DERIVATIVES OF ORBITAL FUNCTIONS, AN EXTENSION OF BEREZIN-GEL'FAND'S THEOREM AND APPLICATIONS

TIN-YAU TAM* AND WILLIAM C. HILL†

Abstract. A generalization of a result of Berezin and Gel'fand in the context of Eaton triples is given. The generalization and its proof are Lie-theoretic free and requires some basic knowledge of nonsmooth analysis. The result is then applied to determine the distance between a point and a G-orbit or its convex hull. We also discuss the derivatives of some orbital functions.

Key words. Berezin-Gel'fand's theorem, subdifferential, Clarke generalized gradient, Lebourg mean value theorem, Eaton triple, reduced triple, finite reflection group

AMS subject classifications. 90C31, 15A18

1. Introduction.

Let us recall a result of Berezin and Gel'fand [3].

THEOREM 1.1. (Berezin-Gel'fand [3]) *Let G be a semisimple Lie group with finite center, whose Lie algebra $\mathfrak{g}$ has Cartan decomposition $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$, where the analytic group of $\mathfrak{k}$ is $K \subset G$. For $x \in \mathfrak{p}$, let $a_+(x)$ denote the unique element of the singleton set $Ad(K)x \cap \mathfrak{a}_+$, where $\mathfrak{a}_+$ is a closed fundamental Weyl chamber. For $y, z \in \mathfrak{p}$, $a_+(z+y) - a_+(z) \in conv\, W a_+(y)$, where conv denotes the convex hull of the underlying set and W denotes the Weyl group of $(\mathfrak{g}, \mathfrak{a})$.*

The result of Berezin-Gel'fand had been known to Lidskii who [24] gave an elementary proof of a special case of Berezin-Gel'fand's theorem, namely, $G = SL(n, \mathbb{R})$, though [24] appeared earlier than [3]. The sketch of the proof of Berezin-Gel'fand's result in [3] is Lie theoretic and a detailed proof, to our best knowledge, is found nowhere. Lidskii's proof is not Lie theoretic but still employs some analytic technique. Wielandt did not fully understand Lidskii's proof and this led him [36] to provide another proof by using minimax property. The result of Lidskii is stated in the following

THEOREM 1.2. (Lidskii [24], Wielandt [36]) *Let A and B be real symmetric (Hermitian, quaternionic Hermitian) matrices. Denote by $\lambda(A) = (\lambda_1(A), \dots, \lambda_n(A))$ the vector of eigenvalues of A with $\lambda_1(A) \ge \cdots \ge \lambda_n(A)$. Then*

$$\lambda(A+B) - \lambda(B) \in conv\, S_n\, \lambda(A),$$

where S_n is the symmetric group. In terms of inequalities, it is equivalent to

$$\max_{1 \le j_1 < \cdots < j_k \le n} \sum_{i=1}^{k} [\lambda_{j_i}(A+B) - \lambda_{j_i}(B)] \le \sum_{i=1}^{k} \lambda_i(A), \qquad k = 1, \dots, n-1,$$
$$\sum_{i=1}^{n} [\lambda_i(A+B) - \lambda_i(B)] = \sum_{i=1}^{n} \lambda_i(A),$$

*Department of Mathematics, 218 Parker Hall, Auburn University, Auburn, AL 36849-5310, USA (email: tamtiny@auburn.edu) homepage: http://www.auburn.edu/∼tamtiny.

†Division of Natural Sciences and Mathematics, Macon State College, 100 College Station Drive, GA 31206, USA (email: whill@mail.maconstate.edu).

and the equality is merely the trace condition.

Later Markus [26] gave another proof of Lidskii's theorem by using an idea of Wielandt [36] but not the minimax property. See three proofs and some historical remarks in [4, 35]. Very recently, Lewis [23] provided a new proof of Lidskii's result via nonsmooth analysis. Though it is not the simplest one, it provides a totally new look to Lidskii's theorem. Inspired by Lewis' approach a generalization of Berezin-Gel'fand's result is given via nonsmooth analysis in Section 4. In order to carry out the approach, the derivatives of some orbital functions are studied and a number of results in [21] are generalized in Section 3. Then we determine the distance between a G-orbit or its convex hull and a given point as applications in Section 5.

The following is a framework for the extension which only requires basic knowledge of linear algebra. Let G be a closed subgroup of the orthogonal group on a finite dimensional real inner product space V. The triple (V, G, F) is an *Eaton triple* if $F \subset V$ is a nonempty closed convex cone such that

(A1) $Gx \cap F$ is nonempty for each $x \in V$.

(A2) $\max_{g\in G}(x, gy) = (x, y)$ for all $x, y \in F$.

The Eaton triple (W, H, F) is called a *reduced* triple of the Eaton triple (V, G, F) if it is an Eaton triple and $W := \operatorname{span} F$ and $H := \{g|_W : g \in G,\ gW = W\} \subset O(W)$, the orthogonal group of W [33]. For $x \in V$, let $F(x)$ denote the unique element of the singleton set $Gx \cap F$. The function (abuse of notation) $F : V \to F$ is idempotent. It is known that H is a finite reflection group [27].

Let us recall some rudiments of finite reflection groups [15]. Let V be a finite dimensional real inner product space. A reflection s_α on V is an element of $O(V)$, which sends some nonzero vector α to its negative and fixes pointwise the hyperplane H_α orthogonal to α, that is, $s_\alpha\lambda := \lambda - 2(\lambda, \alpha)/(\alpha, \alpha)\, \alpha,\ \lambda \in V$. A finite group G generated by reflections is called a *finite reflection group*. A *root system* of G is a finite set of nonzero vectors in V, denoted by Φ, such that $\{s_\alpha : \alpha \in \Phi\}$ generates G, and satisfies

(R1) $\Phi \cap \mathbb{R}\alpha = \{\pm\alpha\}$ for all $\alpha \in \Phi$.

(R2) $s_\alpha\Phi = \Phi$ for all $\alpha \in \Phi$.

The elements of Φ are called *roots*. We do not require that the roots are of equal length. A root system Φ is *crystallographic* if it satisfies the additional requirement:

(R3) $2\frac{(\alpha,\beta)}{(\beta,\beta)} \in \mathbb{Z}$ for all $\alpha, \beta \in \Phi$,

and the group G is known as the *Weyl group* of Φ.

A (open) *chamber* C is a connected component of $V \setminus \cup_{\alpha\in\Phi} H_\alpha$. Given a *total order* $<$ in V [15, p.7], $\lambda \in V$ is said to be *positive* if $0 < \lambda$. Certainly, there is a total order in V: Choose an arbitrary ordered basis $\{v_1, \dots, v_m\}$ of V and say $\mu > \nu$ if the first nonzero number of the sequence $(\lambda, v_1), \dots, (\lambda, v_m)$ is positive, where $\lambda = \mu - \nu$. Now $\Phi^+ \subset \Phi$ is called a *positive system* if it consists of all those roots which are positive relative to a given total order. Of course, $\Phi = \Phi^+ \cup \Phi^-$, where $\Phi^- = -\Phi^+$. Now Φ^+ contains [15, p.8] a unique *simple system* Δ, that is, Δ is a basis for $V_1 := \operatorname{span} \Phi \subset V$, and each $\alpha \in \Phi$ is a linear combination of Δ with coefficients all of the same sign (all nonnegative or all nonpositive). The vectors in Δ are called *simple roots* and the corresponding reflections are called *simple reflections*. The finite reflection group G is generated by the simple reflections. Denote by $\Phi^+(C)$ the positive system obtained by the total order induced by an ordered basis $\{v_1, \ldots, v_m\} \subset C$ of V as described above. Indeed $\Phi^+(C) = \{\alpha \in \Phi : (\lambda, \alpha) > 0 \text{ for all } \lambda \in C\}$. The correspondence $C \mapsto \Phi^+(C)$ is a bijection of the set of all chambers onto the set of all positive systems. The group G acts simply transitively on the sets of positive systems, simple systems and chambers. The closed convex cone $F := \{\lambda \in V : (\lambda, \alpha) \geq 0, \text{ for all } \alpha \in \Delta\}$, that is, $F := C^-$ is the closure of the chamber C which defines Φ^+ and Δ, is called a (closed) *fundamental domain* for the action of G on V associated with Δ. Since G acts transitively on the chambers, given $x \in V$, the set $Gx \cap F$ is a singleton set and its element is denoted by $F(x)$. It is known that (V, G, F) is an Eaton triple (see [27]). Let $V_0 := \{x \in V : gx = x \text{ for all } g \in G\}$ be the set of fixed points in V under the action of G. Let $\Delta = \{\alpha_1, \ldots, \alpha_n\}$, that is, $\dim V_1 = n$, where $V_1 = V_0^\perp$. If $\{\lambda_1, \ldots, \lambda_n\}$ denotes the basis of V_1 dual to the basis $\{\beta_i := 2\alpha_i/(\alpha_i, \alpha_i) : i = 1, \ldots, n\}$, that is, $(\lambda_i, \beta_j) = \delta_{ij}$, then $F = \{\sum_{i=1}^n c_i \lambda_i : c_i \geq 0\} \oplus V_0$. Thus the interior $\operatorname{Int} F = C$ of F is the nonempty set $\{\sum_{i=1}^n c_i \lambda_i : c_i > 0\} \oplus V_0$. The *dual cone* of F in V_1 is the cone

$$\operatorname{dual}_{V_1} F := \{x \in V_1 : (x, u) \geq 0, \text{ for all } u \in F\}$$

induced by F. Notice that $\operatorname{dual}_{V_1} F = \{\sum_{i=1}^n c_i \alpha_i, c_i \geq 0, i = 1, \ldots, n\}$. There is a unique element $\omega \in G$ sending Φ^+ to Φ^- and thus sending F to $-F$. Moreover, the length [15, p.12] of ω is the longest one [15, p.15-16] and thus we call it the *longest element*.

We will present two examples requiring some basic knowledge of Lie theory [18]. Let $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$ be a Cartan decomposition of the Lie algebra $\mathfrak{g}$ of a semisimple Lie group G with finite center. Denote the Killing form of $\mathfrak{g}$ by $B(\cdot, \cdot)$. The Killing form is positive definite on $\mathfrak{p}$ but negative definite on $\mathfrak{k}$. Let K be an analytic subgroup of $\mathfrak{k}$ in the analytic group G of $\mathfrak{g}$. Now $\operatorname{Ad}(K)$ is a subgroup of the orthogonal group on $\mathfrak{p}$ with respect to the restriction of the Killing form on $\mathfrak{p}$ since the Killing form is invariant under $\operatorname{Ad}(K)$. Among the abelian subalgebras of $\mathfrak{g}$ that are contained in $\mathfrak{p}$, choose a maximal one $\mathfrak{a}$ (*maximal abelian subalgebra* in $\mathfrak{p}$). For $\alpha \in \mathfrak{a}^*$ (the dual space of $\mathfrak{a}$), set $\mathfrak{g}_\alpha = \{x \in \mathfrak{g} : [h, x] = \alpha(h)x \text{ for all } h \in \mathfrak{a}\}$. If $0 \neq \alpha \in \mathfrak{a}^*$ and $\mathfrak{g}_\alpha \neq 0$, then α is called a (*restricted*) *root* [18, p.313] of the pair $(\mathfrak{g}, \mathfrak{a})$. The set of roots will be denoted Σ. We have the orthogonal direct sum $\mathfrak{g} = \mathfrak{g}_0 + \sum_{\alpha \in \Sigma} \mathfrak{g}_\alpha$ known as *restricted-root space decomposition* [18, p.313]. We view $\mathfrak{a}$ as a Euclidean space by taking the inner product to be the restriction of B to $\mathfrak{a}$. The map $\mathfrak{a}^* \to \mathfrak{a}$ that assigns to each $\lambda \in \mathfrak{a}^*$ the unique element x_λ of $\mathfrak{a}$ satisfying $\lambda(x) = B(x, x_\lambda)$ for all $x \in \mathfrak{a}$ is a vector space isomorphism. We use this isomorphism to identify $\mathfrak{a}^*$ with $\mathfrak{a}$, allowing us, in particular, to view Σ as a subset of $\mathfrak{a}$. The set $\Phi = \{\alpha \in \Sigma : \frac{1}{2}\alpha \notin \Sigma\}$ generates a finite reflection group W, that is, W is generated by the reflections s_α ($\alpha \in \Sigma$), which is called the *Weyl group* of $(\mathfrak{g}, \mathfrak{a})$, and is a root system of W. It is called the *reduced root system* of the pair $(\mathfrak{g}, \mathfrak{a})$. Now fix a simple system Δ for the root system Φ. Then Δ determines a fundamental domain $\mathfrak{a}_+$ for the action of W on $\mathfrak{a}$. We now describe another way to view the Weyl group W. Use juxtaposition to represent the adjoint action of G on $\mathfrak{g}$, that is, $gx = \operatorname{Ad}(g)x$, $g \in G$, $x \in \mathfrak{g}$. Set $N_K(\mathfrak{a}) = \{k \in K : k\mathfrak{a} \subset \mathfrak{a}\}$ (the normalizer of $\mathfrak{a}$ in K) and $Z_K(\mathfrak{a}) = \{k \in K : kx = x \text{ for all } x \in \mathfrak{a}\}$ (the centralizer of $\mathfrak{a}$ in K). Then the action of K on $\mathfrak{g}$ induces an action of the group $N_K(\mathfrak{a})/Z_K(\mathfrak{a})$ on $\mathfrak{a}$, that is, $[k]x = kx$ for $[k] \in N_K(\mathfrak{a})/Z_K(\mathfrak{a})$.

There exists an isomorphism $\psi : W \to N_K(\mathfrak{a})/Z_K(\mathfrak{a})$ that is, compatible with the two actions on $\mathfrak{a}$, or more precisely, for which $wx = \psi(w)x$, $w \in W$, $x \in \mathfrak{a}$ [18, p.325, p.394]. We use the isomorphism ψ to identify these two groups (in the literature, the Weyl group is usually defined to be $N_K(\mathfrak{a})/Z_K(\mathfrak{a})$). Note in particular that, given $x \in \mathfrak{a}$, we have $Wx = N_K(\mathfrak{a})x \subset Kx$. Since $\mathrm{Ad}\,(k)$ is an automorphism of $\mathfrak{g}$, $N_K(\mathfrak{a}) = \{k \in K : k\mathfrak{a} = \mathfrak{a}\}$. Thus $W = \{\mathrm{Ad}\,(k)|_{\mathfrak{a}} : k \in K, k\mathfrak{a} = \mathfrak{a}\}$. Obviously $\mathfrak{a} = \mathrm{span}\,\mathfrak{a}_+$. A theorem of Cartan asserts that $\mathrm{Ad}\,(K)x \cap \mathfrak{a} \neq \phi$ [18, p.320] for any $x \in \mathfrak{p}$, that is, (A1) is satisfied for $(\mathfrak{p}, \mathrm{Ad}\,(K), \mathfrak{a}_+)$. Indeed $|\mathrm{Ad}\,(K)x \cap \mathfrak{a}_+| = 1$. For verification of (A2), see [23].

Example 1.3. (real semisimple Lie algebras) $(\mathfrak{p}, \mathrm{Ad}\,(K), \mathfrak{a}_+)$ is an Eaton triple with a reduced triple $(\mathfrak{a}, W, \mathfrak{a}_+)$. It is similar for real reductive Lie algebras [11].

Example 1.4. (compact connected Lie groups) Let G be a (real) compact connected Lie group and let $(\cdot,\cdot)$ be a bi-invariant inner product on $\mathfrak{g}$. Now $\mathrm{Ad}\,(G)$ is a subgroup of the orthogonal group on $\mathfrak{g}$ [18, p.196]. Let $\mathfrak{t}_+$ be a fixed (closed) fundamental chamber of the Lie algebra $\mathfrak{t}$ of a maximal torus T of G. Now $(\mathfrak{g}, \mathrm{Ad}\,(G), \mathfrak{t}_+)$ is an Eaton triple with reduced triple $(\mathfrak{t}, W, \mathfrak{t}_+)$, where the Weyl group W of G is often defined as $N(T)/T$, where $N(T)$ is the normalizer of T in G [18, p.201].

2. Some basics of nonsmooth analysis. Let Y be a subset of V which is a finite dimensional real inner product space. A function $f : Y \to \mathbb{R}$ is said to be *Lipschitz* [6, p.25] on Y with Lipschitz constant K if for some $K \geq 0$,

$$|f(y) - f(y')| \leq K\|y - y'\|, \qquad y, y' \in Y, \tag{1}$$

where the norm is induced by the inner product. We say that f is Lipschitz near x if for some $\epsilon > 0$, f satisfies the Lipschitz condition (1) on the set $x + \epsilon B$, where B is the open unit ball.

Let f be Lipschitz near a given $x \in V$ and let $0 \neq v \in V$. The *Clarke directional derivative* [6, p.25] of f at x in the direction v is defined as

$$f^o(x; v) = \limsup_{y \to x \ t \downarrow 0} \frac{f(y + tv) - f(y)}{t}, \tag{2}$$

where $y \in V$ and $t > 0$. The *Clarke generalized gradient* of f at x, denoted by $\partial f(x)$, is defined as

$$\partial f(x) := \{\xi \in V : f^o(x; v) \geq (\xi, v) \text{ for all } v \in V\}. \tag{3}$$

We remark that the definition of $\partial f(x)$ in [6, p.27] is given as a subset of V^*, the dual space of V. By Riesz's representation theorem for V, linear functionals on V are uniquely represented by vectors in V. It is known that $\partial f(x)$ is the convex hull of the set of cluster points of gradients of f at points near x in a set of full Lebesque measure [6, Theorem 2.5.1], that is,

$$\partial f(x) = \mathrm{conv}\,\{\lim_{n\to\infty} \nabla f(x_n) : \ x_n \to x, \ x_n \notin S \cup \Omega_f\}, \tag{4}$$

where S is any fixed set of Lebesque measure 0 in V, Ω_f denotes the set of points at which f fails to be differentiable, and 'conv ' denotes the convex hull of the underlying set. By Rademacher's theorem [6, p.63] Ω_f is of measure zero if f is local Lipschitz. When f is smooth, then $\partial f(x)$ coincides with the usual gradient $\nabla f(x)$, that is, $\partial f(x) = \{\nabla f(x)\}$. Thus the following is a generalization of the classical mean value theorem [6, Theorem 2.3.7].

THEOREM 2.1. *(Lebourg mean value theorem) Let $x, y \in V$ and suppose that f is Lipschitz on an open set containing the closed line segment $\{tx+(1-t)y : 0 \le t \le 1\}$. Then there exists u in the open line segment $\{tx + (1-t)y : 0 < t < 1\}$ such that*

$$f(y) - f(x) \in (\partial f(u), y - x). \tag{5}$$

Suppose that $\varphi : V \to \mathbb{R}$ is a convex function. A vector x^* is said to be a *subgradient* of φ at a point x if

$$\varphi(z) \ge \varphi(x) + (x^*, z - x), \qquad \text{for all } z \in V.$$

The set of subgradients of φ at x is called the *subdifferential* of φ at x and is denoted by $\partial_s\varphi(x)$. It turns out that [29, Theorem 25.1] φ is differentiable at x if and only if $\partial_s\varphi(x)$ is a singleton set. In this event $\partial_s\varphi(x) = \{\nabla\varphi(x)\}$.

3. Derivatives of orbital functions. Throughout this section (V, G, F) is an Eaton triple with reduced triple (W, H, F). By [27, Theorem 3.2], H is a finite reflection group and F is one of the (closed) chambers. Let $W_0 := \{x \in W : hx = x \text{ for all } h \in H\}$ be the set of fixed points in W under the action of H and let $W_1 := W_0^\perp$. Let $\Delta := \{\alpha_1, \ldots, \alpha_n\}$ be a simple system of H such that $F = \{x \in W : (x, \alpha_i) \ge 0, i = 1, \ldots, n\}$. Let $\{\lambda_1, \ldots, \lambda_n\}$ be the basis of W_1 dual to $\{\beta_i := 2\alpha_i/(\alpha_i, \alpha_i) : i = 1, \ldots, n\}$. Thus $F = \{\sum_{i=1}^n c_i\lambda_i : c_i \ge 0\} \oplus W_0$.

The map $F : V \to F$ such that $x \mapsto F(x)$ is positively homogeneous, that is, $F(rv) = rF(v)$ for $r \ge 0$ by using (A1) and (A2). But generally $F(rv) \ne rF(v)$ for $r < 0$.

A subset $U \in W$ is said to be H-invariant if $hU \subset U$ for all $h \in H$. A function f on U is said to be H-invariant if $f(hx) = f(x)$ for all $h \in H$ whenever $x \in U$. Similarly we can define G-invariant sets and functions. In other words, a H-invariant (G-invariant) function is constant on each orbit Hz (Gz) of $z \in W$ ($z \in V$). Thus we call it an *orbital* function.

The results in this section generalize the corresponding indicated results in [21, 28].

LEMMA 3.1. (Compare [28, Lemma 3.2]) *Given $\alpha_m \in \Delta$. If $\mu \in F$ such that $(\mu, \alpha_m) \ne 0$, then*

$$\max\{(\mu, x) : x \in \mathit{conv}\, H\lambda_m\} = (\mu, \lambda_m)$$

and

$$\arg\max\{(\mu, x) : x \in \mathit{conv}\, H\lambda_m\} = \{\lambda_m\}.$$

Proof. Notice that $\max\{(\mu, x) : x \in \operatorname{conv} H\lambda_m\} = \max\{(\mu, x) : x \in H\lambda_m\} = (\mu, \lambda_m)$ by (A2) since $\mu, \lambda_m \in F$. By the definition of F, $(\mu, \alpha_m) > 0$ since $(\mu, \alpha_m) \neq 0$, $\mu \in F$, and $\alpha_m \in \Delta$. It is clear that $\lambda_m \in \arg\max\{(\mu, u) : u \in \operatorname{conv} H\lambda_m\}$. Let $x \in \arg\max\{(\mu, u) : u \in \operatorname{conv} H\lambda_m\} \subset W$. Rewrite

$$x = \sum_{i=1}^{n} \frac{2(x, \lambda_i)}{(\alpha_i, \alpha_i)} \alpha_i + \pi_0(x), \qquad \mu = \sum_{i=1}^{n} \frac{2(\mu, \alpha_i)}{(\alpha_i, \alpha_i)} \lambda_i + \pi_0(\mu),$$

where $\pi_0 : W \to W_0$ is the orthogonal projection. So

$$(\mu, x) = \sum_{i=1}^{n} \frac{4(\mu, \alpha_i)}{(\alpha_i, \alpha_i)^2} (x, \lambda_i) + (\pi_0(\mu), \pi_0(x)),$$

and similarly

$$(\mu, \lambda_m) = \sum_{i=1}^{n} \frac{4(\mu, \alpha_i)}{(\alpha_i, \alpha_i)^2} (\lambda_m, \lambda_i) + (\pi_0(\mu), \pi_0(\lambda_m)).$$

Notice that for any $y \in \operatorname{conv} Hz$, $\pi_0(y) = \pi_0(z)$ if $y, z \in W$ (the same conclusion holds for $y, z \in V$ and $y \in \operatorname{conv} Gz$). It is because if $z = z_1 + \pi_0(z)$, where $z_1 \in W_1$ and $y = \sum_{h \in H} a_h hz$, where $a_h \geq 0$, for all $h \in H$ and $\sum_{h \in H} a_h = 1$, then $y = \sum_{h \in H} a_h hz_1 + \pi_0(z)$ and $\sum_{h \in H} a_h hz_1 \in W_1$. Hence $\pi_0(z) = \pi_0(y)$. So $\pi_0(x) = \pi_0(\lambda_m)$ and thus $(\mu, x) = (\mu, \lambda_m)$ implies

$$\sum_{i=1}^{n} (\mu, \alpha_i)(x, \lambda_i) = \sum_{i=1}^{n} (\mu, \alpha_i)(\lambda_m, \lambda_i).$$

By (A2) again, since $\lambda_i \in F$, $i = 1, \ldots, n$, and $x \in \operatorname{conv} H\lambda_m$, we have $(x, \lambda_i) \leq (\lambda_m, \lambda_i)$, $i = 1, \ldots, n$, since $(\mu, \alpha_i) \geq 0$ as $\mu \in F$ for all i. Thus $(\mu, \alpha_i) \neq 0$ implies

$$(x, \lambda_m) = (\lambda_m, \lambda_m).$$

Write $x = \sum_{i=1}^{k} a_i h_i \lambda_m$, where $\sum_{i=1}^{k} a_i = 1$, $a_i > 0$, and $h_i \in H$ for all $i = 1, \ldots, k$. Thus by (A2)

$$(\lambda_m, \lambda_m) = (x, \lambda_m) = \sum_{i=1}^{k} a_i (h_i \lambda_m, \lambda_m) \leq \sum_{i=1}^{k} a_i (\lambda_m, \lambda_m) = (\lambda_m, \lambda_m).$$

So $(h_i \lambda_m, \lambda_m) = (\lambda_m, \lambda_m)$ for all $i = 1, \ldots, k$. Since $\|h_i \lambda_m\| = \|\lambda_m\|$, it follows that $h_i \lambda_m = \lambda_m$ for all $i = 1, \ldots, k$ by the equality case of Cauchy-Schwarz's inequality $(h_i \lambda_m, \lambda_m) \leq \|h_i \lambda_m\| \|\lambda_m\| = (\lambda_m, \lambda)$ and that h_i is orthogonal. Hence we have the desired $x = \lambda_m$. ☐

THEOREM 3.2. (Compare [21, Theorem 2.1]; also see [28, Lemma 3.3], [13, Corollary 3.10] and [20])) *Let $\lambda \in F$. The function $f_\lambda : V \to \mathbb{R}$ defined by $f_\lambda(z) = (\lambda, F(z))$ is positively homogeneous and convex. Let $\mu \in F$ such that $(\mu, \alpha_m) \neq 0$*

for some $\alpha_m \in \Delta$*, then* f_{λ_m} *is differentiable at* μ *and* $df_{\lambda_m}|_\mu = (\lambda_m, \cdot)$*, that is,* $\nabla f_{\lambda_m}(\mu) = \lambda_m$*.*

Proof. By (A2), if $\lambda \in F$,

$$f_\lambda(z) = \max\{(\lambda, gz) : g \in G\} = \max\{(g\lambda, z) : g \in G\} = \max\{(\xi, z) : \xi \in \operatorname{conv} G\lambda\}.$$

In other words, f_λ is the support function for the compact convex set $\operatorname{conv} G\lambda$ and is therefore positively homogeneous and convex [29, Theorem 13.2]. The subdifferential of the support function f_λ at the point z, denoted by $\partial_s f_\lambda(z)$, consists of the elements of $\operatorname{conv} G\lambda$ attaining the maximum $f_\lambda(z) = (\lambda, F(z))$ [29, Corollary 23.5.3], that is,

$$\partial_s f_\lambda(z) = \arg\max\{(\xi, F(z)) : \xi \in \operatorname{conv} G\lambda\}.$$

Certainly $\lambda \in \operatorname{conv} G\lambda$ and $(\lambda, \mu) = f_\lambda(\mu)$ for any $\mu \in F$ and thus $\lambda \in \partial_s f_\lambda(\mu)$.

Suppose that $\mu \in F$ such that $(\mu, \alpha_m) \neq 0$ for some $\alpha_m \in \Delta$. Let $z \in \partial_s f_{\lambda_m}(\mu) = \arg\max\{(\xi, \mu) : \xi \in \operatorname{conv} G\lambda_m\}$, that is, $(z, \mu) = (\lambda_m, \mu)$ and $z \in \operatorname{conv} G\lambda_m$. Now if $\pi : V \to W$ is the orthogonal projection,

$$(\pi(z), \mu) = (z, \mu) = (\lambda_m, \mu),$$

and by [27, Theorem 3.2], $\pi(z) \in \operatorname{conv} H\lambda_m$. By Lemma 3.1, we have $\pi(z) = \lambda_m$ so that $z = \lambda_m + y$ where $y \in W^\perp$. So

$$\|z\|^2 = \|\lambda_m\|^2 + \|y\|^2.$$

On the other hand, $z \in \operatorname{conv} G\lambda_m$ means $z = \sum_{i=1}^k a_i g_i \lambda_m$, where $\sum_{i=1}^k a_i = 1$, $a_i > 0$, and $g_i \in G$, for all $i = 1, \dots, k$, which implies that

$$\|z\| = \|\sum_{i=1}^k a_i g_i \lambda_m\| \le \sum_{i=1}^k a_i \|g_i \lambda_m\| = \|\lambda_m\|.$$

Thus $y = 0$ and $z = \lambda_m$. Hence $\partial_s f_{\lambda_m}(\mu) = \{\lambda_m\}$ and by [29, Theorem 25.1], the desired result follows. □

EXAMPLE 3.3. The general linear group $GL_n(\mathbb{F})$ is consists of $n \times n$ matrices with nonzero determinant. The Lie algebra is $\mathfrak{gl}_n(\mathbb{F})$, that is, $n \times n$ matrices with elements in $\mathbb{F}$, which is reductive. The Cartan decomposition of $\mathfrak{gl}_n(\mathbb{F})$ is $\mathfrak{gl}_n(\mathbb{F}) = \mathfrak{k} + \mathfrak{p}$ where $\mathfrak{p}$ is the space of real symmetric, Hermitian and quaternionic Hermitian matrices (that is, $A = A^*$ where $A^* = \overline{A}^T$ and $\overline{a_1 + ia_2 + ja_3 + ka_4} = a_1 - ia_2 - ja_3 - ka_4$) when $\mathbb{F} = \mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$ respectively. The group K is $U_n(\mathbb{F})$ and $\mathfrak{k}$ is the algebra of real skew symmetric, skew Hermitian and quaternionic skew Hermitian matrices accordingly; $\mathfrak{a} \subset \mathfrak{p}$ is the subset of real diagonal matrices which will be identified with $\mathbb{R}^n$; $\mathfrak{a}_+ \subset \mathfrak{a}$ can be chosen as the subset of real diagonal matrices with decreasing diagonal entries. Then $F(x) = a_+(x)$ is indeed the vector of eigenvalues of the matrix $x \in \mathfrak{p}$ in descending order. So $(V, G, F) = (\mathfrak{p}, \operatorname{Ad}(U(n)), \mathfrak{a}_+)$ and $(W, H, F) = (\mathfrak{a}, S_n, \mathfrak{a}_+)$ where S_n is the symmetric group of degree n, known as the Weyl group of A_{n-1} type.

Notice that $\mathfrak{a}_0$, the set of fixed points in $\mathfrak{a}$ is the span of e where $e = (1, 1, \ldots, 1)$. The simple roots [15, p.41] of $\mathfrak{a}_1 := \mathfrak{a}_0^\perp$ are

$$\alpha_i = e_i - e_{i+1}, \qquad i = 1, \ldots, n-1,$$

where $\{e_i\}$ is the standard basis of $\mathbb{R}^n$. The corresponding λ_i are

$$\lambda_i = \sum_{k=1}^{i} e_k, \qquad i = 1, \ldots, n-1.$$

Thus $f_{\lambda_m}(z)$ is the sum of the largest m eigenvalues of the matrix $z \in \mathfrak{p}$. So the later part of Theorem 3.2 asserts that if $\mu_1 \geq \cdots \geq \mu_n$ with $\mu_m > \mu_{m+1}$ $(1 \leq m < n)$, Then f_{λ_m} is differentiable at μ and $df_{\lambda_m}(\mu) = (\lambda_m, \cdot) = (\sum_{i=1}^m e_i, \cdot)$ which is exactly the statement of [21, Theorem 2.1].

Example 3.4. Let us consider the a real form of the simple complex Lie algebra $\mathfrak{sl}_n(\mathbb{C})$, namely, $\mathfrak{su}_{p,q}$ $(p+q=n)$, which corresponds to the case of $p \times q$ complex matrices. It is known that

$$\begin{aligned}
\mathfrak{su}_{p,q} &= \{\begin{pmatrix} X_1 & Y \\ Y^* & X_2 \end{pmatrix} : X_1^* = -X_1,\ X_2^* = -X_2,\ \operatorname{tr} X_1 = \operatorname{tr} X_2 = 0,\ Y \in \mathbb{C}_{p\times q}\},\\
K = SU(p,q) &= \{\begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} : U \in U(p), V \in U(q),\ \det U \det V = 1\},\\
\mathfrak{k} &= \mathfrak{su}_{p,q},\ \text{i.e., } Y = 0,\\
\mathfrak{p} &= \{\begin{pmatrix} 0 & Y \\ Y^* & 0 \end{pmatrix} : Y \in \mathbb{C}_{p\times q}\},\\
\mathfrak{a} &= \oplus_{1\leq j\leq p} \mathbb{R}(E_{j,p+j} + E_{p+j,j}),\\
\mathfrak{a}_+ &= \{\oplus_{1\leq j\leq p} a_j(E_{j,p+j} + E_{p+j,j}) : a_1 \geq \cdots \geq a_p \geq 0\}.
\end{aligned}$$

Now the orbit of an element in $\mathfrak{p}$ under the adjoint action of K is

$$\begin{aligned}
&\{\begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}\begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}\begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}^* : U \in U(p), V \in U(q),\ \det U \det V = 1\}\\
&= \{\begin{pmatrix} 0 & UAV^* \\ VA^*U^* & 0 \end{pmatrix} : U \in U(p), V \in U(q),\ \det U \det V = 1\}\\
&= \{\begin{pmatrix} 0 & UAV \\ (UAV)^* & 0 \end{pmatrix} : U \in U(p), V \in U(q)\}.
\end{aligned}$$

The eigenvalues of the matrix

$$\begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} \tag{6}$$

are $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_{\min\{p,q\}} \geq 0 \geq \cdots \geq 0 \geq -\alpha_{\min\{p,q\}} \geq \cdots \geq -\alpha_2 \geq -\alpha_1$ where α's are the singular values of A and there are $p+q-2\min\{p,q\}$ zeros. We may identify $\mathfrak{p}$ with the space of $p\times q$ complex matrices and $\mathfrak{a}$ with $\mathbb{R}^r$ where $r = \min\{p,q\}$. Now we consider the special case $p = q$. The simple roots [15, p.42] are

$$\alpha_i = e_i - e_{i+1}, \qquad i = 1, \ldots, p-1, \qquad \alpha_p = e_p,$$

and the corresponding λ_i are

$$\lambda_i = \sum_{k=1}^{i} e_k, \qquad i = 1, \dots, p-1. \qquad \lambda_p = \frac{1}{2}\sum_{k=1}^{p} e_k.$$

Thus $f_{\lambda_m}(z)$ is the sum of the largest m singular values of the $p \times p$ complex matrix z, that is, Ky Fan's m-norm [4] when $1 \le m \le p-1$ and f_{λ_p} is just one half of Ky Fan's p-norm. So the later part of Theorem 3.2 asserts that if

1. $m = 1, \dots, p-1$ and if $\mu_1 \ge \cdots \ge \mu_p \ge 0$ with $\mu_m > \mu_{m+1}$, then f_{λ_m} is differentiable at μ and $df_{\lambda_m}(\mu) = (\lambda_m, \cdot) = (\sum_{k=1}^{m} e_k, \cdot)$;
2. $m = p$ and $\mu_p > 0$, then f_{λ_p} is differentiable at μ and $df_{\lambda_p}(\mu) = (\lambda_p, \cdot) = (1/2 \sum_{k=1}^{p} e_k, \cdot)$.

Given two vectors β and $\mu \in W$, we say that μ *refines* β if $(\alpha, \beta) = 0$ whenever $(\alpha, \mu) = 0$, $\alpha \in \Delta$, or equivalently, $s_\alpha \beta = \beta$ whenever $s_\alpha \mu = \mu$.

LEMMA 3.5. (Compare [21, Lemma 2.2]) *If $\mu \in F$ refines $\beta \in W$, then the function $f_\beta : V \to \mathbb{R}$ defined by $f_\beta(z) = (\beta, F(z))$ is differentiable at μ with*

$$df_\beta|_\mu = (\beta, \cdot),$$

that is, $\nabla f_\beta(\mu) = \beta$.

Proof. Rewrite $F(z) = \sum_{i=1}^{n} 2(\lambda_i, F(z))/(\alpha_i, \alpha_i)\, \alpha_i + \pi_0(z)$ since $\pi(F(z)) = \pi(z)$ and $\beta = \sum_{i=1}^{n} 2(\alpha_i, \beta)/(\alpha_i, \alpha_i)\, \lambda_i + \pi_0(\beta)$, where $\pi_0 : V \to V_0$ is the orthogonal projection. So

$$\begin{aligned} f_\beta(z) = (\beta, F(z)) &= \sum_{i=1}^{n} \frac{2(\alpha_i, \beta)}{(\alpha_i, \alpha_i)} (\lambda_i, F(z)) + (\pi_0(\beta), \pi_0(z)) \\ &= \sum_{i=1}^{n} \frac{2(\alpha_i, \beta)}{(\alpha_i, \alpha_i)} (\lambda_i, F(z)) + (\pi_0(\beta), z). \end{aligned}$$

Then

$$\begin{aligned} df_\beta|_\mu &= \sum_{i=1}^{n} \frac{2(\alpha_i, \beta)}{(\alpha_i, \alpha_i)} df_{\lambda_i}|_{F(\mu)} + (\pi_0(\beta), \cdot) \\ &= \sum_{i=1}^{n} \frac{2(\alpha_i, \beta)}{(\alpha_i, \alpha_i)} df_{\lambda_i}|_\mu + (\pi_0(\beta), \cdot) \qquad \text{since } \mu \in F \\ &= \sum_{i=1}^{n} \frac{2(\alpha_i, \beta)}{(\alpha_i, \alpha_i)} (\lambda_i, \cdot) + (\pi_0(\beta), \cdot) \qquad \text{by Theorem 3.2 and since } \mu \text{ refines } \beta \\ &= (\beta, \cdot). \end{aligned}$$

□

LEMMA 3.6. (Compare [21, Lemma 2.3]) *Let $U \subset W$ be open and H-invariant. Suppose that the function $f : U \to \mathbb{R}$ is H-invariant and differentiable at $\mu \in F$. Then μ refines $\nabla f(\mu)$ and thus the function $df|_\mu \circ F$ is differentiable at μ with $d(df|_\mu \circ F)|_\mu = (\nabla f(\mu), \cdot)$, that is, $\nabla (df|_\mu \circ F)(\mu) = \nabla f(\mu)$.*

Proof. Let $\alpha \in \Delta$ such that $(\alpha, \mu) = 0$, that is, $s_\alpha \mu = \mu$. Notice that $f(z) = f(s_\alpha z)$ for all $z \in W$ since f is H-invariant. Apply chain rule at $z = \mu$ to have

$$df|_\mu = df|_{s_\alpha \mu} \circ ds_\alpha|_\mu = df|_\mu \circ s_\alpha,$$

since s_α is linear. So $\nabla f(\mu) = s_\alpha \nabla f(\mu)$, that is, μ refines $\nabla f(\mu)$. Since $df|_\mu \circ F = (\nabla f(\mu), F(\cdot))$ and $\nabla f(\mu) \in W$, by Lemma 3.5, the function $df|_\mu \circ F$ is differentiable at μ and $d(df|_\mu \circ F)|_\mu = (\nabla f(\mu), \cdot)$. ∎

LEMMA 3.7. (Compare [21, Theorem 2.4]) *Let $U \subset W$ be open and H-invariant. Suppose that the function $f : U \to \mathbb{R}$ is H-invariant and differentiable at $\mu \in F$, then $f \circ F : V \to \mathbb{R}$ is differentiable and*

$$d(f \circ F)|_\mu = (\nabla f(\mu), \cdot),$$

that is, $\nabla(f \circ F)(\mu) = \nabla f(\mu)$.

Proof. Given any $\epsilon > 0$, since f is differentiable at $\mu \in F$,

$$|f(\gamma) - f(\mu) - df|_\mu(\gamma - \mu)\| \le \epsilon \|\gamma - \mu\|,$$

whenever $\gamma \in W$ is close to μ (the norm is induced by the inner product). It is not hard to see that F is Lipschitz on V with Lipschitz constant 1 because of (A2). Thus for small $y \in V$,

$$|f(F(y + \mu)) - f(\mu) - df|_\mu(F(y + \mu) - \mu)| \le \epsilon \|F(y + \mu) - F(\mu)\| \le \epsilon \|y\|.$$

By Lemma 3.6, for small $y \in V$,

$$\begin{aligned} &|df|_\mu \circ F(y + \mu) - df|_\mu(\mu) - (\nabla f(\mu), y)| \\ = &|df|_\mu \circ F(y + \mu) - df|_\mu \circ F(\mu) - d(f \circ F)|_\mu(y)| \le \epsilon \|y\|. \end{aligned}$$

Adding the two previous inequalities and using triangle inequality, we have

$$|f \circ F(y + \mu) - f(\mu) - (\nabla f(\mu), y)| \le 2\epsilon \|y\|$$

for small y and thus the desired result. ∎

THEOREM 3.8. (Compare [21, Theorem 1.1]) *Let $U \subset W$ be open and H-invariant. Suppose that the function $f : U \to \mathbb{R}$ is H-invariant. Then the function $f \circ F : V \to \mathbb{R}$ is differentiable at $x \in V$ if and only if f is differentiable at $F(x) \in U$. In this case*

$$d(f \circ F)|_x = (g^{-1} \nabla f(F(x)), \cdot),$$

for any $g \in G$ satisfying $gx = F(x)$, that is, $\nabla(f \circ F)(x) = g^{-1} \nabla f(F(x))$.

Proof. It is easy to see that f must be differentiable at $F(x)$ whenever $f \circ F$ is differentiable at x, since we can write $f(y) = (f \circ F)(g^{-1} y)$ with $gx = F(x)$ and apply chain rule at $y = F(x)$, that is,

$$df|_{F(x)} = d(f \circ F)|_x \circ dg^{-1}|_{F(x)} = d(f \circ F)|_x \circ g^{-1}.$$

On the other hand, suppose that f is differentiable at $F(x)$, and let $g \in G$ such that $gx = F(x)$. Now for all $z \in V$, since F is G-invariant,

$$(f \circ F)(z) = (f \circ F)(gz).$$

Applying chain rule at $z = x$ and Lemma 3.7 yields

$$d(f \circ F)|_x = d(f \circ F)|_{gx} \circ g = d(f \circ F)|_{F(x)} \circ g = (\nabla f(F(x)), g(\cdot)) = (g^{-1}\nabla f(F(x)), \cdot),$$

that is, $\nabla(f \circ F)|_x = g^{-1}\nabla f(F(x))$. ∎

The following is an extension of Lemma 3.7.

THEOREM 3.9. (Compare [21, Corollary 2.5]) *Let $U \subset W$ be open and H-invariant. Suppose that the function $f : U \to \mathbb{R}$ is H-invariant and differentiable at $\mu \in U \subset W$. Then $f \circ F : V \to \mathbb{R}$ is differentiable and*

$$d(f \circ F)|_\mu = (\nabla f(\mu), \cdot),$$

that is, $\nabla(f \circ F)(\mu) = \nabla f(\mu)$.

Proof. Let $\mu \in W$ and let $h \in H$ such that $h\mu = F(\mu)$. Since f is H-invariant, $f(h\xi) = f(\xi)$, $\xi \in U$. Applying chain rule at $\xi = \mu$ gives $df|_\mu = df|_{F(\mu)} \circ h$, that is,

$$\nabla f(\mu) = h^{-1}\nabla f(F(\mu)).$$

By Theorem 3.8

$$d(f \circ F)|_\mu = (h^{-1}\nabla f(F(\mu)), \cdot) = (\nabla f(\mu), \cdot).$$

∎

Given $\gamma \in V$, the stablizer of γ in G is the subgroup $G_\gamma = \{k \in G : k\gamma = \gamma\} \subset G$.

THEOREM 3.10. (Compare [21, Theorem 3.3]) *Let $U \subset W$ be open and H-invariant. Suppose that the function $f : U \to \mathbb{R}$ is H-invariant and locally Lipschitz around $\mu \in F$. Then*

$$(f \circ F)^o(\mu; z) = \max\{f^o(\mu; \pi(kz)) : k \in G_\mu\}. \tag{7}$$

Proof. Since V is finite dimensional, we have [6, p.64]

$$(f \circ F)^o(\mu; z) = \limsup_{y \to \mu}\{(\nabla(f \circ F)(y), z) : y \notin S \cup \Omega_{f\circ F}\},$$

where $S \subset V$ is any given set of measure zero and $\Omega_{f\circ F}$ is the set of points at which $f \circ F$ is not differentiable. So there exists a sequence $\{x_n\}$ in $V \setminus (S \cup \Omega_{f\circ F})$ such that $\{x_n\} \to \mu$ (and $F(x_n) \to \mu$ since $v \mapsto F(v)$ is Lipschitz and thus continuous) with

$$(\nabla(f \circ F)(x_n), z) \to (f \circ F)^o(\mu, z).$$

Choose a $g_n \in G$ such that $g_n x_n = F(x_n)$ for each $n = 1, 2, \ldots$. Since G is compact, there is a subsequence $\{g_{n_r}\}$ for which $g_{n_r} \to g_0 \in G$ as $r \to \infty$. Now

$$g_0^{-1}\mu = \lim_{r\to\infty} g_{n_r}^{-1} F(x_{n_r}) = \lim_{r\to\infty} x_{n_r} = \mu,$$

so that $g_0 \in G_\mu$. Hence

$$\begin{aligned}
(f\circ F)^o(\mu; z) &= \lim_{n\to\infty} (\nabla(f\circ F)(x_n), z) \\
&= \lim_{n\to\infty} (g_n^{-1}\nabla f(F(x_n)), z) \qquad \text{by Theorem 3.8} \\
&= \lim_{r\to\infty} (\nabla f(F(x_{n_r})), g_{n_r} z) \\
&= \lim_{r\to\infty} (\nabla f(F(x_{n_r})), \pi(g_{n_r} z)) \qquad \text{since } \nabla f(F(x_{n_r})) \in W \\
&\le \limsup_{\gamma\to\mu} (\nabla f(\gamma)), \pi(g_0 z)) \\
&= f^o(\mu; \pi(g_0 z)),
\end{aligned}$$

where $\pi : V \to W$ denotes the orthogonal projection. Thus we establish '$\le$' in (7).

On the other hand, we have [6, p.64] a sequence $\{\mu_n\} \subset W$ such that $\mu_n \to \mu$ and for all $k \in G_\mu$,

$$\begin{aligned}
f^o(\mu; \pi(kz)) &= \lim_{n\to\infty} (\nabla f(\mu_n), \pi(kz)) \\
&= \lim_{n\to\infty} (\nabla f(\mu_n), kz) \qquad \text{since } \nabla f(\mu_n) \in W \\
&= \lim_{n\to\infty} (\nabla (f\circ F)(\mu_n), kz) \qquad \text{by Theorem 3.9} \\
&\le \limsup_{\gamma\to\mu} (\nabla(f\circ F)(\gamma), kz) \\
&= (f\circ F)^o(\mu; kz).
\end{aligned}$$

Now for any $g \in G$,

$$\begin{aligned}
(f\circ F)^o(g\mu; gz) &= \limsup_{w\to g\mu,\ t\downarrow 0} \frac{f(F(w+tgz)) - f(F(w))}{t} \\
&= \limsup_{y\to\mu,\ t\downarrow 0} \frac{f(F(g(y+tz))) - f(F(gy))}{t} \\
&= \limsup_{y\to\mu,\ t\downarrow 0} \frac{f(F(y+tz)) - f(F(y))}{t} \\
&= (f\circ F)^o(\mu; z).
\end{aligned}$$

Since $k \in G_\mu \subset G$,

$$(f\circ F)^o(\mu; kz) = (f\circ F)^o(k^{-1}\mu; z) = (f\circ F)^o(\mu; z).$$

Hence $f^o(\mu; \pi(kz)) \le (f\circ F)^o(\mu; z)$ for all $k \in G_\mu$. Thus the desired result follows. □

LEMMA 3.11. (Compare [21, Corollary 3.6]) *Let $U \subset W$ be open and H-invariant. Suppose that the function $f : U \to \mathbb{R}$ is H-invariant and locally Lipschitz around $\mu \in F$. Then*

$$\partial(f\circ F)(\mu) = conv\{k\gamma : k \in G_\mu,\ \gamma \in \partial f(\mu)\}.$$

Proof. By (3) or (4) $\partial f(\mu)$ is a compact set in W. Since $G_\mu \subset G$ is a closed subgroup and thus compact, and since the map $(\gamma, k) \mapsto k\gamma$ is continuous, the set

$$D := \{k\gamma : k \in G_\mu,\ \gamma \in \partial f(\mu)\}$$

is compact. So $\operatorname{conv} D$ is a compact convex set. It suffices to show that the support functions of $\operatorname{conv} D$ and of the compact convex set $\partial(f \circ F)(\mu)$ are identical. The support function of $\operatorname{conv} D$, evaluated at the $z \in V$, is

$$\begin{aligned}
&\max\{(z, y) : y \in \operatorname{conv} D\}\\
&= \max\{(z, y) : y \in D\}\\
&= \max\{(z, k\gamma) : k \in G_\mu, \gamma \in \partial f(\mu)\}\\
&= \max\{(kz, \gamma) : k \in G_\mu, \gamma \in \partial f(\mu)\} \qquad \text{since } G_\mu \text{ is a group}\\
&= \max\{(\pi(kz), \gamma) : k \in G_\mu, \gamma \in \partial f(\mu)\} \qquad \text{by } \partial f(\mu) \subset W\\
&= \max\{\max\{(\pi(kz), \gamma) : \gamma \in \partial f(\mu)\} : k \in G_\mu\},
\end{aligned}$$

where $\pi : V \to W$ is the orthogonal projection. By (3) the support function of $\partial(f \circ F)(\mu)$, evaluated at $z \in V$ is the Clarke directional derivative $(f \circ F)^o(\mu; z)$, by Theorem 3.10

$$(f \circ F)^o(\mu; z) = \max\{f^o(\mu; \pi(kz)) : k \in G_\mu\}.$$

Clearly $f^o(\mu; \pi(kz))$ is the support function of ∂f, evaluated at $\pi(kz)$, which is $\max\{(\pi(kz), \gamma) : \gamma \in \partial f(\mu)\}$. ∎

Remark 3.12. In [21, Theorem 3.12] the set $D := \{k\gamma : k \in G_\mu,\ \gamma \in \partial f(\mu)\}$ is proved to be convex for the reductive Lie algebras, $\mathfrak{gl}_n(\mathbb{R})$ and $\mathfrak{gl}_n(\mathbb{C})$ by some argument involving doubly stochastic matrices. Using the fact that D is convex for those two cases, [21, Theorem 1.4] is deduced and is used in [23] to give a new proof of Liskii's theorem which is a special case of Berezin-Gel'fand's theorem. However we are able to bypass that in order to extend Berezin-Gel'fand's theorem, as we will see in the next section. Nevertheless we do not know whether D is convex or not.

4. An extension of Berezin-Gel'fand's theorem. In this section (V, G, F) is an Eaton triple with reduced triple (W, H, F). We will use the notations that we mentioned in Section 1. The following lemma is a slight extension of [33, Theorem 10] (also see [30])). Since the proof is the same, it is omitted.

Lemma 4.1. *Let (V, G, F) be an Eaton triple with a reduced triple (W, H, F). For any $x_1, \ldots, x_k \in V$, $F(\sum_{i=1}^k x_i) \in \operatorname{conv} H(\sum_{i=1}^k F(x_i))$.*

Theorem 4.2. *Let (V, G, F) be an Eaton triple with a reduced triple (W, H, F). For any $y, z \in V$,*

$$F(y + z) - F(z) \in \operatorname{conv} H(F(y)).$$

In terms of inequalities, it amounts to

$$\max_{h \in H}\ (h(F(y + z) - F(z)), \lambda_i) \leq (F(y), \lambda_i) \qquad \text{for all } \ i = 1, \ldots, n.$$

Proof. Let $f_w : W \to \mathbb{R}$ be defined by $f_w(u) = (F(u), w)$, where $w \in W$. It is (globally) Lipschitz on W with Lipschitz constant $\|w\|$ since for any $y, y' \in W$,

$$|f_w(y) - f_w(y')| = |(w, F(y) - F(y')| \leq \|w\| \, \|F(y) - F(y')\| \leq \|w\| \|y - y'\|,$$

where the norm is induced by the inner product. Similarly the function $(f_w \circ F) : V \to \mathbb{R}$ is (globally) Lipschitz on V with Lipschitz constant $\|w\|$. We claim that

$$\partial f_w(u) \subset \operatorname{conv} Hw, \qquad u \in W. \tag{8}$$

The function f_w is differentiable on each (open) chamber. Indeed it is linear on each (open) chamber: Suppose $u \in C \subset W$ where C is an (open) chamber, that is, $\alpha(u) \neq 0$ for all $\alpha \in \Delta$. Then there exists a unique $h_u \in H$ such that $h_u x = F(x)$ for all $x \in C$ because of the simply transitive action of H on the open chambers [15, p.23]. So

$$f_w(x) = (F(x), w) = (h_u x, w) = (x, h_u^{-1} w)$$

for all $x \in C$. Thus f_w behaves linearly in C and clearly $\partial f_w(u) = \{\nabla f_w(u)\} = \{h_u^{-1} w\} \subset \operatorname{conv} Hw$.

On the other hand if $u \in W$ is not regular, that is, u lies in some hyperplane H_α, $\alpha \in \Delta$, then f_w is not differentiable at u and $\Omega_{f_w} = \cup_{\alpha \in \Delta} H_\alpha$ and we choose $S = \Omega_{f_w}$ in (4). By (4), $\partial f_w(u) = \operatorname{conv} H_u^{-1} w$, where $H_u = \{h \in H : hu = F(u)\} \subset H$ is the isotropy group of u in H. So (8) is now established.

Let $x \in V$ and let $g \in G$ such that $g^{-1}x = F(x)$. Given $w \in W$, we consider the composite function $(f_w \circ F) \circ g : V \to \mathbb{R}$ of $f_w \circ F : V \to \mathbb{R}$ and $g : V \to V$. The function $f_w \circ F$ is Lipschitz with Lipschitz constant $\|w\|$ on V and g is an orthongal map. Apply chain rule [6, Theorem 2.3.10] on the composite function at the point $g^{-1}x$ to get

$$\partial(f_w \circ F \circ g)(g^{-1}x) = D_s^* g(g^{-1}x) \partial(f_w \circ F)(x),$$

where $D_s^* g(g^{-1}x)$ is the adjoint of the strict derivative [6, p.30] of g at $g^{-1}x$. Since g is orthogonal, $D_s^* g(g^{-1}x)$ is simply g^{-1}. Hence

$$\partial(f_w \circ F \circ g)(g^{-1}x) = g^{-1} \partial(f_w \circ F)(x)$$

or equivalently,

$$\begin{aligned}
\partial(f_w \circ F)(x) &= g\, \partial(f_w \circ F \circ g)(g^{-1}x) \\
&= g\, \partial(f_w \circ F)(g^{-1}x) \qquad \text{since } F \circ g = F \\
&= g\, \partial(f_w \circ F)(F(x)) \\
&= g \operatorname{conv} \{k\gamma : k \in G_{F(x)},\ \gamma \in \partial f_w(F(x))\} \qquad \text{by Lemma 3.11.}
\end{aligned}$$

By (8) we have

$$\partial(f_w \circ F)(x) \subset g \operatorname{conv} \{k\gamma : k \in G_{F(x)},\ \gamma \in \operatorname{conv} Hw\}, \qquad w \in W. \tag{9}$$

By Lebourg mean value theorem, if $y, z \in V$, there exist $x \in [z, y+z]$ and $v \in \partial(f_w \circ F)(x)$ such that

$$(F(y+z) - F(z), w) = f_w \circ F(y+z) - f_w \circ F(z) = (y, v) \le (F(y), F(v)), \tag{10}$$

for all $w \in W$, where the last inequality follows from (A2). By (9), $v \in \partial(f_w \circ F)(x)$ implies that

$$v = g \sum_{k \in G_{F(x)}} b_k k \,(\sum_{h \in H} a_h^k hw) = \sum_{k \in G_{F(x)},\, h \in H} b_k a_h^k khw, \tag{11}$$

where $a_h^k \ge 0$, $\sum_{h \in H} a_h^k = 1$ for all $k \in G_{F(x)}$, $b_k \ge 0$, $\sum_{k \in G_{F(x)}} b_k = 1$. Since $F(au) = aF(u)$ for all $a \ge 0$, $u \in V$, by Lemma 4.1 and (11),

$$F(v) \in \operatorname{conv} H(\sum_{k \in G_{F(x)},\, h \in H} b_k a_h^k F(khw)) = \operatorname{conv} H(F(w)),$$

that is, $F(v) \in \operatorname{conv} H(F(w))$. Now [33, Lemma 5(2)] states that

$$\text{if } x, y \in F, \text{ then } x \in \operatorname{conv} Hy \text{ if and only if } y - x \in \operatorname{dual}_W F, \tag{12}$$

where $\operatorname{dual}_W F = \{u \in W : (u, x) \ge 0, \text{ for all } x \in F\}$. So $F(w) - F(v) \in \operatorname{dual}_W F$. In particular $(F(y), F(v)) \le (F(y), F(w))$ and thus by (10) we arrive at $(F(y+z) - F(z), w) \le (F(y), F(w))$ for all $w \in W$. This implies

$$(h(F(y+z) - F(z)), x) = (F(y+z) - F(z), h^{-1}x) \le (F(y), x),$$

for all $h \in H$ and $x \in F$. So we conclude

$$F(y) - h(F(y+z) - F(z)) \in \operatorname{dual}_W F, \qquad \text{for all } h \in H.$$

Thus by [33, Lemma 5(1)] $F(y+z) - F(z) \in \operatorname{conv} H(F(y))$.

Now $F(y+z) - F(z) \in \operatorname{conv} H(F(y))$ amounts to $F(y) - h(F(y+z) - F(z)) \in \operatorname{dual}_W F$, for all $h \in H$ by [33, Lemma 5(1)] again, that is,

$$\max_{h \in H} \,(h(F(y+z) - F(z)), \lambda_i) \le (F(y), \lambda_i), \qquad \text{for all } i = 1, \ldots, n.$$

□

Remark 4.3. Lemma 4.1 (when $k = 2$) is now a corollary of Theorem 4.2: by [33, Lemma 5(1)],

$$F(y+z) - F(z) \in \operatorname{conv} HF(y) \Leftrightarrow (F(y+z) - F(z), h^{-1}w) \le (F(y), w),$$

for all $w \in F, h \in H$. So

$$\begin{aligned}(F(y+z), h^{-1}w) &\le (F(y), w) + (F(z), h^{-1}w)\\ &\le (F(y), w) + (F(z), w) \quad \text{by (A2)}\\ &= (F(y) + F(z), w),\end{aligned}$$

for all $w \in F$ and $h \in H$. Thus by [33, Lemma 5(1)]

$$F(y+z) \in \operatorname{conv} H(F(y)+F(z)). \tag{13}$$

We also remark that (13) is symmetric with respect to y and z but Theorem 4.2 is not.

Corollary 4.4. (Wielandt [36], Markus [26]) *Let A and B be $n \times n$ complex matrices. Denote by $s(A) = (s_1(A), \cdots, s_n(A))$ the vector of singular values of A with $s_1(A) \geq \cdots \geq s_n(A) \geq 0$. Then*

$$s(A+B) - s(B) \in conv(S_n \times (\mathbb{Z}/2\mathbb{Z})^n) s(A).$$

In terms of inequalities

$$\max_{1 \leq j_1 < \cdots < j_k \leq n} \sum_{i=1}^{k} |s_{j_i}(A+B) - s_{j_i}(B)| \leq \sum_{i=1}^{k} s_i(A), \quad k = 1, \ldots, n.$$

Proof. Just notice that $a_+(A) = (s_1(A), \ldots, s_n(A))$ under the natural identification where $s_1(A) \geq \cdots \geq s_n(A) \geq 0$ are the singular values of A and $S_n \times (\mathbb{Z}/2\mathbb{Z})^n$ [15, p.42] is the Weyl group for the Example 3.4. □

Let $I_{n,n} = (-I_n) \oplus I_n$. The group $G = SO(n,n)$ is the group of matrices in $SL(2n, \mathbb{R})$ which leaves invariant the quadratic form $-x_1^2 - \cdots - x_n^2 + x_{n+1}^2 + \cdots + x_{2n}^2$. In other words, $SO(n,n) = \{A \in SL_{2n}(\mathbb{R}) : A^T I_{n,n} A = I_{n,n}\}$. It is well known that [18]

$$\begin{aligned}
\mathfrak{so}_{n,n} &= \{\begin{pmatrix} X_1 & Y \\ Y^T & X_2 \end{pmatrix} : X_1^T = -X_1,\ X_2^T = -X_2,\ Y \in \mathbb{R}_{n\times n}\}, \\
K &= SO(n) \times SO(n), \\
\mathfrak{k} &= \mathfrak{so}(n) \oplus \mathfrak{so}(n),\ \text{i.e., } Y = 0, \\
\mathfrak{p} &= \{\begin{pmatrix} 0 & Y \\ Y^T & 0 \end{pmatrix} : Y \in \mathbb{R}_{n\times n}\}, \\
\mathfrak{a} &= \oplus_{1 \leq j \leq n} \mathbb{R}(E_{j,n+j} + E_{n+j,j}),
\end{aligned}$$

where $E_{i,j}$ is the $2n \times 2n$ matrix and 1 at the (i,j) position is the only nonzero entry. The Killing form is

$$B(\begin{pmatrix} 0 & X \\ X^T & 0 \end{pmatrix}, \begin{pmatrix} 0 & Y \\ Y^T & 0 \end{pmatrix}) = 4(n-1) \operatorname{tr} XY^T.$$

Now the adjoint action of K on $\mathfrak{p}$ is given by

$$\begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}^T \begin{pmatrix} 0 & S \\ S^T & 0 \end{pmatrix} \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} = \begin{pmatrix} 0 & U^T S V \\ V^T S^T U & 0 \end{pmatrix},$$

where $U, V \in SO(n)$. We will identify $\mathfrak{p}$ with $\mathbb{R}_{n\times n}$ and thus $\mathfrak{a}$ will then be identified with real diagonal matrices. We may choose $\mathfrak{a}_+ = \{\operatorname{diag}(a_1, \ldots, a_n) : a_1 \geq \cdots \geq$

$a_{n-1} \geq |a_n|\}$. The action of K on $\mathfrak{p}$ is then orthogonal equivalence, that is, $H \mapsto UHV$, where $U, V \in SO(n)$ and $a_+(H) = (s_1(H), \ldots, s_{n-1}(H), [\mathrm{sign}\,(\det H)]\, s_n(H))$, where $s_1(H) \geq \cdots \geq s_n(H)$ are the singular values of H. The action of the Weyl group W on $\mathfrak{a}$ is given by

$$\mathrm{diag}(d_1, \ldots, d_n) \mapsto \mathrm{diag}(\pm d_{\sigma(1)}, \ldots, \pm d_{\sigma(n)}),$$

where $\mathrm{diag}(d_1, \ldots, d_n) \in \mathfrak{a}$, $\sigma \in S_n$ (the symmetric group) and the number of negative signs is even. The simple roots may be taken as $\alpha_i = e_i - e_{i+1}$, $i = 1, \ldots, n-1$ and $\alpha_n = e_{n-1} + e_n$ [15, p.42] and $\lambda_i = e_1 + \ldots + e_i$, $i = 1, \ldots, n-2$, $\lambda_{n-1} = 1/2(e_1 + \ldots + e_{n-1} - e_n)$. $\lambda_n = 1/2(e_1 + \ldots + e_{n-1} + e_n)$. The longest element ω sends $\mathrm{diag}(a_1, \ldots, a_n) \in \mathfrak{a}_+$ to

$$\omega a = \begin{cases} \mathrm{diag}(-a_1, \ldots, -a_{n-1}, a_n) & \text{if } n \text{ is odd} \\ \mathrm{diag}(-a_1, \ldots, -a_n) & \text{if } n \text{ is even.} \end{cases}$$

Applying Theorem 4.2 on the simple Lie algebra $\mathfrak{so}_{n,n}$, we have the following result. Also see [25, 31].

Corollary 4.5. *Let A and B be $n \times n$ real matrices. Denote by $s(A) = (s_1(A), \cdots, s_n(A))$ the vector of singular values of A with $s_1(A) \geq \cdots \geq s_n(A) \geq 0$. If*

$$s'(A) = (s'_1(A), \cdots, s'_n(A)) := (s_1(A), \ldots, s_{n-1}(A), [\mathit{sign}\,\det A]\, s_n(A)),$$

then

$$s'(A+B) - s'(B) \in \mathit{conv}\,(S_n \times (\mathbb{Z}/2\mathbb{Z})^{n-1}) s'(A).$$

In terms of inequalities, if $\#(A,B)$ denotes the number of negative components among $s'(A+B) - s'(A)$ (zero component may be counted either way), then

$$\max_{1 \leq j_1 < \cdots < j_k \leq n} \sum_{i=1}^{k} |s'_{j_i}(A+B) - s'_{j_i}(B)| \leq \sum_{i=1}^{k} s_i(A), \quad k = 1, \ldots, n-2, \tag{14}$$

$$\max_{1 \leq j_1 < \cdots < j_{n-1} \leq n} \sum_{i=1}^{n-1} |s'_{j_i}(A+B) - s'_{j_i}(B)| - (-1)^{\#(A,B)} \min_{1 \leq r \leq n} |s'_r(A+B) - s'_r(B)|$$
$$\leq \sum_{i=1}^{n-1} s_i(A) - [\mathit{sign}\,\det A]\, s_n(A), \tag{15}$$

and

$$\max_{1 \leq j_1 < \cdots < j_{n-1} \leq n} \sum_{i=1}^{n-1} |s'_{j_i}(A+B) - s'_{j_i}(B)| + (-1)^{\#(A,B)} \min_{1 \leq r \leq n} |s'_r(A+B) - s'_r(B)|$$
$$\leq \sum_{i=1}^{n-1} s_i(A) + [\mathit{sign}\,\det A]\, s_n(A). \tag{16}$$

Proof. Notice $\mathfrak{a}_+ = \{\mathrm{diag}(a_1, \dots, a_n) : a_1 \geq \cdots \geq a_{n-1} \geq |a_n| \geq 0\}$. Any real $n \times n$ matrix A is special orthogonally similar to $\mathrm{diag}(a_1, \dots, a_{n-1}, [\mathrm{sign}\,(\det A)]a_n)$ in $\mathfrak{a}_+$, where $a_1 \geq \cdots \geq a_n \geq 0$ are the singular values of A. The Weyl group is $S_n \times (\mathbb{Z}/2\mathbb{Z})^{n-1}$ [15, p.42]. In terms of inequalities By Theorem 4.2 with $\lambda_i = e_1 + \ldots + e_i$, $i = 1, \dots, n-2$, $\lambda_{n-1} = 1/2(e_1 + \ldots + e_{n-1} - e_n)$. $\lambda_n = 1/2(e_1 + \ldots + e_{n-1} + e_n)$, we have the inequalities. □

Remark 4.6. We also have

$$\max_{1 \leq j_1 < \cdots < j_k \leq n} \sum_{i=1}^{k} |s_i(A+B) - s_i(B)| \leq \sum_{i=1}^{k} s_i(A), \quad k = 1, \dots, n-1, n, \tag{17}$$

either by using Corollary 4.4 or by using (15) and (16). That is, adding (15) and (16), we get the second last inequality of (17). Now $\sum_{i=1}^n |s_i(A+B) - s_i(B)|$ is less than or equal to the maximum of the left sides of (15) and (16) and hence not greater than the maximum of the right sides of (15) and (16) which is merely $\sum_{i=1}^n s_i(A)$.

We conclude this section with the following

Remark 4.7. The characterization of the sum of eigenvalues of two Hermitian matrices as well as two real symmetric matrices has been obtained very recently [1, 16, 17, 9, 10, 37] and thus the conjecture of Horn [14] is settled. Namely the complete characterization of three sets of real numbers $\alpha_1 \geq \cdots \geq \alpha_n$, $\beta_1 \geq \cdots \geq \beta_n$, $\gamma_1 \geq \cdots \geq \gamma_n$ which are the eigenvalues of Hermitian (or real symmetric) A, B and $C = A + B$ are obtained. It would be interesting to see how the results are extended to other simple Lie algebras. In our setting of Eaton triple with reduced triple, the even harder question is: what is the necessary and sufficient conditions on α, β and $\gamma \in F$ such that $F(x) = \alpha$, $F(y) = \beta$ and $F(x+y) = \gamma$ for some $x, y \in V$?

A possible generalization of Lidskii's result is asked in [2] associated with hyperbolic polynomials.

5. Distance in terms of G-invariant norm. A norm $\|\cdot\| : V \to \mathbb{R}$ is said to be G-invariant if $\|gx\| = \|x\|$ for all $g \in G$, $x \in V$. We have the following application which extends [32, Theorem 2.1].

Theorem 5.1. *Let (V, G, F) be an Eaton triple with a reduced triple (W, H, F). Let $\|\cdot\| : V \to \mathbb{R}$ be a G-invariant norm. Let $\omega \in H$ be the longest element. If $x, y \in V$, then*

$$\min_{g \in G} \|x - gy\| = \|F(x) - F(y)\|,$$
$$\max_{g \in G} \|x - gy\| = \|F(x) - \omega(F(y))\| = \|F(x) + F(-F(y))\|.$$

Proof. Since $\|\cdot\|$ is G-invariant,

$$\min_{g \in G} \|x - gy\| \leq \|F(x) - F(y)\|,$$
$$\max_{g \in G} \|x - gy\| \geq \|F(x) - \omega(F(y))\|.$$

It is left to prove the reverse inequalities. The *dual norm* $\|\cdot\|^D : V \to \mathbb{R}$ is defined by

$$\|x\|^D = \max_{\|y\| \le 1} (x, y),$$

that is, the dual norm of x is simply the norm of the linear functional induced by x. Let $C = \{F(y) : \|y\|^D \le 1, y \in V\} \subset F$, a compact set. Then by [34, Theorem 2], for any $x \in V$,

$$\|x\| = \max_{\alpha \in C} (F(x), \alpha).$$

For the minimum, by Theorem 4.2, $F(x) - F(gy) \in \operatorname{conv} H(F(x - gy))$ so that for any $g \in G$,

$$\begin{aligned} \|x - gy\| &= \max_{\alpha \in C} (F(x - gy), \alpha) \\ &\ge \max_{\alpha \in C} (F(x) - F(gy), \alpha) \qquad \text{by (12)} \\ &= \max_{\alpha \in C} (F(x) - F(y), \alpha) \\ &= \|F(x) - F(y)\|. \end{aligned}$$

For the maximum, we may assume that $x, y \in W$ or even in F since $\|\cdot\|$ is G-invariant. By Lemma 4.1,

$$\begin{aligned} \|x - gy\| &= \max_{\alpha \in C} (F(x - gy), \alpha) \\ &\le \max_{\alpha \in C} (F(x) + F(-gy), \alpha) \qquad \text{by (12) and Lemma 4.1} \\ &= \max_{\alpha \in C} (F(x) + F(-y), \alpha) \\ &= \max_{\alpha \in C} (F(x) - \omega F(y), \alpha) \\ &= \|F(x) - \omega(F(y))\|, \end{aligned}$$

where the second last equality follows from $F(-y) = -\omega F(y)$. Finally $-\omega F(y) = F(-F(-y))$ by [11, Lemma 2.12]. $\square$

COROLLARY 5.2. *Let $\mathfrak{g}$ be a real semisimple Lie algebra with Cartan decomposition $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$, where the analytic group of $\mathfrak{k}$ is $K \subset G$. For $x \in \mathfrak{p}$, let $a_+(x)$ denote the unique element of the singleton set $Ad(K)x \cap \mathfrak{a}_+$, where $\mathfrak{a}_+$ is a closed fundamental Weyl chamber. Given $x, y \in \mathfrak{p}$, if $z \in Ad(K)y$, then*

$$\begin{aligned} \min_{k \in K} \|x - Ad(k)y\| &= \|a_+(x) - a_+(y)\| \\ \max_{k \in K} \|x - Ad(k)y\| &= \|a_+(x) - \omega a_+(y)\| = \|a_+(x) + a_+(-a_+(y))\|, \end{aligned}$$

where $\|\cdot\|$ is a $Ad(K)$-invariant norm and ω is the longest element of the Weyl group of $(\mathfrak{g}, \mathfrak{a})$.

The following result provides the distance between a point and the convex hull of a G-orbit. The proof is similiar to that in [11] and is omitted.

THEOREM 5.3. *Let (V, G, F) be an Eaton triple with a reduced triple (W, H, F). Let $\|\cdot\| : V \to \mathbb{R}$ be a G-invariant norm. Let $\omega \in H$ be the longest element. If $x, y \in V$, then*

$$\min_{z \in conv\, Gy} \|x - z\| = \|F(x) - F(y)_{F(x)}\|,$$

$$\max_{z\in conv\, Gy} \|x-z\| = \|F(x)-\omega(F(y))\| = \|F(x)+F(-(F(y)))\|,$$

where $F(y)_{F(x)} = F(x) - (F(x)-F(y))^+$ *and* $(F(x)-F(y))^+$ *is given by Algorithm 2.4 in [11].*

Acknowledgement: During the 1998 matrix workshop held in the University of Hong Kong in honor of the retirement of Prof. Y.H. Au-Yeung, Dr. C.M. Cheng has mentioned to the first author the problem of determining the distance between a point and a Ad $(SO(n)\times SO(n))$-orbit for the simple Lie algebra $\mathfrak{so}_{n,n}$. The question provoked the study in this paper. The first author also learned from Prof. A.S. Lewis that he and one of his students have given a proof, via nonsmooth analysis, of Corollary 4.4 which can also be derived by Liskii's result by considering the so called Wielandt matrix (6).

REFERENCES

[1] S. Agnihotri and C. Woodward, Eigenvalues of products of unitary matrices and quantum Schubert calculus *Math. Res. Lett.* 5:817-836, 1998.

[2] H.H. Bauschke, O. Güler, A.S. Lewis and H.S. Sendov, Hyperbolic polynomials and convex analysis, Technical report, University of Waterloo.

[3] F. A. Berezin and I. M. Gel'fand, Some remarks on the theory of spherical functions on symmetric Riemannian manifolds, *Tr. Mosk. Mat. Obshch.*, 5:311-351, 1956. English transl. in *Amer. Math. Soc. Transl.* (2) 21:193-238, 1962.

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

[5] M.T. Chu and K.R. Driessel, The projected gradient method for least squares matrix approximations with spectral constraints, *SIAM J. Numer. Anal.*, 27:1050-1060, 1990.

[6] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.

[7] M.L. Eaton and M.D. Perlman, Reflection groups, generalized Schur functions and the geometry of majorization, *Ann. Probab.* 5:829-860, 1977.

[8] K. Fan, On a theorem of Weyl concerning eigenvalues of linear transformations, *Proc. Nat. Acad. Sci. U.S.A.*, 35:652-655, 1949.

[9] W. Fulton, Eigenvalues of sums of Hermitian matrices (after A. Klyachko), Séminaire Bourbaki 845, June, 1998 *Astérisque* 252:255-269, 1998.

[10] W. Fulton, Eigenvalues, invariant factors, highest weights, and Schubert calculus, *Bull. Amer. Math. Soc. (N.S.)*, 37:209-249 2000.

[11] R.R. Holmes and T.Y. Tam, Distance to the convex hull of an orbit under the action of a compact Lie group, *J. Austral. Math. Soc. Ser. A*, 66:331-357, 1999.

[12] D.R. Jensen, Invariant ordering and order preservation, in *Inequalities in Statistics and Probability*, Y.L. Tong, ed. IMS Lectures Notes, Monograph Series Vol. 5, p26-34, 1984.

[13] J.B. Hiriart-Urruty and D. Ye, Sensitivity analysis of all eigenvalues of a symmetric matrix, *Numer. Math.*, 70:45-72, 1995.

[14] A. Horn, Eigenvalues of sums of Hermitian matrices *Pacific J. Math.*, 12:225-241, 1962.

[15] J.E. Humphreys, *Reflection Groups and Coxeter Groups*, Cambridge University Press, 1990.

[16] A. A. Klyachko, Stable bundles, representation theory and Hermitian operators, *Selecta Math.*, 4:419-445, 1998.

[17] A. Knutson and T. Tao, The honeycomb model of $GL_n(\mathbb{C})$ tensor products I: proof of the saturation conjecture, *J. Amer. Math. Soc.*, 12:1055-1090, 1999.

[18] A.W. Knapp, *Lie Groups Beyond an Introduction*, Birkhäuser, Boston, 1996.

[19] B. Kostant, On convexity, the Weyl group and Iwasawa decomposition, *Ann. Sci. Ecole Norm. Sup.*, (4) 6:413-460, 1973.

[20] A.S. Lewis, Convex analysis on the Hermitian matrices, *SIAM J. Optimization,* 6:164-177, 1996.

[21] A.S. Lewis, Derivatives of spectral functions, *Math. Oper. Res.*, 21:576–588, 1996.

[22] A.S. Lewis, Group invariance and convex matrix analysis, *SIAM J. Matrix Anal. Appl.*, 17:927-949, 1996.

[23] A.S. Lewis, Lidskii's theorem via nonsmooth analysis, *SIAM J. Matrix Anal. Appl.*, 21:379-381, 1999.

[24] V.I. Lidskii, On the proper values of a sum and product of symmetric matrices, *Dokl. Akad. Nauk. SSSR*, 75:769-772, 1950.

[25] H. Miranda and R.C. Thompson, A supplement to the von Neumann trace inequality for singular values, *Linear Algebra Appl.*, 248: 61-66, 1994.

[26] A.S. Markus, The eigen- and singular values of the sum and product of linear operators, *Russian Math. Surveys*, 19:92-120, 1964.

[27] M. Niezgoda, Group majorization and Schur type inequalities, *Linear Algebra Appl.*, 268:9-30, 1998.

[28] M.L. Overton and R.S. Womersley, Optimality conditions and duality theory for minimizing sums of largest eigenvalues of symmetric matrices, *Math. Prog.*, 62:321-357, 1993.

[29] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[30] T.Y. Tam, A unified extension of two result of Ky Fan on the sum of matrices, *Proc. Amer. Math. Soc.* 126:2607-2614, 1998.

[31] T.Y. Tam, A Lie theoretical approach of Thompson's theorems of singular values-diagonal elements and some related results, *J. of London Math. Soc. (2)*, 60:431-448, 1999.

[32] T.Y. Tam, An extension of a result of Lewis, *Electronic J. Linear Algebra* 5:1-10, 1999.

[33] T.Y. Tam, Group majorization, Eaton triples and numerical range, *Linear and Multilinear Algebra*, 47:11-28, 2000

[34] T.Y. Tam and W.C. Hill, On G-invariant norms, *Linear Alg. App.*, 331:101-112, 2001.

[35] R.C. Thompson, Matrix Spectral Inequalities, Johns Hopkins University Lectures Notes, Baltimore, MD, 1988.

[36] H. Wielandt, An extremum property of sums of eigenvalues, *Proc. Amer. Math. Soc.* 6:106-110, 1955.

[37] A. Zelevinsky, Littlewood-Richardson semigroups, in *New Perspectives in Algebraic Combinatorics* (L.J. Billera, A. Björner, C. Greene, R.E. Simion, R.P. Stanley eds), Cambridge University Press (MSRI Publication), p.337-345, 1999.