



## CONVEX SPECTRAL FUNCTIONS OF COMPACT OPERATORS

J. M. BORWEIN\*, J. READ, A. S. LEWIS<sup>†</sup>, AND QIJI ZHU<sup>‡</sup>

**ABSTRACT.** We consider functions on the space of compact self-adjoint Hilbert space operators. Specifically, we study those extended-real functions which depend only on the operators' spectral sequences. Examples include the norms of the Schatten  $p$ -spaces, the Calderón norms, the  $k$ 'th largest eigenvalue, and some infinite-dimensional self-concordant barriers. We show how various convex and nonsmooth-analytic properties of such functions follow from the corresponding properties of the restrictions to the space of diagonal operators, and we derive subdifferential and conjugacy formulas.

### 1. INTRODUCTION

Many optimization problems involve functions of bounded self-adjoint operators on a separable Hilbert space. Often these functions depend only on the spectrum of the operator involved. Many examples have their roots in physical phenomena such as vibrating systems. To optimize a process governed by such a system, we must study the spectra of such operators [14], often asking questions about approximating those spectra. In [6], for example, the author studies extremal eigenvalues arising in the study of the critical mass of a nuclear reactor. In [4] other physical settings lead to the study of the generalized gradient at a multiple eigenvalue. Von Neumann pioneered this area by introducing the concepts of unitarily invariant matrix norms and symmetric gauge functions [19]. This was generalized to the infinite-dimensional setting in the book [7]. More recently, [1] showed that questions in Banach space geometry dealing with the norm of spaces of operators (extreme points, smoothness, (strong) exposedness) reduce to the same questions on the much simpler sequence space where the spectrum lies. (See also [17, 18].)

Lewis [9], motivated by the classical work of von Neumann, considered unitarily invariant convex matrix functions, common in matrix optimization. He extended this work in [10], establishing formulas characterizing Fréchet and limiting subdifferentials of unitarily invariant nonconvex functions in terms of the Fréchet and limiting subdifferentials of corresponding rearrangement-invariant functions. In this paper we take this one step further, examining similar relationships for the convex subdifferentials of unitarily invariant convex functions on spaces of bounded, self-adjoint operators. We also discuss extensions to the limiting subdifferential for nonconvex unitarily invariant Lipschitz functions.

---

1991 *Mathematics Subject Classification.* Primary 49N15, 47A75, Secondary 90C25.

*Key words and phrases.* convexity, eigenvalue, compact operator, unitarily invariant, subdifferential, Fenchel conjugate, spectral function.

\*Research supported by the Natural Sciences and Engineering Research Council and by the Shrum Endowment at Simon Fraser University.

<sup>†</sup>Research supported by the Natural Sciences and Engineering Research Council.

<sup>‡</sup>Research supported by the National Science Foundation under grant DMS-9704203.

The plan of this paper is as follows. In Section 2 we derive an infinite-dimensional version of the von Neumann trace inequality. Section 3 is devoted to a precise discussion of rearrangement and unitary invariance for functions on infinite-dimensional spaces. In Section 4 we discuss conjugacy formulas for convex spectral functions in infinite-dimensional spaces, generalizing corresponding results in [9]. Our main results — the characterization of the subdifferentials of convex spectral functions — are contained in Section 5. In Section 6 we discuss extensions of the main results to locally Lipschitz functions, and we end with some examples.

We finish this section by summarizing some ideas from [7]. All operators we consider are maps from the complex Hilbert space  $\ell_2^{\mathbf{C}}$  to itself. Here,  $\ell_2^{\mathbf{C}}$  is 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  $\sum_{j \in \mathbf{N}} |c_j|^2 < \infty$ . We denote by  $\ell_2$  the standard real normed sequence space (coefficients in  $\mathbf{R}$ ) with canonical basis  $(e^j)_{j=1}^{\infty}$ . Thus the  $i$ th component of the  $j$ th element of this basis is given by the Kronecker delta:  $e_i^j = \delta_{i,j}$ . We also consider the real normed sequence spaces  $\ell_p$  ( $1 \leq p \leq \infty$ ), as well as  $c_0$  (the space of null sequences). (Thus  $\ell_1 \subset \ell_2 \subset c_0 \subset \ell_{\infty}$ , and the norm on  $c_0$  is the same as that on  $\ell_{\infty}$ .)

We next consider the space of bounded self-adjoint operators on  $\ell_2^{\mathbf{C}}$ , which we denote by  $B_{sa}$ . We call an operator  $T \in B_{sa}$  *positive* (denoted  $T \geq 0$ ) if  $\langle Tx|x \rangle \geq 0$  for all  $x$  in  $\ell_2^{\mathbf{C}}$ . To each  $T \in B_{sa}$  one can associate a unique positive operator  $|T| = (T^2)^{\frac{1}{2}} \in B_{sa}$  [12, p. 96]. Now to each positive  $T \in B_{sa}$  one can associate the (possibly infinite) value

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

which we call the *trace* of  $T$ . The trace is actually independent of the orthonormal basis  $(e^j)$  chosen [12, p. 116].

Within  $B_{sa}$  we consider the *trace class operators*, denoted by  $B_1$ , which are those self-adjoint operators  $T$  for which  $\text{tr}(|T|) < \infty$ . 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  $\text{tr}(T) = \text{tr}(T_+) - \text{tr}(T_-)$ . We let  $B_2$  be the self-adjoint *Hilbert-Schmidt* operators, which are those  $T \in B_{sa}$  such that  $T^2 \in B_1$ . Then  $B_1 \subset B_2 \subset B_0 \subset B_{sa}$ , where  $B_0$  are the compact self-adjoint operators. Now, any compact self-adjoint operator  $T$  is *diagonalizable* [12, p. 107]. That is, there exists a unitary operator  $U$  and  $\lambda \in c_0$  such that  $(U^* T U x)_j = \lambda_j x_j$  for all  $j \in \mathbf{N}$  and all  $x \in \ell_2^{\mathbf{C}}$ . This and the fact that  $\text{tr}(ST) = \text{tr}(TS)$  makes *Lidskii's Theorem* [16, p. 45] easy for self-adjoint operators:

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

where  $\{\lambda_i(T)\}$  is any *spectral sequence* of  $T$  (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  $\|T\|_p = (\text{tr}(|T|^p))^{1/p} < \infty$ . When  $T$  is self-adjoint we have (see [7, p. 94])

$$(2) \quad \|T\|_p = \left( \sum_{i=1}^{\infty} |\lambda_i(T)|^p \right)^{1/p}.$$

In this case, for  $p, q \in (1, \infty)$  satisfying  $p^{-1} + q^{-1} = 1$ , the real linear spaces  $B_p$  and  $B_q$  are paired, and the bilinear form  $\langle S, T \rangle := \text{tr}(ST)$  implements the duality on  $B_p \times B_q$ . The space  $B_p$  is the (self-adjoint) *Schatten  $p$ -space* [12, p. 124]. In particular,  $B_1$  is the space of self-adjoint *trace-class* operators, and  $B_2$  is the space of self-adjoint *Hilbert-Schmidt* operators. We can also consider the space  $B_0$  paired with  $B_1$ .

**Remark 1.1.** The corresponding spaces of non-self-adjoint operators are defined somewhat analogously [7]. We need these spaces only in Propositions 5.3, 5.4 and Theorem 5.5.

Now, following [12], for each  $x \in \ell_2^{\mathbb{C}}$  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 *diagonal* operator  $\text{dg } x \in B_{sa}$  pointwise by

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

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

**Definition 1.2.** For  $p \in [1, \infty)$ , we refer to  $\ell_p$  as the spectral sequence space for  $B_p$  and  $c_0$  as the spectral sequence space for  $B_0$ .

**Assumption 1:** We consider paired Banach spaces  $V \times W$  where  $V = \ell_p$  and  $W = \ell_q$ , with  $p \in (1, \infty)$  and  $p^{-1} + q^{-1} = 1$ , or where  $V = c_0$  (with the supremum norm) and  $W = \ell_1$  (or vice versa). We denote the norms on  $V$  and  $W$  by  $\|\cdot\|_V$  and  $\|\cdot\|_W$  respectively. We also consider the corresponding paired Banach spaces  $\mathcal{V} \times \mathcal{W}$  where  $\mathcal{V} = B_p$  and  $\mathcal{W} = B_q$  or  $\mathcal{V} = B_0$  and  $\mathcal{W} = B_1$  (or vice versa). We denote the norms on  $\mathcal{V}$  and  $\mathcal{W}$  by  $\|\cdot\|_{\mathcal{V}}$  and  $\|\cdot\|_{\mathcal{W}}$  respectively. We always take  $V$  to be the spectral sequence space for the operator space  $\mathcal{V}$  and  $W$  that for  $\mathcal{W}$ . In this way fixing  $V \times W$  fixes  $\mathcal{V} \times \mathcal{W}$  and vice versa.

Let  $\mathcal{R}$  be the set of all bijections from  $\mathbf{N}$  to  $\mathbf{N}$ . We sometimes call  $\pi \in \mathcal{R}$  a *rearrangement*. Let  $\mathcal{U}$  be the set of all unitary operators on  $\ell_2^{\mathbb{C}}$ . This brings us to a definition.

**Definition 1.3.** For each  $U \in \mathcal{U}$  we define the bilinear form  $B_U$  on  $V \times W$  by

$$B_U(x, y) = \text{tr}[U^*(\text{dg } x)U(\text{dg } y)].$$

**Lemma 1.4.** Suppose  $U \in \mathcal{U}$ . Define  $w^j := Ue^j$ . Then  $w^j \in \ell_2^{\mathbb{C}}$  and the “infinite matrix”  $(|w_i^j|^2)_{i,j \in \mathbf{N}}$  is doubly stochastic. That is,  $\sum_i |w_i^j|^2 = 1$  for each  $j \in \mathbf{N}$  and  $\sum_j |w_i^j|^2 = 1$  for each  $i \in \mathbf{N}$ . Further,

$$(3) \quad B_U(x, y) = \sum_{i,j} x_i |w_i^j|^2 y_j \leq \|x\|_V \|y\|_W$$

for  $(x, y) \in V \times W$ .

**Proof** It is clear that  $u^j \in \ell_2^{\mathbf{C}}$ . In fact, we know  $(u^j)_{j=1}^{\infty}$  forms an orthonormal basis for  $\ell_2^{\mathbf{C}}$  [12, p. 95]. Thus,  $\sum_i |u_i^j|^2 = 1$  for each  $j \in \mathbf{N}$ . Further, since  $\langle e^j | U^* e^i \rangle = \langle U e^j | e^i \rangle = \langle u^j | e^i \rangle = u_i^j$ , we obtain  $U^* e^i = \sum_j (u_i^j)^* e^j$ . Taking the norm of this equality gives  $\|U^* e^i\| = \sum_j |u_i^j|^2$ . Since  $\|U^* e^i\| = 1$  ( $U$  is unitary), we have  $(|u_i^j|^2)_{i,j \in \mathbf{N}}$  is doubly stochastic. We derive equation (3) by considering the following equalities:

$$\begin{aligned} \operatorname{tr}[U^*(\operatorname{dg} x)U(\operatorname{dg} y)] &= \sum_j \langle U^*(\operatorname{dg} x)U(\operatorname{dg} y) e^j | e^j \rangle \\ &= \sum_j \langle (\operatorname{dg} x)U(y_j e^j) | U e^j \rangle \\ &= \sum_j y_j \langle (\operatorname{dg} x) u^j | u^j \rangle \\ &= \sum_j y_j \left\langle \sum_i x_i u_i^j e^i | u^j \right\rangle \\ &= \sum_{i,j} x_i |u_i^j|^2 y_j. \end{aligned}$$

Now  $V$  and  $W$  are the paired spectral sequence spaces for the paired operator spaces  $\mathcal{V}$  and  $\mathcal{W}$ , and the bilinear form  $\langle S, T \rangle := \operatorname{tr}(ST)$  implements the duality on  $\mathcal{V} \times \mathcal{W}$ . Thus, since  $\|U^*(\operatorname{dg} x)U\|_{\mathcal{V}} = \|\operatorname{dg} x\|_{\mathcal{V}} = \|x\|_{\mathcal{V}}$  for any  $U \in \mathcal{U}$  and any  $x \in V$ , we obtain

$$\begin{aligned} B_U(x, y) &= \langle U^*(\operatorname{dg} x)U, \operatorname{dg} y \rangle \\ &\leq \|U^*(\operatorname{dg} x)U\|_{\mathcal{V}} \|\operatorname{dg} y\|_{\mathcal{W}} \\ &= \|x\|_{\mathcal{V}} \|y\|_{\mathcal{W}}, \end{aligned}$$

and we are done. ♣

## 2. VON NEUMANN TYPE INEQUALITIES

We return to the paired sequence spaces  $V$  and  $W$  of Assumption 1. We now develop an analogue of work originating with von Neumann. The simplest case (a classical inequality) says that if vectors  $x, y \in \mathbf{R}^n$  both have components in decreasing order then

$$(4) \quad x^T P y \leq x^T y$$

for any permutation matrix  $P$  [8, 9]. For general  $x$  in  $\mathbf{R}^n$  we write  $\bar{x} \in \mathbf{R}^n$  to denote that vector derived by writing the components of  $x$  in decreasing order.

**Definition 2.1.** For each rearrangement  $\pi \in \mathcal{R}$  we define the bilinear form  $P_{\pi}$  on  $V \times W$  by

$$P_{\pi}(x, y) = \sum_{i=1}^{\infty} x_{\pi(i)} y_i.$$

**Theorem 2.2. (von Neumann type inequality)** *Each pair  $(x, y) \in V \times W$  satisfies the inequality*

$$\sup_{\pi \in \mathcal{R}} P_{\pi}(x, y) = \sup_{U \in \mathcal{U}} B_U(x, y).$$

**Proof** Given any  $\pi \in \mathcal{R}$  we can define  $U \in \mathcal{U}$  by  $Ue^j = e^{\pi(j)}$  for all  $j \in \mathbf{N}$ . Then  $U^*(\operatorname{dg} x)U = \operatorname{dg}(x_{\pi(j)})_{j=1}^{\infty}$ , so we obtain the inequality

$$(5) \quad \sup_{\pi \in \mathcal{R}} P_{\pi}(x, y) \leq \sup_{U \in \mathcal{U}} B_U(x, y),$$

for  $(x, y) \in V \times W$ . Now let us define two functions on  $V \times W$ . These are

$$\begin{aligned} b(x, y) &:= \sup_{U \in \mathcal{U}} B_U(x, y) \quad \text{and} \\ p(x, y) &:= \sup_{\pi \in \mathcal{R}} P_{\pi}(x, y). \end{aligned}$$

For fixed  $x \in V$  we know that, as a supremum of a family of linear functions, both  $p$  and  $b$  are convex in  $y$ . Lemma 1.4 and (5) together give the inequality  $p(x, y) \leq b(x, y) \leq \|x\|_V \|y\|_W$ , so for fixed  $x \in V$  both  $p$  and  $b$  are everywhere finite, lower semicontinuous (and hence continuous [13, p. 39]) convex functions of  $y$ . The same is true if we hold  $y$  fixed and consider  $b$  and  $p$  as functions of  $x$ .

Consider  $F := \{x \in \ell_{\infty} : x_j = 0 \text{ eventually}\}$ , the set of real, finitely non-zero sequences. We know  $F$  is norm dense in both  $V$  and  $W$ . If we show for fixed  $x \in F$  that  $b(x, y) \leq p(x, y)$  for all  $y \in F$ , then since  $F$  is norm dense in  $W$  and  $b$  and  $p$  are continuous functions (in  $y$  for fixed  $x$ ), we obtain  $b(x, y) \leq p(x, y)$  for all  $y \in W$ . This holds for arbitrary  $x \in F \subset V$ , so  $b(x, y) \leq p(x, y)$  for all  $(x, y) \in F \times W$ . Now fix  $y \in W$ . Since  $b$  and  $p$  are continuous functions in  $x$  the same density arguments give  $b(x, y) \leq p(x, y)$  for all  $x \in V$ . As this  $y$  is arbitrary, we obtain  $b(x, y) \leq p(x, y)$  for all  $(x, y) \in V \times W$ , which together with (5) gives the result. Thus it suffices to show  $b(x, y) \leq p(x, y)$  for any  $(x, y) \in F \times F$ .

Fix  $(x, y) \in F \times F$  and choose  $n \in \mathbf{N}$  such that  $x_i = y_i = 0$  for all  $i \geq n$ . For  $U \in \mathcal{U}$  define the doubly stochastic ‘‘infinite matrix’’ as in Lemma 1.4. Let  $\mathcal{Q}$  be the set of doubly stochastic  $n \times n$  matrices. Define  $Q = (q_{i,j})_{i,j=1}^n$  by

$$\begin{aligned} q_{i,j} &= |w_i^j|^2 & i, j &= 1, \dots, n-1, \\ q_{n,j} &= 1 - \sum_{i=1}^{n-1} q_{i,j} & j &= 1, \dots, n-1, \\ q_{i,n} &= 1 - \sum_{j=1}^{n-1} q_{i,j} & i &= 1, \dots, n-1, \\ q_{n,n} &= 1 - \sum_{j=1}^{n-1} q_{n,j} \quad (= 1 - \sum_{i=1}^{n-1} q_{i,n}). \end{aligned}$$

Clearly,  $Q \in \mathcal{Q}$  and for our  $(x, y) \in F \times F$  we have  $B_U(x, y) = x^T Q y$ , where we abuse notation and interpret  $x$  and  $y$  in  $\mathbf{R}^n$  in the natural way. Let  $\mathcal{P}$  be the set of all  $n \times n$  permutation matrices. Birkhoff’s Theorem [11, p. 117] says that the convex hull of  $\mathcal{P}$  is exactly  $\mathcal{Q}$ . Thus we can write  $Q = \sum_{i=1}^k \lambda_i P_i$  where  $P_i \in \mathcal{P}$  and  $\lambda_i \geq 0$  for  $i = 1, \dots, k$  with  $\sum_{i=1}^k \lambda_i = 1$ . Inequality (4) says  $x^T P y \leq \bar{x}^T \bar{y}$  for any

$P \in \mathcal{P}$ , so that

$$x^T Q y = x^T \left( \sum_{i=1}^k \lambda_i P_i \right) y = \sum_{i=1}^k \lambda_i (x^T P_i y) \leq \sum_{i=1}^k \lambda_i (\bar{x}^T \bar{y}) = \bar{x}^T \bar{y}.$$

If we choose  $\pi \in \mathcal{R}$  (independent of  $U$ ) such that  $P_\pi(x, y) = \bar{x}^T \bar{y}$ , we obtain the inequality  $B_U(x, y) \leq P_\pi(x, y)$ . Take the supremum over  $U \in \mathcal{U}$  to obtain

$$\sup_{U \in \mathcal{U}} B_U(x, y) \leq P_\pi(x, y),$$

which means  $b(x, y) \leq p(x, y)$  for all  $(x, y) \in F \times F$ . ♣

**Remark 2.3.** The suprema in Theorem 2.2 may not be attained. For example, consider the two sequences (in any of our sequence spaces)

$$\begin{aligned} x &= (0, 2^{-1}, 2^{-2}, 2^{-3}, \dots) \\ y &= (2^{-1}, 2^{-2}, 2^{-3}, \dots). \end{aligned}$$

An easy exercise shows the supremum on the left cannot be attained by applying any bijection.

### 3. REARRANGEMENT AND UNITARY INVARIANCE

We return to the setting of Assumption 1, with a sequence space  $V$  and corresponding operator space  $\mathcal{V}$ . We say sequences  $x$  and  $y$  in  $V$  are *rearrangement equivalent* if there is a rearrangement  $\pi$  in  $\mathcal{R}$  such that  $x_{\pi(i)} = y_i$  for all  $i$ . Similarly, operators  $S$  and  $T$  in  $B_{sa}$  are *unitarily equivalent* if there is a unitary  $U$  such that  $U^* T U = S$ .

We say a function  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  is *unitarily invariant* if  $f(U^* T U) = f(T)$  for all  $T \in \mathcal{V}$  and all  $U \in \mathcal{U}$ . Correspondingly, a function  $\Psi : V \rightarrow V$  is *rearrangement invariant* if  $\Psi((x_{\pi(i)})) = \Psi(x)$  for all  $x \in V$  and all  $\pi \in \mathcal{R}$ .

The rearrangement operation  $x \in \mathbf{R}^{\mathbf{n}} \mapsto \bar{x}$  gives an easy tool for understanding rearrangement-invariant functions on  $\mathbf{R}^{\mathbf{n}}$ . Matters are less straightforward on a sequence space  $V$ , but the following construction gives a concrete approach.

For fixed  $x \in V$ , define sets of indices

$$I_{>}(x) := \{i : x_i > 0\}, \quad I_{=}(x) := \{i : x_i = 0\}, \quad \text{and} \quad I_{<}(x) := \{i : x_i < 0\}.$$

Now consider the function  $\Phi : V \rightarrow V$  defined as follows. Given any sequence  $x$  in  $V$ , use the following procedure to construct  $\Phi(x)$ :

(0) Initialize  $j := 1$

- (1) If  $I_{>}(x) \neq \emptyset$ ,
  - (i) choose  $i \in I_{>}(x)$  maximizing  $x_i$ ,
  - (ii) define  $\Phi(x)_j := x_i$
  - (iii) update  $I_{>}(x) := I_{>}(x) \setminus \{i\}$  and  $j := j + 1$ .
- (2) If  $I_{=}(x) \neq \emptyset$ ,
  - (i) choose  $i \in I_{=}(x)$
  - (ii) define  $\Phi(x)_j := 0$
  - (iii) update  $I_{=}(x) := I_{=}(x) \setminus \{i\}$  and  $j := j + 1$ .

- (3) If  $I_{<}(x) \neq \emptyset$ ,
- (i) choose  $i \in I_{<}(x)$  minimizing  $x_i$
  - (ii) define  $\Phi(x)_j := x_i$
  - (iii) update  $I_{<}(x) := I_{<}(x) \setminus \{i\}$  and  $j := j + 1$ .

(4) Go to (1).

Informally, we first select the largest positive component of  $x$ , followed by a zero component and then the smallest negative component and so on. If any of the sets  $I_{>}(x)$ ,  $I_{=}(x)$  or  $I_{<}(x)$  is exhausted we skip the corresponding step. Note that  $\Phi$  has the following two properties:

- (A) Each  $x \in V$  is rearrangement equivalent to  $\Phi(x)$ .
- (B) Any  $x, y \in V$  are rearrangement equivalent if and only if  $\Phi(x) = \Phi(y)$ .

All that is important here is that  $\Phi$  is constant on the equivalence classes of  $V$  that are induced by rearrangements, and that the function  $\Phi$  leaves some canonical element of each equivalence class unchanged. Thus  $\Phi$  also has these two properties:

- (C)  $\Phi^2 = \Phi$ .
- (D)  $\phi : V \rightarrow \bar{\mathbf{R}}$  is rearrangement invariant if and only if  $\phi = \phi \circ \Phi$ .

Define the *eigenvalue function*  $\lambda : \mathcal{V} \rightarrow V$  as follows: for any operator  $T \in \mathcal{V}$ , let  $x \in V$  be any spectral sequence of  $T$ , and define  $\lambda(T) = \Phi(x)$ . Thus  $\lambda(T)$  is a canonical spectral sequence for any given compact self-adjoint operator  $T$ .

**Proposition 3.1.** *The function  $\lambda$  is unitarily invariant, and  $\Phi = \lambda \circ \text{dg}$ .*

The proof of the next result can be found in [12, p. 107].

**Theorem 3.2. (Diagonalization)** *For all  $T \in B_0$  there exists  $U \in \mathcal{U}$  with  $T = U^* \text{dg}(\lambda(T))U$ .*

Note that  $\lambda$  and  $\text{dg}$  act as inverses in the following sense:

- (E)  $(\lambda \circ \text{dg})(x)$  is rearrangement equivalent to  $x$  for all  $x \in V$ .
- (F)  $(\text{dg} \circ \lambda)(T)$  is unitarily equivalent to  $T$  for all  $T \in \mathcal{V}$ .

For any  $\phi : V \rightarrow \bar{\mathbf{R}}$  we have that  $\phi \circ \lambda$  is uniformly invariant, and for any  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  we have that  $f \circ \text{dg}$  is rearrangement invariant. Thus, the maps  $\lambda$  and  $\text{dg}$  allow us to move between rearrangement invariant functions on  $V$  and unitarily invariant functions on  $\mathcal{V}$ , as the easy results below show.

**Theorem 3.3. (Unitary invariance)** *Given  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  the following properties are equivalent:*

- (i)  $f$  is unitarily invariant;
- (ii)  $f = f \circ \text{dg} \circ \lambda$ ;
- (iii)  $f = \phi \circ \lambda$  for some rearrangement invariant  $\phi : V \rightarrow \bar{\mathbf{R}}$

If (iii) holds then  $\phi = f \circ \text{dg}$ .

**Theorem 3.4. (Rearrangement invariance)** *Given  $\phi : V \rightarrow \bar{\mathbf{R}}$  the following are equivalent:*

- (i)  $\phi$  is rearrangement invariant;
- (ii)  $\phi = \phi \circ \lambda \circ \text{dg}$ ;
- (iii)  $\phi = f \circ \text{dg}$  for some unitarily invariant  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$ .

If (iii) holds then  $f = \phi \circ \lambda$ .

Note finally the following useful relations:

$$\begin{aligned}\|\mathrm{dg}(x)\|_{\mathcal{V}} &= \|x\|_V \text{ for } x \in V; \\ \|\lambda(T)\|_V &= \|T\|_{\mathcal{V}} \text{ for } T \in \mathcal{V}.\end{aligned}$$

#### 4. CONJUGATE FORMULAS

Let  $\mathcal{X}$  be any topological vector space. The *convex conjugate* of an arbitrary function  $f : \mathcal{X} \rightarrow \bar{\mathbf{R}}$ , which we denote by  $f^* : \mathcal{X}^* \rightarrow \bar{\mathbf{R}}$ , is the lower semicontinuous convex function

$$f^*(y) = \sup_{x \in \mathcal{X}} \{ \langle x, y \rangle - f(x) \}.$$

If  $\mathcal{X}$  is locally convex and  $f$  is a proper (that is, somewhere finite and never  $-\infty$ ) lower semicontinuous convex function, so is  $f^*$  (see for example [13]). If we go on to consider the second conjugate  $f^{**}$  of  $f$ , then we can consider  $f^{**} : \mathcal{X} \rightarrow \bar{\mathbf{R}}$ . That is, we restrict the domain of  $f^{**}$  to the original space  $\mathcal{X}$  as opposed to considering the space  $\mathcal{X}^{**}$ .

Again we fix two paired spectral sequences spaces  $V$  and  $W$ , as in Assumption 1, with their corresponding paired operator spaces  $\mathcal{V}$  and  $\mathcal{W}$ .

**Theorem 4.1. Conjugacy and diagonals** *Any unitarily invariant function  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  satisfies*

$$f^* \circ \mathrm{dg} = (f \circ \mathrm{dg})^*.$$

**Proof** Choose  $y \in W$ . Then we have

$$\begin{aligned}(f^* \circ \mathrm{dg})(y) &\stackrel{1}{=} f^*(\mathrm{dg} y) \\ &\stackrel{2}{=} \sup \{ \mathrm{tr}[X(\mathrm{dg} y)] - f(X) : X \in \mathcal{V} \} \\ &\stackrel{3}{=} \sup \{ \mathrm{tr}[U^*(\mathrm{dg} x) U(\mathrm{dg} y)] - f(U^*(\mathrm{dg} x)U) : \\ &\quad U \in \mathcal{U}, x \in V \} \\ &\stackrel{4}{=} \sup \{ \sup_{U \in \mathcal{U}} B_U(x, y) - f(\mathrm{dg} x) : x \in V \} \\ &\stackrel{5}{=} \sup \{ \sup_{\pi \in \mathcal{R}} P_{\pi}(x, y) - f(\mathrm{dg} x) : x \in V \} \\ &\stackrel{6}{=} \sup \left\{ \sum_j x_{\pi(j)} y_j - f(\mathrm{dg}(x_{\pi(j)})_{j=1}^{\infty}) : x \in V, \pi \in \mathcal{R} \right\} \\ &\stackrel{7}{=} \sup \{ \langle z, y \rangle - f(\mathrm{dg} z) : z \in V \} \\ &\stackrel{8}{=} (f \circ \mathrm{dg})^*(y).\end{aligned}$$

Equations 1 and 2 follow by definition. Equation 3 follows since we can write any  $X$  as  $U^* \mathrm{dg}(x) U$  for some appropriate  $x \in V$  and  $U \in \mathcal{U}$ . Equation 4 follows by definition of  $B_U$  and the fact that  $f$  is unitarily invariant. Equation 5 follows by Theorem 2.2. Equation 6 follows by definition of  $P_{\pi}$  and the fact that  $f$  is unitarily invariant. Equation 7 follows by replacing  $(x_{\pi(j)})_{j=1}^{\infty}$  with  $z$ . Equation 8 follows by definition. ♣



**Corollary 4.2. (Convexity)** *Assume  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  is unitarily invariant. Then  $f$  is proper, convex, and weakly lower semicontinuous if and only if  $f \circ \text{dg}$  is likewise.*

**Proof** Unitary invariance gives  $f = f^{**}$  if and only if  $f \circ \text{dg} = f^{**} \circ \text{dg}$ . The result follows from the previous theorem, since  $f^{**} \circ \text{dg} = (f \circ \text{dg})^{**}$ . ♣

**Lemma 4.3. (Invariance and conjugacy)** *For any unitarily invariant  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$ , the conjugate  $f^*$  is unitarily invariant. For any rearrangement invariant  $\phi : V \rightarrow \bar{\mathbf{R}}$ , the conjugate  $\phi^*$  is rearrangement invariant.*

**Proof** This lemma follows directly from the definitions. ♣

**Theorem 4.4. (Conjugacy)** *For any rearrangement invariant  $\phi : V \rightarrow \bar{\mathbf{R}}$ , we have*

$$(\phi \circ \lambda)^* = \phi^* \circ \lambda.$$

**Proof** Theorems 3.4 and 4.1 allow us to write

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

If we compose this expression with  $\lambda$  and observe that  $(\text{dg} \circ \lambda)(T)$  is unitarily equivalent to  $T$ , then Lemma 4.3 allows us to write

$$\phi^* \circ \lambda = (\phi \circ \lambda)^* \circ \text{dg} \circ \lambda = (\phi \circ \lambda)^*,$$

and we are done. ♣

The final result of this section synthesizes Theorems 3.3 and 4.4 as well as Theorems 3.4 and 4.1.

**Theorem 4.5.** *Given any unitarily invariant  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  we have  $f = \phi \circ \lambda$  for some rearrangement invariant  $\phi : V \rightarrow \bar{\mathbf{R}}$ . Further, the formula*

$$f^* = \phi^* \circ \lambda$$

*holds. Similarly, given any rearrangement invariant  $\phi : V \rightarrow \bar{\mathbf{R}}$  we have  $\phi = f \circ \text{dg}$  for some unitarily invariant  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$ . Further, the formula*

$$\phi^* = f^* \circ \text{dg}$$

*holds.*

## 5. THE SUBDIFFERENTIAL

We now examine the subdifferential of a unitarily invariant convex function. We begin by recalling some fairly standard results.

**Proposition 5.1.** *Suppose  $R \in B_{sa}$ . Then for  $t \in \mathbf{R}$  we have*

$$U_t := \exp(itR) \in \mathcal{U}.$$

*Further, we have  $\|U_t - I\| \rightarrow 0$  and  $\|t^{-1}(U_t - I) - iR\| \rightarrow 0$  as  $t \rightarrow 0$  (where  $\|\cdot\|$  denotes the uniform operator norm).*

**Proof** It is known that  $U_t \in \mathcal{U}$  [12, p. 214]. We also have

$$\begin{aligned} \|U_t - I\| &= \left\| \sum_{j=1}^{\infty} \left( \frac{(it)^j}{j!} \right) R^j \right\| \\ &\leq \sum_{j=1}^{\infty} \left( \frac{|t|^j}{j!} \right) \|R^j\| \\ &\leq \exp(\|tR\|) - 1. \end{aligned}$$

Thus, we have  $U_t \rightarrow I$  uniformly as  $t \rightarrow 0$ . Further,

$$\begin{aligned} \|t^{-1}(U_t - I) - iR\| &= \left\| t^{-1} \sum_{j=2}^{\infty} \left( \frac{(it)^j}{j!} \right) R^j \right\| \\ &\leq t^{-1} \sum_{j=2}^{\infty} \left( \frac{|t|^j}{j!} \right) \|R^j\| \\ &\leq t^{-1} [\exp(\|tR\|) - (1 + \|tR\|)], \end{aligned}$$

which gives  $t^{-1}(U_t - I) \rightarrow iR$  uniformly as  $t \rightarrow 0$ . ♣

Next we state a lemma, whose proof can be found in [12, p. 98].

**Lemma 5.2.** *If  $\tilde{A}$  is a bounded (not necessarily self-adjoint) operator for which  $\|\tilde{A}\| < 1$ , then there is some  $N > 2$  and  $\{U_i\}_{i=1}^N \subset \mathcal{U}$  such that*

$$\tilde{A} = \frac{1}{N} \sum_{i=1}^N U_i.$$

We use this in the proof of the following result (see Remark 1.1).

**Proposition 5.3.** *Let  $\mathcal{X}$  be any (non-self-adjoint) Schatten  $p$ -space ( $1 \leq p < \infty$ ) or the space of compact or bounded operators. For any bounded operator  $A$  and any  $T \in \mathcal{X}$  we have*

$$\|AT\|_{\mathcal{X}} \leq \|T\|_{\mathcal{X}} \|A\| \geq \|TA\|_{\mathcal{X}}.$$

**Proof** If  $\mathcal{X}$  is the space of compact or bounded operators, then this is immediate, so assume that  $\mathcal{X}$  is a (non-self-adjoint) Schatten  $p$ -space for some  $p \in [1, \infty)$  (see [7]). If  $A = 0$ , the results are also immediate, so assume  $A \neq 0$ . Fix  $\delta \in (0, 1)$ , let  $\tilde{A} = A/(1 + \delta)\|A\|$ , and apply Lemma 5.2 to  $\tilde{A}$ , giving

$$\begin{aligned} \|AT\|_p &= (1 + \delta)\|A\| \cdot \|\tilde{A}T\|_p \\ &= (1 + \delta)\|A\| \cdot \left\| \frac{1}{N} \sum_{i=1}^N U_i T \right\|_p \\ &\leq (1 + \delta)\|A\| \frac{1}{N} \sum_{i=1}^N \|U_i T\|_p \\ &= (1 + \delta)\|A\| \cdot \|T\|_p, \end{aligned}$$

where we use the fact that  $\|UT\|_p = \|T\|_p$  for all  $U \in \mathcal{U}$ . Since this holds for each  $\delta \in (0, 1)$ , we obtain the results we want.  $\clubsuit$

**Proposition 5.4.** *Let the space  $\mathcal{X}$  satisfy the assumption of Proposition 5.3. Let  $A_t$  be a family of bounded operators which converge uniformly to  $A$  as  $t \rightarrow 0$ . Let  $T_t \subset \mathcal{X}$  be a family of operators satisfying  $\lim_{t \rightarrow 0} \|T_t - T\|_{\mathcal{X}} = 0$ . Then*

$$\lim_{t \rightarrow 0} \|A_t T_t - AT\|_{\mathcal{X}} = 0 = \lim_{t \rightarrow 0} \|T_t A_t - TA\|_{\mathcal{X}}.$$

**Proof** If we consider the inequalities

$$\begin{aligned} \|A_t T_t - AT\|_{\mathcal{X}} &= \|A_t(T_t - T) + (A_t - A)T\|_{\mathcal{X}} \\ &\leq \|A_t(T_t - T)\|_{\mathcal{X}} + \|(A_t - A)T\|_{\mathcal{X}} \\ &\leq \|A_t\| \cdot \|T_t - T\|_{\mathcal{X}} + \|A_t - A\| \cdot \|T\|_{\mathcal{X}}, \end{aligned}$$

we obtain the first result by Proposition 5.3. The second is proved similarly.  $\clubsuit$

Again we return to our two paired spectral sequences spaces  $V$  and  $W$ , as in Assumption 1, with their corresponding paired operator spaces  $\mathcal{V}$  and  $\mathcal{W}$ . In the following results,  $\partial$  denotes the usual convex subdifferential.

**Theorem 5.5. (Commutativity)** *Let  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  be a unitarily invariant convex function. Let operators  $S \in \mathcal{V}$  and  $T \in \mathcal{W}$  satisfy  $T \in \partial f(S)$ . Then  $TS = ST$ .*

**Proof** Let  $\mathcal{X}$  be the non-self-adjoint extension of  $\mathcal{V}$ . Define

$$U_t = \exp[-t(ST - TS)] \in \mathcal{U}.$$

That is, in Proposition 5.1 we have  $R = -i(TS - ST) \in B_{sa}$ . Thus,  $U_t \rightarrow I$  uniformly and  $t^{-1}(U_t - I) \rightarrow (ST - TS)$  uniformly as  $t \rightarrow 0$ . By Proposition 5.4 we obtain

$$(6) \quad \|t^{-1}(U_t - I)S + (ST - TS)S\|_{\mathcal{X}} \rightarrow 0.$$

Taking the adjoint of this gives

$$(7) \quad \|t^{-1}S(U_t^* - I) + S(TS - ST)\|_{\mathcal{X}} \rightarrow 0.$$

Now if we apply Proposition 5.4 with the first term of (6) playing the role of  $T_t$  and  $U_t^*$  playing the role of  $A_t$ , we obtain

$$\|[t^{-1}(U_t - I)S]U_t^* + [(ST - TS)S]I\|_{\mathcal{X}} \rightarrow 0,$$

which with (7) gives

$$(8) \quad \|t^{-1}(U_t S U_t^* - S) + (2STS - S^2 T - T S^2)\|_{\mathcal{V}} \rightarrow 0.$$

Hence

$$(9) \quad \text{tr}[t^{-1}(U_t S U_t^* - S)T] \rightarrow -\text{tr}[2STST - S^2 T^2 - T S^2 T].$$

If we examine the right hand side of (9), we see that since  $(S, T) \in \mathcal{V} \times \mathcal{W}$ , we have  $STST, S^2 T^2$  and  $T S^2 T \in B_1$  [12, p. 118]. Further, a simple calculation shows

$$(10) \quad \|TS - ST\|^2 = -\text{tr}[2STST - S^2 T^2 - T S^2 T].$$

If we use the unitary invariance of  $f$ , then for any  $t \in \mathbf{R}$  we obtain the inequality

$$t \cdot \text{tr}[t^{-1}(U_t S U_t^* - S)T] = \langle T, U_t S U_t^* - S \rangle \leq f(U_t S U_t^*) - f(S) = 0,$$

which together with (9) and (10) implies  $\|TS - ST\|_2 = 0$ , or  $TS = ST$ .  $\clubsuit$

The result above allows us to apply the following important tool, which is a well-known consequence of the Spectral Theorem.

**Fact:** Commuting elements of  $B_0$  are simultaneously diagonalizable.

**Lemma 5.6. (Diagonal subgradients)** *Let  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  be a unitarily invariant convex function. For  $x \in V$  we have  $y \in \partial(f \circ \text{dg})(x)$  if and only if  $\text{dg } y \in \partial f(\text{dg } x)$ .*

**Proof** Assume first that  $\text{dg } y \in \partial f(\text{dg } x)$ . Then for each  $z \in V$  we have

$$\langle y, z - x \rangle = \text{tr}[(\text{dg } y)(\text{dg } z - \text{dg } x)] \leq f(\text{dg } z) - f(\text{dg } x)$$

so that  $y \in \partial(f \circ \text{dg})(x)$ .

To see the other implication, given  $Q \in \mathcal{V}$ , choose  $z \in V$  and  $U \in \mathcal{U}$  so that  $Q = U^*(\text{dg } z)U$ . Now we use our von Neumann type inequality (Theorem 2.2), the assumption that  $y \in \partial(f \circ \text{dg})(x)$  and the unitary invariance of  $f$  (twice) to obtain

$$\begin{aligned} \text{tr}[(\text{dg } y)(Q - \text{dg } x)] &= \text{tr}[(\text{dg } y)(U^* \text{dg } z U)] - \langle y, x \rangle \\ &= B_U(z, y) - \langle y, x \rangle \\ &\leq \sup_{\pi \in \mathcal{R}} P_\pi(z, y) - \langle y, x \rangle \\ &= \sup\{\langle y, (z_{\pi(j)}) - x \rangle : \pi \in \mathcal{R}\} \\ &\leq \sup\{f(\text{dg } (z_{\pi(j)})_{j=1}^\infty) - f(\text{dg } x) : \pi \in \mathcal{R}\} \\ &= f(\text{dg } z) - f(\text{dg } x) \\ &= f(Q) - f(\text{dg } x), \end{aligned}$$

which gives us  $\text{dg } y \in \partial f(\text{dg } x)$ .  $\clubsuit$

The next lemma is elementary.

**Lemma 5.7.** *Let  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  be a unitarily invariant convex function. For  $S \in \mathcal{V}$ ,  $T \in \mathcal{W}$ , and  $U \in \mathcal{U}$ , we have  $T \in \partial f(S)$  if and only if  $U^*TU \in \partial f(U^*SU)$ .*

**Theorem 5.8. (Convex subgradients)** *Let  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  be a unitarily invariant convex function. For  $S \in \mathcal{V}$  and  $T \in \mathcal{W}$  we have  $T \in \partial f(S)$  if and only if there exists  $U \in \mathcal{U}$ ,  $x \in V$  and  $y \in W$  with  $S = U^*(\text{dg } x)U$ ,  $T = U^*(\text{dg } y)U$  and  $y \in \partial(f \circ \text{dg})(x)$ .*

**Proof** Assume first the existence of  $U$ ,  $x$  and  $y$  with the stated properties. Since  $y \in \partial(f \circ \text{dg})(x)$ , Lemma 5.6 tells us that  $\text{dg } y \in \partial f(\text{dg } x)$ . The unitary invariance of  $f$  gives  $T = U^*(\text{dg } y)U \in \partial f(U^* \text{dg } x U) = \partial f(S)$ .

To see the reverse implication, note that  $T \in \partial f(S)$  implies  $TS = ST$  by Theorem 5.5, which means  $S$  and  $T$  are simultaneously diagonalizable. Thus, there exists  $U \in \mathcal{U}$ ,  $x \in V$  and  $y \in W$  with  $S = U^*(\text{dg } x)U$ , and  $T = U^*(\text{dg } y)U$ , so that  $U^*(\text{dg } y)U \in \partial f(U^*(\text{dg } x)U)$ . Now the unitary invariance allows us to rewrite this as  $\text{dg } y \in \partial f(\text{dg } x)$ , which by Lemma 5.6 implies  $y \in \partial(f \circ \text{dg})(x)$ , as required.  $\clubsuit$

Our next theorem shows that the differentiability of a convex unitarily invariant function  $f$  is characterized by that of  $f \circ \text{dg}$ .

**Theorem 5.9. (Convexity and differentiability)** *Let  $f : \mathcal{V} \rightarrow \bar{R}$  be a unitarily invariant lower semicontinuous convex function. Then  $f$  is Gâteaux differentiable at  $A \in \mathcal{V}$  if and only if  $f \circ \text{dg}$  is Gâteaux differentiable at  $\lambda(A) \in V$ .*

**Proof** Note that a lower semicontinuous convex function is Gâteaux differentiable at a point if and only if its subdifferential is a singleton at that point. Assume that  $f$  is Gâteaux differentiable at  $A$ . Let  $y, z \in \partial(f \circ \text{dg})(\lambda(A))$  and let  $U \in \mathcal{U}$  be such that  $U^*\lambda(A)U = A$ . By Theorem 5.8,

$$U^*(\text{dg } y)U \text{ and } U^*(\text{dg } z)U \in \partial f(A).$$

Since  $f$  is Gâteaux differentiable at  $A$ , we know  $U^*(\text{dg } y)U = U^*(\text{dg } z)U$ , which implies that  $y = z$ . Thus  $\partial(f \circ \text{dg})(\lambda(A))$  is a singleton and  $f \circ \text{dg}$  is Gâteaux differentiable at  $\lambda(A)$ .

To prove the converse we assume that  $f \circ \text{dg}$  is Gâteaux differentiable at  $\lambda := \lambda(A)$ , so  $\partial(f \circ \text{dg})(\lambda)$  is a singleton  $\{\mu\}$ .

First we observe that if  $\lambda_i = \lambda_j$  then  $\mu_i = \mu_j$ . In fact, when  $\lambda_i = \lambda_j$  the permutation invariance and the Gâteaux differentiability of  $f \circ \text{dg}$  implies that

$$\mu_i t + o(t) = f \circ \text{dg}(\lambda + te^i) - f \circ \text{dg}(\lambda) = f \circ \text{dg}(\lambda + te^j) - f \circ \text{dg}(\lambda) = \mu_j t + o(t).$$

Therefore,  $\mu_i = \mu_j$ .

Next we show that if  $B \in \mathcal{U}$  commutes with  $\text{dg } \lambda$ , that is

$$(11) \quad B(\text{dg } \lambda) = (\text{dg } \lambda)B,$$

then  $B$  also commutes with  $\text{dg } \mu$ . Obviously it suffices to show that, for arbitrary natural numbers  $i$  and  $j$ ,

$$\langle e^i, B(\text{dg } \mu)e^j \rangle = \langle e^i, (\text{dg } \mu)Be^j \rangle,$$

or equivalently,

$$(12) \quad \mu_j \langle e^i, Be^j \rangle = \mu_i \langle e^i, Be^j \rangle.$$

When  $\lambda_i = \lambda_j$  we have  $\mu_i = \mu_j$  which immediately leads to (12). It remains to consider the case when  $\lambda_i \neq \lambda_j$ . Note that it follows from (11) that

$$\lambda_j \langle e^i, Be^j \rangle = \langle e^i, B(\text{dg } \lambda)e^j \rangle = \langle e^i, (\text{dg } \lambda)Be^j \rangle = \lambda_i \langle e^i, Be^j \rangle,$$

which implies that  $\langle e^i, Be^j \rangle = 0$ . Thus, both sides of (12) are 0.

Now consider any two elements  $S$  and  $T$  of  $\partial f(A)$ . By Theorem 5.8 (Convex subgradients) there exist unitary operators  $U, V$  satisfying

$$(13) \quad U^*(\text{dg } \lambda)U = V^*(\text{dg } \lambda)V = A$$

such that  $S = U^*(\text{dg } \mu)U$  and  $T = V^*(\text{dg } \mu)V$ . It follows from (13) that  $B := VU^*$  commutes with  $\text{dg } \lambda$ . Therefore,  $B$  also commutes with  $\text{dg } \mu$ . Then we have

$$S = U^*(\text{dg } \mu)U = V^*B(\text{dg } \mu)U = V^*(\text{dg } \mu)BU = V^*(\text{dg } \mu)V = T.$$

Thus  $\partial f(A)$  is a singleton, so  $f$  is Gâteaux differentiable at  $A$ . ♣

## 6. NONCONVEX FUNCTIONS

Again we fix two paired spectral sequences spaces  $V$  and  $W$ , as in Assumption 1, with their corresponding paired operator spaces  $\mathcal{V}$  and  $\mathcal{W}$ .

Our main result, Theorem 5.8 (Convex subgradients), extends to nonconvex Lipschitz functions with the convex subdifferential replaced by the limiting subdifferential. To state this result precisely we recall the corresponding concepts first. Let  $\mathcal{X}$  be a Banach space and let  $f : \mathcal{X} \rightarrow \mathbf{R}$  be a Lipschitz function. The *Gâteaux subdifferential*  $D_G f(x)$  of  $f$  at  $x \in X$  is defined by

$$D_G f(x) := \left\{ x^* \in X^* : \liminf_{t \rightarrow 0} \frac{f(x + th) - f(x)}{t} \geq \langle x^*, h \rangle, \forall h \in X \right\},$$

and the *limiting subdifferential*  $\partial_L f(x)$  of  $f$  at  $x \in X$  is defined by

$$\partial_L f(x) := \{x^* \in X^* : x^* = w^* - \lim_{n \rightarrow \infty} x_n^*, x_n^* \in D_G f(x_n), x_n \rightarrow x\}.$$

The proof of the extension of Theorem 5.8 follows the same strategy as the previous section. First we extend Theorem 5.5 (Commutativity).

**Theorem 6.1. (Commutativity)** *Let  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  be a unitarily invariant locally Lipschitz function. Let  $S \in \mathcal{V}$  and  $T \in \mathcal{W}$  satisfy  $T \in \partial_L f(S)$ . Then  $TS = ST$ .*

**Proof** We need only prove the case  $T \in D_G f(S)$ . The conclusion for the limiting subdifferential follows directly from a limiting process.

Define, as in the proof of Theorem 5.5,  $U_t = \exp[-t(ST - TS)] \in \mathcal{U}$ . Then it follows from (8) that

$$U_t S U_t^* = S - t(2STS - S^2T - TS^2) + o(t).$$

Since  $f$  is unitarily invariant and locally Lipschitz, and  $T \in D_G f(S)$ , we have

$$\begin{aligned} 0 &= t^{-1}[f(U_t S U_t^*) - f(S)] \\ &= t^{-1}[f(S - t(2STS - S^2T - TS^2)) - f(S)] + o(1) \\ &\geq -\langle T, 2STS - S^2T - TS^2 \rangle + o(1) \\ &= \|TS - ST\|^2 + o(1). \end{aligned}$$

Taking limits as  $t \rightarrow 0$  yields  $\|TS - ST\| = 0$ . ♣

**Lemma 6.2.** *Let  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  be a unitarily invariant locally Lipschitz function. For  $x \in V$  we have  $y \in D_G(f \circ \text{dg})(x)$  if and only if  $\text{dg } y \in D_G f(\text{dg } x)$ . Furthermore, if  $y \in \partial_L(f \circ \text{dg})(x)$  then  $\text{dg } y \in \partial_L f(\text{dg } x)$ .*

**Proof** The idea is to approximate operators and functions on  $\mathcal{V}$  by their finite-dimensional restrictions. It is not hard to check that Lewis's results for the Gâteaux subdifferential [10] apply to these restrictions. Then the general result is derived by taking limits.

It suffices to prove the lemma for the Gâteaux subdifferential and the result for the limiting subdifferential follows from a limiting process. Observe that the 'if' part is easy. We concentrate on the 'only if' part. We need the following notation and simple facts.

Define a subspace  $E_n$  of  $\ell_2^{\mathbf{C}}$  by  $E_n := \text{span}\{e^1, e^2, \dots, e^n\}$ . Let  $H_n$  denote the Hermitian operators on the space  $E_n$ , define the eigenvalue map  $\lambda_n : H_n \rightarrow \mathbf{R}^n$

mapping operators to their eigenvalue sequence in decreasing order, and define the diagonal map  $\text{dg}_n : \mathbf{R}^n \rightarrow H_n$  in the natural way: for  $y \in \mathbf{R}^n$  and  $\mu \in \mathbf{C}^n$ ,

$$(\text{dg}_n y) \sum_{i=1}^n \mu_i e^i = \sum_{i=1}^n \mu_i y_i e^i.$$

Let  $P_n : \ell_2^{\mathbf{C}} \rightarrow E_n$  denote the orthogonal projection: it is easy to check that the adjoint  $P_n^* : E_n \rightarrow \ell_2^{\mathbf{C}}$  is just inclusion. Furthermore, the operator  $P_n^* P_n : \ell_2^{\mathbf{C}} \rightarrow \ell_2^{\mathbf{C}}$  is just the orthogonal projection onto  $E_n$ . Given any operator  $Z \in \mathcal{V}$  we have, relative to the decomposition  $\ell_2^{\mathbf{C}} = E_n \oplus E_n^\perp$ , the block decomposition

$$P_n^* P_n Z P_n^* P_n = \begin{pmatrix} P_n Z P_n^* & 0 \\ 0 & 0 \end{pmatrix}.$$

Furthermore, by [7, Thm 6.3, p. 90], we have

$$\lim_{n \rightarrow \infty} \|Z - P_n^* P_n Z P_n^* P_n\|_{\mathcal{V}} = 0.$$

Finally, let  $P_n^V : V \rightarrow \mathbf{R}^n$  and  $P_n^W : W \rightarrow \mathbf{R}^n$  denote the natural truncation maps, and note the identity

$$\text{dg}_n(P_n^V z) = P_n(\text{dg} z)P_n^*, \text{ for } z \in V.$$

The analogous identity also holds in  $W$ .

Suppose, then, that elements  $x \in V$  and  $y \in W$  satisfy  $y \in D_G(f \circ \text{dg})(x)$ . Define a permutation-invariant locally Lipschitz function  $\phi : \mathbf{R}^n \rightarrow \mathbf{R}$  by

$$\phi(z) := f(\text{dg}(z_1, z_2, \dots, z_n, x_{n+1}, x_{n+2}, \dots)).$$

It is straightforward to check  $P_n^W y \in D_G \phi(P_n^V x)$ , from which it follows, by [10, Cor 5.14], that  $\text{dg}_n P_n^W y \in D_G(\phi \circ \lambda_n)(\text{dg}_n P_n^V x)$ . Denote the Lipschitz constant for  $f$  by  $L$ . Then for any operator  $Z$  in the space  $\mathcal{V}$  and small real  $t > 0$  we have

$$\begin{aligned} f(\text{dg} x + tZ) - f(\text{dg} x) &\geq \\ &-Lt\|Z - P_n^* P_n Z P_n^* P_n\|_{\mathcal{V}} + f(\text{dg} x + tP_n^* P_n Z P_n^* P_n) - f(\text{dg} x). \end{aligned}$$

By using the above block decomposition we know

$$\begin{aligned} &f(\text{dg} x + tP_n^* P_n Z P_n^* P_n) - f(\text{dg} x) \\ &= \phi(\lambda_n(\text{dg}_n(P_n^V x) + tP_n Z P_n^*)) - \phi(\lambda_n(\text{dg}_n(P_n^V x))) \\ &\geq t\langle \text{dg}_n P_n^W y, P_n Z P_n^* \rangle + \|Z\|_{\mathcal{V}} o(t) \\ &\geq t\langle P_n(\text{dg} y)P_n^*, P_n Z P_n^* \rangle + \|Z\|_{\mathcal{V}} o(t) \\ &\geq t\langle \text{dg} y, Z \rangle - t\|y\|_{\mathcal{W}} \|Z - P_n^* P_n Z P_n^* P_n\|_{\mathcal{V}} + \|Z\|_{\mathcal{V}} o(t). \end{aligned}$$

Hence we deduce

$$\begin{aligned} &f(\text{dg} x + tZ) - f(\text{dg} x) \geq \\ &t\langle \text{dg} y, Z \rangle - t(L + \|y\|_{\mathcal{W}})\|Z - P_n^* P_n Z P_n^* P_n\|_{\mathcal{V}} + \|Z\|_{\mathcal{V}} o(t). \end{aligned}$$

Taking the limit as  $n \rightarrow \infty$  shows

$$\liminf_{t \rightarrow 0} \frac{f(\text{dg} x + tZ) - f(\text{dg} x)}{t} \geq \langle \text{dg} y, Z \rangle,$$

and since  $Z$  was arbitrary we deduce  $\text{dg} y \in D_G f(\text{dg} x)$ , as required. ♣

The next lemma is elementary.

**Lemma 6.3.** *Let  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  be a unitarily invariant locally Lipschitz function. For  $S \in \mathcal{V}$ ,  $T \in \mathcal{W}$ , and  $U \in \mathcal{U}$ , we have  $T \in \partial_L f(S)$  (or  $T \in D_G f(S)$ ) if and only if  $U^*TU \in \partial_L f(U^*SU)$  (or  $U^*TU \in D_G f(U^*SU)$  respectively).*

**Theorem 6.4. (Gâteaux and limiting subdifferentials)** *Let  $f : \mathcal{V} \rightarrow \bar{\mathbf{R}}$  be a unitarily invariant locally Lipschitz function. If there exist  $U \in \mathcal{U}$ ,  $x \in V$  and  $y \in W$  with  $S = U^*(\text{dg } x)U$ ,  $T = U^*(\text{dg } y)U$  and  $y \in \partial_L(f \circ \text{dg})(x)$  (or  $y \in D_G(f \circ \text{dg})(x)$ ) then  $T \in \partial_L f(S)$  (or  $T \in D_G f(S)$  respectively). The converse is also true for the Gâteaux subdifferential.*

**Proof** We prove this theorem for the Gâteaux subdifferential. The proof for the limiting subdifferential is similar.

Assume first the existence of  $U$ ,  $x$  and  $y$  with the stated properties. Since  $y \in D_G(f \circ \text{dg})(x)$ , Lemma 6.2 tells us that  $\text{dg } y \in D_G f(\text{dg } x)$ . The unitary invariance of  $f$  gives  $T = U^*(\text{dg } y)U \in D_G f(U^*(\text{dg } x)U) = D_G f(S)$ .

To see the reverse implication note that  $T \in D_G f(S)$  implies  $TS = ST$  by Theorem 6.1, which means  $T$  and  $S$  are simultaneously diagonalizable. Thus, there exists  $U \in \mathcal{U}$ ,  $x \in V$  and  $y \in W$  with  $S = U^*(\text{dg } x)U$ , and  $T = U^*(\text{dg } y)U$ , so that  $U^*(\text{dg } y)U \in D_G f(U^*(\text{dg } x)U)$ . Now the unitary invariance allows us to rewrite this as  $\text{dg } y \in D_G f(\text{dg } x)$ , which by Lemma 6.2 implies  $y \in D_G(f \circ \text{dg})(x)$ , as required.  $\clubsuit$

## 7. EXAMPLES

We illustrate some of the results in this paper with a number of examples.

**Example 7.1.** We know that the norm in  $\ell_p$  for  $p \in (1, \infty)$  is strictly differentiable away from the origin. As an immediate consequence of Theorem 5.9, so is the norm of  $B_p$ .

Another interesting classical norm to consider appears in [7, p. 139].

**Example 7.2.** For an element  $x$  of  $c_0$ , we rearrange the components  $|x_i|$  into decreasing order to obtain a new element  $x^*$  of  $c_0$ . Now for a fixed decreasing sequence  $\{\pi_i\}$  in  $\mathbf{R}_+$  (not identically zero), the norms

$$x \in V \mapsto \sup_n \left\{ \frac{\sum_1^n x_i^*}{\sum_1^n \pi_i} \right\},$$

and

$$y \in W \mapsto \sum_1^\infty \pi_i y_i^*$$

are dual to each other, and hence, by Theorem 4.5, so are their compositions with  $\lambda$  on the spaces  $\mathcal{V}$  and  $\mathcal{W}$ . The case  $\pi_i = i^{1/q} - (i-1)^{1/q}$  ( $1 < q < \infty$ ) gives the Calderón norms [16, p. 13].

Now let us consider an example of some recent interest.



**Example 7.3.** In [5] the author looks at self-concordant barriers, which play a central role in many types of optimization problems. The theory developed thus far gives an appealing setting. The function  $g : \mathbf{R} \rightarrow \bar{\mathbf{R}}$  given by

$$g(t) = \begin{cases} t - \ln(1+t) & (t > -1) \\ +\infty & (t \leq -1) \end{cases}$$

is convex. Further,  $g$  has conjugate  $g^*(s) = g(-s)$ . Now we define the function  $\phi$  on  $\ell_2$  by writing

$$\phi(x) = \sum_{i=1}^{\infty} g(x_i).$$

Clearly,  $\phi$  is finite and continuous at 0 since  $0 \leq g(t) \leq t^2$  for small  $t$ . Using our techniques, we obtain a function  $f$  on  $B_2$  in the obvious way,

$$f = \phi \circ \lambda,$$

and Theorem 4.5 shows  $f^*(T) = f(-T)$ . Notice that if  $T \in B_2$  and  $I + T \geq 0$ , then

$$f(T) = \text{tr}(T) - \ln(\det(I + T)).$$

Our last example concerns the ordered eigenvalues of a positive operator.

**Example 7.4.** The  $k$ 'th largest eigenvalue of a positive self-adjoint operator  $T \in \mathcal{V}$ , which we denote  $\mu_k(T)$ , arises naturally in many applications. We can extend  $\mu_k$  to the whole space  $\mathcal{V}$  by defining  $\mu_k(T)$  to be the  $k$ 'th largest eigenvalue of  $T$  if  $T$  has at least  $k$  nonnegative eigenvalues, and otherwise  $\mu_k(T) = 0$ . To see why this definition is natural, consider the rearrangement invariant convex function  $\sigma_k : V \rightarrow \mathbf{R}$  defined by

$$\sigma_k(x) = \sup \left\{ \sum_{i \in I} x_i : |I| = k \right\},$$

for  $k \geq 1$  and  $\sigma_0 \equiv 0$ . If we define a locally Lipschitz rearrangement invariant function  $\phi_k = \sigma_k - \sigma_{k-1}$  then  $\mu_k = \phi_k \circ \lambda$ . Clearly  $\phi_k(x)$  is the  $k$ 'th largest component of  $x$  if  $x$  has at least  $k$  nonnegative components, and otherwise  $\phi_k(x) = 0$ .

Following the technique of [10], we arrive at the following result.

**Theorem 7.5. ( $k$ 'th largest component)** *Suppose the element  $x \in V$  has at least  $k$  strictly positive components: that is  $\phi_k(x) > 0$ . Let*

$$\Delta = \{e^i : x_i = \phi_k(x)\}.$$

Then

$$D_G \phi_k(x) = \begin{cases} \text{conv } \Delta, & \text{if } \phi_{k-1}(x) > \phi_k(x), \\ \emptyset, & \text{otherwise,} \end{cases}$$

and  $y \in \partial_L \phi_k(x)$  if and only if  $y$  is a convex combination of at most

$$1 - k + |\{j : x_j \geq \phi_k(x)\}|$$

vectors from  $\Delta$ .

Applying Theorem 6.4 gives the following result.

**Theorem 7.6. (*k*'th largest eigenvalue)** *Suppose the operator  $T \in \mathcal{V}$  has at least  $k$  strictly positive eigenvalues: that is,  $\mu_k(T) > 0$ . Let*

$$\Gamma = \text{conv} \{u \odot u : u \in \ell_2, \|u\| = 1, Tu = \mu_k(T)u\}.$$

*Then*

$$D_G \mu_k(T) = \begin{cases} \Gamma, & \text{if } \mu_{k-1}(T) > \mu_k(T), \\ \emptyset, & \text{otherwise,} \end{cases}$$

*and if  $S \in \Gamma$  and*

$$\text{rank } S \leq 1 - k + |\{j : \mu_j(T) \geq \mu_k(T)\}|,$$

*then  $S \in \partial_L \mu_k(T)$ .*

#### 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] J.M. Borwein and Q.J. Zhu, A survey of subdifferentials and their applications, CECM Research Report 98-105 (1998), to appear in *Nonlinear Analysis, TMA*.
- [3] 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).
- [4] S.J. Cox, The generalized gradient at a multiple eigenvalue, *Journal of Functional Analysis* **133** (1995) 30-40.
- [5] 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.
- [6] S. Friedland and P. Nowosad, Extremal Eigenvalue Problems with Indefinite Kernels, *Adv. in Math.* **40** (1981) 128-154.
- [7] 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.
- [8] G.H. Hardy, J.E. Littlewood and G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, U. K. 1952.
- [9] A.S. Lewis, Convex analysis on the Hermitian matrices, *SIAM J. Optim.* **6**, 164-177 (1996).
- [10] A.S. Lewis, Nonsmooth analysis of eigenvalues, *Mathematical Programming*, **84**, 1-24 (1999).
- [11] H. Minc, *Nonnegative matrices*, John Wiley & Sons, New York, 1988.
- [12] G. Pederson, *Analysis Now*, Springer Verlag, Berlin, 1989.
- [13] 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).
- [14] J. Read, The Approximation of Optimal Control of Vibrations - A Geometrical Method, *Mathematical Methods in the Applied Sciences*, **19**, 87-129 (1996)
- [15] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, N.J. 1970.
- [16] B. Simon, *Trace Ideals and Their Applications*, Cambridge University Press, 1979.
- [17] N. Tomczak-Jaegerman, The moduli of smoothness and convexity and the Rademacher averages of trace classes  $S_p$  ( $1 < p < \infty$ ), *Studia Math.* **50** 163-182 (1974).
- [18] N. Tomczak-Jaegerman, On the differentiability of the norm in trace classes  $S_p$ , *Seminaire Maurey-Schwartz 1974-1975*, Exp. XXII, Centre Math. Polytech. Paris, 1975.
- [19] J. von Neumann, Some matrix inequalities and metrization of matrix-space, *Tomsk University Review*, **1** (1937) 286-300. In *Collected Works*, Pergamon, Oxford, 1962, Volume IV, 205-218.

J. M. BORWEIN AND J. READ

Centre for Experimental and Constructive Mathematics, Department of Mathematics and Statistics, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6

*E-mail address:* `jborwein@cecm.sfu.ca`, `read@cecm.sfu.ca`

A. S. LEWIS

Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

*E-mail address:* `aslewis@orion.uwaterloo.ca`

QIJI ZHU

Department of Mathematics and Statistics, Western Michigan University, Kalamazoo, MI 49008, USA

*E-mail address:* `zhu@math-stat.wmich.edu`