Abstract

It was shown in Part I of this work that the Gateaux differentiability of a convex unitarily invariant function is characterized by that of a similar induced rearrangement invariant function on the corresponding spectral space. A natural question is then whether this is also the case for Fréchet differentibility. In this paper we show the answer is positive. Although the result appears very natural, the proof turns out to be quite technically involved.

Convex Spectral Functions of Compact Operators, Part II: Lower Semicontinuity and Rearrangement Invariance

J.M. Borwein*
Centre for Experimental and Constructive Mathematics
Department of Mathematics and Statistics
Simon Fraser University
Burnaby, British Columbia V5A 1S6, CANADA
e-mail: jborwein@cecm.sfu.ca

A.S. Lewis†
Department of Combinatorics and Optimization
University of Waterloo
Waterloo, Ontario N2L 3G1, CANADA
e-mail: aslewis@orion.uwaterloo.ca

— and —

Q.J. Zhu‡
Department of Mathematics and Statistics
Western Michigan University
Kalamazoo, MI 49008, USA
e-mail: zhu@math-stat.wmich.edu

March 27, 2002

1 Introduction

Unitarily invariant functions of bounded self-adjoint operators on a separable Hilbert space appear natural in many analysis and optimization problems. A well known (finite dimensional) example is the maximum eigenvalue function for symmetric matrices. In various settings (generalized) differentials of such functions play a crucial role. On reflection, it becomes increasingly clear that the (generalized) differentiability of such a function is intrinsically related to that of an induced, rearrangement invariant, function on the space of the spectra of the operators. Research along this line was pioneered by von Neumann who introduced and exploited the concepts of unitarily invariant matrix norms and symmetric gauge functions [21]. Needless to say, a large operator theoretic literature exists on the subject.

Von Neumann's work was extended in another direction by Lewis [12] to arbitrary unitarily invariant functions of Hermitian matrices. The key pattern established in [12] is that Fréchet derivatives, Fréchet subdifferentials and limiting subdifferentials of a unitarily invariant function f of symmetric matrices are characterized by that of the induced function $f \circ \mathrm{dg}$, where dg is the canonical mapping that maps an n dimensional vector to an n by n diagonal matrix.

Recently, in [3], such a relationship was extended to convex subdifferentials of convex unitarily invariant functions on spaces of compact, self-adjoint operators. Moreover, it was shown in [3] that the Gâteaux differentiability of a convex unitarily invariant function is characterized by that of a similar induced rearrangement invariant function on the corresponding spectral space. A natural question is then whether this is also the case for Fréchet differentiability. In this paper we show the answer is positive. Although the result appears very natural, the proof turns out to be quite technically involved.

We now sketch a rough account of the difficulties we will have to overcome along with an outline of the paper. We observe that the proof of the finite dimensional case given in [12] relies on several nice properties (in

*Research was supported by NSERC and by the Shrum Endowment at Simon Fraser University.

†Research was supported by NSERC.

‡Research was supported by the National Science Foundation under grant DMS-9704203.

particular, the nonexpansivity) of the canonical eigenvalue mapping which sends a symmetric or Hermitian matrix to the vector of its eigenvalues arranged in the decreasing order (see Theorem 2.1 below). We start by examining the possible extensions of this fact. In an infinite dimensional space even the definition of such an eigenvalue mapping is less immediate. One superficially attractive approach was introduced in [3]. Informally, if λ is the eigenvalue mapping in [3] and A an appropriate compact operator, then $\lambda(A)$ is a vector whose components are the eigenvalues of A arranged by rotating among available positive eigenvalues, zero eigenvalues and negative eigenvalues of A, in the order of decreasing magnitude (see [3] for the precise definition). However, this eigenvalue mapping λ does not have any nonexpansivity property. In fact, it is not even continuous as shown by the following simple example: consider 'diagonal' operators A_n with diagonal elements $(-1, -1/2, -1/4, ..., -2^{1-n}, 0, -2^{-n}, ...)$ and A with diagonal elements $(-1, -1/2, -1/4, ...)$. Then A_n converges in various common operator norms to A (see definitions of these norms below). However, $\lambda(A_n) = (0, -1, -1/2, -1/4, ...)$ does not converge to $\lambda(A) = (-1, -1/2, -1/4, ...)$ in any corresponding norm of the spectral space. This led us to use a mapping $\hat{\lambda}$ derived from λ by removing all the zeroes in between a positive and a negative component of the λ and inserting a zero between components with the same sign. In Sections 2 and 3, we show that $\hat{\lambda}$ does have desired properties. Using $\hat{\lambda}$ creates a new problem. It is not an 'honest' eigenvalue mapping: zeroes may well be added or deleted in the process of forming $\hat{\lambda}$. Fortunately, for a *lower semicontinuous* rearrangement invariant function, the addition or deletion of zeroes to or from its variable does not change its value, continuity and differentiability at the given point. These facts are established in Section 4. Finally in Section 5 we prove our main results.

Our notation is similar to that of [3]. Let $\ell_2^{\mathbf{C}}$ be the space of sequences $(c_j)_{j\in\mathbf{N}}$ such that $c_j \in \mathbf{C}$ for each $j \in \mathbf{N}$ and such that $\sum_{j\in\mathbf{N}} |c_j|^2 < \infty$. All operators which we consider in this paper are maps from the <u>complex</u> Hilbert space $\ell_2^{\mathbf{C}}$ to itself. We denote by ℓ_2 the standard <u>real</u> normed sequence space (coefficients in $\mathbf{R}$) with canonical basis $(e^j)_{j=1}^{\infty}$. Thus, the ith component of the jth element of this basis is given by the Kronecker delta: $e_i^j = \delta_{i,j}$. Likewise, we define the <u>real</u> normed sequence spaces ℓ_1 and ℓ_∞, as well as c_0 (the space of null sequences). This means $\ell_1 \subset \ell_2 \subset c_0 \subset \ell_\infty$, and the norm on c_0 is the same as on ℓ_∞.

We also consider the space of bounded self-adjoint operators on $\ell_2^{\mathbf{C}}$, which we denote by B_{sa}. It is common knowledge [16, p. 96] that to each bounded operator T on $\ell_2{}^{C\prime}$ one can associate a unique positive operator $|T| = (T^*T)^{\frac{1}{2}} \in B_{sa}$. Now to each positive $T \in B_{sa}$ one can associate the (possibly infinite) value

$$\mathrm{tr}(T) := \sum_{j\in\mathbf{N}} \langle Te^j | e^j \rangle, \tag{1}$$

which we call the trace of T. The trace is actually independent of the orthonormal basis (e^j) chosen [16, p. 116].

We warn the reader that we chose our *trace class operators*, denoted by B_1, to be those <u>self-adjoint</u> operators T for which $\mathrm{tr}(|T|) < \infty$. Since any self-adjoint operator T can be decomposed as $T = T_+ - T_-$ where $T_+ \geq 0$ and $T_- \geq 0$ the trace operator can be extended to any $T \in B_1$ by $\mathrm{tr}(T) = \mathrm{tr}(T_+) - \mathrm{tr}(T_-)$. We let B_2 be the self-adjoint Hilbert-Schmidt operators, which are those to each bounded operator T on $\ell_2{}^{C\prime}$ such that $T^2 = (T^*T) \in B_1$. This gives $B_1 \subset B_2 \subset B_0 \subset B_{sa}$, where B_0 are the compact, self-adjoint operators. Now any compact, self-adjoint operator T is diagonalizable [16, p. 107]. That is, there exists a unitary operator U and $\lambda \in c_0$ such that $(U^*TU\ x)_j = \lambda_j x_j$ for all $j \in \mathbf{N}$ and all $x \in B_0$. This and the fact that $\mathrm{tr}(ST) = \mathrm{tr}(TS)$ makes proving *Lidskii's trace theorem* (difficult in general) easy for self-adjoint operators. Lidskii's trace theorem (see [20, p.45]) states

$$\mathrm{tr}(T) = \sum_{i=1}^{\infty} \lambda_i(T)$$

where $\{\lambda_i(T)\}$ is any <u>spectral sequence</u> of T, which is any sequence of eigenvalues of T (counted with multiplicities). Define $B_p \subset B_0$ for $p \in [1, \infty)$ by writing $T \in B_p$ if $\|T\|_p = (\mathrm{tr}(|T|^p))^{1/p} < \infty$. When T is self-adjoint we have (see [8, p.94])

$$\|T\|_p = (\mathrm{tr}(|T|^p))^{1/p} = (\sum_{i=1}^{\infty} |\lambda_i(T)|^p)^{1/p}. \tag{2}$$

In this case for $p, q \in (1, \infty)$ and $p^{-1} + q^{-1} = 1$, we get that B_p and B_q are paired, and the sesquilinear form $\langle S, T \rangle := \mathrm{tr}(ST)$ implements the duality on $B_p \times B_q$. These spaces are the *Schatten p-spaces* [16, p. 124]. We also consider the space B_0 paired with B_1.

Now, following [16] for each $x \in \ell_2^{\mathbf{C}}$ define the operator $x \odot x \in B_1$ by

$$(x \odot x)y = \langle x|y\rangle x.$$

For each $x \in \ell_\infty$ we define the operator $\operatorname{dg} x \in B_{sa}$ pointwise by

$$\operatorname{dg} x := \sum_i x_i(e^i \odot e^i),$$

For $p \in [1, \infty)$ if we have $x \in \ell_p$, then $\operatorname{dg} x \in B_p$ and $\|\operatorname{dg} x\|_p = \|x\|_p$. If we have $x \in c_0$, then $\operatorname{dg} x \in B_0$ and $\|\operatorname{dg} x\|_\infty = \|x\|_\infty$. This motivates

Definition 1.1 *For $p \in [1, \infty)$ we say ℓ_p is the spectral sequence space for B_p and c_0 is the spectral sequence space for B_0.*

2 Nonexpansivity of the Eigenvalue Map

Denote the vector space of $m \times m$ Hermitian matrices by H^m and consider the map $\lambda : H^m \to R^m$ which sends a Hermitian matrix to the vector with components its eigenvalues, arranged in decreasing order. We seek to generalize the following nonexpansive property.

Theorem 2.1 (Finite-Dimensional Nonexpansivity) *Any matrices S and T in H^m satisfy the inequality*

$$\|\lambda(S) - \lambda(T)\|_p \leq \|\lambda(S - T)\|_p, \forall 1 \leq p \leq \infty.$$

Proof. By Lidskii's majorization theorem (see [2, 13]),

$$\lambda(S) - \lambda(T) \in \text{conv } P^m \lambda(S - T),$$

where P^m denotes the group of $m \times m$ permutation matrices. The result now follows easily. ♣

A suitable infinite-dimensional version of Lidskii's majorization theorem exists (see [14, Thm 5.1]), and indeed gives the generalization we seek. In this section, rather than simply stating the result, we use ideas analogous to [14] to break down the proof into some simple steps, which we re-use later.

The normed space V is the real sequence space ℓ_p $(1 \leq p < \infty)$ or c_0. We consider the corresponding self-adjoint norms $\|\cdot\|_p$ $(1 \leq p < \infty)$ and $\|\cdot\|_\infty$ (the operator norm) respectively. Our next result generalizes Theorem 2.1 to finite rank operators.

For any vector $x \in R^n$, let $\bar{x}$ denote the vector in R^n obtained by reordering the components into decreasing order. We can define an analogous map on the *positive cone*

$$V^+ = \{x \in V : x_i \geq 0, \forall i\}$$

as follows. Given an element $x \in V^+$, set $j := 1$ and $I = N$ and repeat:

1. Choose i solving $\max\{x_i : i \in I\}$,
2. Set $\bar{x}_j := x_i, j := j+1, I := I/\{i\}$.

Thus, $^-$ reorders the nonzero components of an element of V^+ into decreasing order.

We need an analogous canonical reordering operation of the whole space V. We proceed as follows. For $x \in V$, define the *positive* and *negative parts* x^+ and $x^- \in V^+$ by

$$x_i^+ = \max(x_i, 0) \text{ and } x_i^- = \max(-x_i, 0).$$

Thus, $x = x^+ - x^-$. Now, for any elements y and $z \in V$, define $y \oplus z \in V$ by

$$(y \oplus z)_{2i-1} = y_i \text{ and } (y \oplus z)_{2i} = z_i$$

for $i \in N$. We define a map $\widehat{\ }: V \to V$ by

$$\widehat{x} = \overline{x^+} \oplus -\overline{x^-}, \text{ for } x \in V.$$

Thus, $\widehat{\ }$ reorders the nonzero components of an element of V, by alternately selecting positive and negative components, decreasing in size. Notice the components of $\widehat{x}$ are not typically a rearrangement of the components of x: zero components may be either added or deleted.

Lemma 2.2 *Given vectors s and t in R^n, define*

$$(s;0_n) = (s_1, s_2, ..., s_n, 0, 0, ..., 0) \in R^{2n}$$

and $(t, 0_n)$ analogously, and define

$$(s;0) = (s_1, s_2, ..., s_n, 0, 0, ...) \in V$$

and $(t;0)$ analogously. Then

$$\|\overline{(s;0_n)} - \overline{(t;0_n)}\|_p = \|\widehat{(s;0)} - \widehat{(t;0)}\|_p,$$

for all $1 \le p \le \infty$.

Proof. Let $\overline{(s;0_n)} = (\bar{s}_1, \bar{s}_2, ..., \bar{s}_{2n})$ and $\overline{(t;0_n)} = (\bar{t}_1, \bar{t}_2, ..., \bar{t}_{2n})$. Then

$$\widehat{(s;0)} = (\bar{s}_1, \bar{s}_{2n}, \bar{s}_2, \bar{s}_{2n-1}, ..., \bar{s}_n, \bar{s}_{n+1}, 0, 0, ...)$$

and

$$\widehat{(t;0)} = (\bar{t}_1, \bar{t}_{2n}, \bar{t}_2, \bar{t}_{2n-1}, ..., \bar{t}_n, \bar{t}_{n+1}, 0, 0, ...).$$

It suffices to observe that $\overline{(s;0_n)} - \overline{(t;0_n)}$ and $\widehat{(s;0)} - \widehat{(t;0)}$ have the same nonzero components. ♣

Theorem 2.3 (Finite Rank Nonexpansivity) *If the sequences x and y in V are finitely nonzero and the map U is unitary then*

$$\|U^*(\mathrm{dg}\, x)U - \mathrm{dg}\, y\|_p \ge \|\hat{x} - \hat{y}\|_p,$$

for all $1 \le p \le \infty$.

Proof. Choose $n \in N$ and vectors s and t in R^n so that $x = (s;0)$ and $y = (t;0)$ (in the notation of Lemma 2.2). Now choose any $2n$-dimensional subspace

$$L = \mathrm{span}\{e^i, U^*e^i : i = 1, 2, ..., n\}$$

and write the Hilbert space ℓ_2^C as $L \oplus L^\perp$. Since both the self-adjoint operators $U^*(\mathrm{dg}\, x)U$ and $\mathrm{dg}\, y$ have range in L we can decompose them correspondingly:

$$U^*(\mathrm{dg}\, x)U = S \oplus O \text{ and } \mathrm{dg}\, y = T \oplus O,$$

where the operators S and T are Hermitian on L. Since $S \oplus O$ is unitarily similar to $\mathrm{dg}\, x$, it has the same eigenvalues so

$$\lambda(S) = \overline{(s; 0_n)} \text{ and } \lambda(T) = \overline{(t; 0_n)}.$$

Returning to the inequality we wish to prove,

$$\begin{aligned} \|U^*(\mathrm{dg}\, x)U - \mathrm{dg}\, y\|_p &= \|\lambda(S-T)\|_p \\ &\geq \|\lambda(S) - \lambda(T)\|_p \\ &= \|\overline{(s;0_n)} - \overline{(t;0_n)}\|_p \\ &= \|\widehat{(s;0)} - \widehat{(t;0)}\|_p \\ &= \|\hat{x} - \hat{y}\|_p, \end{aligned}$$

by Theorem 2.1 (Finite-Dimensional Nonexpansivity) and Lemma 2.2. ♣

To proceed to the general case, we make finite rank approximations. Given any element x in V, we define the truncation x^n for $n \in N$ by

$$x_i^n = \begin{cases} x_i & \text{if } i \leq n, \\ 0 & \text{if } i > n. \end{cases}$$

Clearly $x^n \to x$ as $n \to \infty$.

Lemma 2.4 *For any element x of V^+,*

$$\overline{x^n} \to \overline{x} \text{ as } n \to \infty.$$

Proof. For any $m \in N$, providing $n \in N$ is sufficiently large we have

$$(\overline{x^n})_i = \overline{x}_i, \ (i \leq m),$$

(or in other words $(\overline{x^n})^m = \overline{x}^m$) and hence, the nonzero entries in the sequence $(\overline{x^n})_i$ $(i = m+1, m+2, ...)$ all correspond to nonzero entries in the sequence $\overline{x}_i$ $(i = m+1, m+2, ...)$. We deduce

$$\|\overline{x^n} - (\overline{x})^m\| \leq \|\overline{x} - (\overline{x})^m\|,$$

so

$$\|\overline{x^n} - \overline{x}\| \leq \|\overline{x^n} - (\overline{x})^m\| + \|(\overline{x})^m - \overline{x}\| \leq 2\|(\overline{x})^m - \overline{x}\|.$$

Since the right hand side can be made arbitrarily small, the result follows. ♣

Lemma 2.5 *For any element x of V,*

$$\widehat{x^n} \to \widehat{x} \text{ as } n \to \infty.$$

Proof. We prove the case $V = \ell_p$ $(1 \leq p < \infty)$: the case $V = c_0$ is similar. We have

$$\begin{aligned} \|\widehat{x^n} - \widehat{x}\|_p^p &= \|(\overline{(x^n)^+} \oplus -\overline{(x^n)^-}) - (\overline{x^+} \oplus -\overline{x^-})\|_p^p \\ &= \|\overline{(x^n)^+} - \overline{x^+}\|_p^p - \|\overline{(x^n)^-} \oplus -\overline{x^-}\|_p^p \\ &= \|\overline{(x^+)^n} - \overline{x^+}\|_p^p - \|\overline{(x^-)^n} \oplus -\overline{x^-}\|_p^p \to 0 \end{aligned}$$

by Lemma 2.4. ♣

Consider now the eigenvalue map $\widehat{\lambda} : B_p \to \ell_p$ $(1 \leq p < \infty)$ (or $B_0 \to c_0$) which maps an operator with eigenvalues $\{x_1, x_2, ...\}$ to $\widehat{x}$. Notice our definition of the eigenvalue map is essentially equivalent to that in [14]. The main result of this section shows $\widehat{\lambda}$ is nonexpansive.

Theorem 2.6 (Nonexpansivity) *If the operators X and Y belong to the space B_p $(1 \leq p < \infty)$ then they satisfy the inequality*

$$\|X - Y\|_p \geq \|\widehat{\lambda}(X) - \widehat{\lambda}(Y)\|_p.$$

A corresponding inequality holds in B_0.

Proof. It suffices to prove, for arbitrary elements x and y in V and unitary U, the inequality

$$\|U^*(\mathrm{dg}\, x)U - \mathrm{dg}\, y\| \geq \|\widehat{x} - \widehat{y}\|.$$

But by Theorem 2.3 (Finite Rank Nonexpansivity) we know that

$$\|U^*(\mathrm{dg}\, x^n U) - \mathrm{dg}\, y^n\| \geq \|\widehat{x^n} - \widehat{y^n}\|$$

for all $n \in N$, and the result follows by taking limits and applying Lemma 2.5. ♣

Corollary 2.7 *The map $x \mapsto \widehat{x}$ on V is nonexpansive.*

Proof. Restrict to diagonal operators in the previous result. ♣

3 Von Neumann-type Inequalities

Any matrices S and T in H^n satisfy the inequality

$$\mathrm{tr}(ST) \leq \langle \lambda(S), \lambda(T) \rangle.$$

The special case for diagonal matrices is the inequality

$$\langle s, t \rangle \leq \langle \bar{s}, \bar{t} \rangle, \text{ for } s, t, \in R^n. \tag{3}$$

In this section we generalize these inequalities to infinite dimensions. We consider the paired spaces

$$(V, W) = (\ell_p, \ell_q), (1 < p < \infty) \text{ or } (\ell_1, c_0) \text{ or } (c_0, \ell_1).$$

Lemma 3.1 *Any elements $x \in V^+$ and $y \in W^+$ satisfy the inequality*

$$\langle x, y \rangle \leq \langle \bar{x}, \bar{y} \rangle.$$

Proof. Notice the truncation x^n satisfies $(\overline{x^n})_i \leq \overline{x}_i$ for all i, so by inequality (3) we obtain

$$\langle \bar{x}, \bar{y} \rangle \geq \langle \overline{x^n}, \overline{y^n} \rangle \geq \langle x^n, y^n \rangle \to \langle x, y \rangle$$

as $n \to \infty$. ♣

Theorem 3.2 *Any elements x of V and y of W satisfy the inequality*

$$\langle x, y \rangle \leq \langle \widehat{x}, \widehat{y} \rangle.$$

Proof. We have by the previous lemma,

$$\begin{aligned}
\langle \widehat{x}, \widehat{y} \rangle &= \langle \overline{x^+} \oplus -\overline{x^-}, \overline{y^+} \oplus -\overline{y^-} \rangle \\
&= \langle \overline{x^+}, \overline{y^+} \rangle + \langle \overline{x^-}, \overline{y^-} \rangle \\
&\geq \langle x^+, y^+ \rangle + \langle x^-, y^- \rangle \\
&\geq \langle x^+ - x^-, y^+ - y^- \rangle = \langle x, y \rangle.
\end{aligned}$$

♣

The next result illuminates the infinite-dimensional von Neumann-type inequality proved in Part I of this paper (see [3, Thm 2.2]).

Theorem 3.3 *Any elements x of V and y of W satisfy*

$$\sup_{\pi \in \mathcal{R}} \sum_{i=1}^{\infty} x_{\pi(i)} y_i = \langle \widehat{x}, \widehat{y} \rangle. \tag{4}$$

Here, $\mathcal{R}$ is the set of all permutations of the sequence of natural numbers.

Proof. We first observe that the left-hand-side of (4) is continuous in x and y separably, and so is the right-hand-side, since $\widehat{\ }$ is nonexpansive. Hence as in the proof of [3, Theorem 2.2], it suffices to prove the result for finitely nonzero x and y.

Suppose, therefore, without loss of generality

$$x_1 \geq x_2 \geq ... \geq x_p > 0 > x_{p+1} \geq x_{p+2} \geq ... \geq x_{p+q}$$

and $x_i = 0$ $(i > p + q)$, and

$$y_1 \geq y_2 \geq ... \geq y_r > 0 > y_{r+1} \geq y_{r+2} \geq ... \geq y_{r+s}$$

and $y_i = 0$ $(i > r + s)$. The left-hand-side of equation (4) considers a supremum over possible matching of components of x with components of y, and the possible values of the sum will not change if we consider only those rearrangements π satisfying

$$\pi(i) = i \ (i > p + q + r + s).$$

Hence, by inequality (3), the supremum is attained by

$$\langle (x_1, x_2, ..., x_p, 0, ..., 0, x_{p+1}, x_{p+2}, ..., x_{p+q}), \\ (y_1, y_2, ..., y_r, 0, ..., 0, y_{r+1}, y_{r+2}, ..., y_{r+s}) \rangle$$

which is exactly the right-hand-side. ♣

Combining this result with [3, Theorem 2.2] gives the following elegant inequality.

Theorem 3.4 *Let $(\mathcal{V}, \mathcal{W})$ be the paired spaces (B_p, B_q), $(1 < p < \infty)$ or (B_1, B_0) or (B_0, B_1). Any operators X in $\mathcal{V}$ and Y in $\mathcal{W}$ satisfy the inequality*

$$\mathrm{tr}(XY) \leq \langle \widehat{\lambda}(X), \widehat{\lambda}(Y) \rangle.$$

4 Lower Semicontinuity and Rearrangements

We consider again the paired spaces

$$(V, W) = (\ell_p, \ell_q),\ (1 < p < \infty) \text{ or } (\ell_1, c_0) \text{ or } (c_0, \ell_1).$$

In this section we consider the consequences of lower semicontinuity for a rearrangement invariant function $\phi : V \to \bar{R} := R \cup \{+\infty\}$.

Recall from our previous notation, that for any element x of V we have

$$(x \oplus 0)_{2i-1} = x_i \text{ and } (x \oplus 0)_{2i} = 0.$$

The map $\Lambda : V \to V$ is the unilateral shift

$$(\Lambda x)_i = \begin{cases} x_{i-1} & (i > 1), \\ 0 & (i = 1). \end{cases}$$

Theorem 4.1 (Insignificance of Zeroes) *If the function $\phi : V \to \bar{R}$ is rearrangement invariant and norm lower semicontinuous, then the value $\phi(x)$ for any element x of V is unchanged by the addition or deletion of arbitrary many zero components to or from x. In particular,*

$$\phi(x) = \phi(\Lambda x) = \phi(x \oplus 0) = \phi(\widehat{x}).$$

Proof. Consider the convergent sequence

$$z^n := (x_1, x_2, ..., x_{n-1}, 0, x_n, x_{n+1}, ...) \to x$$

as $n \to \infty$. By the lower semicontinuity and rearrangement invariance we have

$$\phi(x) \leq \liminf \phi(z^n) = \phi(\Lambda x).$$

On the other hand, a similar argument with the sequence

$$(x_n, x_1, x_2, ..., x_{n-1}, x_{n+1}, x_{n+2} ...) \to \Lambda x$$

shows $\phi(x) \geq \phi(\Lambda x)$, so, in fact, $\phi(x) = \phi(\Lambda x)$. The fact that $\phi(x) = \phi(x \oplus 0)$ follows similarly by considering the sequence

$$(0, x_1, x_2, ..., x_n, 0, x_{n+1}, 0, x_{n+2}, 0, ...) \to \Lambda x$$

and (using $\phi(x) = \phi(\Lambda^n x)$ for all positive integers n) the sequence

$$(x_1, 0, x_2, 0, x_3, ..., 0, x_n, 0, x_{n+1}, x_{n+2}, x_{n+3}, ...) \to x \oplus 0.$$

The rest of the results follows. ♣

In passing we note that the following curiosity. The 'weak' topology on V means the $\sigma(V, W)$ topology.

Proposition 4.2 *If the function $\phi : V \to \bar{R}$ is rearrangement invariant and weakly lower semicontinuous then*

$$\phi(a, x_1, x_2, ...) \geq \phi(x), \;\; a \in R,$$

$$\phi(x \oplus z) \geq \phi(x) \;\; (z \in V),$$

and ϕ attains its infimum at 0.

Proof. For the first parts, consider the weakly convergent sequences

$$(x_1, x_2, ..., x_n, a, x_{n+1}, x_{n+2}, ...) \to x$$

$$(x_1, x_2, ..., x_n, z_1, x_{n+1}, z_2, x_{n+2}, ...) \to x$$

and for the second claim, observe $\Lambda^n x$ converges weakly to 0. ♣

Theorem 4.3 *If the function $\phi : V \to \bar{R}$ is convex, norm lower semicontinuous and rearrangement invariant then any elements $x \in V$ and $y \in W$ satisfy*

$$y \in \partial\phi(x) \Rightarrow \hat{y} \in \partial\phi(\hat{x}).$$

Proof. Let us denote the Fenchel conjugate of ϕ by ϕ^*. If $y \in \partial\phi(x)$ holds then

$$\phi(\hat{x}) + \phi^*(\hat{y}) = \phi(x) + \phi^*(y) = \langle x, y\rangle \leq \langle \hat{x}, \hat{y}\rangle,$$

using Theorems 2.3 and 4.1 and the fact that ϕ^* is lower semicontinuous and rearrangement invariant. ♣

Just as the values of a lower semicontinuous, rearrangement invariant function agree at a point x and $\hat{x}$, so do several of its analytic properties. The following is an example.

Theorem 4.4 (Continuity and Differentiability) *For any norm lower semicontinuous, rearrangement invariant function $\phi : V \to \bar{R}$ and any element x of V, norm continuity of ϕ at the points x and $\widehat{x}$ are equivalent. The analogous equivalence for Fréchet differentiability also holds.*

Proof. Clearly, continuity of ϕ at x and at $(x_{\pi(i)})$ for any rearrangement π are equivalent (as are Fréchet differentiability). Since the nonzero components of $\widehat{x}$ are simply a rearrangement of the nonzero components of x, we can derive $\widehat{x}$ from a rearrangement of x by adding or deleting (possibly infinite many) zeros. Observing that adding or deleting zeros to x can always be done by applying operations Λx and $x \oplus 0$ a finite number of times, the theorem follows from the following two lemmas.

Lemma 4.5 *For any norm lower semicontinuous, rearrangement invariant function $\phi : V \to \bar{R}$ and any element x of V, norm continuity of ϕ at the points x and Λx are equivalent. The analogous equivalence for Fréchet differentiability also holds.*

Proof. First we show that continuity at x is equivalent to continuity at Λx. Suppose ϕ is continuous at Λx. For any sequence $z^r \to x$, we know $\Lambda z^r \to \Lambda x$, so by Theorem 4.1 (Insignificance of Zeroes)

$$\phi(z^r) = \phi(\Lambda z^r) \to \phi(\Lambda x) = \phi(x).$$

Thus ϕ is continuous at x.

Conversely, if ϕ is continuous at x and $z^r \to \Lambda x$ then we can write

$$z^r = a_r e^1 + \Lambda y^r$$

where $a_r \to 0 \in R$ and $y^r \to x$, and then if we define a sequence

$$\tilde{y}^r := (y_1^r, y_2^r, ..., y_n^r, a_r, y_{n+1}^r, y_{n+2}^r, ...) \to x.$$

We see

$$\phi(z^r) = \phi(\tilde{y}^r) \to \phi(x) = \phi(\Lambda x)$$

so ϕ is continuous at Λx.

Next suppose ϕ has Fréchet derivative $\phi'(\Lambda x)$ at Λx. Then we know

$$\begin{aligned}\phi(z^r) &= \phi(\Lambda z^r) \\ &= \phi(\Lambda x) + \langle \phi'(\Lambda x), \Lambda z^r - \Lambda x\rangle + o(\|z^r - x\|) \\ &= \phi(x) + \langle \Lambda^* \phi'(\Lambda x), z^r - x\rangle + o(\|z^r - x\|)\end{aligned}$$

so ϕ has Fréchet derivative $\Lambda^*\phi'(\Lambda x)$ at x.

Conversely, suppose ϕ has Fréchet derivative $\phi'(x)$ at x. Then

$$\begin{aligned}\phi(z^r) &= \phi(\tilde{y}^r) \\ &= \phi(x) + \langle \phi'(x), \tilde{y}^r - x\rangle + o(\|\tilde{y}^r - x\|) \\ &= \phi(\Lambda x) + \langle \Lambda\phi'(x), z^r - \Lambda x\rangle + o(\|z^r - \Lambda x\|)\end{aligned}$$

so ϕ has Fréchet derivative $\Lambda\phi'(x)$ at Λx. ♣

Lemma 4.6 *For any norm lower semicontinuous, rearrangement invariant function $\phi : V \to \bar{R}$ and any element x of V, norm continuity of ϕ at the points x and $x \oplus 0$ are equivalent. The analogous equivalence for Fréchet differentiability also holds.*

Proof. Again we deal with the equivalence of continuity first. Suppose ϕ is continuous at $x \oplus 0$. For any sequence $z^r \to x$, we have $z^r \oplus 0 \to x \oplus 0$, so

$$\phi(z^r) = \phi(z^r \oplus 0) \to \phi(x \oplus 0) = \phi(x).$$

Thus, ϕ is continuous at x.

Conversely, if ϕ is continuous at x and $z^r \to x \oplus 0$ then we can write $z^r = y^r \oplus u^r$ where $y^r \to x$ and $u^r \to 0$, and then if we define a sequence

$$\tilde{y}^r = (y_1^r, y_2^r, ..., y_n^r, u_1^r, y_{n+1}^r, u_2^r, y_{n+2}^r, ...) \to x$$

we see

$$\phi(z^r) = \phi(\tilde{y}^r) \to \phi(x) = \phi(x \oplus 0),$$

so ϕ is continuous at $x \oplus 0$.

If ϕ has Fréchet derivative $\phi'(x \oplus 0)$ at $x \oplus 0$ then

$$\begin{aligned}\phi(z^r) &= \phi(z^r \oplus 0)\\ &= \phi(x \oplus 0) + \langle \phi'(x \oplus 0), (z^r - x) \oplus 0\rangle + o(\|z^r - x\|)\end{aligned}$$

so ϕ has Fréchet derivative $(\phi(x \oplus 0)_{2i-1})_{i=1}^{\infty}$ at x.

Finally, suppose ϕ has Fréchet derivative $\phi'(x)$ at x. Then

$$\begin{aligned}\phi(z^r) &= \phi(\tilde{y}^r)\\ &= \phi(x) + \langle \phi'(x), \tilde{y}^r - x\rangle + o(\|\tilde{y}^r - x\|)\\ &= \phi(x \oplus 0) + \langle \phi'(x) \oplus 0, z^r - (x \oplus 0)\rangle + o(\|z^r - (x \oplus 0)\|)\end{aligned}$$

so ϕ has Fréchet derivative $\phi'(x) \oplus 0$ at $x \oplus 0$. The result now follows, since the nonzero components of $\widehat{x}$ are simply a rearrangement of the nonzero components of x. ♣

Note in the case $(V, W) = (\ell_1, c_0)$ we do not necessarily require the Fréchet derivative to lie in c_0 in the above result, in keeping with the Gâteaux case. Moreover, if one derivative lies in c_0, clearly so must the other. However, the Fréchet derivative of a w*-closed convex function necessarily lies in the predual.

5 Lower Semicontinuity and Unitary Invariance

We are now ready to prove our main results. We begin with a characterization of lower semicontinuous unitarily invariant convex functions analogous to Theorem 3.3 in [3]. We again consider the paired sequence spaces

$$(V, W) = (\ell_p, \ell_q)\ (1 < p < \infty) \text{ or } (\ell_1, c_0) \text{ or } (c_0, \ell_1)$$

and the corresponding paired spaces

$$(\mathcal{V}, \mathcal{W}) = (B_p, B_q)\ (1 < p < \infty) \text{ or } (B_1, B_0) \text{ or } (B_0, B_1).$$

Theorem 5.1 (Unitary Invariance) *A function $f : \mathcal{V} \to \bar{R}$ is unitarily invariant and norm lower semicontinuous if and only if it can be written in the form $\phi \circ \widehat{\lambda}$ for some rearrangement invariant, norm lower semicontinuous function $\phi : V \to \bar{R}$.*

Proof. Assume first $f = \phi \circ \widehat{\lambda}$ as above. For any sequence $X_n \to X$ in $\mathcal{V}$, the nonexpansivity of $\widehat{\lambda}$ implies $\widehat{\lambda}(X_n) \to \widehat{\lambda}(X)$, so

$$\liminf_{n\to\infty} \phi(\widehat{\lambda}(X_n)) \geq \phi(\widehat{\lambda}(X)),$$

since ϕ is lower semicontinuous. Clearly f is unitarily invariant since $\widehat{\lambda}$ is.

Conversely, if f is unitarily invariant and norm lower semicontinuous, define $\phi = f \circ \mathrm{dg}$. Then ϕ is rearrangement invariant and norm lower semicontinuous, and, by Theorem 3.3 of [3] and Theorem 4.1 (Insignificance of Zeroes) we have

$$f = \phi \circ \lambda = \phi \circ \widehat{\lambda}$$

as required. ♣

Notice that for any rearrangement invariant, norm lower semicontinuous function $\phi : V \to \bar{R}$, we have the following Fenchel conjugate formula, by [3, Thm 4.4]:

$$(\phi \circ \widehat{\lambda})^* = (\phi^* \circ \widehat{\lambda}).$$

Theorem 5.2 (Convex Subgradients) *If the function $\phi : V \to \bar{R}$ is rearrangement invariant, norm lower semicontinuous and convex, then for any operators X in $\mathcal{V}$ and Y in $\mathcal{W}$ we have*

$$Y \in \partial(\phi \circ \widehat{\lambda})(X) \Rightarrow \widehat{\lambda}(Y) \in \partial\phi(\widehat{\lambda}(X)).$$

Proof. If $Y \in \partial(\phi \circ \widehat{\lambda})(X)$ holds then

$$\begin{aligned} \phi(\widehat{\lambda}(X)) + \phi^*(\widehat{\lambda}(Y)) &= (\phi \circ \widehat{\lambda})(X) + (\phi^* \circ \widehat{\lambda})(Y) \\ &= \mathrm{tr}(XY) \leq \langle \widehat{\lambda}(X), \widehat{\lambda}(Y) \rangle \end{aligned}$$

by Theorem 2.6. ♣

Theorem 5.3 (Continuity) *If the function $\phi : V \to \bar{R}$ is rearrangement invariant and norm lower semicontinuous then $\phi \circ \widehat{\lambda}$ is norm continuous at the point X in $\mathcal{V}$ if and only if ϕ is norm continuous at $\widehat{\lambda}(X)$.*

Proof. If ϕ is norm continuous at $\widehat{\lambda}(X)$ then for any sequence $X_n \to X$ in $\mathcal{V}$ we know $\widehat{\lambda}(X_n) \to \widehat{\lambda}(X)$ since $\widehat{\lambda}$ is nonexpansive, so $\phi(\widehat{\lambda}(X_n)) \to \phi(\widehat{\lambda}(X))$. Thus $\phi \circ \widehat{\lambda}$ is continuous, at X.

Conversely, if $\phi \circ \widehat{\lambda}$ is continuous at X, then by the unitary invariance of $\widehat{\lambda}$, $\phi \circ \widehat{\lambda}$ is continuous at $\mathrm{dg}\,\lambda(X)$. Then for any sequence $z^r \to \lambda(X)$ in V we know $\mathrm{dg}\, z^r \to \mathrm{dg}\,\lambda(X)$, so

$$\phi(z^r) = \phi(\widehat{z^r}) = \phi(\widehat{\lambda}(\mathrm{dg}\, z^r)) \to \phi(\widehat{\lambda}(\mathrm{dg}\,\lambda(X))) = \phi(\widehat{\lambda}(X)) = \phi(\lambda(X))$$

by Theorem 4.1 (Insignificance of Zeroes). Hence ϕ is continuous at $\lambda(X)$, and hence at $\widehat{\lambda}(X)$ by Theorem 4.4 (Continuity). ♣

Theorem 5.4 (Fréchet Differentiability) *Suppose $V = \ell_p$ $(1 < p < \infty)$ or c_0 and consider a convex, lower semicontinuous, rearrangement invariant function $\phi : V \to \bar{R}$. Then the function $\phi \circ \widehat{\lambda}$ is Fréchet differentiable at the point $X \in \mathcal{V}$ if and only if ϕ is Fréchet differentiable at $\widehat{\lambda}(X)$.*

Proof. Suppose first $\phi \circ \widehat{\lambda}$ is differentiable at X, and hence, by unitary invariace at $\mathrm{dg}\,\lambda(X)$. Then for any sequence $z^r \to \lambda(X)$ on V we know $\mathrm{dg}\, z^r \to \mathrm{dg}\,\lambda(X)$, so

$$\begin{aligned}
\phi(z^r) &= \phi(\widehat{z^r}) = \phi(\widehat{\lambda}(\mathrm{dg}\, z^r)) \\
&= \phi(\widehat{\lambda}(\mathrm{dg}\,\lambda(X))) + \\
&\qquad \langle (\phi \circ \widehat{\lambda})'(\mathrm{dg}\,\lambda(X)), \mathrm{dg}\,(z^r - \lambda(X)) \rangle + o(\|z^r - \lambda(X)\|) \\
&= \phi(\lambda(X)) + \\
&\qquad \langle \mathrm{dg}^*(\phi \circ \widehat{\lambda})'(\mathrm{dg}\,\lambda(X)), z^r - \lambda(X) \rangle + o(\|z^r - \lambda(X)\|)
\end{aligned}$$

so ϕ is differentiable at $\lambda(X)$ and hence at $\widehat{\lambda}(X)$ as required, by Theorem 4.4 (Continuity and Differentiability).

Conversely, suppose ϕ is differentiable at $\widehat{\lambda}(X)$ and hence, by Theorem 4.4, at $\lambda(X)$ so the convex function ϕ is Gâteaux differentiable at $\lambda(X)$ and hence $\phi \circ \widehat{\lambda} = \phi \circ \lambda$ is Gâteaux differentiable at X:

$$\partial(\phi \circ \widehat{\lambda})(X) = \{Y\}$$

for some operator Y in $\mathcal{W}$. It now suffices to prove the subdifferential $\partial(\phi \circ \widehat{\lambda})$ is norm-norm upper semicontinuous at X, so suppose we have sequences

$$X_n \to X \text{ in } \mathcal{V}, Y_n \in \partial(\phi \circ \widehat{\lambda})(X_n) \ (n \in \mathbf{N})$$

We wish to show $Y_n \to Y$ in $\mathcal{W}$.

Since the subdifferential is norm-weak* upper semicontinuous, we know Y_n weak* converges to Y. But the norm on $\mathcal{W}(= B_q$ or $B_1)$ has the weak* – Kadec property: in B_q this follows from local uniform rotundity [1], and in B_0 this is a result of [10]. Hence it suffices to prove $\|Y_n\| \to \|Y\|$.

To see this, note by Theorem 5.2 (Convex Subgradients) we have

$$\widehat{\lambda}(Y_n) \in \partial\phi(\widehat{\lambda}(X_n)),$$

and

$$\widehat{\lambda}(Y) = \phi'(\widehat{\lambda}(X)),$$

and since $\widehat{\lambda}$ is nonexpansive we know

$$\widehat{\lambda}(X_n) \to \widehat{\lambda}(X).$$

But ϕ is differentiable at $\widehat{\lambda}(X)$ so $\partial\phi$ is norm–norm upper semicontinuous there. hence

$$\widehat{\lambda}(Y_n) \to \phi'(\widehat{\lambda}(X)) = \widehat{\lambda}(Y),$$

so

$$\|Y_n\| = \|\widehat{\lambda}(Y_n)\| \to \|\widehat{\lambda}(Y)\| = \|Y\|,$$

as required. ♣

We conclude with some examples.

Example 5.5 *Notice that the l_p norm is Fréchet smooth for $p > 1$. We immediately recover the fact that the norm on B_p is Fréchet differentiable fpr $p > 1$.*

Another easy yet useful example is:

Example 5.6 *Consider the maximum eigevalue function for operators in $B_p, p > 1$. Then it is Fréchet differentiable if and only if its multiplicity is 1.*

Now let us consider an example which has been the subject of much interest in recent years.

Example 5.7 In [6] the author looks at self-concordant barriers, which play a central role in many types of optimization problems. A number of the details here can be put in a quite appealing setting by the theory developed thus far. We start with the Taylor expansion

$$\begin{aligned} \ln(1+t) &= t - \frac{1}{2}t^2 + \frac{1}{3}t^3 - \frac{1}{4}t^4 + \ldots \\ &= -\sum_{n=1}^{\infty} \frac{(-1)^n}{n} t^n. \end{aligned}$$

Thus, the function $g : \mathbf{R} \to \mathbf{R}$ given by

$$g(t) = \ln(1+t) - t$$

is concave, since $g'(t) = -t/(1+t)$ and $g''(t) = -(1+t)^{-2} < 0$. We write the concave conjugate of g as

$$\begin{aligned} g_*(s) &= \inf_{t\in\mathbf{R}} \{st - g(t)\} \\ &= g(s) \end{aligned}$$

Further, $-g$ has the conjugate $(-g)^*(s) = -g(-s)$. Now we define the function f on ℓ_2 by writing

$$f(x) = \sum_i g(x_i).$$

Clearly, f is finite and continuous on ℓ_2 since $|g(t)| \le kt^2$ if $|t| \le 1/2$. Using this f and the λ notation defined in [3, Section 3], we can derive a function F on B_2 in the obvious way

$$F = f \circ \lambda.$$

For general concave h we have $h_*(y) = -(-h)^*(-y)$, so that since $f_* = f$ we get $F_* = F$.

Now, if $T \in B_1$ and T has its spectrum in $(-1,1)$, then $f(\lambda(T)) = \sum \ln(1+\lambda_i(T)) - \sum \lambda_i(T)$ is well defined so that we get the formula

$$F(T) = \ln(\det(I+T)) - \mathrm{tr}(T).$$

Now we show that F is Fréchet differentiable in the interior of its domain. Observe that $F = f \circ \lambda = f \circ \hat{\lambda}$. By Theorem 5.4 we need only to show that f is Fréchet differentiable in the interior of its domain. For any x in the in the interior of the domain of f, set

$$\xi = \sum_{i=1}^{\infty} \frac{x_i}{1+x_i} e^i.$$

We have, for h in a small neighborhood of 0,

$$\begin{aligned} & |f(x+h) - f(x) - \rangle \xi, h \langle | \\ = & |\sum_{i=1}^{\infty} \frac{h_i}{(1+\eta_i)^2}, \eta_i \in (x_i, x_i + h_i) \\ \leq & \|h\|^2. \end{aligned}$$

Thus, f is Fréchet differentiable at x and $f'(x) = \xi$.

References

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

[2] R. Bhatia, *Matrix Analysis*, Springer, New York, 1997.

[3] J. M. Borwein, J. Read, A. S. Lewis and Q. J. Zhu, Convex spectral functions of compact operators, *International J. of Nonlinear and Convex Analysis* **1** (1999), 17–35.

[4] J. M. Borwein and Q. J. Zhu, A survey of subdifferentials and their applications, CECM Research Report 98-105 (1998), *Nonlinear Analysis, TMA*, **38** (1999), 687-773.

[5] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983, Russian edition MIR, Moscow, (1988). Reprinted as Vol. 5 of the series Classics in Applied Mathematics, SIAM, Philadelphia, (1990).

[6] L. Faybusovich, Infinite-dimensional semidefinite programming: regularized determinants and self-concordant barriers, *Topics in semidefinite and interior-point methods (Toronto, ON, 1996)*, 39-49, Fields Inst. Commun., **18**, Amer. Math. Soc. Providence, RI, 1998.

[7] S. Friedland and P. Nowosad, Extremal Eigenvalue Problems with Indefinite Kernals, *Adv. in Math.* **40** (1981), 128–154.

[8] I. C. Gohberg and M. G. Kreĭn, *Introduction to the Theory of Linear Nonselfadjoint Operators*, Translations of Mathematical Monographs, Vol. 18, Amer. Math. Soc. Providence, RI, 1969.

[9] G.H. Hardy, J. E. Littlewood and G. Pólya, Inequalities, Cambridge University Press, Cambridge, U. K. 1952.

[10] C. Lennard, $\mathcal{C}_1$ is uniformly Kadec-Klee. Proc. Amer. Math. Soc. **109** (1990), no. 1, 71–77.

[11] A. S. Lewis, Convex analysis on the Hermitian matrices, *SIAM J. Optim.* **6** (1996), 164–177.

[12] A. S. Lewis, Nonsmooth analysis of eigenvalues, *Mathematical Programming* **84** (1999), 1–24.

[13] A. S. Lewis, Lidskii's theorem via nonsmooth analysis, *SIAM J. Matrix An.* (1999), to appear.

[14] A.S. Markus, The eigen- and singular values of the sum and product of linear operators, *Uspehi Mat. Nauk* **9** (1964) 91–120.

[15] Henryk Minc, *Nonnegative matrices*, John Wiley & Sons, New York, 1988.

[16] G. Pederson, *Analysis Now*, Springer Verlag, Berlin, 1989.

[17] R. R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, Lecture Notes in Mathematics, No. 1364, Springer Verlag, N.Y., Berlin, Tokyo, (1988), Second Edition (1993).

[18] J. Read, The Approximation of Optimal Control of Vibrations - A Geometrical Method, *Mathematical Methods in the Applied Sciences*, **19** (1996), 87–129.

[19] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, N.J. 1970.

[20] B. Simon, *Trace Ideals and Their Applications*, Cambridge University Press, 1979.

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