}$$

This shows that $\{(d^k, Z^k) \in \operatorname{conv} \mathcal{D}(A,B)\} \to (d, Z)$, which gives the desired inclusion concluding the proof. □

As a simple consequence of the bi-conjugacy relationship in (5), we obtain the following corollary.

Corollary 4.4. *For $A \in \mathbb{R}^{p \times n}$ and $B \in \mathbb{R}^{p \times m}$ with $\operatorname{rge} B \in \operatorname{rge} A$, the conjugate of the support functional $\sigma_{\mathcal{D}(A,B)}$ for the set $\mathcal{D}(A,B)$ from (4) is given by*

$$\sigma^*_{\mathcal{D}(A,B)} = \delta_{\overline{\operatorname{conv}}\, \mathcal{D}(A,B)},$$

where $\overline{\operatorname{conv}}\, \mathcal{D}(A,B)$ is provided by Proposition 4.3.

To describe the subdifferential of $\sigma_{D(A,B)}$, we first compute the normal cone of the set $\mathcal{F}(A,B)$ defined in (9). Our approach uses the following preparatory lemmas.

Lemma 4.5. *Let $\mathcal{F} := \mathcal{F}(0,0)$ be defined through (9) and let $(d, Z) \in \mathcal{F}$. Then*

$$N_{\mathcal{F}}(d,Z) = \left[\operatorname{span} d + \bigtimes_{i=1}^{\kappa+1} \begin{cases} \{0\} & \text{if} \quad d_i > 0, \\ \mathbb{R}_-. & \text{if} \quad d_i = 0, \end{cases} \right] \times \{0\},$$

and $\mathcal{F}$ is regular.

Proof. Let $(d, Z) \in \mathcal{F} = \left(\mathcal{S}_\kappa \cap \mathbb{R}_+^{\kappa+1}\right) \times \mathbb{R}^{n \times m(\kappa+1)}$. Due to [16, Prop. 6.41], we have

$$N_{\mathcal{F}}(d,Z) = N_{\mathcal{S}_\kappa \cap \mathbb{R}_+^{\kappa+1}}(d) \times N_{\mathbb{R}^{n \times m(\kappa+1)}}(Z).$$

Clearly, $N_{\mathbb{R}^{n \times m(\kappa+1)}}(Z) = \{0\}$, and by invoking [16, Ex. 6.8], we find that $\mathcal{S}_\kappa$ is regular with

$$N_{\mathcal{S}_\kappa}(d) = \operatorname{span} d.$$

Again by [16, Prop. 6.41] in conjunction with [16, Ex. 6.10], we have

$$N_{\mathbb{R}_+^{\kappa+1}}(d) = \bigtimes_{i=1}^{\kappa+1} N_{[0,\infty)}(d_i) = \bigtimes_{i=1}^{\kappa+1} \begin{cases} \{0\} & \text{if} \quad d_i > 0, \\ \mathbb{R}_-, & \text{if} \quad d_i = 0. \end{cases}$$

In particular, we have the following implication:

$$[v \in N_{\mathcal{S}_\kappa}(d), \quad w \in N_{\mathbb{R}_+^{\kappa+1}}(d): \quad v + w = 0] \quad \Longrightarrow \quad v = w = 0.$$

The desired expression for $N_{\mathcal{S}_\kappa \cap \mathbb{R}_+^{\kappa+1}}$ now follows from [16, Th. 6.42], since $\mathcal{S}_\kappa$ is regular (being a smooth manifold) and so is $\mathbb{R}_+^{\kappa+1}$ (as a convex set). This result also tells us that $\mathcal{S}_\kappa \cap \mathbb{R}_+^{\kappa+1}$ is regular; hence $\mathcal{F}$ is also regular (cf. [16, Prop. 6.41]). □

Lemma 4.6. *For $A \in \mathbb{R}^{p\times n}$ and $B \in \mathbb{R}^{p\times m}$ consider the linear operator $J : \mathbb{R}^{\kappa+1} \times \mathbb{R}^{n\times m(\kappa+1)} \to \mathbb{R}^{p\times m(\kappa+1)}$ defined by*

$$J(d, Z) := (AZ_1 - d_1 B, \dots, AZ_{\kappa+1} - d_{\kappa+1} B).$$

The adjoint $J^* : \mathbb{R}^{p\times m(\kappa+1)} \to \mathbb{R}^{\kappa+1} \times \mathbb{R}^{n\times m(\kappa+1)}$ (w.r.t. the inner products introduced on matrix product-spaces in Section 2.2) is given by

$$J^*(W) = \left[-\begin{pmatrix} \mathrm{tr}\,(B^T W_1) \\ \vdots \\ \mathrm{tr}\,(B^T W_{\kappa+1}) \end{pmatrix}, \left[A^T W_1, \dots, A^T W_{\kappa+1}\right] \right].$$

Proof. Let (d, Z) and W be arbitrary elements of $\mathbb{R}^{\kappa+1} \times \mathbb{R}^{n\times m(\kappa+1)}$ and $\mathbb{R}^{p\times m(\kappa+1)}$, respectively, having conformal decompositions $d \in \mathbb{R}^{\kappa+1}$, $Z = [Z_1, \dots, Z_{\kappa+1}]$ with $Z_i \in \mathbb{R}^{n\times m}$, $i = 1, \dots, \kappa+1$, and $W = [W_1, \dots, W_{\kappa+1}]$ with $W_i \in \mathbb{R}^{p\times m}$, $i = 1, \dots, \kappa+1$. Then

$$\begin{aligned} \langle W, J(d, Z)\rangle &= \sum_{i=1}^{\kappa+1} \mathrm{tr}\,\left(W_i^T (AZ_i - d_i B)\right) \\ &= \sum_{i=1}^{\kappa+1} \mathrm{tr}\,\left((A^T W_i) Z_i\right) - \sum_{i=1}^{\kappa+1} d_i \mathrm{tr}\,(B^T W_i) \\ &= \left\langle [-(\langle B, W_1\rangle, \dots, \langle B, W_{\kappa+1}\rangle)^T, [A^T W_1, \dots, A^T W_{\kappa+1}], (d, Z) \right\rangle, \end{aligned}$$

which proves the result. □

We now establish a normal cone formula for the set $\mathcal{F}(A, B)$.

Proposition 4.7. *Let $A \in \mathbb{R}^{p\times n}$ and $B \in \mathbb{R}^{p\times m}$ such that $\mathrm{rge}\, B \subset \mathrm{rge}\, A$. Then*

$$N_{\mathcal{F}(A,B)}(d, Z) = N_{\mathcal{F}}(d, Z) + \mathrm{rge}\, J^* \quad \forall (d, Z) \in \mathcal{F}(A, B),$$

with $\mathcal{F}$ and J given by Lemmas 4.5 and 4.6, respectively.

Proof. We first note that

$$\mathcal{F}(A, B) = \mathcal{F} \cap \ker J.$$

Moreover, $\mathcal{F}$ is regular by Lemma 4.5 and so is $\ker J$ as a subspace. The assertion will follow from [16, Th. 6.42] as soon as we establish the implication

$$(r, R) + (s, S) = 0 \;\Rightarrow\; (r, R) = (s, S) = 0, \tag{12}$$

for all $(r, R) \in N_{\mathcal{F}}(d, Z)$ and $(s, S) \in N_{\ker T}(d, Z)$. For these purposes, first note that $N_{\ker J}(d, Z) = \mathrm{rge}\, J^*$, see [16, Ex. 6.8]. Now, let $(r, R) \in N_{\mathcal{F}}(d, Z)$ and

$(s,S)\in N_{\ker T}(d,Z)$ such that $(r,R)+(s,S)=0$: By the representation of J^* from Lemma 4.6, we know that there exists $W\in R^{p\times m(\kappa+1)}$ such that

$$s=-\begin{pmatrix}\mathrm{tr}\,(B^TW_1)\\ \vdots\\ \mathrm{tr}\,(B^TW_{\kappa+1})\end{pmatrix}\quad\text{and}\quad S=\left[A^TW_1,\dots,A^TW_{\kappa+1}\right].$$

Moreover, from the representation of $N_{\mathcal{F}}(d,Z)$ from Lemma 4.5 we know that $R=0$, hence $S=0$, i.e. $A^TW_i=0$ for all $i=1,\dots,\kappa+1$. This, in turn, implies that

$$\mathrm{rge}\,W_i\subset\ker A^T\subset\ker B^T\quad\forall i=1,\cdots\kappa+1,$$

thus $B^TW_i=0$ for all $i=1,\dots,\kappa+1$. This immediately implies $s=0$, hence $r=0$, which proves the implication in (12). This concludes the proof. □

A formula for the the subdifferential $\partial\sigma_{\mathcal{D}(A,B)}$ follows.

Theorem 4.8. *Let $A\in\mathbb{R}^{p\times n}$ and $B\in\mathbb{R}^{p\times m}$ such that $\mathrm{rge}\,B\subset\mathrm{rge}\,A$. Then, for all $(X,V)\in\mathrm{dom}\,\sigma_{D(A,B)}$, the subdifferential of $\sigma_{D(A,B)}$ at (X,V) is given by*

$$\left\{\left(\sum_{i=1}^{\kappa+1}\bar d_i\bar Z_i,-\frac12\sum_{i=1}^{\kappa+1}\bar Z_i\bar Z_i^T\right)\left|\begin{array}{l}\begin{pmatrix}\bar Z_i\\ \bar W_i\end{pmatrix}=\bar d_iM(V)^\dagger\begin{pmatrix}X\\ B\end{pmatrix}+\begin{pmatrix}N_{1i}\\ N_{2i}\end{pmatrix},\ i=1,2,\dots,\kappa+1\\ \text{for some }\bar W,N_2\in\mathbb{R}^{p\times m(\kappa+1)},\ N_1\in\mathbb{R}^{n\times m(\kappa+1)},\\ \text{with }(\bar d,\bar Z)\in\mathcal{F}(A,B)\text{ and }\mathrm{rge}\begin{pmatrix}N_1\\ N_2\end{pmatrix}\subset\ker M(V)\end{array}\right.\right\}.$$

Moreover, $\sigma_{\mathcal{D}(A,B)}$ is continuously differentiable on $\mathrm{int}\,(\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)})$.

Proof. Let $(X,V)\in\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}$, i.e., $\mathrm{rge}\binom{X}{B}\subset\mathrm{rge}\,M(V)$ and V is positive semidefinite on $\ker A$. Then

$$\begin{aligned}
&(\bar Y,\bar W)\in\partial\sigma_{\mathcal{D}(A,B)}(X,V)\\
&\Leftrightarrow(\bar Y,\bar W)\in\overline{\mathrm{conv}}\,\mathcal{D}(A,B)\quad\text{and}\quad(X,V)\in N_{\overline{\mathrm{conv}}\,\mathcal{D}(A,B)}(\bar Y,\bar W)\\
&\Leftrightarrow(\bar Y,\bar W)\in\overline{\mathrm{conv}}\,\mathcal{D}(A,B)\quad\text{and}\\
&\qquad 0\ge\langle(X,V),(Y,W)-(\bar Y,\bar W)\rangle\quad\forall(Y,W)\in\overline{\mathrm{conv}}\,\mathcal{D}(A,B)\\
&\Leftrightarrow\exists(\bar d,\bar Z)\in\mathcal{F}(A,B):\ \bar Y=\sum_{i=1}^{\kappa+1}\bar d_i\bar Z_i,\ \bar W=-\frac12\sum_{i=1}^{\kappa+1}\bar Z_i\bar Z_i^T,\ \text{and}\\
&\qquad 0\ge\sum_{i=1}^{\kappa+1}\mathrm{tr}\left(X^T(d_iZ_i-\bar d_i\bar Z_i)\right)-\frac12\sum_{i=1}^{\kappa+1}\mathrm{tr}\left(V(Z_iZ_i^T-\bar Z_i\bar Z_i^T)\right)\ \forall\,(d,Z)\in\mathcal{F}(A,B)\\
&\Leftrightarrow\exists(\bar d,\bar Z)\in\underset{(d,Z)\in\mathcal{F}(A,B)}{\mathrm{argmin}}\ f(d,Z):\ \bar Y=\sum_{i=1}^{\kappa+1}\bar d_i\bar Z_i,\ \bar W=-\frac12\sum_{i=1}^{\kappa+1}\bar Z_i\bar Z_i^T,
\end{aligned}$$

where $f:\mathbb{R}^{\kappa+1}\times\mathbb{R}^{n\times m(\kappa+1)}\to\mathbb{R}$ is defined by

$$f(d,V):=\frac12\sum_{i=1}^{\kappa+1}\mathrm{tr}\,(Z_i^TVZ_i)-\sum_{i=1}^{\kappa+1}d_i\mathrm{tr}\,(X^TZ_i).$$

Here, the first equivalence uses [16, Cor. 8.25], and the third one exploits the description of $\overline{\mathrm{conv}}\,\mathcal{D}(A,B)$ from Proposition 4.3.

The necessary optimality conditions for $(\bar{d}, \bar{Z})$ to be a minimizer of f over $\mathcal{F}(A, B)$ are

$$0 \in f'(\bar{d}, \bar{Z}) + N_{\mathcal{F}(A,B)}(\bar{d}, \bar{Z}) \quad \text{with} \quad (\bar{d}, \bar{Z}) \in \mathcal{F}(A, B), \tag{13}$$

which, invoking Proposition 4.7 (as well as Lemma 4.5 and Lemma 4.6), reads

$$\exists \bar{\mu} \in \mathbb{R}, \bar{W} \in \mathbb{R}^{p \times m(\kappa+1)} : \quad \begin{array}{ll} \bar{d}_i \mathrm{tr}\left(\binom{X}{B}^T \binom{\bar{Z}_i}{\bar{W}_i}\right) = \bar{\mu}\bar{d}_i & (\bar{d}_i > 0), \\ \mathrm{tr}\left(\binom{X}{B}^T \binom{\bar{Z}_i}{\bar{W}_i}\right) \leq 0 & (\bar{d}_i = 0), \\ M(V)\binom{\bar{Z}_i}{\bar{W}_i} = \bar{d}_i \binom{X}{B} & (i = 1, \ldots, \kappa+1), \\ d \in \mathcal{S}_\kappa \cap \mathbb{R}^{\kappa+1}_+ . & \end{array} \tag{14}$$

The third condition gives

$$\begin{pmatrix} \bar{Z}_i \\ \bar{W}_i \end{pmatrix} = \bar{d}_i M(V)^\dagger \begin{pmatrix} X \\ B \end{pmatrix} + \begin{pmatrix} N_{1i} \\ N_{2i} \end{pmatrix} \quad \text{with} \quad \mathrm{rge}\begin{pmatrix} N_{1i} \\ N_{2i} \end{pmatrix} \subset \ker M(V), \tag{15}$$

for some $N_{1i} \in \mathbb{R}^{n \times m}$ and $N_{2i} \in \mathbb{R}^{p \times m}$ $(i = 1, \ldots, \kappa + 1)$. In particular, as $\ker M(V) \subset \ker \binom{X}{B}^T$, we get

$$\mathrm{tr}\left(\begin{pmatrix} X \\ B \end{pmatrix}^T \begin{pmatrix} \bar{Z}_i \\ \bar{W}_i \end{pmatrix}\right) = \bar{d}_i \mathrm{tr}\left(\begin{pmatrix} X \\ B \end{pmatrix}^T M(V)^\dagger \begin{pmatrix} X \\ B \end{pmatrix}\right) \ (i = 1, \ldots, \kappa + 1),$$

which tells us that the first condition in (14) is generically fulfilled with

$$\bar{\mu} := \mathrm{tr}\left(\begin{pmatrix} X \\ B \end{pmatrix}^T M(V)^\dagger \begin{pmatrix} X \\ B \end{pmatrix}\right).$$

Moreover, if $\bar{d}_i = 0$, we get from the third condition in (14) that

$$\mathrm{rge}\begin{pmatrix} \bar{Z}_i \\ \bar{W}_i \end{pmatrix} \subset \ker M(V) \subset \ker \begin{pmatrix} X \\ B \end{pmatrix} = \left(\mathrm{rge}\begin{pmatrix} X \\ B \end{pmatrix}^T\right)^\perp,$$

and hence

$$\mathrm{tr}\left(\begin{pmatrix} X \\ B \end{pmatrix}^T \begin{pmatrix} \bar{Z}_i \\ \bar{W}_i \end{pmatrix}\right) = 0 \quad \forall i : \bar{d}_i = 0,$$

i.e., the second condition in (14) is also generically fulfilled. This means that the set of all critical points of f over $\mathcal{F}(A, B)$ is given by

$$\mathcal{C} := \left\{(\bar{d}, \bar{Z}) \in \mathcal{F}(A, B) \,\middle|\, \exists\, \bar{W} \in \mathbb{R}^{p \times m(\kappa+1)} : \mathrm{rge}\begin{pmatrix} \bar{Z}_i \\ \bar{W}_i \end{pmatrix} \subset \bar{d}_i M(V)^\dagger \begin{pmatrix} X \\ B \end{pmatrix} + \ker M(V)\right\}.$$

Now, we pick an arbitrary critical point $(\bar{d},\bar{Z})\in\mathcal{C}$ and plug it into the objective function f. This gives

$$
\begin{aligned}
f(\bar{d},\bar{Z}) &= \frac{1}{2}\sum_{i=1}^{\kappa+1}\mathrm{tr}\,(\bar{Z}_i(\underbrace{V\bar{Z}_i-\bar{d}_iX}_{=-A^T\bar{W}_i})) - \frac{1}{2}\sum_{i=1}^{\kappa+1}\bar{d}_i\mathrm{tr}\,\left(X^T\bar{Z}_i\right)\\
&= -\frac{1}{2}\sum_{i=1}^{\kappa+1}\mathrm{tr}\,((\underbrace{A\bar{Z}_i}_{=\bar{d}_iB})^T\bar{W}_i) - \frac{1}{2}\sum_{i=1}^{\kappa+1}\bar{d}_i\mathrm{tr}\,\left(X^T\bar{Z}_i\right)\\
&= -\frac{1}{2}\sum_{i=1}^{\kappa+1}\bar{d}_i\left[\mathrm{tr}\,(B^T\bar{W}_i)+\mathrm{tr}\,(\bar{Z}_i^TX)\right]\\
&= -\frac{1}{2}\sum_{i=1}^{\kappa+1}\mathrm{tr}\,\left(\begin{pmatrix}X\\B\end{pmatrix}^T\begin{pmatrix}\bar{Z}_i\\\bar{W}_i\end{pmatrix}\right)\\
&= -\frac{1}{2}\sum_{i=1}^{\kappa+1}\bar{d}_i^2\mathrm{tr}\,\left(\begin{pmatrix}X\\B\end{pmatrix}^T M(V)^\dagger\begin{pmatrix}X\\B\end{pmatrix}\right)\\
&= -\frac{1}{2}\mathrm{tr}\,\left(\begin{pmatrix}X\\B\end{pmatrix}^T M(V)^\dagger\begin{pmatrix}X\\B\end{pmatrix}\right).
\end{aligned}
$$

Hence, every critical point has the same function value. Theorem 4.1 tells us that $\partial\sigma_{\mathcal{D}(A,B)}(X,V)\neq\emptyset$ for $(X,V)\in\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}$. In view of the chain of equivalences at the beginning of this proof, we know that there is a minimizer of f over $\mathcal{F}(A,B)$, which is in particular a critical point. Therefore, all critical points are minimizers of f over $\mathcal{F}(A,B)$, and the expression for the subdifferential follows.

In order to prove the smoothness assertion, we first note that, clearly, the maximally possible set of continuous differentiability of $\sigma_{\mathcal{D}(A,B)}$ is $\mathrm{int}\,(\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)})$. Moreover, we recall the well-known fact, that a proper, lsc, convex function is continuously differentiable at a point in the interior of its domain if and only if the subdifferential is a singleton, which hence comprises its gradient (cf., e.g., [16, Th. 9.18]). Furthermore, we have the following chain of equivalences:

$$
\begin{aligned}
\partial\sigma_{\mathcal{D}(A,B)}(V,X)\text{ is a singleton} &\iff \exists\, U\subset\mathbb{R}^p:\ \ker M(V)=\{0\}^n\times U\\
&\iff V\succ_{\ker A}0\\
&\iff (X,V)\in\mathrm{int}\,\left(\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}\right).
\end{aligned}
$$

Here, the first equivalence is obtained from the representation of $\partial\sigma_{\mathcal{D}(A,B)}$ that was proven above. The second equivalence is due to Corollary 3.5 b), bearing in mind that $(X,V)\in\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}$, hence $V\succeq_{\ker A}0$. The third equivalence simply uses the representation of $\mathrm{int}\,(\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)})$ given in Theorem 4.1. □

Corollary 4.9. *Let $A\in\mathbb{R}^{p\times n}$ and $B\in\mathbb{R}^{p\times m}$ such that $\mathrm{rge}\,B\subset\mathrm{rge}\,A$ and $\mathrm{rank}\,A=p$. Then $\sigma_{\mathcal{D}(A,B)}$ is continuously differentiable on $\mathrm{int}\,(\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)})$ with*

$$\sigma'_{\mathcal{D}(A,B)}(X,V)=\left(Y(X,V),-\frac{1}{2}Y(X,V)Y(X,V)^T\right)\ \forall\,(X,V)\in\mathrm{int}\,(\mathrm{dom}\,\sigma_{\mathcal{D}(A,B)}),$$

where

$$Y(X,V):=A^\dagger B+P(P^TVP)^{-1}P^T[X-VA^\dagger B]$$

and $P\in\mathbb{R}^{n\times(n-p)}$ is such that its columns form an orthonormal basis of $\ker A$.

Proof. Apply the inversion formula in Proposition 3.3 to the subdifferential of $\sigma_{\mathcal{D}(A,B)}$ from Theorem 4.8, bearing in mind, that $\sum_{i=1}^{\kappa+1}\bar{d}_i^2 = 1$ for $(\bar{d},\bar{Z}) \in \mathcal{F}(A,B)$. □

5 Applications

5.1 The optimal value function v

Equation (8) in the proof of Theorem 4.1 tells us that the optimal value function v defined in (2) satisfies $v = -\sigma_{\mathcal{D}(A,b)}$. As a consequence we have the following theorem.

Theorem 5.1. *Let $v : \mathbb{R}^n \times \mathbb{S}^n \to \bar{\mathbb{R}}$ be the optimal value function for the constrained quadratic optimization problem*

$$v(x,V) := \inf_{u\in\mathbb{R}^n}\left\{\frac{1}{2}u^TVu - x^Tu \mid Au = b\right\},$$

where $A \in \mathbb{R}^{p\times n}$ and $b \in \mathbb{R}^p$ are such that $b \in \operatorname{rge} A$. Then

$$\begin{aligned} v(x,V) &= -\sigma_{\mathcal{D}(A,b)}(x,V) \\ &= \begin{cases} -\frac{1}{2}\binom{x}{b}^T M(V)^\dagger \binom{x}{b} & \text{if } V \succeq_{\ker A} 0,\ \binom{x}{b} \in \operatorname{rge} M(V), \\ -\infty & \text{else.}\end{cases} \end{aligned} \tag{16}$$

That is, v is concave and upper semicontinuous. In addition,

$$\operatorname{dom} v = \operatorname{dom}\partial v = \left\{(x,V) \in \mathbb{R}^n \times \mathbb{S}^n \;\middle|\; \begin{pmatrix} x \\ b\end{pmatrix} \in \operatorname{rge} M(V) \text{ and } V \succeq_{\ker A} 0\right\},$$

where ∂v is the concave subdifferential of v. The function v is continuously differentiable on

$$\operatorname{int}(\operatorname{dom} v) = \left\{(x,V) \in \mathbb{R}^n \times \mathbb{S}^n \;\middle|\; \begin{pmatrix} x \\ b\end{pmatrix} \in \operatorname{rge} M(V) \text{ and } V \succ_{\ker A} 0\right\},$$

and, if $\operatorname{rank} A = p$*, then, for all $(x,V) \in \operatorname{int}(\operatorname{dom} v)$, $v'(x,V) = (-y, \frac{1}{2}yy^T)$ with $y = A^\dagger b + P(P^TVP)^{-1}(x - VA^\dagger b)$ and $P \in \mathbb{R}^{n\times(n-p)}$ is any matrix whose columns form an orthonormal basis for* $\ker A$.

Remark 5.2. *Theorem 5.1 refers to the* concave *subdifferential of a concave function ϕ which is given by $\partial\phi(\bar{u}) := \{g \mid \phi(u) \le \phi(\bar{u}) + \langle g, u-\bar{u}\rangle\ \forall u\} = -\partial(-\phi)(\bar{u})$ where $\partial(-\phi)$ is the convex subdifferential of the convex function $-\phi$.*

Proof. The equivalence (16) is an immediate consequence of (8). Consequently, the remaining statements in the theorem follow from Theorem 4.1 and Corollary 4.9. □

5.2 The matrix-fractional function and its generalization

Consider the function $\gamma : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ given by

$$\gamma(X,V) := \begin{cases} \frac{1}{2}\operatorname{tr}\left(X^TV^\dagger X\right) & \text{if} \quad V \succeq 0,\ \operatorname{rge} X \subset \operatorname{rge} V, \\ +\infty & \text{else.}\end{cases} \tag{17}$$

For the case $m = 1$ (recall that $\mathbb{E} = \mathbb{R}^{n\times m} \times \mathbb{S}^n$), this function has been investigated under the moniker *pseudo matrix-fractional function* [7, Ex. 3.5.0.0.2]. We refer to γ as the *generalized matrix-fractional function* when $m > 1$. It is related to the *general fractional-quadratic matrix inequality* introduced in [1, p. 155].

Theorem 4.1 tells us that γ is the support functional $\sigma_{\mathcal{D}(0,0)}$. This yields a series of immediate consequences.

Theorem 5.3. *The function* γ *defined in* (17) *is the support functional of the set* $\mathcal{D} := \mathcal{D}(0,0)$ *defined by* (4)*, i.e,* $\gamma = \sigma_{\mathcal{D}}$*. In particular,* γ *is sublinear (hence convex) and lsc. Moreover,*

$$\operatorname{dom}\gamma = \operatorname{dom}\partial\gamma = \{(X,V)\in\mathbb{E} \mid V\succeq 0,\ \operatorname{rge}X\subset\operatorname{rge}V\},$$

which is closed with

$$\operatorname{int}(\operatorname{dom}\gamma) = \{(X,V)\in\mathbb{E}\mid V\succ 0\}.$$

In addition, for $(X,V)\in\operatorname{dom}\gamma$*, we have*

$$\partial\gamma(X,V) = \left\{\left(\sum_{i=1}^{\kappa+1}\bar d_i\bar Z_i, -\frac{1}{2}\sum_{i=1}\bar Z_i\bar Z_i^T\right)\ \middle|\ \begin{array}{l}(\bar d,\bar Z)\in\mathcal{F},\ N\in\mathbb{R}^{n\times m(\kappa+1)}\\ \textit{with}\ \operatorname{rge}N\subset\ker V\\ \textit{such that}\quad \bar Z_i = \bar d_i V^\dagger X + N_i\end{array}\right\},$$

and γ *is continuously differentiable exactly on* $\operatorname{int}(\operatorname{dom}\gamma)$ *with*

$$\gamma'(X,V) = \left[V^{-1}X,\ -\frac{1}{2}V^{-1}XX^TV^{-1}\right]\quad \forall (X,V)\in\operatorname{int}(\operatorname{dom}\gamma).$$

Proof. Set $A=0$ and $B=0$ in Theorem 4.1 and Theorem 4.8, respectively, to get all assertions other than the explicit formula for the derivative. In order to prove the latter, let $(X,V)\in\operatorname{int}(\operatorname{dom}\gamma)$. In particular, this implies that $\ker V=\{0\}$ and $V^\dagger = V^{-1}$. Hence, for $\bar d\in\mathcal{S}_\kappa\cap\mathbb{R}^{\kappa+1}_+$ and $\bar Z_i = \bar d_i V^{-1}X$, we have

$$\sum_{i=1}^{\kappa+1}\bar d_i(\bar d_i V^{-1}X) = V^{-1}X$$

and

$$-\frac{1}{2}\sum_{i=1}^{\kappa+1}(\bar d_i V^{-1}X)(\bar d_i V^{-1}X)^T = -\frac{1}{2}V^{-1}XX^TV^{-1}.$$

Therefore,

$$\partial\gamma(X,V) = \left\{\left(V^{-1}X, -\frac{1}{2}V^{-1}XX^TV^{-1}\right)\right\} = \{\gamma'(X,V)\},$$

which concludes the proof. □

Next we consider the function $\phi:\mathbb{E}\to\mathbb{R}\cup\{+\infty\}$ defined by

$$\phi(X,V) := \begin{cases} \frac{1}{2}\operatorname{tr}(X^TV^{-1}X) & \text{if}\quad V\succ 0,\\ +\infty & \text{else.}\end{cases} \tag{18}$$

Clearly, this function is closely related to the function γ defined above with $\operatorname{dom}\phi = \mathbb{R}^{n\times m}\times\mathbb{S}^n_{++} = \operatorname{int}(\operatorname{dom}\gamma)$ and $\gamma_{|\operatorname{dom}\phi} = \phi_{|\operatorname{dom}\phi}$. In the case where $m=1$, ϕ is called the matrix-fractional function, e.g., see [2, Ex. 3.4] or [7, Ex. 3.5.0.0.4]. In the remainder of this section, we show that $\gamma = \operatorname{cl}\phi$. We begin with the following well-known result which can be found for example in [2, App. A.5.5] or [10, Th. 16.1] with the latter containing a proof.

Lemma 5.4 (Schur complement)**.** *Let* $S\in\mathbb{S}^n, T\in\mathbb{S}^m, R\in\mathbb{R}^{n\times m}$*. Then*

$$\begin{pmatrix} S & R\\ R^T & T\end{pmatrix}\succeq 0 \quad\Longleftrightarrow\quad [S\succeq 0,\quad \operatorname{rge}R\subset\operatorname{rge}S,\quad T-R^TS^\dagger R\succeq 0]$$

Proposition 5.5. *For the functions* γ *from* (17) *and* ϕ *from* (18) *we have:*

a) $\mathrm{epi}\,\gamma = \left\{(X,V,\alpha) \,|\, \exists\, Y \in \mathbb{S}^m \,:\, \begin{pmatrix} V & X \\ X^T & Y \end{pmatrix} \succeq 0,\ \frac{1}{2}\mathrm{tr}\,(Y) \le \alpha \right\}.$

b) $\mathrm{epi}\,\phi = \left\{(X,V,\alpha) \,|\, \exists\, Y \in \mathbb{S}^m \,:\, \begin{pmatrix} V & X \\ X^T & Y \end{pmatrix} \succeq 0,\ V \succ 0,\ \frac{1}{2}\mathrm{tr}\,(Y) \le \alpha \right\}.$

Moreover, $\mathrm{epi}\,\gamma = \mathrm{cl}\,(\mathrm{epi}\,\phi)$*, or equivalently,* $\gamma = \mathrm{cl}\,\phi$*. Therefore,* $\mathrm{epi}\,\phi$ *is a convex cone, i.e.,* ϕ *is sublinear.*

Remark 5.6. *An immediate consequence of part a) of this proposition is that* $\mathrm{epi}\,\gamma$ *is* semidefinite representable*, e.g., see* [1, p. 144].

Proof. a) First note that

$$Y \succeq X^T V^\dagger X \ \Rightarrow\ \mathrm{tr}\,(Y) \ge \mathrm{tr}\,(X^T V^\dagger X). \tag{19}$$

We have

$$\begin{aligned} (X,V,\alpha) \in \mathrm{epi}\,\gamma \ &\Leftrightarrow\ V \succeq 0,\ \mathrm{rge}\,X \subset \mathrm{rge}\,V,\ \frac{1}{2}\mathrm{tr}\,(X^T V^\dagger X) \le \alpha \\ &\Leftrightarrow\ \exists\, Y \in \mathbb{S}^m \,:\, V \succeq 0,\ \mathrm{rge}\,X \subset \mathrm{rge}\,V,\ Y \succeq X^T V^\dagger X,\ \frac{1}{2}\mathrm{tr}\,(Y) \le \alpha \\ &\Leftrightarrow\ \exists\, Y \in \mathbb{S}^m \,:\, \begin{pmatrix} V & X \\ X^T & Y \end{pmatrix} \succeq 0,\ \frac{1}{2}\mathrm{tr}\,(Y) \le \alpha, \end{aligned}$$

where the second equivalence follows from (19) and the fact that we may take $Y = X^T V^\dagger X$, and the final equivalence follows from Lemma 5.4.

b) Analogous to a) with the additional condition that $V \succ 0$ in each step.

In order to show that $\mathrm{cl}\,(\mathrm{epi}\,\phi) = \mathrm{epi}\,\gamma$, we first note that, as γ is lsc, $\mathrm{epi}\,\gamma$ is closed. Moreover, since obviously $\mathrm{epi}\,\phi \subset \mathrm{epi}\,\gamma$, the inclusion $\mathrm{cl}\,(\mathrm{epi}\,\phi) \subset \mathrm{epi}\,\gamma$ follows immediately.

In order to see the converse inclusion, let $(X,V,\alpha) \in \mathrm{epi}\,\gamma$. Now, set $V_k := V + \frac{1}{k} I_n \succ 0$ and put $\alpha_k := \alpha + \gamma(X,V_k) - \gamma(X,V)$ for all $k \in \mathbb{N}$. Then, by definition, we have

$$\alpha_k = \alpha + \gamma(X,V_k) - \gamma(X,V) \ge \gamma(X,V_k) = \phi(X,V_k),$$

i.e., $(X,V_k,\alpha_k) \in \mathrm{epi}\,\phi$ for all $k \in \mathbb{N}$. Clearly, $V_k \to V$ with V_k nonsingular, and so $V_k^{-1} \to V^+$ (e.g., see [15, p. 153]). Consequently, $\phi(X,V_k) = \gamma(X,V_k) \to \phi(X,V_k)$ so that $\alpha_k \to \alpha$. Therefore, $(X,V_k,\alpha_k) \to (X,V,\alpha)$, and so $\mathrm{epi}\,\gamma = \mathrm{cl}\,(\mathrm{epi}\,\phi)$, i.e., $\gamma = \mathrm{cl}\,\phi$.

Also, since γ is a support functional (by Theorem 5.3), hence, in particular, sublinear and lsc, $\mathrm{epi}\,\gamma$ is a closed, convex cone. Moreover, as

$$\mathrm{epi}\,\phi = \mathrm{epi}\,\gamma \cap \mathcal{K},$$

with $\mathcal{K} := \mathbb{R}^{n\times m} \times \mathbb{S}^n_{++} \times \mathbb{R}$ being a convex cone, also $\mathrm{epi}\,\phi$ is a convex cone. Hence, ϕ is also sublinear.

□

5.3 The relationship between $\sigma_{\mathcal{D}(A,B)}$ and the nuclear norm

There is a fascinating relationship between $\sigma_{\mathcal{D}(A,B)}$ and the nuclear norm which was first suggested to us by results in [13] for the case $(A,B) = (0,0)$ where it is used to smooth the nuclear norm. The key idea is to consider optimization problems of the form

$$\min_{V\in\mathbb{S}^n} \left\langle V, \widehat{W}\right\rangle + \sigma_{\mathcal{D}(A,B)}(X,V), \tag{20}$$

where $X \in \mathbb{R}^{n\times m}$ and $\widehat{W} \in \mathbb{S}^n$ are given and fixed. In this section we consider one possible choice for the matrix $\widehat{W}$ (described below).

To begin with, let $X \in \mathbb{R}^{n\times m}$, $A \in \mathbb{R}^{p\times n}$ and $B \in \mathbb{R}^{p\times m}$ be given with $\mathrm{rge}\, B \subset \mathrm{rge}\, A$ and $m \geq n$. Set $k := \dim(\ker A)$, and define $P \in \mathbb{R}^{n\times k}$ to be any matrix whose columns form an orthonormal basis for $\ker A$ so that PP^T is the orthogonal projection onto $\ker A$. Let PP^TX have singular value decomposition

$$[U_1, U_2, U_3]\begin{bmatrix} \Sigma & 0_{t\times(k-t)} & 0_{t\times(n-k)} & 0_{t\times(m-n)} \\ 0_{(k-t)\times t} & 0_{(k-t)\times(k-t)} & 0_{(k-t)\times(n-k)} & 0_{(k-t)\times(m-n)} \\ 0_{(n-k)\times t} & 0_{(n-k)\times(k-t)} & 0_{(n-k)\times(n-k)} & 0_{(n-k)\times(m-n)} \end{bmatrix}\begin{bmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \\ Q_4^T \end{bmatrix} = U_1\Sigma Q_1^T, \tag{21}$$

where $t := \mathrm{rank}\,(PP^TX)$, $\Sigma := \mathrm{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_t)$ with $\sigma_1, \sigma_2, \ldots, \sigma_t$ the singular values of PP^TX, $U := [U_1, U_2, U_3]$ and $Q := [Q_1, Q_2, Q_3, Q_4]$ are orthogonal, and $\widehat{U} := [U_1, U_2]$ is chosen so that its columns form an orthonormal basis for the kernel of A. Set $\widehat{P} := \widehat{U}\widehat{U}^T$, so that

$$PP^T = \widehat{P}\widehat{P}^T = U_1U_1^T + U_2U_2^T. \tag{22}$$

Finally, define

$$\widehat{W} := \frac{1}{2}(A^\dagger B + \widehat{U}\widehat{Q}^T)(A^\dagger B + \widehat{U}\widehat{Q}^T)^T, \tag{23}$$

where $\widehat{Q} := [Q_1, Q_2]$. The matrix $\widehat{U}\widehat{Q}^T$ establishes an isometry between the k-dimensional subspaces $\ker A$ and $\mathrm{rge}\,\widehat{Q}$, where $\ker A = \mathrm{rge}\,\widehat{U} = \mathrm{rge}\, P$. We have the following result.

Theorem 5.7. *Let X, A, B, P, Σ, U, Q, and $\widehat{W}$ be as above with $m \geq n$. Then the optimal value in* (20) *is*

$$\left\langle X, A^\dagger B\right\rangle + \left\|PP^TX\right\|_n, \tag{24}$$

and is attained at the matrix $\widehat{V} := U_1\Sigma U_1$. In particular, if $\mathrm{rge}\, X \subset \ker A$, then the optimal value is $\|X\|_n$; while, on the other hand, if $\mathrm{rge}\, X \subset \mathrm{rge}\, A^T = (\ker A)^\perp$, then the optimal value is $\left\langle X, A^\dagger B\right\rangle$.

Proof. We begin by showing that the matrix $\widehat{V} := U_1\Sigma U_1^T$ solves (20). For this it is helpful to keep in mind certain key relationships between the matrices defined above. First, recall that $AA^\dagger$ is the orthogonal projection onto $\mathrm{rge}\, A$ and $A^\dagger A$ is the orthogonal projection onto $\mathrm{rge}\, A^T$, i.e., $A^\dagger A = I - PP^T$. Moreover, $\mathrm{rge}\, A^\dagger = (\ker A)^\perp$ which, in particular, implies that $P^TA^\dagger = 0$, $U_1^TA^\dagger = 0$, and $U_2^TA^\dagger = 0$. In addition, we use the fact that $\widehat{V}^\dagger = U_1\Sigma^{-1}U_1^T$ (see discussion following (6)), as well as the relations $\widehat{V} = PP^T\widehat{V}PP^T$ and $\widehat{V}^\dagger = PP^T\widehat{V}^\dagger PP^T$ which follow from the equivalence $U_1 = PP^TU_1$.

We claim that

$$M(\widehat{V})^\dagger = K := \begin{bmatrix} \widehat{V}^\dagger & A^\dagger \\ (A^\dagger)^T & 0 \end{bmatrix}.$$

By (6), we need to show that

$$M(\widehat{V})KM(\widehat{V}) = M(\widehat{V}) \quad \text{and} \quad KM(\widehat{V})K = K \tag{25}$$

and

$$(M(\widehat{V})K)^T = M(\widehat{V})K \quad \text{and} \quad (KM(\widehat{V}))^T = KM(\widehat{V}). \tag{26}$$

First observe that

$$\begin{aligned}
M(\widehat{V})K &= \begin{bmatrix} U_1\Sigma U_1^T U_1\Sigma^{-1}U_1^T + A^T(A^\dagger)^T & U_1\Sigma U_1^T A^\dagger \\ AU_1\Sigma U_1^T & AA^\dagger \end{bmatrix} \\
&= \begin{bmatrix} U_1U_1^T + A^\dagger A & 0 \\ 0 & AA^\dagger \end{bmatrix} \\
&= \begin{bmatrix} I - U_2U_2^T & 0 \\ 0 & AA^\dagger \end{bmatrix},
\end{aligned}$$

and

$$\begin{aligned}
KM(\widehat{V}) &= \begin{bmatrix} U_1\Sigma^{-1}U_1^T U_1\Sigma U_1^T + A^\dagger A & U_1\Sigma^{-1}U_1^T A^\dagger \\ (A^\dagger)^T U_1\Sigma^{-1}U_1^T & (A^\dagger)^T A^T \end{bmatrix} \\
&= \begin{bmatrix} U_1U_1^T + A^\dagger A & 0 \\ 0 & AA^\dagger \end{bmatrix} \\
&= \begin{bmatrix} I - U_2U_2^T & 0 \\ 0 & AA^\dagger \end{bmatrix}.
\end{aligned}$$

This establishes (26). To see (25), write

$$\begin{aligned}
KM(\widehat{V})K &= \begin{bmatrix} U_1\Sigma^{-1}U_1^T(I - U_2U_2^T) & A^\dagger AA^\dagger \\ (A^\dagger)^T(I - U_2U_2^T) & 0 \end{bmatrix} \\
&= \begin{bmatrix} \widehat{V}^\dagger & A^\dagger \\ (A^\dagger)^T & 0 \end{bmatrix} \\
&= K
\end{aligned}$$

and

$$\begin{aligned}
M(\widehat{V})KM(\widehat{V}) &= \begin{bmatrix} (I - U_2U_2^T)\widehat{V} & (I - U_2U_2^T)A^T \\ AA^\dagger A & 0 \end{bmatrix} \\
&= \begin{bmatrix} \widehat{V} & A^T \\ A & 0 \end{bmatrix} \\
&= M(\widehat{V}).
\end{aligned}$$

Since the problem (20) is convex, one need only show that $\widehat{V}$ satisfies the first-order necessary condition for optimality, i.e., $0 \in \widehat{W} + \partial_V \sigma_{\mathcal{D}(A,B)}(X, \widehat{V})$. By Theorem 4.8 and [16, Cor. 10.11], these conditions read

$$\begin{aligned}
&\exists\, \bar{W} \in \mathbb{R}^{p\times m(\kappa+1)},\ (\bar{d}, \bar{Z}) \in \mathcal{F}(A,B), \quad \text{and} \\
&\begin{pmatrix} \bar{N}_1 \\ \bar{N}_2 \end{pmatrix} \in \mathbb{R}^{n\times m(\kappa+1)} \times \mathbb{R}^{p\times m(\kappa+1)} \text{ with } \operatorname{rge}\begin{pmatrix} \bar{N}_1 \\ \bar{N}_2 \end{pmatrix} \subset \ker M(V)
\end{aligned} \tag{27}$$

such that

$$\begin{pmatrix} \bar{Z}_i \\ \bar{W}_i \end{pmatrix} = \bar{d}_i M(V)^\dagger \begin{pmatrix} X \\ B \end{pmatrix} + \begin{pmatrix} \bar{N}_{1i} \\ \bar{N}_{2i} \end{pmatrix} \quad (i = 1, 2, \dots, \kappa+1) \tag{28}$$

and

$$(A^\dagger B + \widehat{U}\widehat{Q}^T)(A^\dagger B + \widehat{U}\widehat{Q}^T)^T = \sum_{i=1}^{\kappa+1} \bar{Z}_i \bar{Z}_i^T. \tag{29}$$

We now construct a suitable $\bar{W}$, $(\bar{d}, \bar{Z})$, and $\binom{\bar{N}_1}{\bar{N}_2}$. Set

$$\bar{d}_1 = 1,\ \bar{d}_i = 0,\ \begin{pmatrix} N_{11} \\ N_{21} \end{pmatrix} := \begin{pmatrix} U_2 Q_2^T \\ 0 \end{pmatrix},\ \begin{pmatrix} N_{1i} \\ N_{2i} \end{pmatrix} := \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (i = 2, 3, \dots, \kappa + 1),$$

and define $\binom{\bar{Z}}{\bar{W}}$ using (28). This gives

$$\begin{pmatrix} \bar{Z}_1 \\ \bar{W}_1 \end{pmatrix} = M(V)^\dagger \begin{pmatrix} X \\ B \end{pmatrix} + \begin{pmatrix} U_2 Q_2^T \\ 0 \end{pmatrix} = \begin{pmatrix} \widehat{V}^\dagger X + A^\dagger B + U_2 Q_2^T \\ (A^\dagger)^T X \end{pmatrix} = \begin{pmatrix} A^\dagger B + \widehat{U}\widehat{Q}^T \\ (A^\dagger)^T X \end{pmatrix}$$

and

$$\begin{pmatrix} \bar{Z}_i \\ \bar{W}_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (i = 2, 3, \dots, \kappa + 1).$$

In particular, this implies that (29) holds. Clearly, $0 \le \bar{d}$ and $\|\bar{d}\| = 1$ and $A\bar{Z}_1 = AA^\dagger B = B$, so that $(\bar{d}, \bar{Z}) \in \mathcal{F}(A, B)$ and (27) is satisfied. Hence $\widehat{V}$ is indeed a solution to (20).

We now compute the optimal value in (20). This is given by

$$\left\langle \widehat{Y},\, X \right\rangle + \left\langle \widehat{W},\, \widehat{V} \right\rangle + \frac{1}{2}\mathrm{tr}\left(\begin{pmatrix} X \\ B \end{pmatrix}^T M(\widehat{V})^\dagger \begin{pmatrix} X \\ B \end{pmatrix} \right).$$

We have

$$\begin{aligned}
\left\langle \widehat{W},\, \widehat{V} \right\rangle &= \frac{1}{2}\mathrm{tr}\left((A^\dagger B + \widehat{U}\widehat{Q}^T)(A^\dagger B + \widehat{U}\widehat{Q}^T)^T U_1 \Sigma U_1^T \right) \\
&= \frac{1}{2}\mathrm{tr}\left((A^\dagger B + \widehat{U}\widehat{Q}^T)(B^T (A^\dagger)^T U_1 \Sigma U_1^T + Q_1 \Sigma U_1^T) \right) \\
&= \frac{1}{2}\mathrm{tr}\left((A^\dagger B + \widehat{U}\widehat{Q}^T) Q_1 \Sigma U_1^T \right) \\
&= \frac{1}{2}\left[\mathrm{tr}\left(A^\dagger B Q_1 \Sigma U_1^T \right) + \mathrm{tr}\left(U_1 \Sigma U_1^T \right) \right] \\
&= \frac{1}{2}\mathrm{tr}\left(U_1^T A^\dagger B Q_1 \Sigma \right) + \frac{1}{2}\mathrm{tr}\,(\Sigma) \\
&= \frac{1}{2}\left\| PP^T X \right\|_n
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{tr}\left(\begin{pmatrix} X \\ B \end{pmatrix}^T M(\widehat{V})^\dagger \begin{pmatrix} X \\ B \end{pmatrix} \right) &= \mathrm{tr}\left(\begin{pmatrix} X \\ B \end{pmatrix}^T \begin{pmatrix} \widehat{V}^\dagger X + A^\dagger B \\ (A^\dagger)^T X \end{pmatrix} \right) \\
&= \mathrm{tr}\,(X^T \widehat{V}^\dagger X) + 2\mathrm{tr}\,(X^T A^\dagger B) \\
&= \mathrm{tr}\,(X^T PP^T \widehat{V}^\dagger PP^T X) + 2\mathrm{tr}\,(X^T A^\dagger B) \\
&= \mathrm{tr}\,(Q_1 \Sigma Q_1^T) + 2\mathrm{tr}\,(X^T A^\dagger B) \\
&= \left\| PP^T X \right\|_n + 2\mathrm{tr}\,(X^T (I - PP^T) A^\dagger B) \\
&= \left\| PP^T X \right\|_n + 2\mathrm{tr}\,(X^T A^\dagger B)\ .
\end{aligned}$$

Hence the optimal value is given by (24) when $\widehat{Y} := 0$. Finally, note that if $\mathrm{rge}\, X \subset \ker A$, then $PP^T X = X$ so $\left\langle X,\, A^\dagger B \right\rangle = \left\langle PP^T X,\, A^\dagger B \right\rangle = \left\langle X,\, PP^T A^\dagger B \right\rangle = 0.$ □

By applying Theorem 5.7 to $\gamma = \sigma_{\mathcal{D}(0,0)}$ with $\widehat{W} = \frac{1}{2}I$, we obtain the following result.

Corollary 5.8. [13, Lemma 1] *Let $X \in \mathbb{R}^{n\times m}$ with $m \geq n$, and let $\gamma : \mathbb{E} \to \mathbb{R}\cup\{+\infty\}$ be defined in* (17)*. Then*

$$\|X\|_n = \min_{V\in\mathbb{S}^n} \frac{1}{2}\mathrm{tr}\,(V) + \gamma(X,V),$$

with the minimum value attained at $\hat{V} := U_1\Sigma U_1^T$, where $X = U_1\Sigma Q_1^T$ is the reduced SVD for X.

This representation of the nuclear norm is used in [13] as a means to introduce a smoothing term for optimization problems involving a nuclear norm regularizer. Indeed, Theorem 5.3 shows that, on the interior of its domain, γ is a smooth function with an easily computed derivative. Using the more general formulation developed here, it is possible to derive smoothers for more general types of regularizers. We illustrate one such possibility in the following corollary.

Corollary 5.9. *Let $X \in \mathbb{R}^{p\times p}$ and let $P \in \mathbb{R}^{p\times k}$ be any matrix having orthonormal columns. Define $A := I - PP^T$ and $B := \lambda(I - PP^T)X$. With this specification, the matrix $\widehat{W}$ in* (23) *is given by $\hat{W} = \frac{1}{2}(I+B)(I+B)^T$ and the optimal value in* (20) *with $\widehat{Y} := 0$ is*

$$\lambda\left\|(I-PP^T)X\right\|_F^2 + \left\|PP^TX\right\|_n.$$

Proof. Simply observe that $A^\dagger = A$ and $A^\dagger B = \lambda B$, and apply Theorem 5.7 with $n = p = m$. □

6 Final Remarks

We have introduced a new class of support functionals $\sigma_{\mathcal{D}(A,B)}$, where

$$\mathcal{D}(A,B) = \left\{\left(Y, -\frac{1}{2}YY^T\right) \mid AY = B\right\}.$$

It is shown that these support functionals have a natural connection to the optimal value function for quadratics over affine manifolds. Connections are also made to the matrix-fractional function and the pseudo matrix-fractional function as well as their generalizations. It is remarkable that all of these objects can be represented by the support functional of $\mathcal{D}(A,B)$ for an appropriate choice of A and B. These support functional representations had not previously been observed. In addition, we have computed the the closed convex hull of $\mathcal{D}(A,B)$ which allowed us to compute the subdifferential of $\sigma_{\mathcal{D}(A,B)}$. This computation made use of nontrivial techniques from nonconvex variational analysis. Finally, we obtained a general representation theorem that can be used to give alternative representation for the nuclear norm and related regularizers in matrix optimization. These representations are particularly useful in that they can be used to develop smoothers for the regularization terms.

References

[1] A. Ben-Tal and A. Nemirovski *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications* MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2001.

[2] S. Boyd and L. Vandenbergh: *Convex Optimization.* Cambridge University Press, 2004.

[3] J.M. Borwein and A.S. Lewis: *Convex Analysis and Nonlinear Optimization. Theory and Examples.* CMS Books in Mathematics, Springer-Verlag, New York, 2000.

[4] J.V. Burke: *Nonlinear Optimization.* Lecture Notes, Math 408, University of Washington, Spring 2014.

[5] J.V. Burke and A.Y Aravkin: *Smoothing dynamic systems with state-dependent covariance matrices.* To appear in the Conference on Decision and Control (CDC), 2014 (6 pages).

[6] Y. Chabrillac and J.-P. Crouzeix: *Definiteness and semidefiniteness of quadratic forms.* Linear Algebra and its Applications 63, 1984, pp. 283–292.

[7] J. Dattorro: *Convex Optimization & Euclidean Distance Geometry.* $M\varepsilon\beta oo$ Publishing USA, Version 2014.04.08, 2005 .

[8] M. Faliva and M.G. Zoia: *On a partitioned inversion formula having useful applications in econometrics.* Econometric Theory 18, 2002, pp. 525–530.

[9] P. Finsler: *Über das Vorkommen definiter und semidefiniter Formen und Scharen quadratischer Formen.* Commentarii Mathematici Helvetici 9, 1937, pp. 188–192.

[10] J. Gallier: *Geometric Methods and Applications: For Computer Science and Engineering.* Texts in Applied Mathematics, Springer New York, Dordrecht, London, Heidelberg, 2011.

[11] G.H. Golub and C.F van Loan: *Matrix Computations.* John Hopkins Series in Mathematicals Sciences, The John Hopkins University Press, Baltimore, Maryland, 1983.

[12] R. A. Horn and C. R. Johnson: *Matrix Analysis.* Cambridge University Press, New York, N.Y., 1985.

[13] C.-J. Hsieh and P. Olsen: *Nuclear Norm Minimization via Active Subspace Selection.* JMLR W&CP 32 (1) :575-583, 2014.

[14] S.-J. Kim and S. Boyd: *A Minimax Theorem with Applications to Machine Learning, Signal Processing, and Finance.* SIAM Journal on Optimization, 19(3):13441367, 2008.

[15] J.R. Magnus and H. Neudecker *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley & Sons, New York, NY, 1988.

[16] R.T. Rockafellar and R.J.-B. Wets: *Variational Analysis.* A Series of Comprehensive Studies in Mathematics, Vol. 317, Springer, Berlin, Heidelberg, 1998.

University of Washington, Department of Mathematics, Box 354350 Seattle, Washington 98195-4350

E-mail address: burke@math.washington.edu

University of Würzburg, Institute of Mathematics, Campus Hubland Nord, Emil-Fischer-Strasse 30, 97074 Würzburg, Germany

E-mail address: hoheisel@mathematik.uni-wuerzburg.de