CONVEX GEOMETRY OF THE GENERALIZED MATRIX-FRACTIONAL FUNCTION

JAMES V. BURKE*, YUAN GAO†, AND TIM HOHEISEL‡

Abstract. Generalized matrix-fractional (GMF) functions are a class of matrix support functions introduced by Burke and Hoheisel as a tool for unifying a range of seemingly divergent matrix optimization problems associated with inverse problems, regularization and learning. In this paper we dramatically simplify the support function representation for GMF functions as well as the representation of their subdifferentials. These new representations allow the ready computation of a range of important related geometric objects whose formulations were previously unavailable.

Key words. matrix optimization, matrix-fractional function, support function, gauge function

AMS subject classifications. 68Q25, 68R10, 68U05

1. Introduction.

Generalized matrix-fractional (GMF) functions were introduced in [3] as a means to unify a range of seemingly divergent tools in matrix optimization related to inverse problems, regularization and machine learning. Somewhat surprisingly GMF functions coincide with the negative of the optimal value function for affinely constrained quadratic programs, and are representable as support functions on the matrix space $\mathbb{E} := \mathbb{R}^{n\times m} \times \mathbb{S}^n$, where $\mathbb{R}^{n\times m}$ and $\mathbb{S}^n$ are the vector spaces of real $n \times m$ and symmetric $n \times n$ matrices, respectively. The most significant challenge in [3] is the derivation of an expression for the closed convex set associated with the support function representation. Unfortunately, the representation given in [3] is exceedingly complicated. The main contribution of this paper is to provide a simple, elegant, and intuitive representation for this set. We then use this representation to provide a simplified expression for the subdifferential of a GMF function and to compute various related geometric objects that were previously unavailable. These representations dramatically simplify the use of these tools to a wide range of applications [4]. Before proceeding, we review the definition of a GMF function.

Given $(A, B) \in \mathbb{R}^{p\times n} \times \mathbb{R}^{p\times m}$ with $\operatorname{rge} B \subset \operatorname{rge} A$, the graph of the matrix valued mapping $Y \mapsto -\frac{1}{2}YY^T$ over an affine manifold $\{Y \in \mathbb{R}^{n\times m} \mid AY = B\}$ is given by

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

The associated GMF function is the support function of the set $\mathcal{D}(A, B)$:

$$\sigma_{\mathcal{D}(A,B)}(X,V) = \sup_{(Y,W)\in\mathcal{D}(A,B)} \langle (X,V),\, (Y,W)\rangle,$$

where we use the Frobenius inner product on $\mathbb{E}$,

$$\langle (Y,W),\, (X,V)\rangle = \operatorname{tr}(Y^TX) + \operatorname{tr} WV = \operatorname{tr}(XY^T + WV).$$

*Department of Mathematics, University of Washington, Seattle, WA 98195 (jvburke@uw.edu).
Research is supported in part by the National Science Foundation under grant number DMS-1514559.
†Department of Applied Mathematics, University of Washington, Seattle, WA 98195 (yuangao@uw.edu).
‡McGill University, 805 Sherbrooke St West, Room1114, Montréal, Québec, Canada H3A 0B9 tim.hoheisel@mcgill.ca)

In [3, Theorem 4.1], it is shown that

$$(2)\quad \sigma_{\mathcal{D}(A,B)}(X,V)=\begin{cases}\frac{1}{2}\mathrm{tr}\left(\binom{X}{B}^T M(V)^\dagger\binom{X}{B}\right) & \text{if }\ \mathrm{rge}\binom{X}{B}\subset \mathrm{rge}\, M(V),\ V\in\mathcal{K}_A,\\ +\infty & \text{else,}\end{cases}$$

where $\mathcal{K}_A := \left\{V\in\mathbb{S}^n \;\middle|\; u^TVu\geq 0\ (u\in\ker A)\right\}$ and $M(V)^\dagger$ is the Moore-Penrose pseudo inverse of the matrix

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

In particular, this implies that

$$(3)\quad \begin{aligned}\mathrm{dom}\,\sigma_{D(A,B)} &= \mathrm{dom}\,\partial\sigma_{D(A,B)}\\ &= \left\{(X,V)\in\mathbb{R}^{n\times m}\times\mathbb{S}^n \;\middle|\; \mathrm{rge}\begin{pmatrix}X\\ B\end{pmatrix}\subset\mathrm{rge}\,M(V),\ V\in\mathcal{K}_A\right\}.\end{aligned}$$

Note that $\mathrm{dom}\,\sigma_{D(A,B)}$ is clearly not a closed set. To see this consider the case $A=B=0$ and $V=\eta I$ so that any $X\neq 0$ has $\mathrm{rge}\,X\in\mathrm{rge}\,V$. But as $\eta\downarrow 0$ it is not the case that $\mathrm{rge}\,X\subset\mathrm{rge}\,0$. Consequently, the statement in [3, Theorem 4.1] that this domain is closed is clearly false. This error does not affect the validity of the other results in [3] since none of them require that the set $\mathrm{dom}\,\sigma_{D(A,B)}$ be closed.

The representation (2) is the basis for the name *generalized matrix-fractional function* since the matrix-fractional functions [2, 5, 9, 10] are obtained when the matrices A and B are both taken to be zero.

The paper is organized as follows: Section 2 begins with a study of the cones $\mathcal{K}_A$ defined in (6) and their polars. This is immediately followed by deriving our new representation of the set $\Omega(A,B):=\overline{\mathrm{conv}}\,\mathcal{D}(A,B)$ in Theorem 2. With this representation in hand, we derive new simplified descriptions for the normal cone $N_{\Omega(A,B)}$ and the subdifferential $\partial\sigma_{\Omega(A,B)}$ in Section 3. In Section 4 we explore the convex geometry of the set $\Omega(A,B)$, and conclude in Section 5 with the important special case where $B=0$ and $\sigma_{\Omega(A,0)}$ is a gauge function.

Notation: Let $\mathcal{E}$ be a finite dimensional Euclidean space with inner product denoted by $\langle\cdot,\cdot\rangle$ and the induced norm $\|\cdot\|:=\sqrt{\langle\cdot,\cdot\rangle}$ with the closed ϵ-ball about a point $x\in\mathcal{E}$ denoted by $B_\epsilon(x)$. Let $S\subset\mathcal{E}$ be nonempty. The (topological) *closure* and *interior* of S are denoted by $\mathrm{cl}\,S$ and $\mathrm{int}\,S$, respectively. The *(linear) span* of S will be denoted by $\mathrm{span}\,S$.

The *convex hull* of S is the set of all convex combinations of elements of S and is denoted by $\mathrm{conv}\,S$. Its closure (the *closed convex hull*) is $\overline{\mathrm{conv}}\,S:=\mathrm{cl}\,(\mathrm{conv}\,S)$. The *conical hull* of S is the set

$$\mathbb{R}_+S:=\{\lambda x\mid x\in S,\ \lambda\geq 0\}.$$

The *convex conical hull* of S is

$$\mathrm{cone}\,S:=\left\{\sum_{i=1}^r\lambda_i x_i\mid r\in\mathbb{N},\ x_i\in S,\ \lambda_i\geq 0\right\}.$$

It is easily seen that $\mathrm{cone}\,S=\mathbb{R}_+(\mathrm{conv}\,S)=\mathrm{conv}\,(\mathbb{R}_+S)$. The closure of the latter is $\overline{\mathrm{cone}}\,S:=\mathrm{cl}\,(\mathrm{cone}\,S)$. The *affine hull* of S, denoted by $\mathrm{aff}\,S$, is the smallest affine space that contains S.

The *relative interior* of a convex set $C \subset \mathcal{E}$ is its interior in the relative topology with respect to the affine hull, i.e.

$$\operatorname{rint} C = \{x \in C \mid \exists \varepsilon > 0 : \ B_\varepsilon(x) \cap \operatorname{aff} C \subset C\}.$$

It is well known, see e.g. [1, Section 6.2], that the points $x \in \operatorname{rint} C$ are characterized through

$$\mathbb{R}_+(C - x) = \operatorname{span}(C - x), \tag{4}$$

where the latter is the (unique) subspace parallel to $\operatorname{aff} C$. In particular, we have $\mathbb{R}_+ C = \operatorname{aff} C = \operatorname{span} C$ if and only if $0 \in \operatorname{rint} C$.

The *polar set* of S is defined by

$$S^\circ := \{v \in \mathcal{E} \mid \langle v, x\rangle \le 1 \ (x \in S)\}.$$

Moreover, we define the *bipolar set* of S by $S^{\circ\circ} := (S^\circ)^\circ$. It is well known that $S^{\circ\circ} = \overline{\operatorname{cone}}(S \cup \{0\})$. If $K \subset \mathcal{E}$ is a cone (i.e. $\mathbb{R}_+ K \subset K$) it can be seen by a homogeneity argument that

$$K^\circ = \{v \in \mathcal{E} \mid \langle v, x\rangle \le 0 \ (x \in K)\},$$

and if $\mathcal{S} \subset \mathcal{E}$ is a subspace, $\mathcal{S}^\circ$ is the orthogonal subspace $\mathcal{S}^\perp$. The *horizon cone* of S is the set

$$S^\infty := \{v \in \mathcal{E} \mid \exists \{\lambda_k\} \downarrow 0, \ \{x_k \in S\} : \ \lambda_k x_k \to v\}$$

which is always a closed cone. For a convex set $C \subset \mathcal{E}$, C^∞ coincides with the *recession cone* of the closure of C, i.e.

$$C^\infty = \{v \mid x + tv \in \operatorname{cl} C \ (t \ge 0, \ x \in C)\} = \{y \mid C + y \subset C\}.$$

For $f : \mathcal{E} \to \mathbb{R} \cup \{+\infty\}$ its *domain* and *epigraph* are given by

$$\operatorname{dom} f := \{x \in \mathcal{E} \mid f(x) < +\infty\} \quad \text{and} \quad \operatorname{epi} f := \{(x, \alpha) \in \mathcal{E} \times \mathbb{R} \mid f(x) \le \alpha\}.$$

We call f *convex* if its epigraph $\operatorname{epi} f$ is a convex set.

For a convex function $f : \mathcal{E} \to \mathbb{R} \cup \{+\infty\}$ its *subdifferential* at a point $\bar{x} \in \operatorname{dom} f$ is given by

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

Given a nonempty set $S \subset \mathcal{E}$, its *indicator function* $\delta_S : \mathcal{E} \to \mathbb{R} \cup \{+\infty\}$ is given by

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

The indicator of S is convex if and only if S is a convex set, in which case the *normal cone* of S at $\bar{x} \in S$ is given by

$$N_S(\bar{x}) := \partial \delta_S(\bar{x}) = \{v \in \mathcal{E} \mid \langle v, x - \bar{x}\rangle \le 0 \ (x \in S)\}.$$

The *support function* $\sigma_S : \mathcal{E} \to \mathbb{R} \cup \{+\infty\}$ and the *gauge function* $\gamma_S : \mathcal{E} \to \mathbb{R} \cup \{+\infty\}$ of a nonempty set $S \subset \mathcal{E}$ are given by

$$\sigma_S(x) := \sup_{v \in S} \langle v, x\rangle \quad \text{and} \quad \gamma_S(x) := \inf\{t \ge 0 \mid x \in tS\},$$

respectively. Here we use the standard convention that $\inf \emptyset = +\infty$. It is easy to see that

$$\sigma_S = \sigma_{\overline{\operatorname{conv}} S}. \tag{5}$$

2. New Representation of $\overline{\text{conv}}\,\mathcal{D}(A,B)$. In view of (5), in order to obtain a complete understanding of the variational properties of σ_S, it is critical to have a useful description of the closed convex hull $\overline{\text{conv}}\,S$. This is often a non-trivial task. In [3, Proposition 4.3], a representation for $\overline{\text{conv}}\,\mathcal{D}(A,B)$ is obtained after great effort, and this representation is arduous. Although it is successfully used in [3, Section 5] in several important situations, the representation is an obstacle to a deeper understanding of the function $\sigma_{\mathcal{D}(A,B)}$ as well as its ease of use in applications. The focus of this section is to provide a new and intuitively appealing representation that dramatically facilitates the use of $\sigma_{\mathcal{D}(A,B)}$. The key to this new representation is the class of cones

$$\mathcal{K}_{\mathcal{S}} := \left\{ V \in \mathbb{S}^n \;\middle|\; u^T V u \geq 0,\ (u \in \mathcal{S}) \right\}, \tag{6}$$

where $\mathcal{S}$ is a subspace of $\mathbb{R}^n$, that is, $\mathcal{K}_{\mathcal{S}}$ is the set of all symmetric matrices that are positive definite with respect to the given subspace $\mathcal{S}$. Observe that if $P \in \mathbb{S}^n$ is the orthogonal projection onto $\mathcal{S}$, then

$$\mathcal{K}_{\mathcal{S}} = \{ V \in \mathbb{S}^n \mid PVP \succeq 0 \}. \tag{7}$$

Clearly, $\mathcal{K}_{\mathcal{S}}$ is a convex cone, and, for $\mathcal{S} = \mathbb{R}^n$, it reduces to $\mathbb{S}^n_+$. Given a matrix $A \in \mathbb{R}^{p\times n}$, the cones $\mathcal{K}_{\ker A}$ play a special role in our analysis. For this reason, we simply write $\mathcal{K}_A$ to denote $\mathcal{K}_{\ker A}$, i.e. $\mathcal{K}_A := \mathcal{K}_{\ker A}$.

Proposition 1 ($\mathcal{K}_{\mathcal{S}}$ and its polar). *Let $\mathcal{S}$ be a nonempty subspace of $\mathbb{R}^n$ and let P be the orthogonal projection onto $\mathcal{S}$. Then the following hold:*

a) $\mathcal{K}_{\mathcal{S}}^\circ = \text{cone}\,\left\{ -vv^T \mid v \in \mathcal{S} \right\} = \{ W \in \mathbb{S}^n \mid W = PWP \preceq 0 \}$.

b) $\text{int}\,\mathcal{K}_{\mathcal{S}} = \left\{ V \in \mathbb{S}^n \;\middle|\; u^T V u > 0\ (u \in \mathcal{S} \setminus \{0\}) \right\}$.

c) $\text{aff}\,(\mathcal{K}_{\mathcal{S}}^\circ) = \text{span}\,\left\{ vv^T \mid v \in \mathcal{S} \right\} = \{ W \in \mathbb{S}^n \mid \text{rge}\, W \subset \mathcal{S} \}$.

d) $\text{rint}\,(\mathcal{K}_{\mathcal{S}}^\circ) = \left\{ W \in \mathcal{K}_{\mathcal{S}}^\circ \;\middle|\; u^T W u < 0\ \ (u \in \mathcal{S} \setminus \{0\}) \right\}$ *when* $\mathcal{S} \neq \{0\}$ *and* $\text{rint}\,(\mathcal{K}_{\{0\}}^\circ) = \{0\}$ *(since* $\mathcal{K}_{\{0\}} = \mathbb{S}^n$*).*

Proof.

a) Put $B := \left\{ -ss^T \mid s \in \mathcal{S} \right\} \subset \mathbb{S}^n_-$ and observe that

$$\text{cone}\, B = \left\{ -\sum_{i=1}^r \lambda_i s_i s_i^T \mid r \in \mathbb{N}, s_i \in \mathcal{S}, \lambda_i \geq 0\ (i = 1, \dots, r) \right\}.$$

We have cone $B = \left\{ W \in \mathbb{S}^n_- \mid W = PWP \right\}$: To see this, first note that cone $B \subset \left\{ W \in \mathbb{S}^n_- \mid W = PWP \right\}$. The reverse inclusion invokes the spectral decomposition of $W = \sum_{i=1}^n \lambda_i q_i q_i^T$ for $\lambda_1, \dots, \lambda_n \leq 0$. In particular, this representation of cone B shows that it is closed. We now prove the first equality in a): To this end, observe that

$$\begin{aligned} \mathcal{K}_{\mathcal{S}} &= \left\{ V \in \mathbb{S}^n \;\middle|\; s^T V s \geq 0\ (s \in \mathcal{S}) \right\} \\ &= \left\{ V \in \mathbb{S}^n \;\middle|\; \left\langle V, -ss^T \right\rangle \leq 0\ (s \in \mathcal{S}) \right\} \\ &= (\text{cone}\, B)^\circ, \end{aligned}$$

where the third equality uses simply the linearity of the inner product in the second argument. Polarization then gives

$$\mathcal{K}_{\mathcal{S}}^\circ = (\text{cone}\, B)^{\circ\circ} = \overline{\text{cone}}\, B = \text{cone}\, B.$$

b) The proof is straightforward and follows the pattern of proof for $\operatorname{int}\mathbb{S}^n_+ = \mathbb{S}^n_{++}$.

c) With B as defined above, observe that

$$\operatorname{aff}\mathcal{K}_{\mathcal{S}}^\circ = \operatorname{span}\mathcal{K}_{\mathcal{S}}^\circ = \operatorname{span} B,$$

since $0 \in \mathcal{K}_{\mathcal{S}}^\circ$, which shows the first equality. It is hence obvious that $\operatorname{aff}\mathcal{K}_{\mathcal{S}} \subset \{W \in \mathbb{S}^n \mid \operatorname{rge} W \subset \mathcal{S}\}$. On the other hand, every $W \in \mathbb{S}^n$ such that $\operatorname{rge} W \subset \mathcal{S}$ has a decomposition $W = \sum_{i=1}^{\operatorname{rank} W} \lambda_i q_i q_i^T$ where $\lambda_i \neq 0$ and $q_i \in \operatorname{rge} W \subset \mathcal{S}$ for all $i = 1, \dots, \operatorname{rank} W$, i.e. $W \in \operatorname{span} B = \operatorname{aff}\mathcal{K}_{\mathcal{S}}^\circ$.

d) Set $R := \left\{W \in \mathcal{K}_{\mathcal{S}}^\circ \;\middle|\; u^T W u < 0 \;\; (u \in \mathcal{S}\setminus\{0\})\right\}$ and let $W \in \operatorname{rint}(\mathcal{K}_{\mathcal{S}}^\circ)\setminus R \subset \mathcal{K}_{\mathcal{S}}^\circ$. Then there exists $u \in \mathcal{S}$ with $\|u\| = 1$ such that $u^T W u = 0$. Then for every $\varepsilon > 0$ we have $u^T(W + \varepsilon uu^T)u = \varepsilon > 0$. Therefore $W + \varepsilon uu^T \in (B_\varepsilon(W) \cap \operatorname{aff}(\mathcal{K}_{\mathcal{S}}^\circ)) \setminus \mathcal{K}_{\mathcal{S}}^\circ$ for all $\varepsilon > 0$, and hence $W \notin \operatorname{rint}(\mathcal{K}_{\mathcal{S}}^\circ)$, which contradicts our assumption. Hence, $\operatorname{rint}(\mathcal{K}_{\mathcal{S}}^\circ) \subset R$.

To see the reverse implication assume there were $W \in R \setminus \operatorname{rint}(\mathcal{K}_{\mathcal{S}}^\circ)$, i.e. for all $k \in \mathbb{N}$ there exists $W_k \in B_{\frac{1}{k}}(W) \cap \operatorname{aff}(\mathcal{K}_{\mathcal{S}}^\circ) \setminus \mathcal{K}_{\mathcal{S}}^\circ$. In particular, there exists $\{u_k \in \mathcal{S} \mid \|u_k\| = 1\}$ such that $u_k^T W_k u_k \geq 0$ for all $k \in \mathbb{N}$. W.l.o.g. we can assume that $u_k \to u \in \mathcal{S}\setminus\{0\}$. Letting $k \to \infty$, we find that $u^T W u \geq 0$ since $W_k \to W$. This contradicts the fact that $W \in R$. ∎

We are now in a position to prove the main result of this paper which gives a new, simplified description of the closed convex hull of $\Omega(A, B)$.

THEOREM 2. *Let $\mathcal{D}(A, B)$ be as given by (1), then $\overline{\operatorname{conv}}\,\mathcal{D}(A, B) = \Omega(A, B)$, where*

$$\Omega(A, B) := \left\{(Y, W) \in \mathbb{E} \;\middle|\; AY = B \text{ and } \frac{1}{2}YY^T + W \in \mathcal{K}_A^\circ\right\}. \tag{8}$$

Proof. We first show that $\Omega(A, B)$ is itself a closed convex set. Obviously, $\Omega(A, B)$ is closed since $\mathcal{K}_A^\circ$ is closed and the mappings $Y \mapsto AY$ and $(Y, W) \mapsto \frac{1}{2}YY^T + W$ are continuous.

So we need only show that $\Omega(A, B)$ is convex: To this end, let $(Y_i, W_i) \in \Omega(A, B)$, $i = 1, 2$ and $0 \leq \lambda \leq 1$. Then there exist $M_i \in \mathcal{K}_A^\circ$, $i = 1, 2$ such that $W_i = -\frac{1}{2}Y_iY_i^T + M_i$. Observe that $A((1-\lambda)Y_1 + \lambda Y_2) = B$. Moreover, we compute that

$$\begin{aligned}
&\frac{1}{2}((1-\lambda)Y_1 + \lambda Y_2)((1-\lambda)Y_1 + \lambda Y_2)^T + ((1-\lambda)W_1 + \lambda W_2)\\
=&\frac{1}{2}((1-\lambda)Y_1 + \lambda Y_2)((1-\lambda)Y_1 + \lambda Y_2)^T + \left((1-\lambda)(-\frac{1}{2}Y_1Y_1^T + M_1) + \lambda(-\frac{1}{2}Y_2Y_2^T + M_2)\right)\\
=&\frac{1}{2}\lambda(1-\lambda)(-Y_1Y_1^T + Y_1Y_2^T + Y_2Y_1^T - Y_2Y_2^T) + (1-\lambda)M_1 + \lambda M_2\\
=&\lambda(1-\lambda)\left(-\frac{1}{2}(Y_1 - Y_2)(Y_1 - Y_2)^T\right) + (1-\lambda)M_1 + \lambda M_2.
\end{aligned}$$

Since $\operatorname{rge}(Y_1 - Y_2) \subset \ker A$, this shows $\lambda(1-\lambda)\left(-\frac{1}{2}(Y_1 - Y_2)(Y_1 - Y_2)^T\right) + (1-\lambda)M_1 + \lambda M_2 \in \mathcal{K}_A^\circ$. Consequently, $\Omega(A, B)$ is a closed convex set.

Next note that if $(Y, -\frac{1}{2}YY^T) \in \mathcal{D}(A, B)$, then $(Y, -\frac{1}{2}YY^T) \in \Omega(A, B)$ since $0 \in \mathcal{K}_A^\circ$. Hence, $\overline{\operatorname{conv}}\,\mathcal{D}(A, B) \subset \Omega(A, B)$.

It therefore remains to establish the reverse inclusion: For these purposes, let $(Y, W) \in \Omega(A, B)$. By Carathéodory's theorem, there exist $\mu_i \geq 0, v_i \in \ker A$ $(i =$

$1,\dots,N)$ such that

$$W = -\frac{1}{2}YY^T - \sum_{i=1}^{N}\mu_i v_i v_i^T,$$

where $N = \frac{n(n+1)}{2}+1$. Let $0<\epsilon<1$. Set $\lambda_1 := 1-\epsilon$ and $\lambda_2 = \dots = \lambda_{N+1} = \lambda := \epsilon/N$. Denote $Y_1 := Y/\sqrt{1-\epsilon}$. Take $Z_i \in \mathbb{R}^{n\times m},\ i=1,\dots,N$ such that $AZ_i = B$. Finally, set

$$V_i = \left[\sqrt{\frac{2\mu_i}{\lambda}}v_i, 0,\dots,0\right] \in \mathbb{R}^{n\times m} \quad \text{and} \quad Y_{i+1} = Z_i + V_i,\ (i=1,\dots,N).$$

Observe that

$$\sum_{i=1}^{N+1}\lambda_i Y_i = \sqrt{1-\epsilon}Y + \frac{\epsilon}{N}\sum_{i=2}^{N+1} Y_i = \sqrt{1-\epsilon}Y + \frac{\epsilon}{N}\sum_{i=1}^{N} Z_i + \sqrt{\frac{\epsilon}{N}}\sum_{i=1}^{N}\bar{V}_i,$$

where $\bar{V}_i = [\sqrt{2\mu_i}v_i, 0,\dots,0],\ i=1,\dots,N$, and

$$\begin{aligned} -\frac{1}{2}\sum_{i=1}^{N+1}\lambda_i Y_i Y_i^T &= -\frac{1}{2}YY^T - \frac{1}{2}\sum_{i=1}^{N}\frac{\epsilon}{N}\left(Z_iZ_i^T + Z_iV_i^T + V_iZ_i^T\right) - \sum_{i=1}^{N}\mu_i v_i v_i^T \\ &= W - \sum_{i=1}^{N}\frac{1}{2}\left(\frac{\epsilon}{N}Z_iZ_i^T + \sqrt{\frac{\epsilon}{N}}Z_i\bar{V}_i^T + \sqrt{\frac{\epsilon}{N}}\bar{V}_iZ_i^T\right), \end{aligned}.$$

Therefore

$$\begin{aligned} &\left(\sqrt{1-\epsilon}Y + \frac{\epsilon}{N}\sum_{i=1}^{N}Z_i + \sqrt{\frac{\epsilon}{N}}\sum_{i=1}^{N}\bar{V}_i,\quad W - \sum_{i=1}^{N}\frac{1}{2}\left(\frac{\epsilon}{N}Z_iZ_i^T + \sqrt{\frac{\epsilon}{N}}Z_i\bar{V}_i^T + \sqrt{\frac{\epsilon}{N}}\bar{V}_iZ_i^T\right)\right) \\ &= \left(\sum_{i=1}^{N+1}\lambda_i Y_i,\quad -\frac{1}{2}\sum_{i=1}^{N+1}\lambda_i Y_i Y_i^T\right). \end{aligned} \tag{9}$$

Set $\kappa := \dim\mathbb{E}$. By Carathéodory's theorem,

$$\operatorname{conv}\mathcal{D}(A,B) = \left\{\left(\sum_{i=1}^{\kappa+1}\lambda_i Y_i, -\frac{1}{2}\sum_{i=1}^{\kappa+1}\lambda_i Y_i Y_i^T\right) \,\middle|\, \begin{array}{c}\lambda\in\mathbb{R}_+^{\kappa+1}, \sum_{i=1}^{\kappa+1}\lambda_i = 1,\ Y_i\in\mathbb{R}^{n\times m} \\ AY_i = B\ (i=1,\dots,\kappa+1)\end{array}\right\}.$$

By letting $\epsilon\downarrow 0$ in (9), we find $(Y,W)\in\overline{\operatorname{conv}}\,\mathcal{D}(A,B)$ thereby concluding the proof. □

3. Normal cone of $\Omega(A,B)$ and the subdifferential of $\sigma_{\mathcal{D}(A,B)}$.

The new representation for $\overline{\operatorname{conv}}\,\mathcal{D}(A,B)$ allows us to dramatically simplify the representation for the subdifferential of $\sigma_{\mathcal{D}(A,B)}$ given in [3, Theorem 4.8]. For this we use the well-established relation

$$\partial\sigma_C(x) = \left\{z\in\overline{\operatorname{conv}}\,C \mid x\in N_{\overline{\operatorname{conv}}\,C}(z)\right\}, \tag{10}$$

where $C\subset\mathbb{E}$ is nonempty and convex.

Proposition 3 (The normal cone to $\Omega(A,B)$). *Let $\Omega(A,B)$ be as given by (8) and let $(Y,W)\in\Omega(A,B)$. Then*

$$N_{\Omega(A,B)}(Y,W) = \left\{(X,V)\in\mathbb{E} \,\middle|\, \begin{array}{c} V\in\mathcal{K}_A,\ \left\langle V, \frac{1}{2}YY^T + W\right\rangle = 0 \\ \textit{and}\ \operatorname{rge}(X-VY)\subset(\ker A)^\perp \end{array}\right\}$$

Proof. Observe that $\Omega(A,B) = C_1 \cap C_2 \subset \mathbb{E}$ where

$$C_1 := \left\{ Y \in \mathbb{R}^{n\times m} \mid AY = B \right\} \times \mathbb{S}^n \quad \text{and} \quad C_2 := \{(Y,W) \mid F(Y,W) \in \mathcal{K}_A^\circ \},$$

with $F(Y,W) := \frac{1}{2}YY^T + W$. Clearly, C_1 is affine, hence convex, and C_2 is also convex, which can be seen by an analogous reasoning as for the convexity of $\Omega(A,B)$ (cf. the proof of Theorem 2). Therefore, [11, Corollary 23.8.1] tells us that

$$N_{\Omega(A,B)}(Y,W) = N_{C_1}(Y,W) + N_{C_2}(Y,W), \tag{11}$$

where

$$N_{C_1}(Y,W) = \left\{ R \in \mathbb{R}^{n\times m} \;\middle|\; \operatorname{rge} R \subset (\ker A)^\perp \right\} \times \{0\}.$$

We now compute $N_{C_2}((Y,W))$. First recall that for any nonempty closed convex cone $C \subset \mathcal{E}$, we have $N_C(x) = \{ z \in C^\circ \mid \langle z, x \rangle = 0 \}$ for all $x \in C$. Next, note that

$$\nabla F(Y,W)^* U = (UY,\ U) \quad (U \in \mathbb{S}^n),$$

so that $\nabla F(Y,W)^*U = 0$ if and only if $U = 0$. Hence, by [12, Exercise 10.26 Part (d)],

$$N_{C_2}(Y,W) = \left\{ (VY, V) \;\middle|\; V \in \mathcal{K}_A,\ \left\langle V,\, \frac{1}{2}YY^T + W \right\rangle = 0 \right\}.$$

Therefore, by (11), $N_{\Omega(A,B)}(Y,W)$ is given by

$$\left\{ (X,V) \;\middle|\; \operatorname{rge}(X - VY) \subset (\ker A)^\perp,\ V \in \mathcal{K}_A,\ \left\langle V,\, \frac{1}{2}YY^T + W \right\rangle = 0 \right\},$$

which proves the result. ∎

By combining (10) and Proposition 3 we obtain a simplified representation of the subdifferential of the support function $\sigma_{\mathcal{D}}(A,B)$.

COROLLARY 4 (The subdifferential of $\sigma_{\mathcal{D}(A,B)}$). *Let $\mathcal{D}(A,B)$ be as given in (1). Then, for all $(X,V) \in \operatorname{dom} \sigma_{\mathcal{D}(A,B)}$ (see (3)) we have*

$$\partial \sigma_{\mathcal{D}(A,B)}(X,V) = \left\{ (Y,W) \in \Omega(A,B) \;\middle|\; \begin{array}{l} \exists Z \in \mathbb{R}^{p\times m}:\ X = VY + A^T Z, \\ \left\langle V,\, \frac{1}{2}YY^T + W \right\rangle = 0 \end{array} \right\}.$$

Proof. This follows directly from the normal cone description in Proposition 3 and the relation (10). ∎

4. The geometry of $\boldsymbol{\Omega(A,B)}$. We first compute the relative interior and the affine hull of $\Omega(A,B)$. For these purposes, we recall an established result on the relative interior of a convex set in a product space.

PROPOSITION 5 ([11, Theorem 6.8]). *Let $C \subset \mathbb{E}_1 \times \mathbb{E}_2$. For each $y \in \mathbb{E}_1$ we define $C_y := \{ z \in \mathbb{E}_2 \mid (y,z) \in C \}$ and $D := \{ y \mid C_y \neq \emptyset \}$. Then*

$$\operatorname{rint} C = \{ (y,z) \mid y \in \operatorname{rint} D,\ z \in \operatorname{rint} C_y \}.$$

We use this result to get a representation for the relative interior of $\Omega(A,B)$ directly, and then mimic its technique of proof to tackle the affine hull.

LEMMA 6. *Let $A, B \subset \mathbb{E}$ be convex with* $\operatorname{rint} A \cap \operatorname{rint} B \neq \emptyset$. *Then* $\operatorname{aff}(A \cap B) = \operatorname{aff} A \cap \operatorname{aff} B$.

Proof. The inclusion $\operatorname{aff}(A\cap B)\subset \operatorname{aff} A\cap \operatorname{aff} B$ is clear since the latter set is affine and contains $A\cap B$.

For proving the reverse inclusion, we can assume w.l.o.g. that $0\in \operatorname{rint} A\cap \operatorname{rint} B = \operatorname{rint}(A\cap B)$, where for the latter equality we refer to [11, Theorem 6.5]. In particular we have

$$\operatorname{aff} A=\mathbb{R}_+A,\ \operatorname{aff} B=\mathbb{R}_+B \ \text{ and } \ \operatorname{aff}(A\cap B)=\mathbb{R}_+(A\cap B), \tag{12}$$

see (4) and the discussion afterwards. Now, let $x\in \operatorname{aff} A\cap \operatorname{aff} B$. If $x=0$ there is nothing to prove. If $x\neq 0$, by (12), we have $x=\lambda a=\mu b$ for some $\lambda,\mu>0$ and $a\in A, b\in B$. W.l.o.g we have $\lambda>\mu$, and hence, by convexity of B, we have

$$a=\left(1-\frac{\mu}{\lambda}\right)0+\frac{\mu}{\lambda}b\in B.$$

Therefore $x=\lambda a\in\mathbb{R}_+(A\cap B)=\operatorname{aff}(A\cap B)$, see (12). □

We now prove a result analogous to Proposition 5.

Proposition 7. *In addition to the assumptions of Proposition 5 assume that D is affine. Then $(y,z)\in \operatorname{aff} C$ if and only if $y\in D$ and $z\in \operatorname{aff} C_y$.*

Proof. We imitate the proof of [11, Theorem 6.8]: Let $L:(y,z)\mapsto z$. Since D is assumed to be affine (hence $D=\operatorname{aff} D=\operatorname{rint} D$), we have

$$D=L(C)=L(\operatorname{rint} C)=L(\operatorname{aff} C), \tag{13}$$

where we invoke the fact that linear mappings commute with the relative interior and the affine hull, see [11, Theorem 6.7 and p. 8].

Now fix $y\in D=\operatorname{rint} D$ and define the affine set $M_y:=\{(y,z)\mid z\in\mathbb{E}_2\}=\{y\}\times\mathbb{E}_2$. Then, by (13), there exists $z\in\mathbb{E}_2$ such that $y=L(y,z)$ and $(y,z)\in\operatorname{rint} C$. Hence, $\operatorname{rint} M_y\cap \operatorname{rint} C\neq\emptyset$ and we can invoke Lemma 6 to obtain

$$\operatorname{aff} M_y\cap \operatorname{aff} C=\operatorname{aff}(M_y\cap C)=\operatorname{aff}(\{y\}\times C_y)=\{y\}\times \operatorname{aff} C_y.$$

Hence, if $y\in D, z\in\operatorname{aff} C_y$, we have $(y,z)\in\{y\}\times\operatorname{aff} C_y=M_y\cap\operatorname{aff} C\subset \operatorname{aff} C$.

In turn, for $(y,z)\in C$, we have $(y,z)\in M_y\cap \operatorname{aff} C=\{y\}\times C_y$, hence $z\in C_y\neq\emptyset$, so $y\in D$. □

We are now in a position to prove the desired result on the relative interior and the affine hull of $\Omega(A,B)$.

Proposition 8. *For $\Omega(A,B)$ given by (8) the following hold:*

a) $\operatorname{rint}\Omega(A,B)=\left\{(Y,W)\in\mathbb{E}\;\middle|\; AY=B \text{ and } \tfrac12 YY^T+W\in\operatorname{rint}(\mathcal{K}_A^\circ)\right\}.$

b) $\operatorname{aff}\Omega(A,B)=\left\{(Y,W)\in\mathbb{E}\;\middle|\; AY=B \text{ and } \tfrac12 YY^T+W\in\operatorname{span}\mathcal{K}_A^\circ\right\},$ *where* $\operatorname{span}\mathcal{K}_A^\circ=\operatorname{span}\left\{vv^T\mid v\in\ker A\right\}$.

Proof. We apply the format of Proposition 5 and 7, respectively, for $C:=\Omega(A,B)$. Then

$$D=\{Y\mid AY=B\}\quad\text{and}\quad C_y=\begin{cases}\mathcal{K}_A^\circ-\frac12 YY^T, & \text{if } \ AY=B,\\ \emptyset, & \text{else.}\end{cases}\qquad (Y\in\mathbb{R}^{n\times m}),$$

a) Apply Proposition 5 and observe that $\operatorname{rint}(\mathcal{K}_A^\circ-\frac12YY^T)=\operatorname{rint}(\mathcal{K}_A^\circ)-\frac12YY^T$.

b) Apply Proposition 7 and observe that D is affine, and that $\operatorname{aff}(\mathcal{K}_A^\circ-\frac12YY^T)=\operatorname{aff}(\mathcal{K}_A^\circ)-\frac12YY^T$. □

As a direct consequence of Propositions 1 and 8, we obtain the following result for the special case $(A, B) = (0, 0)$.

COROLLARY 9. *It holds that*

$$\overline{\mathrm{conv}}\left\{(Y, -\frac{1}{2}YY^T) \mid Y \in \mathbb{R}^{n\times m}\right\} = \left\{(Y, W) \in \mathbb{E} \,\middle|\, W + \frac{1}{2}YY^T \preceq 0\right\},$$

and

$$\mathrm{int}\left(\overline{\mathrm{conv}}\left\{(Y, -\frac{1}{2}YY^T) \mid Y \in \mathbb{R}^{n\times m}\right\}\right) = \left\{(Y, W) \in \mathbb{E} \,\middle|\, W + \frac{1}{2}YY^T \prec 0\right\}.$$

We conclude this section by giving representations for the horizon cone and polar of $\Omega(A, B)$.

PROPOSITION 10 (The polar of $\Omega(A, B)$). *Let $\Omega(A, B)$ be as given in (8). Then*

$$\Omega(A, B)^\circ = \left\{(X, V) \,\middle|\, \begin{array}{c} \mathrm{rge}\binom{X}{B} \subset \mathrm{rge}\, M(V),\ V \in \mathcal{K}_A, \\ \frac{1}{2}\mathrm{tr}\left(\binom{X}{B}^T M(V)^\dagger \binom{X}{B}\right) \le 1 \end{array}\right\}.$$

Moreover,

$$\Omega(A, B)^\infty = \{0_{n\times m}\} \times \mathcal{K}_A^\circ \tag{14}$$

and

$$(\Omega(A, B)^\circ)^\infty = \left\{(X, V) \,\middle|\, \begin{array}{c} \mathrm{rge}\binom{X}{B} \subset \mathrm{rge}\, M(V),\ V \in \mathcal{K}_A, \\ \frac{1}{2}\mathrm{tr}\left(\binom{X}{B}^T M(V)^\dagger \binom{X}{B}\right) \le 0 \end{array}\right\}. \tag{15}$$

Proof. Given any nonempty closed convex set $C \subset \mathbb{E}$, it is easily seen that $C^\circ = \{z \mid \sigma_C(z) \le 1\}$. Consequently, our expression for $\Omega(A, B)^\circ$ follows from (2).

To see (14), let $(Y, W) \in \Omega(A, B)$ and recall that $(S, T) \in \Omega(A, B)^\infty$ if and only if $(Y + tS, W + tT) \in \Omega(A, B)$ for all $t \ge 0$. In particular, for $(S, T) \in \Omega(A, B)^\infty$, we have $A(Y + tS) = B$ and

$$\frac{1}{2}\left[YY^T + t(SY^T + YS^T) + \frac{t^2}{2}SS^T\right] + (W + tT) \in \mathcal{K}_A^\circ \quad (t > 0). \tag{16}$$

Consequently, $AS = 0$ and, if we divide (16) by t^2 and let $t \uparrow \infty$, we see that $SS^T \in \mathcal{K}_A^\circ$. But $SS^T \in \mathcal{K}_A$ since $\mathrm{rge}\, S \subset \ker A$, so we must have $S = 0$. If we now divide (16) by t and let $t \uparrow \infty$, we find that $T \in \mathcal{K}_A^\circ$. Hence the set on the left-hand side of (14) is contained in the one on the right. To see the reverse inclusion, simply recall that $\mathcal{K}_A^\circ$ is a closed convex cone so that $\mathcal{K}_A^\circ + \mathcal{K}_A^\circ \subset \mathcal{K}_A^\circ$.

Finally, we show (15). Since $(0, 0) \in \Omega(A, B)^\circ$, we have $(S, T) \in (\Omega(A, B)^\circ)^\infty$ if and only if $(tS, tT) \in \Omega(A, B)^\circ$ for all $t > 0$, or equivalently, for all $t > 0$,

$$tT \in \mathcal{K}_A \ \text{ and } \ \exists\, (Y_t, Z_t) \in \mathbb{R}^{n\times m} \times \mathbb{R}^{p\times m} \ \text{ s.t. } \begin{pmatrix} tS \\ B \end{pmatrix} = M(tT)\begin{pmatrix} Y_t \\ Z_t \end{pmatrix}$$

$$\text{with } \frac{1}{2}\mathrm{tr}\left(\begin{pmatrix} Y_t \\ Z_t \end{pmatrix}^T M(tT)\begin{pmatrix} Y_t \\ Z_t \end{pmatrix}\right) \le 1,$$

or equivalently, by taking $\widehat{Z}_t := t^{-1}Z_t$,

$$T \in \mathcal{K}_A \quad\text{and}\quad \exists\,(Y_t,\widehat{Z}_t)\in\mathbb{R}^{n\times m}\times\mathbb{R}^{p\times m}\ \text{ s.t. } \begin{pmatrix} S\\ B\end{pmatrix} = M(T)\begin{pmatrix} Y_t\\ \widehat{Z}_t\end{pmatrix}$$
$$\text{with }\ \frac{t}{2}\mathrm{tr}\left(\begin{pmatrix} Y_t\\ \widehat{Z}_t\end{pmatrix}^T M(T)\begin{pmatrix} Y_t\\ \widehat{Z}_t\end{pmatrix}\right)\le 1.$$

If we take $\binom{Y_t}{\widehat{Z}_t} := M(T)^\dagger\binom{S}{B}$, we find that $(S,T)\in(\Omega(A,B)^\circ)^\infty$ if and only if

$$T \in \mathcal{K}_A \quad\text{and}\quad \frac{t}{2}\mathrm{tr}\left(\begin{pmatrix} S\\ B\end{pmatrix}^T M(T)^\dagger\begin{pmatrix} S\\ B\end{pmatrix}\right)\le 1 \quad (t>0),$$

which proves the result. □

5. $\boldsymbol{\sigma_{\Omega(A,0)}}$ as a gauge. Note that the origin is an element of $\Omega(A,B)$ if and only if $B=0$. In this case the support function of $\Omega(A,0)$ equals the gauge of $\Omega(A,0)^\circ$. Gauges are important in a number of applications and they posses their own duality theory [6, 7, 8]. An explicit representation for both $\gamma_{\Omega(A,0)^\circ}$ and $\gamma_{\Omega(A,0)}$ will be given in the following theorem.

THEOREM 11 ($\sigma_{\mathcal{D}(A,0)}$ is a gauge). *Let $\Omega(A,B)$ be as given in (8). Then*

$$\sigma_{\Omega(A,0)}(X,V) = \gamma_{\Omega(A,0)^\circ}(X,V) = \gamma^\circ_{\Omega(A,0)}(X,V), \tag{17}$$

and

$$\begin{aligned}\gamma_{\Omega(A,0)}(Y,W) &= \sigma_{\Omega(A,0)^\circ}(Y,W)\\ &= \begin{cases}\frac{1}{2}\sigma_{\min}^{-1}(-Y^\dagger W(Y^\dagger)^T) & \text{if } \mathrm{rge}\,Y\subset\ker A\cap\mathrm{rge}\,W,\ W\in\mathcal{K}_A^\circ,\\ +\infty & \text{else,}\end{cases}\end{aligned} \tag{18}$$

where $\sigma_{\min}(-Y^\dagger W(Y^\dagger)^T)$ is the smallest nonzero singular-value of $-Y^\dagger W(Y^\dagger)^T$ when such an eigenvalue exists and $+\infty$ otherwise, e.g. when $Y=0$. Here we interpret $\frac{1}{\infty}$ as 0 ($0=\frac{1}{\infty}$), and so, in particular, $\gamma_{\Omega(A,0)}(0,W)=\delta_{\mathcal{K}_A^\circ}(W)$.

Proof. The expression (17) follows from [11, Theorem 14.5]. To show (18), first observe that

$$t\,\Omega(A,0) = \left\{(Y,W)\ \middle|\ AY=0 \ \text{ and }\ \frac{1}{2}YY^T+tW\in\mathcal{K}_A^\circ\right\}, \tag{19}$$

whose straightforward proof is left to the reader.

Given $\bar t\ge 0$, by (19), $(Y,W)\in t\Omega(A,0)$ for all $t>\bar t$ if and only if $AY=0$ and $\frac{1}{2}YY^T+tW\in\mathcal{K}_A^\circ$ for all $t>\bar t$. By Proposition 1 a), this is equivalent to $AY=0$ and

$$\frac{1}{2}YY^T+tW = P\left(\frac{1}{2}YY^T+tW\right)P\preceq 0 \quad (t>\bar t), \tag{20}$$

where, again, P is the orthogonal projection onto $\ker A$. Dividing this inequality by t and taking the limit as $t\uparrow\infty$ tells us that $W=PWP\preceq 0$. Since YY^T is positive semidefinite, inequality (20) also tells us that $\ker W\subset\ker Y^T$, i.e. $\mathrm{rge}\,Y\subset\mathrm{rge}\,W$. Consequently,

$$\mathrm{dom}\,\gamma_{\Omega(A,0)}\subset\{(Y,W)\mid \mathrm{rge}\,Y\subset\ker A\cap\mathrm{rge}\,W,\ W\in\mathcal{K}_A^\circ\}.$$

Now suppose $(Y, W) \in \operatorname{dom} \gamma_{\Omega(A,0)}$. Let $Y = U\Sigma V^T$ be the reduced singular-value decomposition of Y where Σ is an invertible diagonal matrix and U, V have orthonormal columns. Since $\operatorname{rge} Y \subset \operatorname{rge} W = (\ker W)^\perp$, we know that $U^T W U$ is negative definite, and so $\Sigma^{-1} U^T W U \Sigma^{-1}$ is also negative definite. Multiplying (20) on the left by $\Sigma^{-1} U^T$ and on the right by $U\Sigma^{-1}$ gives

$$\mu I \preceq -2\Sigma^{-1} U^T W U \Sigma^{-1} \quad (0 < \mu \leq \bar{\mu}),$$

where $\bar{\mu} = \bar{t}^{-1}$. The largest $\bar{\mu}$ satisfying this inequality is

$$\sigma_{\min}(-2Y^\dagger W (Y^\dagger)^T) = \sigma_{\min}(-2\Sigma^{-1} U^T W U \Sigma^{-1}) > 0,$$

or equivalently, the smallest possible $\bar{t}$ in (20) is $1/\sigma_{\min}(-2Y^\dagger W (Y^\dagger)^T)$, which proves the result. ∎

6. Conclusions.

The representation $\Omega(A, B)$ for the closed convex hull of the set $\mathcal{D}(A, B)$ in Theorem 2 is a dramatic simplification of the one given in [3]. As a consequence, we also obtain simplified expressions for both the normal cone to $\Omega(A, B)$ and the subdifferential for generalized matrix-fractional functions in Section 3. In addition, representations for several important geometric objects related to the set $\Omega(A, B)$ are computed in Section 4. These results provide the key to the applications discussed in [3], and open the door to the numerous further applications discussed in [4].

Acknowledgments. The authors thank the Department of Mathematics and Statistics at McGill University for supporting this research project.

REFERENCES

[1] H. H. Bauschke and P. L. Combettes, *Convex analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer-Verlag, 2011.
[2] S. Boyd and L. Vandenbergh: *Convex Optimization.* Cambridge University Press, 2004.
[3] J. V. Burke and T. Hoheisel, *Matrix support functionals for inverse problems, regularization, and learning*, SIAM Journal on Optimization, 25 (2015), pp. 1135–1159.
[4] J. V. Burke, Y. Gao and T. Hoheisel: *Infimal projections of the generalized matrix-fractional function.* Preprint, University of Washington, 2017.
[5] J. Dattorro: *Convex Optimization & Euclidean Distance Geometry.* M$\varepsilon\beta oo$ Publishing USA, Version 2014.04.08, 2005 .
[6] R. M. Freund. Dual gauge programs, with applications to quadratic programming and the minimum-norm problem. *Mathematical Programming*, 38(1):47–67, 1987.
[7] M. P. Friedlander and I. Macêdo. Low-rank spectral optimization via gauge duality. *SIAM Journal on Scientific Computing*, 28(3):A1616–A1638, 2016.
[8] M. P. Friedlander, I. Macedo, and T. K. Pong. Gauge optimization and duality. *SIAM Journal on Optimization*, 24(4):1999–2022, 2014.
[9] J. Gallier: *Geometric Methods and Applications: For Computer Science and Engineering.* Texts in Applied Mathematics, Springer New York, Dordrecht, London, Heidelberg, 2011.
[10] C.-J. Hsieh and P. Olsen: *Nuclear Norm Minimization via Active Subspace Selection.* JMLR W&CP 32 (1) :575-583, 2014.
[11] R. T. Rockafellar, *Convex analysis*, Princeton University Press, 1970.
[12] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317, Springer, 1998.