Derivatives of orbital function and an extension of Berezin-Gel’fand’s theorem

DOI 10.1515/spma-2016-0033
Received June 5, 2016; accepted November 28, 2016

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.

Keywords: Berezin-Gel’fand’s theorem, subdifferential, Clarke generalized gradient, Lebourg mean value theorem, Eaton triple, reduced triple, finite reflection group

MSC: 90C31, 15A18

1 Introduction

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

**Theorem 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 $\text{Ad}(K)x \cap a_+$, where $a_+$ is a closed fundamental Weyl chamber. For $y, z \in \mathfrak{p}$, $a_+(z + y) - a_+(z) \in \text{conv} W a_+(y)$, where $W$ denotes the Weyl group of $(\mathfrak{g}, a)$, $Wa_+(y)$ denotes the orbit of $a_+(y)$ under the action of $W$, and conv denotes the convex hull.

The result of Berezin-Gel’fand had been known to Lidskii who [29] gave an elementary proof of a special case of Berezin-Gel’fand’s theorem, namely, when $G = SL(n, \mathbb{R})$, though [29] 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 [42] to provide another proof by using minimax property. The result of Lidskii is stated in the following

**Theorem 2.** (Lidskii [29], Wielandt [42]) Let $A$ and $B$ be real symmetric (Hermitian, quaternionic Hermitian) matrices. Denote by $\lambda(A) = (\lambda_1(A), \ldots, \lambda_n(A))$ the vector of eigenvalues of $A$ in nonincreasing order $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$. Then

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

*Corresponding Author: Tin-Yau Tam: Department of Mathematics and Statistics, 221 Parker Hall, Auburn University, Auburn, AL 36849-5310, USA, E-mail: tamtiny@auburn.edu
William C. Hill: Department of Mathematics, Middle Georgia State University, 100 University Pkwy, Macon, GA 31206, USA, E-mail: chris.hill@mga.edu
where $S_n$ is the full symmetric group and $S_n\lambda(A)$ denotes the orbit of $\lambda(A)$ under the action of $S_n$. In terms of inequalities, it is equivalent to

\[
\max_{1 \leq j_1 < \cdots < j_k \leq n} \sum_{i=1}^{k} [\lambda_{j_i}(A + B) - \lambda_{j_i}(B)] \leq \sum_{i=1}^{k} \lambda_{i}(A), \quad k = 1, \ldots, n - 1,
\]

and the equality is merely the trace condition.

Later Markus [31] gave another proof of Lidskii’s theorem by using an idea of Wielandt [42] but not the min-max property. See three proofs and some historical remarks in [4, 40]. Recently, Lewis [25] provided an entirely 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 5. Since the framework of the generalization is Eaton triple, Section 2 contains some fundamental concepts of Eaton triple and finite reflection group. We give some basic notions of nonsmooth analysis in Section 3. In order to carry out the approach, the derivatives of some orbital functions are studied and a number of results in [23] are generalized in Section 4. Then we determine the distance between a $G$-orbit or its convex hull and a given point as applications in Section 6.

## 2 Eaton triple and finite reflection group

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 inner product is denoted by $(\cdot, \cdot)$. 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 [38] of the Eaton triple $(V, G, F)$ if it is an Eaton triple where $W := \text{span} F$ and $H := \{g|_W : g \in G, gW = W\} \subset O(W)$, the orthogonal group of $W$. It is known that $Gx \cap F$ is a singleton set [32, p.14]. 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 [32, Theorem 3.2].

Let us recall some rudiments of finite reflection groups [17, 21]. Let $V$ be a finite dimensional real inner product space. A reflection $s_a$ on $V$ is an element of $O(V)$, which sends some nonzero vector $a$ to its negative and fixes pointwise the hyperplane $H_a$ orthogonal to $a$, that is, $s_a\lambda := \lambda - 2\frac{(\lambda, a)}{(a, a)} a, \lambda \in V$. 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 $\Phi$, such that $\{s_a : a \in \Phi\}$ generates $G$, and satisfies

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

(R2) $s_a\Phi = \Phi$ for all $a \in \Phi$.

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

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

and the group $G$ is known as the Weyl group of $\Phi$. 

Download Date | 8/25/18 7:08 PM
A (open) chamber $C$ is a connected component of $V \setminus \cup_{a \in \Phi} H_a$. Given a total order $\prec$ in $V$ [17, 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 $\{v_1, \ldots, v_m\}$ of $V$ and say $\mu \prec \nu$ if the first nonzero number of the sequence $\langle \lambda, v_1 \rangle, \ldots, \langle \lambda, v_m \rangle$ 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 $\Phi^+$ contains [17, p.8] a unique simple system $\Delta$, that is, $\Delta$ is a basis for $V_1 := \text{span} \Phi \subset V$, and each $\alpha \in \Phi$ is a linear combination of $\Delta$ with coefficients all of the same sign (all nonnegative or all nonpositive). The vectors in $\Delta$ 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 $\Phi^+(C) = \{ \alpha \in \Phi : \langle \lambda, \alpha \rangle > 0 \text{ for all } \lambda \in C \}$. The correspondence $C \mapsto \Phi^+(C)$ is a bijection of the set of all chambers onto the set of all positive systems [41, p.372]. The group $G$ acts simply transitively on the sets of positive systems, simple systems and chambers. The closed convex cone $F := \{ \lambda \in V : \langle \lambda, \alpha \rangle \geq 0, \text{ for all } \alpha \in \Delta \}$, that is, $F := C^-$ is the closure of the chamber $C$ which defines $\Phi^+$ and $\Delta$, is called a (closed) fundamental domain for the action of $G$ on $V$ associated with $\Delta$. 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 due to a result of Eaton and Perlman [7, (3.5) and Lemma 4.1] (also see [32]). 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 $V_1 := V_0$. If $\{\alpha_1, \ldots, \alpha_n\}$ denotes the basis of $V_1$ dual to the basis $\beta_i := \frac{x_{\alpha_i}}{\alpha_i} : i = 1, \ldots, n$, that is, $\langle \lambda, \beta \rangle = \delta_{ij}$, then $F = \{ \sum_{i=1}^n c_i \alpha_i : c_i \geq 0 \} \oplus V_0$. Thus the interior $\text{Int} F = C$ of $F$ is the nonempty set $\{ \sum_{i=1}^n c_i \alpha_i : c_i > 0 \} \oplus V_0$. The dual cone of $F$ in $V_1$ is the cone

\[
\text{dual } V_1 F := \{ x \in V_1 : \langle x, u \rangle > 0, \text{ for all } u \in F \}
\]

induced by $F$. Notice that

\[
\text{dual } V_2 F = \{ \sum_{i=1}^n c_i \alpha_i, c_i \geq 0, i = 1, \ldots, n \},
\]

the cone generated by $\Delta$. There is a unique element $\omega \in G$ sending $\Phi^+$ to $\Phi^-$ and thus sending $F$ to $-F$.

Moreover, the length [17, p.12] of $\omega$ is the longest one [17, p.15-16] and thus we call it the longest element.

We will present two examples requiring some basic knowledge of Lie theory [20]. Let $g = t + p$ be a Cartan decomposition of the Lie algebra $g$ of a semisimple Lie group $G$ with finite center. Denote the Killing form of $g$ by $B(\cdot, \cdot)$. The Killing form is positive definite on $p$ but negative definite on $t$. Let $K$ be an analytic subgroup of $t$ in the analytic group $G$ of $g$. Now $Ad(K)$ is a subgroup of the orthogonal group on $p$ with respect to the restriction of the Killing form on $p$ since the Killing form is invariant under $Ad(K)$. Among the abelian subalgebras of $g$ that are contained in $p$, choose a maximal one $a$ (maximal abelian subalgebra in $p$). For $a \in a^*$ (the dual space of $a$), set $g_a = \{ x \in g : [h, x] = a(h)x \text{ for all } h \in a \}$. If $0 \neq a \in a^*$ and $g_a \neq 0$, then $a$ is called a (restricted) root [20, p.313] of the pair $(g, a)$. The set of roots will be denoted by $\Sigma$. We have the orthogonal direct sum $g = g_0 + \sum_{a \in \Sigma} g_a$ known as the restricted-root space decomposition [20, p.313]. We view $a$ as an Euclidean space by taking the inner product to be the restriction of $B$ to $a$. The map $a^* \to a$ that assigns to each $\lambda \in a^*$ the unique element $x_\lambda$ of a satisfying $\lambda(x) = B(x, x_\lambda)$ for all $x \in a$ is a vector space isomorphism.

We use this isomorphism to identify $a^*$ with $a$, allowing us, in particular, to view $\Sigma$ as a subset of $a$. 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 $s_\alpha$ ($\alpha \in \Sigma$), which is called the Weyl group of $(g, a)$, and is a root system of $W$. It is called the reduced root system of the pair $(g, a)$. Now fix a simple system $\Delta$ for the root system $\Phi$. Then $\Delta$ determines a fundamental domain for the action of $W$. We now describe another way to view the Weyl group $W$. Use juxtaposition to represent the adjoint action of $G$ on $g$, that is, $gx = Ad(g)x, g \in G, x \in g$. Set $N_K(a) = \{ k \in K : ka \subset a \}$ (the normalizer of $a$ in $K$) and $Z_K(a) = \{ k \in K : xk = x \text{ for all } x \in a \}$ (the centralizer of $a$ in $K$). Then the action of $K$ on $g$ induces an action of the group $N_K(a)/Z_K(a)$ on $a$, that is, $\{ kx = kxk \} \subset N_K(a)/Z_K(a)$. There exists an isomorphism $\psi : W \to N_K(a)/Z_K(a)$ that is, compatible with the two actions on $a$, or more precisely, for which $wx = \psi(w)x, w \in W, x \in a$ [20, p.325, p.394]. We use the isomorphism $\psi$ to identify these two groups (in the literature, the Weyl group is usually defined to be $N_K(a)/Z_K(a)$). Note in particular that, given $x \in a$, we have $Wx = N_K(a)x \subset Kx$. Since $Ad(K)$ is an automorphism of $g$, $N_K(a) = \{ k \in K : ka = a \}$. 

Unauthenticated
3 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 $K > 0$,

\[ |f(y) - f(y')| \leq K\|y - y'\|, \quad y, y' \in Y, \quad (1) \]

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

\[ f^0(x; v) = \limsup_{y \to x, t \downarrow 0} \frac{f(y + tv) - f(y)}{t}, \quad (2) \]

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

\[ \partial f(x) := \{ \xi \in V : f^0(x; v) \geq (\xi, v) \text{ for all } v \in V \}. \quad (3) \]

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,

\[ \partial f(x) = \text{conv} \{ \lim_{n \to \infty} \nabla f(x_n) : x_n \to x, x_n \notin S \cup \Omega_f \}, \quad (4) \]

where $S$ is any fixed set of Lebesque measure 0 in $V$, $\Omega_f$ 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] $\Omega_f$ 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 5. (Lebourg mean value theorem) Let $x, y \in V$ and suppose that $f$ is Lipschitz on an open set containing the closed line segment $\{tx + (1 - t)y : 0 \leq t \leq 1\}$. Then there exists $u$ in the open line segment $\{tx + (1 - t)y : 0 < t < 1\}$ such that

\[ f(y) - f(x) \in (\partial f(u), y - x). \quad (5) \]

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

\[ \varphi(z) \geq \varphi(x) + (x^*, z - x), \quad \text{ for all } z \in V. \]

The set of subgradients of $\varphi$ at $x$ is called the subdifferential of $\varphi$ at $x$ and is denoted by $\partial \varphi(x)$. It turns out that [34, Theorem 25.1] $\varphi$ is differentiable at $x$ if and only if $\partial \varphi(x)$ is a singleton set. In this event $\partial \varphi(x) = \{ \nabla \varphi(x) \}$.
4 Derivatives of orbital functions

Throughout this section \((V, G, F)\) is an Eaton triple with reduced triple \((W, H, F)\). By [32, 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 \(W_1 := W_0^U\). Let \(A := \{a_1, \ldots, a_n\}\) be a simple system of \(H\) such that \(F = \{x \in W : (a_i, x) \geq 0, i = 1, \ldots, n\}\). Let \(\lambda_1, \ldots, \lambda_n\) be the basis of \(W_1\) dual to \(\beta_i := 2a_i/(a_i, a_i) : i = 1, \ldots, n\). Thus \(F = \{\sum_{i=1}^n c_i \beta_i : c_i \geq 0\} \oplus W_0\).

The results in this section generalize the corresponding indicated results in [23, 33].

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

A subset \(U \subset W\) 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, an \(H\)-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 [23, 33].

**Lemma 6.** (Compare [33, Lemma 3.2]) Given \(a_m \in \Delta\). If \(\mu \in F\) such that \((\mu, a_m) \neq 0\), then
\[
\max\{\mu, x\} : x \in \conv H\lambda_m = (\mu, \lambda_m)
\]

and
\[
\arg\max\{\mu, x\} : x \in \conv H\lambda_m = \{\lambda_m\}.
\]

**Proof.** Notice that \(\max\{\mu, x\} : x \in \conv H\lambda_m = \max\{\mu, x\} : x \in H\lambda_m = (\mu, \lambda_m)\) by (A2) since \(\mu, \lambda_m \in F\). By the definition of \(F\), \((\mu, a_m) > 0\) since \((\mu, a_m) \neq 0\), \(\mu \in F\), and \(a_m \in \Delta\). It is clear that \(\lambda_m \in \arg\max\{\mu, u\} : u \in \conv H\lambda_m\). Let \(x \in \arg\max\{\mu, u\} : u \in \conv H\lambda_m\) \(\subset W\). Rewrite
\[
x = \sum_{i=1}^n 2(x, \lambda_i)_{(a_i, a_i)} a_i + \pi_0(x), \quad \mu = \sum_{i=1}^n 2(\mu, a_i)_{(a_i, a_i)} \lambda_i + \pi_0(\mu),
\]

where \(\pi_0 : W \rightarrow W_0\) is the orthogonal projection. So
\[
(\mu, x) = \sum_{i=1}^n 2(\mu, a_i)_{(a_i, a_i)} (x, \lambda_i) + (\pi_0(\mu), \pi_0(x)),
\]

and similarly
\[
(\mu, \lambda_m) = \sum_{i=1}^n 2(\mu, a_i)_{(a_i, a_i)} (\lambda_m, \lambda_i) + (\pi_0(\mu), \pi_0(\lambda_m)).
\]

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

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

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)
\[
(\lambda_m, \lambda_m) = (x, \lambda_m) = \sum_{i=1}^k a_i (h_i \lambda_m, \lambda_m) \leq \sum_{i=1}^k a_i (\lambda_m, \lambda_m) = (\lambda_m, \lambda_m).
\]
Thus $\langle h_i \lambda_m, \lambda_m \rangle = (\lambda_m, \lambda_m)$ 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 $(h_i \lambda_m, \lambda_m) \leq \|h_i \lambda_m\| \|\lambda_m\| = (\lambda_m, \lambda_m)$ and that $h_i$ is orthogonal. Hence we have the desired $x = \lambda_m$. □

**Theorem 7.** (Compare [23, Theorem 2.1]; also see [33, Lemma 3.3], [15, Corollary 3.10] and [22])) 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 $f_{\lambda_m}$ is differentiable at $\mu$ and $df_{\lambda_m}|_{\mu} = (\lambda_m, \cdot)$, that is, $\nabla f_{\lambda_m}(\mu) = \lambda_m$.

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

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

In other words, $f_\lambda$ is the support function for the compact convex set $\text{conv } G\lambda$ and is therefore positively homogeneous and convex [34, Theorem 13.2]. The subdifferential of the support function $f_\lambda$ at the point $z$, denoted by $\partial f_\lambda(z)$, consists of the elements of $\text{conv } G\lambda$ attaining the maximum $f_\lambda(z) = (\lambda, F(z))$ [34, Corollary 23.5.3], that is,

$$\partial f_\lambda(z) = \arg \max\{\langle \xi, z \rangle : \xi \in \text{conv } G\lambda\}.$$

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

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

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

and by [32, Theorem 3.2], $\pi(z) \in \text{conv } H\lambda_m$. By Lemma 6, 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, $z \in \text{conv } G\lambda_m$ means $z = \sum_{i=1}^{k} a_i \lambda_m$, where $\sum_{i=1}^{k} a_i = 1$, $a_i > 0$, and $g_i \in G$, for all $i = 1, \ldots, k$, which implies that

$$\|z\|^2 = \|\sum_{i=1}^{k} a_i \lambda_m\| \leq \sum_{i=1}^{k} a_i \|g_i \lambda_m\| = \|\lambda_m\|.$$

Thus $y = 0$ and $z = \lambda_m$. Hence $\partial f_{\lambda_m}(\mu) = (\lambda_m)$ and by [34, Theorem 25.1], the desired result follows. □

**Example 8.** The general linear group $GL_n(\mathbb{F})$ consists of $n \times n$ matrices with nonzero determinant. The Lie algebra is $gl_n(\mathbb{F})$, that is, $n \times n$ matrices with elements in $\mathbb{F}$, which is reductive. The Cartan decomposition of $gl_n(\mathbb{F}) = \mathfrak{t} + \mathfrak{p}$ with $\mathfrak{t}$ is the space of real symmetric, Hermitian and quaternionic Hermitian matrices (that is, $A = A^\ast$ where $A^\ast = \overline{A}^T$ and $a_1 = a_2 = \cdots = a_n = a$) and $\mathfrak{p}$ is the space of the matrix $x \in \mathbb{F}$ in descending order. Then $F(x) = a_0(x)$ is indeed the vector of eigenvalues of the matrix $x \in \mathfrak{p}$ in descending order. So $(V, G, F) = (\mathfrak{p}, \text{Ad}(U(n)), a_0)$ and $(W, H, F) = (a, S_n, a_0)$ where $S_n$ is known as the symmetric group of degree $n$, known as the Weyl group of $A_{n-1}$ type. Notice that $a_0$, the set of fixed points in $a$ is the span of $e$ where $e = (1, 1, \ldots, 1)$. The simple roots [17, p.41] of $a_0 := a_0^\perp$ are

$$a_i = e_i - e_{i+1}, \quad i = 1, \ldots, n - 1,$$

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

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

Thus $f_{\lambda_m}(z)$ is the sum of the largest $n$ eigenvalues of the matrix $z \in \mathfrak{p}$. So the later part of Theorem 7 asserts that if $\mu_1 \geq \cdots \geq \mu_n$ with $\mu_m > \mu_{m+1}$ (1 ≤ $m$ < $n$), Then $f_{\lambda_m}$ is differentiable at $\mu$ and $df_{\lambda_m}|_{\mu} = (\lambda_m, \cdot)$ which is exactly the statement of [23, Theorem 2.1].
Lemma 10. (Compare [23, Lemma 2.2]) Let \( \mu \in F \) refines \( \beta \in W \), then the function \( f_\beta : V \to \mathbb{R} \) defined by \( f_\beta(z) = (\beta, F(z)) \) is differentiable at \( \mu \) with \[
\frac{\partial f_\beta}{\partial \mu}(\mu) = (\beta, \cdot),
\]
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 $\pi_0(F(z)) = \pi_0(z)$ and $\beta = \sum_{i=1}^{n} 2(\alpha_i, \beta)/(\alpha_i, \alpha_i) \lambda_i + \pi_0(\beta)$, where $\pi_0 : V \to W_0$ is the orthogonal projection. So

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

Then

$$df_\beta|_\mu = \sum_{i=1}^{n} 2(\alpha_i, \beta) (\alpha_i, \alpha_i) df_{\lambda_i}|_{F(\mu)} + (\pi_0(\beta), \cdot)$$

$$= \sum_{i=1}^{n} 2(\alpha_i, \beta) (\alpha_i, \alpha_i) df_{\lambda_i}|_{\mu} + (\pi_0(\beta), \cdot) \quad \text{since } \mu \in F$$

$$= \sum_{i=1}^{n} 2(\alpha_i, \beta) (\alpha_i, \cdot) + (\pi_0(\beta), \cdot) \quad \text{by Theorem 7 and since } \mu \text{ refines } \beta$$

$$= (\beta, \cdot).$$

\[
\]

Lemma 11. (Compare [23, Lemma 2.3]) Let $U \subseteq W$ be open and $H$-invariant. Suppose that the function $f : U \to \mathbb{R}$ is $H$-invariant and differentiable at $\mu \in F$. Then $\mu$ refines $\nabla f(\mu)$ and thus the function $df|_\mu \circ F$ is differentiable at $\mu$ with $d(df|_\mu \circ F)|_\mu = (\nabla f(\mu), \cdot)$, that is, $\nabla(df|_\mu \circ F)(\mu) = \nabla f(\mu)$.

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

$$df|_\mu = df|_{s_a \mu} \circ ds_a|_\mu = df|_\mu \circ s_a,$$

since $s_a$ is linear. So $\nabla f(\mu) = s_a \nabla f(\mu)$, that is, $\mu$ refines $\nabla f(\mu)$. Since $df|_\mu \circ F = (\nabla f(\mu), F(\cdot))$ and $\nabla f(\mu) \in W$, by Lemma 10, the function $df|_\mu \circ F$ is differentiable at $\mu$ and $d(df|_\mu \circ F)|_\mu = (\nabla f(\mu), \cdot)$. \qed

Lemma 12. (Compare [23, Theorem 2A]) Let $U \subseteq W$ 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 $\mu \in F$,

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

whenever $\gamma \in W$ is close to $\mu$ (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 $y \in V$,

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

By Lemma 11, for small $y \in V$,

$$|df|_\mu \circ F(y + \mu) - df|_\mu(\mu) - (\nabla f(\mu), y)| = |df|_\mu \circ F(y + \mu) - df|_\mu \circ F(\mu) - d(f \circ F)|_\mu(y)| \leq \epsilon ||y||.$$

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

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

for small $y$ and thus the desired result. \qed
Theorem 13. (Compare [23, 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,

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

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 12 yields

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

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

The following is an extension of Lemma 12.

Theorem 14. (Compare [23, 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 13

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

$\square$

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

Theorem 15. (Compare [23, 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

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

Proof. Since $V$ is finite dimensional, we have [6, p.66]

$$f \circ F \in C^1(V, W),$$

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

$$(\nabla f \circ F)(x_n, z) \to (f \circ F)^0(\mu, z).$$
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
\[
\lim_{r \to \infty} g_{n_r}^{-1} F(x_{n_r}) = \lim_{r \to \infty} x_{n_r} = \mu,
\]
so that \( g_0 \in G_\mu \). Hence
\[
(f \circ F)^0(\mu; z) = \lim_{n \to \infty} (\nabla (f \circ F)(x_n), z)
= \lim_{n \to \infty} (g_{n_r}^{-1} \nabla F(x_n), z)
= \lim_{n \to \infty} (\nabla F(x_n), g_{n_r} z)
= \lim_{r \to \infty} (\nabla F(x_n), \pi(g_{n_r} z)) \quad \text{since } \nabla F(x_n) \in W
\leq \lim \sup_{r \to \infty}(\nabla F(\gamma), \pi(g_0 z))
= f^0(\mu; \pi(g_0 z)),
\]
where \( \pi : V \to W \) denotes the orthogonal projection. Thus we establish ‘s’ 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 \( k \in G_\mu \),
\[
f^0(\mu; \pi(k z)) = \lim_{n \to \infty} (\nabla f(\mu_n), \pi(k z))
= \lim_{n \to \infty} (\nabla f(\mu_n), k z)
\quad \text{since } \nabla f(\mu_n) \in W
= \lim_{n \to \infty} (\nabla (f \circ F)(\mu_n), k z)
\quad \text{by Theorem 14}
\leq \lim \sup_{\gamma \to \mu}(\nabla (f \circ F)(\gamma), k z)
= (f \circ F)^0(\mu; k z).
\]

Now for any \( g \in G \),
\[
(f \circ F)^0(\mu; g z) = \lim_{w \to g \mu, t \downarrow 0} \frac{f(F(w + tz)) - f(F(w))}{t}
= \lim_{y \to \mu, t \downarrow 0} \frac{f(F(g(y + tz))) - f(F(g y))}{t}
= \lim_{y \to \mu, t \downarrow 0} \frac{f(F(y + tz)) - f(F(y))}{t}
= (f \circ F)^0(\mu; z).
\]

Since \( k \in G_\mu \subset G \),
\[
(f \circ F)^0(\mu; k z) = (f \circ F)^0(k^{-1} \mu; z) = (f \circ F)^0(\mu; z).
\]
Hence \( f^0(\mu; \pi(k z)) \leq (f \circ F)^0(\mu; z) \) for all \( k \in G_\mu \). Thus the desired result follows. \( \square \)

**Lemma 16.** (Compare [23, 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
\[
\delta(f \circ F)(\mu) = \text{conv} \{k \gamma : k \in G_\mu, \gamma \in \partial f(\mu)\}.
\]

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

\[
\max \{ (z, y) : y \in \text{conv} \, D \} = \max \{ (z, y) : y \in D \} = \max \{ (z, k\gamma) : k \in G_\mu, \gamma \in \partial f(\mu) \} = \max \{ (kz, \gamma) : k \in G_\mu, \gamma \in \partial f(\mu) \} \quad \text{since } G_\mu \text{ is a group}
\]

\[
= \max \{ (\pi(kz), \gamma) : k \in G_\mu, \gamma \in \partial f(\mu) \} \quad \text{by } \partial f(\mu) \subset W = \max \{ \max \{ (\pi(kz), \gamma) : \gamma \in \partial f(\mu) \} : k \in G_\mu \},
\]

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 \( (f \circ F)'(\mu; z) \), by Theorem 15

\[
(f \circ F)'(\mu; z) = \max \{ f'(\mu; \pi(kz)) : k \in G_\mu \}.
\]

Clearly \( f'(\mu; \pi(kz)) \) is the support function of \( \partial f \), evaluated at \( \pi(kz) \), which is \( \max \{ (\pi(kz), \gamma) : \gamma \in \partial f(\mu) \} \). \( \square \)

**Remark 17.** In [23, 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, \( gl_n(\mathbb{R}) \) and \( gl_n(\mathbb{C}) \) by some argument involving doubly stochastic matrices. Using the fact that \( D \) is convex for those two cases, [23, Theorem 1.4] is deduced and is used in [25] to give a new proof of Liskin’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.

## 5. 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 [38, Theorem 10] (also see [35])). Since the proof is the same, it is omitted.

**Lemma 18.** Let \((V, G, F)\) be an Eaton triple with a reduced triple \((W, H, F)\). For any \( x_1, \ldots, x_k \in V \), \( F(\sum_{i=1}^k x_i) \in \text{conv} \, H(\sum_{i=1}^k F(x_i)) \).

**Theorem 19.** Let \((V, G, F)\) be an Eaton triple with a reduced triple \((W, H, F)\). For any \( y, z \in V \),

\[
F(y + z) - F(z) \in \text{conv} \, H(F(y)).
\]

In terms of inequalities, it amounts to

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

**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

\[
\partial f_w(u) \subset \text{conv} \, Hw, \quad u \in W.
\] (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, \( a(u) \neq 0 \) for all \( a \in \Delta \). 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 [17, 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 $\partial f_w(u) = \{ \nabla f_w(u) \} = \{ h_u^{-1} w \} \subset \text{conv } \mathcal{H}_w$.

On the other hand if $u \in W$ is not regular, that is, $u$ lies in some hyperplane $H_{\alpha}, \alpha \in \Delta$, then $f_w$ is not differentiable at $u$ and $\Omega f_w = \bigcup_{\alpha \in \Delta} H_{\alpha}$ and we choose $S = \Omega f_w$ in (4). By (4), $\partial f_w(u) = \text{conv } H_{u}^{-1} w$, where $H_u = \{ h \in H : h u = F(u) \} \subset H$ 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 $(f_w \circ F) \circ g : V \rightarrow \mathbb{R}$ of $f_w \circ F : V \rightarrow \mathbb{R}$ and $g : V \rightarrow V$. The function $f_w \circ F$ is Lipschitz with Lipschitz constant $\| w \|$ on $V$ and $g$ is an orthogonal map. Apply chain rule [6, Theorem 2.3.10] on the composite function at the point $g^{-1} x$ to get

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

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

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

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

$$= g \partial (f_w \circ F)(F(x))$$

$$= g \text{conv } \{ k \gamma : k \in G_{F(x)}, \gamma \in \partial f_w(F(x)) \} \quad \text{by Lemma 16.}$$

By (8) we have

$$\partial (f_w \circ F)(x) \subset g \text{conv } \{ k \gamma : k \in G_{F(x)}, \gamma \in \text{conv } \mathcal{H}_w \}, \quad w \in W. \quad (9)$$

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

$$(F(y + z) - F(z), w) = f_w \circ F(y + z) - f_w \circ F(z) = (y, v) \leq (F(y), F(v)), \quad (10)$$

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

$$v = g \sum_{k \in G_{F(x)}} b_k k \left( \sum_{h \in H} a_k^h h w \right) = \sum_{k \in G_{F(x)}, h \in H} b_k a_k^h h w, \quad (11)$$

where $a_k^h \geq 0$, $\sum_{h \in H} a_k^h = 1$ for all $k \in G_{F(x)}$, $b_k \geq 0$, $\sum_{k \in G_{F(x)}} b_k = 1$. Since $F(a)u = aF(u)$ for all $a \geq 0, u \in V$, by Lemma 18 and (11),

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

that is, $F(v) \in \text{conv } H(F(w))$. Now [38, Lemma 5(2)] states that

$$\text{if } x, y \in F, \text{ then } x \in \text{conv } H y \text{ if and only if } y - x \in \text{dual } F,$$

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

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

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

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

Thus by [38, Lemma 5(1)] $F(y + z) - F(z) \in \text{conv } H(F(y))$. 

Unauthenticated
Download Date | 8/25/18 7:08 PM
Now $F(y + z) - F(z) \in \text{conv } H(F(y))$ amounts to $F(y) - h(F(y + z) - F(z)) \in \text{dual } wF$, for all $h \in H$ by [38, Lemma 5(1)] again, that is,
\[
\max_{h \in H} (h(F(y + z) - F(z)), \lambda_i) \leq (F(y), \lambda_i), \quad \text{for all } i = 1, \ldots, n.
\]

\[\square\]

**Remark 20.** Lemma 18 (when $k = 2$) is now a corollary of Theorem 19: by [38, Lemma 5(1)],
\[
F(y + z) - F(z) \in \text{conv } HF(y) \iff (F(y + z) - F(z), h^{-1}w) \leq (F(y), w),
\]
for all $w \in F, h \in H$. So
\[
(F(y + z), h^{-1}w) \leq (F(y), w) + (F(z), h^{-1}w) \\
\leq (F(y), w) + (F(z), w) \quad \text{by (A2)}
\]
\[
= (F(y) + F(z), w),
\]
for all $w \in F$ and $h \in H$. Thus by [38, Lemma 5(1)]
\[
F(y + z) \in \text{conv } H(F(y) + F(z)).
\]

(13)

We also remark that (13) is symmetric with respect to $y$ and $z$ but Theorem 19 is not.

**Corollary 21.** (Wielandt [42], Markus [31]) 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
\[
s(A + B) - s(B) \in \text{conv } (S_n \times (\mathbb{Z}/2\mathbb{Z})^n)s(A).
\]

In terms of inequalities
\[
\max_{1 \leq i < \cdots < j \leq n} \sum_{i=1}^{k} |s_j(A + B) - s_j(B)| \leq \sum_{i=1}^{k} s_i(A), \quad k = 1, \ldots, n.
\]

**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 $S_n \times (\mathbb{Z}/2\mathbb{Z})^n$ [17, p.42] is the Weyl group for the Example 9. \[\square\]

We remark that via nonsmooth analysis Lewis and Sendov [27, 28] gave a proof of Corollary 21 which can also be derived by Liskii’s result by considering the so-called Wielandt matrix (6).

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 [20]
\[
\mathfrak{so}_{n,n} = \left\{ \begin{pmatrix} X & Y \\ Y^T & X_2 \end{pmatrix} : X_1 = -X_1, \ X_2 = -X_2, \ X_1, X_2, Y \in \mathbb{R}^{n \times n} \right\},
\]
\[
\mathfrak{k} = \text{SO}(n) \times \text{SO}(n),
\]
\[
\mathfrak{t} = \left\{ \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} : X_1 = -X_1, \ X_2 = -X_2, \ X_1, X_2, \in \mathbb{R}^{n \times n} \right\},
\]
\[
\mathfrak{p} = \left\{ \begin{pmatrix} 0 & Y \\ Y^T & 0 \end{pmatrix} : Y \in \mathbb{R}^{n \times n} \right\},
\]
\[
\mathfrak{a} = \oplus_{1 \leq i \neq j \leq n} \mathbb{R}(E_{i,n+j} + E_{n+j,i}),
\]

where $E_{ij}$ is the $2n \times 2n$ matrix and 1 at the $(i, j)$ position is the only nonzero entry. The Killing form is
\[
\mathfrak{g} = \left\{ \begin{pmatrix} 0 & X \\ X^T & 0 \end{pmatrix}, \ \begin{pmatrix} 0 & Y \\ Y^T & 0 \end{pmatrix} \right\} = 4(n - 1) \text{tr } XY^T.
\]
Now the adjoint action of $K$ on $p$ is given by
\[
\begin{pmatrix} U & 0 \\ V & 0 \end{pmatrix}^T \begin{pmatrix} 0 & S \\ S^T & 0 \end{pmatrix} \begin{pmatrix} U & 0 \\ V & 0 \end{pmatrix} = \begin{pmatrix} 0 & U^T S V \\ V^T S^T U & 0 \end{pmatrix},
\]
where $U, V \in \text{SO}(n)$. We will identify $p$ with $\mathbb{R}^{n\times n}$ and thus $a_+$ will then be identified with real diagonal matrices. We may choose $a_+ = \{ \text{diag}(a_1, \ldots, a_n) : a_1 \geq \cdots \geq a_{n-1} \geq |a_n| \}$. The action of $K$ on $p$ is then orthogonal equivalence, that is, $H \mapsto UHV$, where $U, V \in \text{SO}(n)$ and $a_-(H) = (s_1(H), \ldots, s_{n-1}(H), [\text{sign} \det H] s_n(H))$, where $s_1(H) \geq \cdots \geq s_n(H)$ are the singular values of $H$. The action of the Weyl group $W$ on $a_+$ is given by
\[
\text{diag}(d_1, \ldots, d_n) \mapsto \text{diag}(\pm d_{a(1)}, \ldots, \pm d_{a(n)}),
\]
where $\text{diag}(d_1, \ldots, d_n) \in a_+ \subseteq S_n$ (the symmetric group) and the number of negative signs is even. The simple roots may be taken as $a_i = e_1 - e_{i+1}, i = 1, \ldots, n-1$ and $a_n = e_n - e_{n-1}$ [17, p.42] and $\lambda_i = e_1 + \cdots + e_i$, $i = 1, \ldots, n-2$. $\lambda_{n-1} = 1/2(e_1 + \cdots + e_{n-1} - e_n)$. $\lambda_n = 1/2(e_1 + \cdots + e_{n-1} + e_n)$. The longest element $\omega$ sends $\text{diag}(a_1, \ldots, a_n) \in a_+$ to
\[
\omega a = \begin{cases} 
\text{diag}(-a_1, \ldots, -a_{n-1}, a_n) & \text{if } n \text{ is odd} \\
\text{diag}(-a_1, \ldots, -a_n) & \text{if } n \text{ is even}.
\end{cases}
\]
Applying Theorem 19 on the simple Lie algebra $so_{n,n}$, we have the following result. Also see [30, 36].

**Corollary 22.** 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
\[
s'(A) = (s'_1(A), \cdots, s'_n(A)) := (s_1(A), \ldots, s_{n-1}(A), [\text{sign} \det A] s_n(A)),
\]
then
\[
s'(A + B) - s'(B) \in \text{conv}(S_n \times (\mathbb{Z}/2\mathbb{Z})^{n-1}) s'(A).
\]
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
\[
\max_{1 \leq j_1 < \cdots < j_k \leq n} \sum_{i=1}^{k} |s'_{j_i}(A + B) - s'_{j_i}(B)| \leq \sum_{i=1}^{k} s_i(A), \quad k = 1, \ldots, n - 2, \quad (14)
\]
\[
\max_{1 \leq j_1 < \cdots < j_k \leq n} \sum_{i=1}^{n-1} |s'_{j_i}(A + B) - s'_{j_i}(B)| - (-1)^{\#(A, B)} \min_{1 \leq j \leq n} \sum_{i=1}^{n-1} s_i(A) \leq \sum_{i=1}^{n-1} s_i(A) - [\text{sign} \det A] s_n(A), \quad (15)
\]
and
\[
\max_{1 \leq j_1 < \cdots < j_k \leq n} \sum_{i=1}^{n-1} |s'_{j_i}(A + B) - s'_{j_i}(B)| + (-1)^{\#(A, B)} \min_{1 \leq j \leq n} \sum_{i=1}^{n-1} s_i(A) \leq [\text{sign} \det A] s_n(A). \quad (16)
\]

**Proof.** Notice $a_+ = \{ \text{diag}(a_1, \ldots, a_n) : a_1 \geq \cdots \geq a_{n-1} \geq a_n \geq 0 \}$. Any real $n \times n$ matrix $A$ is special orthogonally similar to $\text{diag}(a_1, \ldots, a_{n-1}, [\text{sign} \det A] a_n)$ in $a_+$, where $a_1 \geq \cdots \geq a_{n-1} \geq 0$ are the singular values of $A$. The Weyl group is $S_n \times (\mathbb{Z}/2\mathbb{Z})^{n-1}$ [17, p.42]. By Theorem 19 with $\lambda_i = e_1 + \cdots + e_i, i = 1, \ldots, n-2$, $\lambda_{n-1} = 1/2(e_1 + \cdots + e_{n-1} - e_n)$, $\lambda_n = 1/2(e_1 + \cdots + e_{n-1} + e_n)$, we have the inequalities. \hfill $\Box$

**Remark 23.** We also have
\[
\max_{1 \leq j_1 < \cdots < j_k \leq n} \sum_{i=1}^{k} |s_i(A + B) - s_i(B)| \leq \sum_{i=1}^{k} s_i(A), \quad k = 1, \ldots, n - 1, n, \quad (17)
\]
either by using Corollary 21 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 24.** The characterization of the sum of eigenvalues of two Hermitian matrices as well as two real symmetric matrices has been obtained recently [1, 9, 10, 18, 19, 43] and thus the conjecture of Horn [16] is settled. Namely the complete characterization of three sets of real numbers \( a_1 \geq \cdots \geq a_n, \beta_1 \geq \cdots \geq \beta_n, \gamma_1 \geq \cdots \geq \gamma_n \) 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 \( \alpha, \beta \) 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.

### 6 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 [37, Theorem 2.1].

**Theorem 25.** 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

\[
\begin{align*}
\min_{g \in G} \|x - gy\| &= \|F(x) - F(y)\|, \\
\max_{g \in G} \|x - gy\| &= \|F(x) - \omega(F(y))\| = \|F(x) + F(-F(y))\|.
\end{align*}
\]

**Proof.** Since \( \| \cdot \| \) is \( G \)-invariant,

\[
\begin{align*}
\min_{g \in G} \|x - gy\| &\leq \|F(x) - F(y)\|, \\
\max_{g \in G} \|x - gy\| &\geq \|F(x) - \omega(F(y))\|.
\end{align*}
\]

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

\[
\|x\|^D = \max_{\|y\| \leq 1} \langle x, y \rangle,
\]

that is, the dual norm of \( x \) is simply the norm of the linear functional induced by \( x \). Let \( C = \{F(y) : \|y\|^D \leq 1, y \in V\} \subset F \), a compact set. Then by [39, Theorem 2], for any \( x \in V \),

\[
\|x\| = \max_{\alpha \in C} \langle F(x), \alpha \rangle.
\]

For the minimum, by Theorem 19, \( F(x) - F(gy) \in \text{conv} H(F(x - gy)) \) so that for any \( g \in G \),

\[
\begin{align*}
\|x - gy\| &= \max_{\alpha \in C} \langle F(x - gy), \alpha \rangle \\
&\geq \max_{\alpha \in C} \langle F(x) - F(gy), \alpha \rangle \quad \text{by (12)} \\
&= \max_{\alpha \in C} \langle F(x) - F(y), \alpha \rangle \\
&= \|F(x) - F(y)\|.
\end{align*}
\]
For the maximum, we may assume that \( x, y \in W \) or even in \( F \) since \( \| \cdot \| \) is \( G \)-invariant. By Lemma 18,

\[
\| x - gy \| = \max_{a \in C} (F(x - gy), a) \\
\leq \max_{a \in C} (F(x) + F(-gy), a) \quad \text{by (12) and Lemma 18} \\
= \max_{a \in C} (F(x) + F(-y), a) \\
= \max_{a \in C} (F(x) - \omega F(y), a) \\
= \| F(x) - \omega F(y) \|,
\]

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

**Corollary 26.** Let \( g \) be a real semisimple Lie algebra with Cartan decomposition \( g = \mathfrak{k} + \mathfrak{p} \), where the analytic group of \( \mathfrak{t} \) is \( K \subset G \). For \( x \in \mathfrak{p} \), let \( a_\cdot(x) \) denote the unique element of the singleton set \( \text{Ad}(K)x \cap a_\cdot \), where \( a_\cdot \) is a closed fundamental Weyl chamber. Given \( x, y \in \mathfrak{p} \), if \( z \in \text{Ad}(K)y \), then

\[
\min_{k \in K} \| x - \text{Ad}(k)y \| = \| a_\cdot(x) - a_\cdot(y) \| \\
\max_{k \in K} \| x - \text{Ad}(k)y \| = \| a_\cdot(x) - \omega a_\cdot(y) \| = \| a_\cdot(x) + a_\cdot(-a_\cdot(y)) \|,
\]

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

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

**Theorem 27.** 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

\[
\min_{z \in \text{Conv}_G y} \| x - z \| = \| F(x) - F(F(y)) \|, \\
\max_{z \in \text{Conv}_G y} \| x - z \| = \| F(x) - \omega F(y) \| = \| F(x) + F(-F(y)) \|,
\]

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 \( \text{Ad}(SO(n) \times SO(n)) \)-orbit for the simple Lie algebra \( so_{n,n} \). The question provoked the study in this paper.

**Remark:** Most results in this paper are contained in William C. Hill’s 2001 Auburn University PhD dissertation “\( G \)-invariant norm, an extension of Berezin-Gelfand’s theorem via nonsmooth analysis and applications”.

**References**


