Journal of Convex Analysis
Volume 2 (1995), No.1/2, 173–183

The Convex Analysis of Unitarily Invariant Matrix Functions

A.S. Lewis[1]
Department of Combinatorics and Optimization,
University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1.
email: aslewis@orion.uwaterloo.ca

Received 12 July 1994
Revised manuscript received December 1995

Dedicated to R. T. Rockafellar on his 60th Birthday

A fundamental result of von Neumann's identifies unitarily invariant matrix norms as symmetric gauge functions of the singular values. Identifying the subdifferential of such a norm is important in matrix approximation algorithms, and in studying the geometry of the corresponding unit ball. We show how to reduce many convex-analytic questions of this kind to questions about the underlying gauge function, via an elegant Fenchel conjugacy formula. This approach also allows such results to be extended to more general unitarily invariant matrix functions.

Keywords : unitarily invariant norm, matrix function, Fenchel conjugacy, subdifferential, matrix optimization, extreme point, exposed point.

1991 Mathematics Subject Classification: Primary: 15A60 49J52, Secondary: 90C25 65F35

1. Introduction

Consider the vector space of $m \times n$ matrices, either real or complex. To be concrete we shall restrict attention to the complex case $\mathbb{C}^{m\times n}$ throughout: the parallel development for the real case will be immediately apparent. Many of the most important examples of norms $\|\cdot\|$ on $\mathbb{C}^{m\times n}$ are *unitarily invariant*:

$$\|VXU\| = \|X\| \ \text{ for any } X \in \mathbb{C}^{m\times n},\ U \in \mathcal{U}_n,\ V \in \mathcal{U}_m$$

(where $\mathcal{U}_n$ denotes the set of $n \times n$ unitary matrices).
A famous fundamental result of von Neumann's [9] (see for example [5]) states that such matrix norms can be characterized as composite functions of the form $f(\sigma(\cdot)) = f \circ \sigma$, where the function $\sigma : \mathbb{C}^{m\times n} \to \mathbb{R}^q$ (with $q = \min\{m, n\}$) has components $\sigma_1(X) \geq \sigma_2(X) \geq \ldots \geq \sigma_q(X) \geq 0$, the singular values of the matrix X, and the function $f : \mathbb{R}^q \to \mathbb{R}$ is a *symmetric gauge function* (a norm on $\mathbb{R}^q$ which is invariant under sign-changes and permutations of the components). Von Neumann's proof revolves around the

1 Research partially supported by the Natural Sciences and Engineering Research Council of Canada

ISSN 0944-6532 / \$ 2.50

elegant idea that

$$(f \circ \sigma)^D = f^D \circ \sigma \tag{1.1}$$

(where $\|\cdot\|^D$ denotes the polar, or dual norm).

The most common examples fall into two families. The *Schatten p norms* correspond to $f = \|\cdot\|_p$ (for $1 \le p \le \infty$), special cases being the *trace norm* ($p = 1$), the *Frobenius norm* ($p = 2$), and the *spectral norm* ($p = \infty$). The other family are the *Ky Fan k norms* (for $1 \le k \le q$), which correspond to the symmetric gauge function

$$f(\gamma) = \text{sum of } k \text{ largest of } \{|\gamma_1|, |\gamma_2|, \ldots, |\gamma_q|\},$$

the special case $k = 1$ giving the spectral norm once again, and $k = q$ giving the trace norm. Such norms have been the focus of recent interest in matrix approximation algorithms (see for example [11]), and in a variety of investigations aiming to analyze the geometry of the unit ball in the matrix space, $B_{f\circ\sigma}$, in terms of the geometry of the underlying ball B_f in $\mathbb{R}^q$ (see [1], [12], [13], [3], [4]). In each case a central problem is the investigation of the *subdifferential* of $f \circ \sigma$ at a matrix X,

$$\begin{aligned}\partial(f\circ\sigma)(X) = \{&Y \in \mathbb{C}^{m\times n} \mid \\ &\mathrm{Re}(\mathrm{trace} ZY^*) \le f(\sigma(X+Z)) - f(\sigma(X)) \;\forall\, Z \in \mathbb{C}^{m\times n}\}.\end{aligned} \tag{1.2}$$

The aim of this paper is to give a simple, self-contained approach to this problem, giving back the subdifferential formula for (1.2) in [13] for example. Our idea will be to generalize von Neumann's result somewhat by asking which convex *functions* (rather than simply norms) are unitarily invariant: appropriately, the key idea will be a Fenchel conjugacy formula analogous to von Neumann's polarity formula (1.1):

$$(f \circ \sigma)^* = f^* \circ \sigma.$$

(For all ideas and notation from convex analysis we refer the reader to [8].) Since the subdifferential (1.2) can be identified (directly from the definition of the conjugate) as

$$\partial(f\circ\sigma)(X) = \{Y \mid (f\circ\sigma)(X) + (f\circ\sigma)^*(Y) = \mathrm{Re}(\mathrm{trace} XY^*)\},$$

it becomes easy to calculate $\partial(f\circ\sigma)$ in terms of ∂f. In this fashion we can link smoothness (and, dually, strict convexity) properties of the matrix function $f \circ \sigma$ with those of the underlying function f: in the norm case, such results appear in [1] and independently in [12]. Furthermore, the identification of smooth, extreme and exposed points of the matrix unit ball $B_{f\circ\sigma}$ in terms of those of the underlying unit ball B_f is also straightforward with these tools (c.f. [1], [12]).

As we note, the results we derive in the norm case are not new. On the other hand, this new approach has twofold appeal. First, the assumptions on the underlying function f are weakened (essentially by dropping the requirement of positive homogeneity). Secondly, and perhaps more significantly, all the matrix analysis required is hidden in one key variational result of von Neumann's (Theorem 2.1): we avoid having to use analytic parametrizations of the singular values (as in [10], for example), or any partitioning (as in [13], [3]).

We remark finally that an analogous development for *weakly unitarily invariant* convex functions of Hermitian matrices (functions satisfying $F(X) = F(U^*XU)$ for unitary U) may be found in [6]. Such functions have become rather important in optimization: an example is the *logarithmic barrier*, $-\log\det X$ (for positive definite X).

2. Conjugates and subdifferentials of unitarily invariant functions

By $\mathbb{C}^{m\times n}$ we shall mean the *real* inner product space of $m \times n$ complex matrices, with the inner product $\langle X, Y\rangle = \mathrm{Re}(\mathrm{trace} XY^*)$. (The results we present can be immediately translated to the real case simply by replacing 'unitary' with 'orthogonal' throughout.) We shall also use $\langle\cdot,\cdot\rangle$ to denote the standard inner product on $\mathbb{R}^q$. For a vector γ in $\mathbb{R}^q$, the entries of the $m \times n$ diagonal matrix $\mathrm{Diag}\,\gamma$ are given by $(\mathrm{Diag}\,\gamma)_{ij} = \gamma_i$ if $i = j$ and 0 otherwise. The following result is essentially due to von Neumann (see the discussion in [3]).

Theorem 2.1. **(von Neumann)** *For any $m \times n$ matrices X, Y and Z,*

$$\max\{\langle VZU, Y\rangle \mid U \in \mathcal{U}_n,\ V \in \mathcal{U}_m\} = \langle\sigma(Z), \sigma(Y)\rangle, \tag{2.2}$$

and hence

$$\langle X, Y\rangle \le \langle\sigma(X), \sigma(Y)\rangle. \tag{2.3}$$

Equality holds in (2.3) *if and only if there exists a simultaneous singular value decomposition of X and Y of the form*

$$X = V(\mathrm{Diag}\,\sigma(X))U, \ \textit{and}$$
$$Y = V(\mathrm{Diag}\,\sigma(Y))U,$$

with U in $\mathcal{U}_n$ and V in $\mathcal{U}_m$.

Given any vector γ in $\mathbb{R}^q$, we write $\hat{\gamma}$ for the vector with components $|\gamma_i|$ arranged in nonincreasing order. Notice that $\sigma(\mathrm{Diag}\,\gamma) = \hat{\gamma}$. The special case of inequality (2.3) for diagonal matrices shows that for any vectors γ and μ in $\mathbb{R}^q$,

$$\langle\gamma, \mu\rangle \le \langle\hat{\gamma}, \hat{\mu}\rangle. \tag{2.4}$$

This is also easy to see directly.

A function $f : \mathbb{R}^q \to [-\infty, +\infty]$ is *absolutely symmetric* if $f(\gamma) = f(\hat{\gamma})$ for any vector γ in $\mathbb{R}^q$. Let us call a $q \times q$ matrix a *generalized permutation matrix* if it has exactly one nonzero entry in each row and each column, that entry being ± 1: we will denote the set of such matrices by Λ_q. Clearly f is absolutely symmetric if and only if

$$f(Q\gamma) = f(\gamma), \quad \text{for all } \gamma \in \mathbb{R}^q \text{ and } Q \in \Lambda_q. \tag{2.5}$$

A function $F : \mathbb{C}^{m\times n} \to [-\infty, +\infty]$ is *(strongly) unitarily invariant* if

$$F(VXU) = F(X) \ \text{ for all } X \text{ in } \mathbb{C}^{m\times n},\ U \text{ in } \mathcal{U}_n \text{ and } V \text{ in } \mathcal{U}_m.$$

The following simple result characterizes unitarily invariant functions on $\mathbb{C}^{m\times n}$ as functions of the form $f \circ \sigma$, with f absolutely symmetric on $\mathbb{R}^q$.

Proposition 2.2. *Unitarily invariant functions on* $\mathbb{C}^{m\times n}$ *are in one-to-one correspondence with absolutely symmetric functions on* $\mathbb{R}^q$*. Specifically, if the function* $F : \mathbb{C}^{m\times n} \to [-\infty, +\infty]$ *is unitarily invariant then the function* $f : \mathbb{R}^q \to [-\infty, +\infty]$ *defined by*

$$f(\gamma) = F(\operatorname{Diag}\gamma) \tag{2.7}$$

is absolutely symmetric, with $F = f \circ \sigma$*. Conversely, for any given function* $f : \mathbb{R}^q \to [-\infty, +\infty]$*, the function* $f \circ \sigma$ *is unitarily invariant on* $\mathbb{C}^{m\times n}$*.*

Proof. Suppose that F is unitarily invariant, and define f by (2.7). Given any vector γ in $\mathbb{R}^q$ we can choose $m \times m$ and $n \times n$ generalized permutation matrices Q_m and Q_n so that $Q_m(\operatorname{Diag}\gamma)Q_n = \operatorname{Diag}\hat{\gamma}$. Then

$$f(\hat{\gamma}) = F(\operatorname{Diag}\hat{\gamma}) = F(Q_m(\operatorname{Diag}\gamma)Q_n) = F(\operatorname{Diag}\gamma) = f(\gamma),$$

whence f is absolutely symmetric.
Now fix any $m \times n$ matrix X, and choose a singular value decomposition $X = V(\operatorname{Diag}\sigma(X))U$, with U in $\mathcal{U}_n$ and V in $\mathcal{U}_m$. Then

$$F(X) = F(V(\operatorname{Diag}\sigma(X))U) = F(\operatorname{Diag}\sigma(X)) = f(\sigma(X)),$$

whence $F = f \circ \sigma$. The converse is immediate since the set of singular values is unitarily invariant. □

If $(\mathcal{X}, \langle\cdot,\cdot\rangle)$ is a finite-dimensional, real inner product space, then given a function $g : \mathcal{X} \to [-\infty, +\infty]$, the *(Fenchel) conjugate* of g is the function $g^* : \mathcal{X} \to [-\infty, +\infty]$ defined by

$$g^*(y) = \sup\{\langle x, y\rangle - g(x) \mid x \in \mathcal{X}\}.$$

It is always convex and lower semicontinuous. The function g is *proper* if it never takes the value $-\infty$ and the set $\operatorname{dom} g = \{x \mid g(x) < +\infty\}$ is nonempty. In this case, for points x in $\operatorname{dom} g$ we can define the *subdifferential*

$$\partial g(x) = \{y \in \mathcal{X} \mid g(x) + g^*(y) = \langle x, y\rangle\},$$

whose elements are called *subgradients*. When g is proper and convex, g^* is also proper, with $g^{**} = g$ providing that g is also lower semicontinuous. The standard reference for all these ideas is [8]. The following collection of results is an easy exercise, using inequality (2.4), (2.5) and the Chain Rule.

Lemma 2.3. *Suppose that the function* $f : \mathbb{R}^q \to [-\infty, +\infty]$ *is absolutely symmetric. Then:*
(a) *The function* f *is differentiable at a vector* γ *if and only if it is differentiable at* $\hat{\gamma}$*.*
(b) *The conjugate function* f^* *is absolutely symmetric.*
(c) *For any vector* μ*,*

$$f^*(\mu) = \sup\{\langle\gamma, \hat{\mu}\rangle - f(\gamma) \mid \gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_q \geq 0\}.$$

(d) *If* μ *lies in* $\partial f(\gamma)$ *then* $\hat{\mu}$ *lies in* $\partial f(\hat{\gamma})$*.*

The next result is our key tool: a simple formula for the conjugate of any unitarily invariant function.

Theorem 2.4. **(Conjugacy Formula)** *If the function $f : \mathbb{R}^q \to [-\infty, +\infty]$ is absolutely symmetric then $(f \circ \sigma)^* = f^* \circ \sigma$.*

Proof. For any $m \times n$ matrix Y we have (using von Neumann's Theorem (2.1)):

$$\begin{aligned}
&(f \circ \sigma)^*(Y) \\
&= \sup\{\langle X, Y\rangle - f(\sigma(X)) \mid X \in \mathbb{C}^{m\times n}\} \\
&= \sup\{\langle VZU, Y\rangle - f(\sigma(Z)) \mid Z \in \mathbb{C}^{m\times n},\ U \in \mathcal{U}_n,\ V \in \mathcal{U}_m\} \\
&= \sup\{\sup\{\langle VZU, Y\rangle \mid U \in \mathcal{U}_n,\ V \in \mathcal{U}_m\} - f(\sigma(Z)) \mid Z \in \mathbb{C}^{m\times n}\} \\
&= \sup\{\langle \sigma(Z), \sigma(Y)\rangle - f(\sigma(Z)) \mid Z \in \mathbb{C}^{m\times n}\} \\
&= \sup\{\langle \gamma, \sigma(Y)\rangle - f(\gamma) \mid \gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_q \geq 0\} \\
&= f^*(\sigma(Y)),
\end{aligned}$$

the last step being a consequence of Lemma 2.3. □

As a consequence we obtain a simple formula for the subdifferential of a unitarily invariant function.

Corollary 2.5. **(Characterization of Subgradients)** *Let us suppose that the function $f : \mathbb{R}^q \to (-\infty, +\infty]$ is absolutely symmetric, and that the $m \times n$ matrix X has $\sigma(X)$ in $\mathrm{dom} f$. Then the $m \times n$ matrix Y lies in $\partial(f \circ \sigma)(X)$ if and only if $\sigma(Y)$ lies in $\partial f(\sigma(X))$ and there exists a simultaneous singular value decomposition of the form*

$$\begin{aligned}
X &= V(\mathrm{Diag}\,\sigma(X))U, \ \ \textit{and} \\
Y &= V(\mathrm{Diag}\,\sigma(Y))U,
\end{aligned}$$

with U in $\mathcal{U}_n$ and V in $\mathcal{U}_m$. In fact

$$\begin{aligned}
&\partial(f \circ \sigma)(X) = \\
&\quad \{V(\mathrm{Diag}\,\mu)U \mid \mu \in \partial f(\sigma(X)),\ U \in \mathcal{U}_n,\ V \in \mathcal{U}_m,\ X = V(\mathrm{Diag}\,\sigma(X))U\}.
\end{aligned}$$

Proof. The matrix Y lies in $\partial(f \circ \sigma)(X)$ if and only if

$$f(\sigma(X)) + f^*(\sigma(Y)) = (f \circ \sigma)(X) + (f \circ \sigma)^*(Y) = \langle X, Y\rangle,$$

by the Conjugacy Formula (Theorem 2.4). Since von Neumann's Theorem (2.1) implies that for any X and Y,

$$f(\sigma(X)) + f^*(\sigma(Y)) \ \geq\ \langle \sigma(X), \sigma(Y)\rangle \ \geq\ \langle X, Y\rangle,$$

we deduce that Y lies in $\partial(f \circ \sigma)(X)$ if and only if equality holds throughout. The first part now follows from the equality condition in von Neumann's Theorem.
One inclusion in the last formula is also now clear. To see the opposite inclusion, note that if μ lies in $\partial f(\sigma(X))$ and $X = V(\mathrm{Diag}\,\sigma(X))U$ for some unitary U and V then

$$\begin{aligned}
\langle X, V(\mathrm{Diag}\,\mu)U\rangle &\leq (f \circ \sigma)(X) + (f \circ \sigma)^*(V(\mathrm{Diag}\,\mu)U) \\
&= f(\sigma(X)) + f^*(\hat{\mu}) \\
&= f(\sigma(X)) + f^*(\mu) \\
&= \langle \sigma(X), \mu\rangle \\
&= \langle V(\mathrm{Diag}\,\sigma(X))U, V(\mathrm{Diag}\,\mu)U\rangle,
\end{aligned}$$

using the Conjugacy Formula and Lemma 2.3. Hence $V(\operatorname{Diag}\mu)U$ lies in $\partial(f \circ \sigma)(X)$ as required. □

We can also now easily characterize those unitarily invariant functions which are convex and lower semicontinuous.

Corollary 2.6. **(Characterization of Convexity)** *Suppose that the function $f : \mathbb{R}^q \to (-\infty, +\infty]$ is absolutely symmetric. Then the corresponding unitarily invariant function $f \circ \sigma$ is convex and lower semicontinuous on $\mathbb{C}^{m\times n}$ if and only if f is convex and lower semicontinuous.*

Proof. Suppose that f is convex and lower semicontinuous. Then since $f = f^{**}$ we have by the Conjugacy Formula (Theorem 2.4),

$$f \circ \sigma = f^{**} \circ \sigma = (f^* \circ \sigma)^*,$$

and so $f\circ\sigma$, being a conjugate function, is convex and lower semicontinuous. The converse is immediate, by restricting attention to diagonal matrices. □

A *symmetric gauge function* on $\mathbb{R}^q$ is a norm which is absolutely symmetric. Thus an absolutely symmetric convex function $f : \mathbb{R}^q \to (-\infty, +\infty]$ is a symmetric gauge function exactly when it is *positively homogeneous* ($f(k\gamma) = kf(\gamma)$ for all γ in $\mathbb{R}^q$ and scalar $k \geq 0$), and finite and strictly positive except at 0. A famous result of von Neumann's [9] states that unitarily invariant norms on $\mathbb{C}^{m\times n}$ can be characterized as functions of the form $f \circ \sigma$ where f is a symmetric gauge function. Evidently this result is a special case of the above corollary. The Characterization of Subgradients (Corollary 2.5), in the special case where f is a symmetric gauge function, subsumes the characterizations of subdifferentials of unitarily invariant norms in the recent papers [10], [13]. We illustrate with some well-known examples.

Example 2.7. The c-spectral norms [7].
For any vector μ in $\mathbb{R}^q$ with $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_q \geq 0$, we can define a symmetric gauge function by $f(\gamma) = \langle \mu, \hat{\gamma}\rangle$: only convexity needs checking, and this follows by writing f as a maximum of linear functionals:

$$f(\gamma) = \max\{\langle \mu, Q\gamma\rangle \mid Q \text{ a generalized permutation matrix}\}$$

(using inequality (2.4)). Hence the function $\langle \mu, \sigma(\cdot)\rangle$ is a unitarily invariant norm. The case where $\mu_i = 1$ for $i = 1, 2, \ldots, k$ and 0 otherwise gives the Ky Fan k norm.

Example 2.8. The spectral norm.
As we have seen, in the case where the function f is $\|\cdot\|_\infty$ we obtain the spectral norm. A straightforward calculation with the Characterization of Subgradients (Corollary 2.5) shows that the subdifferential of this norm at the matrix X is just the convex hull of matrices of the form $(v^1)(u^1)^*$ for column vectors u^1 in $\mathbb{C}^n$ and v^1 in $\mathbb{C}^m$ appearing in some singular value decomposition of X of the form $X = \sum_{i=1}^q \sigma_i(X)(v^i)(u^i)^*$ (c.f. Example 1 in [10], and also [2], [13]).

3. Differentiability and strict convexity

Given a finite-dimensional inner product space $\mathcal{X}$, then a convex function $g : \mathcal{X} \to (-\infty, +\infty]$ is differentiable at a point x in dom$g$ if and only if the subdifferential $\partial g(x)$ is a singleton (in which case $\partial g(x) = \{\nabla g(x)\}$) [8, Thm 25.1]. Our Characterization of Subgradients (Corollary 2.5) now makes it easy to identify when convex, unitarily invariant functions are differentiable (c.f. [4, Cor 5.2]).

Theorem 3.1. (Gradient Formula) *If a function $f : \mathbb{R}^q \to (-\infty, +\infty]$ is convex and absolutely symmetric then the corresponding convex, unitarily invariant function $f \circ \sigma$ is differentiable at the $m \times n$ matrix X if and only if f is differentiable at $\sigma(X)$. In this case*

$$\nabla(f \circ \sigma)(X) = V(\mathrm{Diag}\, \nabla f(\sigma(X)))U, \tag{3.2}$$

for any matrices U in $\mathcal{U}_n$ and V in $\mathcal{U}_m$ with $X = V(\mathrm{Diag}\, \sigma(X))U$, and furthermore

$$\sigma(\nabla(f \circ \sigma)(X)) = \nabla f(\sigma(X)). \tag{3.3}$$

Proof. In one direction the first part is clear (without convexity) simply by restricting attention to matrices of the form $V(\mathrm{Diag}\, \gamma)U$ for vectors γ in $\mathbb{R}^q$. Conversely then, suppose that f is differentiable at $\sigma(X)$ (which therefore lies in the interior of dom$f$). Since f must agree with f^{**} on a neighbourhood of $\sigma(X)$ [8, Thms 7.4 and 12.2], and consequently $f \circ \sigma$ agrees with $f^{**} \circ \sigma$ on a neighbourhood of X (since σ is continuous), we may as well assume that f is lower semicontinuous (by replacing f with the lower semicontinuous function f^{**} if necessary). Hence $f \circ \sigma$ is convex by the Characterization of Convexity (Corollary 2.6), and, by the Characterization of Subgradients (Corollary 2.5), any subgradient Y has $\sigma(Y) = \nabla f(\sigma(X))$. The subdifferential is convex, and the Frobenius norm of each element is constant ($\|\nabla f(\sigma(X))\|_2$). Thus the set is a singleton (because the Frobenius norm is strict), and the result follows. $\square$

Given a function $g : \mathcal{X} \to (-\infty, +\infty]$ which is lower semicontinuous. proper and convex, we say that g is *essentially smooth* if its subdifferential is a singleton whenever it is nonempty: in particular, a finite convex function on $\mathcal{X}$ is essentially smooth exactly when it is everywhere differentiable. We say that the function g is *essentially strictly convex* if it is strictly convex on any convex set on which the subdifferential is everywhere nonempty: in particular, a finite convex function on $\mathcal{X}$ is essentially strictly convex exactly when it is strictly convex. Essential smoothness and essential strict convexity are dual notions: the function g is essentially smooth if and only if the conjugate g^* is essentially strictly convex (and vice versa). All these ideas may be found in [8, §26]. For the special case of the following result for norms, see [1].

Corollary 3.2. (Smoothness and Strict Convexity) *Suppose that the function $f : \mathbb{R}^q \to (-\infty, +\infty]$ is proper, convex, lower semicontinuous and absolutely symmetric. Then f is essentially smooth (respectively essentially strictly convex) if and only if the corresponding unitarily invariant function $f \circ \sigma$ is essentially smooth (respectively essentially strictly convex).*

Proof. Suppose first that $f \circ \sigma$ is essentially smooth. If $\partial f(\mu)$ is nonempty then so is $\partial f(\hat{\mu})$ by Lemma 2.3, and hence so is $\partial(f \circ \sigma)(\mathrm{Diag}\, \hat{\mu})$ by the Characterization of Subgradients (Corollary 2.5). Since $f \circ \sigma$ is essentially smooth it must be differentiable at $\mathrm{Diag}\,\hat{\mu}$, whence f is differentiable at $\hat{\mu}$ by the Gradient Formula (Theorem 3.1), and hence at μ by Lemma 2.3.

Conversely, suppose that f is essentially smooth. If $\partial(f \circ \sigma)(X)$ is nonempty then so is $\partial f(\sigma(X))$ by the Characterization of Subgradients. Hence f is differentiable at $\sigma(X)$, and therefore $f \circ \sigma$ is differentiable at X by the Gradient Formula.

The strict case follows by taking conjugates using the Conjugacy Formula (Theorem 2.4). □

The example that we choose to illustrate this development is the special case of norms. The results in this case are known. However, as we remarked in the introduction, one of our principal concerns is to show how these may be derived in this framework in a concise, unified way, without further matrix analysis.

Given a norm $\|\cdot\|$ on $\mathcal{X}$, the dual norm $\|\cdot\|_*$ on $\mathcal{X}$ is defined by

$$\|y\|_* = \max\{\langle x, y\rangle \mid \|x\| = 1\}.$$

It is a straightforward computation to check that

$$(\|\cdot\|^2/2)^* = \|\cdot\|_*^2/2. \tag{3.5}$$

The norm $\|\cdot\|$ is *smooth* when $\|\cdot\|^2/2$ is everywhere differentiable, and is *strict* when $\|\cdot\|^2/2$ is strictly convex.

Example 3.3. (Smooth and strict norms) Suppose that the norm $\|\cdot\|$ is a symmetric gauge function on $\mathbb{R}^q$. Then the corresponding unitarily invariant norm $\|\sigma(\cdot)\|$ is smooth (respectively strict) if and only if $\|\cdot\|$ is smooth (respectively strict), and the dual norm is $\|\sigma(\cdot)\|_*$. To see this, note first that the discussion after the Characterization of Convexity (Corollary 2.6) ensures that $|||\cdot||| = \|\sigma(\cdot)\|$ is indeed a unitarily invariant norm. By (3.5), the dual norm is determined by

$$|||\cdot|||_*^2/2 = (|||\cdot|||^2/2)^* = (\|\sigma(\cdot)\|^2/2)^* = \|\sigma(\cdot)\|_*^2/2,$$

using the Conjugacy Formula on the absolutely symmetric function $\|\sigma(\cdot)\|^2/2$. It follows that $|||\cdot|||_* = \|\sigma(\cdot)\|_*$. Now as we observed, $\|\sigma(\cdot)\|$ is smooth if and only if $\|\sigma(\cdot)\|^2/2$ is essentially smooth. Applying the previous result shows that this is equivalent to the essential smoothness of $\|\cdot\|^2/2$, and hence to the smoothness of $\|\cdot\|$. The strict case follows by taking conjugates. These characterizations first appeared in [1] and independently in [12].

The above example shows how we can use to advantage the fact that our analysis applies to a somewhat more general class of functions than symmetric gauge functions (even when these may be of primary interest): the most significant feature is that we do not require positive homogeneity. For example, if $\|\cdot\|$ is a symmetric gauge function and $h : \mathbb{R}_+ \to (-\infty, +\infty]$ is lower semicontinuous, convex, nondecreasing, and finite at zero, then the function $h(\|\cdot\|)$ satisfies the conditions for our various characterizations (2.4, 2.5, 2.6, 3.1, 3.2), as well as any finite sum or supremum of such functions. Of course, such examples are not completely distinct from the symmetric gauge case, being decomposable into building blocks depending on some collection of symmetric gauge functions. However, it may not be transparent how to decompose a given function into this form, and there does not appear to be any analytic advantage to working with the function expressed in terms of symmetric gauge functions.

We continue by studying an example of the above form, $h(\|\cdot\|)$, to derive further known results in this framework. Given a norm $\|\cdot\|$ on $\mathcal{X}$, we denote the (closed) unit ball by

$$B_{\|\cdot\|} = \{x \in \mathcal{X} \mid \|x\| \leq 1\},$$

and the *indicator function* of this set is defined by

$$\delta_{B_{\|\cdot\|}}(x) = \begin{cases} 0, & \text{if } \|x\| \leq 1, \\ +\infty, & \text{otherwise.} \end{cases}$$

If $\|x\| = 1$, the *normal cone* at x is defined by

$$N(x|B_{\|\cdot\|}) = \{y \in \mathcal{X} \mid \|y\|_* = \langle x, y\rangle\},$$

and it is easily checked that

$$N(x|B_{\|\cdot\|}) = \partial\delta_{B_{\|\cdot\|}}(x) \tag{3.6}$$

(c.f. [8, pp. 215, 226]). Suppose now that $\|\cdot\|$ is a symmetric gauge function on $\mathbb{R}^q$. If we apply the Characterization of Subgradients (Corollary 2.5) to the function $\delta_{B_{\|\cdot\|}}$ at a matrix X with $\|\sigma(X)\| = 1$, we obtain the following normal cone formula:

$$N(X|B_{\|\sigma(\cdot)\|}) = \\ \{V(\operatorname{Diag}\mu)U \mid \mu \in N(\sigma(X)|B_{\|\cdot\|}),\ U \in \mathcal{U}_n,\ V \in \mathcal{U}_m,\ X = V(\operatorname{Diag}\sigma(X))U\}.$$

This formula is analogous to the 'dual set' formulae in [3], or [13, Thm 3.2]).

We conclude by showing how results on the extremal structure of the unit ball of a unitarily invariant norm appearing in [1], [12], for example, can be easily derived in this framework. Given a norm $\|\cdot\|$ on $\mathcal{X}$, and a point x with norm 1, we say that x is an *extreme point* (of the unit ball $B_{\|\cdot\|}$) if $x = \alpha y + (1-\alpha)z$ for points y and z in the unit ball and α in $(0, 1)$ implies that $x = y = z$, and x is a *smooth point* if the norm is differentiable at x. We say that x is an *exposed point* if there is a vector y such that the functional $\langle\cdot, y\rangle$ is maximized over the unit ball uniquely at x. Equivalently [8, Cor 25.1.3],

$$\nabla\|y\|_* = x, \quad \text{for some point } y. \tag{3.7}$$

Theorem 3.4. **(Extremal Structure)** *Suppose that $\|\cdot\|$ is a symmetric gauge function. Then a matrix X is exposed (respectively smooth, extreme) in the unit ball of the corresponding unitarily invariant norm $\|\sigma(\cdot)\|$ if and only if $\sigma(X)$ is exposed (respectively smooth, extreme) in the unit ball for $\|\cdot\|$.*

Proof. Suppose first that the vector $\sigma(X)$ is exposed, so that $\nabla\|\mu\|_* = \sigma(X)$ for some vector μ. By Lemma 2.3 we also have $\nabla\|\hat{\mu}\|_* = \sigma(X)$. Choose a singular value decomposition, $X = V(\operatorname{Diag}\sigma(X))U$ for some unitary U and V. Then the gradient of the dual norm $\|\sigma(\cdot)\|_*$ (see the previous example) at the matrix $V(\operatorname{Diag}\hat{\mu})U$ is exactly X, by the Gradient Formula (Theorem 3.1), and hence X is exposed.

Conversely, suppose that X is exposed, so that X is the gradient of $\|\sigma(\cdot)\|_*$ at some matrix Y. Then by (2.7), $\nabla\|\sigma(Y)\|_* = \sigma(X)$, whence $\sigma(X)$ is exposed.
The smooth case is a direct consequence of the Gradient Formula (Theorem 3.1).
The final case is that of extreme points. First, if the vector $\sigma(X)$ is not extreme in the unit ball $B_{\|\cdot\|}$ then there are vectors γ and μ distinct from $\sigma(X)$ in the unit ball and an α in $(0,1)$ with $\sigma(X) = \alpha\gamma + (1-\alpha)\mu$. Choose a singular value decomposition $X = V(\mathrm{Diag}\,\sigma(X))U$, with U in $\mathcal{U}_n$ and V in $\mathcal{U}_m$, and define matrices $Y = V(\mathrm{Diag}\,\gamma)U$ and $Z = V(\mathrm{Diag}\,\mu)U$. Then clearly Y and Z are distinct from X in $B_{\|\sigma(\cdot)\|}$, with $X = \alpha Y + (1-\alpha)Z$, and hence X is not extreme.
Conversely, suppose that $\sigma(X)$ is extreme, and yet that

$$X = \alpha_1 Y_1 + (1-\alpha_1)Z_1, \quad \text{for some } \alpha_1 \in (0,1),$$

where Y_1 and Z_1 are distinct from X in $B_{\|\sigma(\cdot)\|}$. Let us define

$$Y = \begin{cases} \frac{1}{2}(X+Y_1), & \text{if } \sigma(X) = \sigma(Y_1), \\ Y_1, & \text{otherwise.} \end{cases}$$

Then if $\sigma(X) = \sigma(Y_1)$, since the Frobenius norm is strict we have that

$$\|\sigma(Y)\|_2 < (\|\sigma(X)\|_2 + \|\sigma(Y_1)\|_2)/2 = \|\sigma(X)\|_2.$$

Thus in either case $\sigma(Y) \neq \sigma(X)$. By defining Z similarly, we can write $X = \alpha Y + (1-\alpha)Z$ with α in $(0,1)$, and $\sigma(Y)$ and $\sigma(Z)$ distinct from $\sigma(Z)$ in $B_{\|\cdot\|}$.
Clearly $\sigma(X)$ is distinct from $Q\sigma(Y)$ and $Q\sigma(Z)$ for any generalized permutation matrix Q. Hence, since $\sigma(X)$ is extreme it must be disjoint from the convex hull of the set

$$\{Q\sigma(Y), Q\sigma(Z) \mid Q \in \Lambda_q\}.$$

Via a separating hyperplane we see the existence of a vector μ in $\mathbb{R}^q$ with

$$\langle \mu, \sigma(X)\rangle > \langle \mu, Q\sigma(Y)\rangle \text{ and } \langle \mu, Q\sigma(Z)\rangle$$

for all Q in Λ_q. Choosing Q so that $Q\hat{\mu} = \mu$ shows that

$$\langle \hat{\mu}, \sigma(Y)\rangle \text{ and } \langle \hat{\mu}, \sigma(Z)\rangle < \langle \mu, \sigma(X)\rangle \leq \langle \hat{\mu}, \sigma(X)\rangle,$$

by (2.4). However this is in contradiction to the fact that, since the function $\langle \hat{\mu}, \sigma(\cdot)\rangle$ is convex (by the example after Corollary 2.6),

$$\langle \hat{\mu}, \sigma(X)\rangle = \langle \hat{\mu}, \sigma(\alpha Y + (1-\alpha)Z)\rangle \leq \alpha\langle \hat{\mu}, \sigma(Y)\rangle + (1-\alpha)\langle \hat{\mu}, \sigma(Z)\rangle.$$

This completes the proof. □

More generally, suppose that the set $C \subset \mathbb{R}^q$ is closed, convex, and *absolutely symmetric* (which is to say that a point γ lies in C if and only if $\hat{\gamma}$ lies in C). Then a matrix X is exposed (respectively extreme) in the set $\{Z \mid \sigma(Z) \in C\}$ if and only if $\sigma(X)$ is exposed (respectively extreme) in C. The proof is identical, with $\|\cdot\|_*$ replaced by δ^*_C.

References

[1] J. Arazy: On the geometry of the unit ball of unitary matrix spaces. Integral Equations and Operator Theory, 4:151–171, 1981.

[2] H. Berens and M. Finzel: A continuous selection of the metric projection in matrix spaces. In L. Collatz, G. Meinardus, and G. Nurnberger, editors, Numerical Methods of Approximation Theory, Volume 8. Birkhäser, Basel, 1987.

[3] E.M. de Sá: Exposed faces and duality for symmetric and unitarily invariant norms. Linear Algebra and its Applications, 197,198:429–450, 1994.

[4] E.M. de Sá: Faces of the unit ball of a unitarily invariant norm. Linear Algebra and its Applications, 197,198:451–493, 1994.

[5] R.A. Horn and C. Johnson: Matrix analysis. Cambridge University Press, Cambridge, U.K., 1985.

[6] A.S. Lewis: Convex analysis on the Hermitian matrices. Technical Report CORR 93-33, University of Waterloo, 1993. Accepted by SIAM Journal on Optimization.

[7] C.K. Li and N.K. Tsing: Duality between some linear preserver problems, II. Isometries with respect to c-spectral norms and matrices with fixed singular values. Linear Algebra and its Applications, 110:181–212, 1988.

[8] R.T. Rockafellar: Convex Analysis. Princeton University Press, Princeton, N.J., 1970.

[9] J. von Neumann: Some matrix inequalities and metrization of matric-space. Tomsk University Review, 1:286–300, 1937. In: Collected Works, Pergamon, Oxford, 1962, Volume IV, 205-218.

[10] G.A. Watson: Characterization of the subdifferential of some matrix norms. Linear Algebra and its Applications, 170:33–45, 1992.

[11] G.A. Watson: On matrix approximation problems with Ky Fan k norms. Numerical Algorithms, 5:263–272, 1993.

[12] K. Ziętak: On the characterization of the extremal points of the unit sphere of matrices. Linear Algebra and its Applications, 106:57–75, 1988.

[13] K. Ziętak: Subdifferentials, faces and dual matrices. Linear Algebra and its Applications, 185:125–141, 1993.

HIER :

Leere Seite
184