Gradient Flows, Convexity, and Adjoint Orbits

by

Xuhua Liu

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 4, 2012

Keywords: Lie group, Lie algebra, semisimple, reductive, adjoint orbit, gradient flow, numerical range, Hessian, connectedness, convexity.

Approved by

Tin-Yau Tam, Chair, Professor of Mathematics
Randall R. Holmes, Professor of Mathematics
Huajun Huang, Associate Professor of Mathematics

Abstract

This dissertation studies some matrix results and gives their generalizations in the context of semisimple Lie groups. The adjoint orbit is the primary object in our study.

The dissertation consists of four chapters. Chapter 1 is a brief introduction about the interplay between matrix theory and Lie theory.

In Chapter 2 we introduce some structure theory of semisimple Lie groups and Lie algebras. It involves the root space decompositions for complex and real semisimple Lie algebras, Cartan decomposition and Iwasawa decomposition for real semisimple Lie algebras and Lie groups. They play significant roles in our generalizations.

In Chapter 3 we introduce a famous problem on Hermitian matrices proposed by H. Weyl in 1912, which has been completely solved. Motivated by a recent paper of Li et al. [34] we consider a generalized problem in the context of semisimple as well as reductive Lie groups. We give the gradient flow of a function corresponding to the generalized problem. This provides a unified approach to deriving several results in [34].

Chapter 4 is essentially a brief survey on some generalized numerical ranges associated with Lie algebras. The classical numerical range of an $n \times n$ complex matrix is the image of the unit sphere in $\mathbb{C}^n$ under the quadratic form. One of the most beautiful properties is that the numerical range of a matrix is always convex. It is known as the Toeplitz-Hausdorff theorem. We give another proof of the convexity of some generalized numerical range associated with a compact Lie group. The Toeplitz-Hausdorff theorem becomes a special case.

Acknowledgments

This dissertation arose out of my five years of graduate study and research at Auburn University, where I have worked with many nice people whose contributions to the completion of this work deserved special mention. It is a pleasure to convey my gratitude to them all in this acknowledgment.

First of all, my deepest gratitude goes to the Most High God, Who is my Creator and Savior. He also gives me knowledge and strength to complete this work: "(Psalm 127:1) Unless the Lord builds the house, They labor in vain who build it; Unless the Lord guards the city, The watchman stays awake in vain."

I am grateful in every possible way to Professor Tin-Yau Tam for his supervision, advice, and guidance as well as giving me extraordinary experiences of research throughout the five years. He is an outstanding graduate mentor, and I have been particularly impressed by his foresight on success for his students. His mathematical intuition exceptionally inspired and enriched my growth as a student and a researcher. Above all, he provided me steadfast encouragement and support in various ways, not only in academic but also in personal life. He is my spiritual mentor as well. I am indebted to him much more than he knows.

I gratefully acknowledge Professor Randall R. Holmes for his supervision and many efforts and valuable comments on the first completed manuscript of my dissertation. He taught me extraordinarily excellent classes on group theory and representation theory, where he nourished my mathematical maturity that I will benefit from for a lifetime. His rigorous and meticulous style of mathematical writing has been inspiring me in preparing research papers. He is among the best teachers in my student life.

Many thanks go to Dr. Huajun Huang for his valuable advice and supervision. I learned a lot from the Lie group class taught by him. Our discussions in the Lie theory seminar were interesting and helpful.

My parents deserve special mention for their support and encouragement. My father, Xuedun Liu, is the person who cultivated my learning character. My mother, Ziying Guo raised me with her caring love and contributed to my tough character. She would have been very proud to see the completion of my PhD study.

I thank my wife, Yaqin Xiao, for her unselfish love since we knew each other. Her persistent confidence in me was a great source of encouragement, which drove me to move forward steadfastly. Our lovely children Hannah and Joseph have been inspiring me to work hard and efficiently.

Finally, I would like to thank everybody who was important to the successful completion of this dissertation.

Table of Contents

Abstract ii

Acknowledgments iii

1 Introduction 1

2 Structure Theory of Semisimple Lie Groups and Lie Algebras 3

2.1 Smooth Manifolds 3

2.2 Lie Groups and Their Lie Algebras 5

2.3 Complex Semisimple Lie Algebras 9

2.4 Real Forms 10

2.5 Cartan Decomposition 11

2.6 Root Space Decomposition and Iwasawa Decomposition 13

2.7 Weyl Groups 15

3 Gradient Flows for the Minimum Distance to the Sum of Adjoint Orbits 17

3.1 Introduction 17

3.2 Formulation 18

3.3 Gradient Flow 20

3.4 Examples 25

3.5 Special Case: $A \in \mathfrak{p}$, $B \in \mathfrak{k}$ 29

3.6 Global Extrema 41

4 Convexity of Generalized Numerical Ranges Associated with Lie Algebras 43

4.1 Classical Numerical Range 43

4.2 Generalized Numerical Ranges 44

4.3 Generalized Numerical Ranges Associated with Lie Algebras 45

4.3.1 Compact Case 45

4.3.2 Complex Semisimple Case 47

4.3.3 Real Semisimple Case 49

Bibliography 57

Chapter 1

Introduction

This dissertation is a study of some matrix results and their generalizations in the context of semisimple Lie groups. It is universally believed that matrix theory has many applications in various branches of mathematics and sciences. Matrix theory has a close relationship with the theory of Lie groups. For example, the general linear groups $\mathrm{GL}_n(\mathbb{C})$ and $\mathrm{GL}_n(\mathbb{R})$ are Lie groups. Roughly speaking, a Lie group is a smooth manifold which is also a group and in which the group operations are smooth. The tangent space at the identity of a Lie group has a Lie algebra structure, which captures most information about the Lie group via the exponential map. The classical Lie groups are matrix groups. This close connection between matrix theory and Lie theory is beneficial to both fields: matrix theory provides various results which may lead to new developments of Lie theory and, in turn, Lie theory often provides unified approaches to matrix results and thus inspires deeper understanding of them.

The main tools in this dissertation are some important decompositions of Lie algebras and their counterparts for Lie groups. These decompositions are Cartan decomposition and Iwasawa decomposition which are corresponding to Hermitian decomposition (algebra level), polar decomposition (group level), and QR decomposition in matrix theory. They exist for semisimple Lie algebras and Lie groups as well as for reductive Lie algebras and Lie groups.

These decompositions reflect the rich structures of Lie groups. For example, if G is a connected semisimple Lie group with Lie algebra $\mathfrak{g}$ and if θ is a Cartan involution of $\mathfrak{g}$, then $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ is a Cartan decomposition, where $\mathfrak{k}$ and $\mathfrak{p}$ are the $+1$ and -1 eigenspaces of θ, respectively, and the Killing form B on $\mathfrak{g}$ induces a positive definite symmetric bilinear form B_θ on $\mathfrak{g}$ defined by $B_\theta(X, Y) = -B(X, \theta Y)$. The bilinear form B_θ, together with adjoint orbits in $\mathfrak{g}$, enables one to do fruitful analysis on the Lie group G via the exponential map from $\mathfrak{g}$ to G.

In matrix theory, any complex $n \times n$ matrix is an element of $\mathfrak{gl}_n(\mathbb{C})$, the Lie algebra of the general linear group $\mathrm{GL}_n(\mathbb{C})$ and the Lie bracket is given by $[X, Y] = XY - YX$. If we define a Cartan involution θ of $\mathfrak{g}$ by $\theta(X) = -X^*$ for all $X \in \mathfrak{g}$, the corresponding Cartan decomposition is just the Hermitian decomposition and B_θ is (up to a positive scalar multiple) the usual inner product given by $B_\theta(X, Y) = \mathrm{tr}\, XY^*$.

The following is the structure of this dissertation. In Chapter 2, we introduce some structures of semisimple Lie groups and Lie algebras for future reference. They are root space decompositions for complex and real semisimple Lie algebras, Cartan and Iwasawa decomposition for real semisimple Lie algebras and Lie groups. In Chapter 3, we consider a generalization of a famous problem on sum of Hermitian matrices proposed by Weyl in the context of semisimple as well as reductive Lie groups, where we will derive the gradient flow and provide a unified approach to several results in [34]. Chapter 4 is essentially a brief survey on some generalized numerical ranges associated with Lie algebras. The classical numerical range of a complex square matrix is the image of the unit sphere under the quadratic form. One of the most beautiful properties is that the numerical range of a matrix is always convex. We give another proof of the convexity of a generalized numerical range associated with a compact Lie group via a connectedness argument. The main tools are a connectedness result of Atiyah [1] and a Hessian index result of Duistermaat, Kolk and Varadarajan [16].

Chapter 2

Structure Theory of Semisimple Lie Groups and Lie Algebras

In this chapter, we introduce most notations in the dissertation and summarize some structures of semisimple Lie groups and Lie algebras for later reference. Since a Lie group is simultaneously a smooth manifold and a group such that the group operations are smooth, we begin with smooth manifolds, for which our main references are [32] and [50].

2.1 Smooth Manifolds

A *topological manifold* of dimension n is a second countable Hausdorff topological space of which every point has an open neighborhood that is homeomorphic to an open subset of $\mathbb{R}^n$. Let M be a topological manifold of dimension n. A *chart* on M is a pair (U, φ), where $U \subset M$ is open and φ is a homeomorphism of U onto an open subset of $\mathbb{R}^n$. Recall that a map $F : U \to V$, where U and V are open subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, is said to be C^∞ or *smooth* if each of the component functions of F has continuous partial derivatives of all orders. A *smooth structure* on M is a collection of charts $\{(U_\alpha, \varphi_\alpha) : \alpha \in I\}$ such that

(1) $\displaystyle\bigcup_{\alpha \in I} U_\alpha = M$,

(2) $\varphi_\alpha \circ \varphi_\beta^{-1}$ is C^∞ for all $\alpha, \beta \in I$, and

(3) the collection is maximal with respect to (2).

A topological manifold with a smooth structure is called a *smooth manifold*, or simply *manifold* unless otherwise specified. A chart on a manifold is said to be *smooth* if it is an element of its smooth structure.

Let M and N be manifolds. A continuous map $F : M \to N$ is said to be *smooth* if for every $p \in M$, there exist smooth charts (U, φ) containing p and (V, ϕ) containing $F(p)$ such that $F(U) \subset V$ and the composite map $\phi \circ F \circ \varphi^{-1}$ is C^∞ from $\varphi(U)$ to $\phi(V)$. If $N = \mathbb{R}$, F is called a *smooth function* on M if for every $p \in M$, there exists a smooth chart (U, φ) containing p such that $F \circ \varphi^{-1}$ is C^∞. Let $C^\infty(M)$ denote the set of all smooth functions on M. A linear map $v : C^\infty(M) \to \mathbb{R}$ is called a *derivation* at $p \in M$ if it satisfies

$$v(fg) = f(p)v(g) + g(p)v(f), \qquad \forall f, g \in C^\infty(M).$$

The set $T_p(M)$ of all derivations of $C^\infty(M)$ at p forms a vector space called the *tangent space* to M at p. Elements of $T_p(M)$ are called *tangent vectors* at p.

Let $F : M \to N$ be a smooth map and let $p \in M$. The differential of F at p is the linear map $dF_p : T_p(M) \to T_{F(p)}(N)$ defined by

$$dF_p(v)(f) = v(f \circ F), \qquad \forall v \in T_p(M), \forall f \in C^\infty(N).$$

The *rank* of F at $p \in M$ is the rank of dF_p. The smooth map F is an *immersion* if $\operatorname{rank} F = \dim M$ at every $p \in M$. A *submanifold* of M is a subset $S \subset M$ endowed with a manifold topology and a smooth structure, i.e., S is a smooth manifold in its own right, such that the inclusion map $\iota : S \to M$ is an immersion.

The *tangent bundle* $T(M)$ of M is the disjoint union of the tangent spaces at all points of M. The projection map $\pi : T(M) \to M$ is defined by sending each vector in $T_p(M)$ to $p \in M$. The tangent bundle has a natural topology and smooth structure that make it into a manifold such that $\pi : T(M) \to M$ is a smooth map. A *vector field* on M is a map $X : M \to T(M)$ such that $X_p := X(p) \in T_p(M)$ for all $p \in M$. The set of smooth vector fields on M forms in the obvious way a vector space over $\mathbb{R}$; it is also a module over the ring $C^\infty(M)$: if X is a vector field on M and $f \in C^\infty(M)$, then $Xf \in C^\infty(M)$ is defined by $Xf(p) = X_p(f)$. Note that a vector field X on M is $\mathbb{R}$-linear on $C^\infty(M)$ and satisfies

$$X(f \cdot g) = (Xf) \cdot g + f \cdot Xg, \qquad \forall f, g \in C^\infty(M).$$

In other words, X acts as a derivation of the $\mathbb{R}$-algebra $C^\infty(M)$. In fact, derivations of $C^\infty(M)$ can be identified with smooth vector fields: A function $\mathcal{X} : C^\infty(M) \to C^\infty(M)$ is a derivation if and only if it is of the form $\mathcal{X}(f) = Xf$ for some smooth vector field X on M [32, p.86]. If X and Y are smooth vector fields on M, then $X \circ Y : C^\infty(M) \to C^\infty(M)$ need not be a smooth vector field in general, but the Lie bracket $[X, Y] := X \circ Y - Y \circ X$ always is. The space of smooth vector fields on a manifold has the structure of a Lie algebra over $\mathbb{R}$.

2.2 Lie Groups and Their Lie Algebras

A vector space $\mathfrak{g}$ over a field $\mathbb{F}$ with a product $\mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, denoted by $(X, Y) \mapsto [X, Y]$ and called the *Lie bracket* of X and Y, is called a *Lie algebra* over $\mathbb{F}$ if the following three conditions are satisfied:

(1) The Lie bracket is bilinear.

(2) $[X, X] = 0$ for all $X \in \mathfrak{g}$.

(3) The Jacobi identity $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$ holds for all $X, Y, Z \in \mathfrak{g}$.

An example of a Lie algebra is the *general linear algebra* $\mathfrak{gl}(V)$ consisting of all linear operators on a vector space V with the Lie bracket defined by

$$[X, Y] = XY - YX, \quad \forall X, Y \in \mathfrak{gl}(V).$$

Let $\mathfrak{g}$ and $\mathfrak{h}$ be Lie algebras. A linear transformation $\varphi : \mathfrak{g} \to \mathfrak{h}$ is called a *homomorphism* if

$$\varphi([X, Y]) = [\varphi(X), \varphi(Y)], \quad \forall X, Y \in \mathfrak{g}.$$

It follows from the bilinearity and the Jacobi identity that the linear transformation

$$\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$$

given by $\operatorname{ad} X(Y) = [X, Y]$ for all $X, Y \in \mathfrak{g}$ is a Lie algebra homomorphism, called the *adjoint representation* of $\mathfrak{g}$. A subspace $\mathfrak{s}$ of $\mathfrak{g}$ is called a *subalgebra* if $[X, Y] \in \mathfrak{s}$ for all $X, Y \in \mathfrak{s}$; it is called an *ideal* if $[X, Y] \in \mathfrak{s}$ for all $X \in \mathfrak{g}$ and $Y \in \mathfrak{s}$.

A *Lie group* G is both a smooth manifold and a group such that the maps $m : G \times G \to G$ and $i : G \to G$ defined by multiplication and inversion are smooth. For example, the set of all nonsingular complex matrices forms a Lie group, called the *general linear group* and denoted by $\mathrm{GL}_n(\mathbb{C})$. Any closed subgroup of $\mathrm{GL}_n(\mathbb{C})$ is a Lie group, called a *closed linear group.*

Let G be a Lie group. For each $g \in G$, the left translation $L_g : G \to G$ defined by $L_g(h) = gh$ is a diffeomorphism of G. A smooth vector field X on G is *left-invariant* if X is L_g-related to itself for every $g \in G$, i.e.,

$$X \circ L_g = dL_g \circ X, \qquad \forall g \in G.$$

If we regard X as a derivation, left-invariance is expressed by

$$(Xf) \circ L_g = X(f \circ L_g), \qquad \forall f \in C^\infty(G), \forall g \in G.$$

The space of left-invariant smooth vector fields on G is closed under the Lie bracket, and is therefore a Lie algebra $\mathfrak{g}$, called *the Lie algebra of* G. The map $X \mapsto X_e$ is a vector space isomorphism of $\mathfrak{g}$ onto $T_e(G)$. If $X_e, Y_e \in T_e(G)$, let $[X_e, Y_e]$ denote the tangent vector $[X, Y]_e$. The vector space $T_e(G)$, with the composition rule $(X_e, Y_e) \mapsto [X_e, Y_e]$, forms a Lie algebra which is identified with $\mathfrak{g}$.

A smooth map $\varphi : G \to H$ between Lie groups G and H is called a *smooth homomorphism* if it is also a group homomorphism. The differential $d\varphi : \mathfrak{g} \to \mathfrak{h}$ between the corresponding Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$ is a Lie algebra homomorphism, called the *derived homomorphism* of φ.

Let G be a Lie group with Lie algebra $\mathfrak{g}$. A *one-parameter subgroup* of G is a smooth homomorphism $\phi : \mathbb{R} \to G$. It is a consequence of the theorem of existence and uniqueness of solutions of linear ordinary differential equations that the map $\phi \mapsto d\phi(0)$ is a bijection of the set of one-parameter subgroups of G onto $\mathfrak{g}$ [23, p.103]. For each $X \in \mathfrak{g}$, let ϕ_X be the one-parameter subgroup corresponding to X and define the exponential map $\exp : \mathfrak{g} \to G$ by $\exp(X) = \phi_X(1)$. It follows that $\phi_X(t) = \exp(tX)$ and that the one-parameter subgroups are the maps $t \mapsto \exp tX$ for $X \in \mathfrak{g}$. The exponential map for a closed linear group is given by the matrix exponential function [29, p.76]. An important property of the exponential map is its naturality: if $\varphi : G \to H$ is a smooth homomorphism, then $\varphi \circ \exp_{\mathfrak{g}} = \exp_{\mathfrak{h}} \circ d\varphi$.

A submanifold H of G is called a *Lie subgroup* if H is a Lie group with binary operation being that induced by the binary operation on G. A Lie subgroup of G is called a *closed subgroup* if it is a closed subset of G. The following theorem shows a one-to-one correspondence between connected Lie subgroups of a Lie group and subalgebras of its Lie algebra [23, p.112, p.115, p.118].

Theorem 2.1. *Let G be a Lie group with Lie algebra $\mathfrak{g}$. If H is a Lie subgroup of G, then the Lie algebra $\mathfrak{h}$ of H is a subalgebra of $\mathfrak{g}$. Moreover, $\mathfrak{h} = \{X \in \mathfrak{g} : \exp tX \in H \text{ for all } t \in \mathbb{R}\}$. Each subalgebra of $\mathfrak{g}$ is the Lie algebra of exactly one connected Lie subgroup of G.*

For each $g \in G$, let I_g be the inner automorphism of G defined by $x \mapsto gxg^{-1}$. The derived homomorphism of I_g, denoted by $\mathrm{Ad}\, g$, is an automorphism of $\mathfrak{g}$. We thus have

$$\exp(Ad(g)X) = g(\exp X)g^{-1}, \qquad \forall g \in G, \forall X \in \mathfrak{g}.$$

In the special case that G is a closed linear group, we have $\mathrm{Ad}\,(g)X = gXg^{-1}$. Since exp has a smooth inverse in a neighborhood of $e \in G$, if X is small $\mathrm{Ad}\,(g)X$ is smooth as a function from a neighborhood of e to $\mathfrak{g}$. That is, $g \mapsto \mathrm{Ad}\, g$ is smooth from a neighborhood of e into $\mathrm{GL}(\mathfrak{g})$. Moreover $\mathrm{Ad}\, g \circ \mathrm{Ad}\, h = \mathrm{Ad}\,(gh)$ because $I_g \circ I_h = I_{gh}$. Thus the smoothness is valid everywhere on G. Therefore $\mathrm{Ad} : G \to \mathrm{GL}(\mathfrak{g})$ is a smooth homomorphism, called the *adjoint representation* of G. The derived homomorphism of Ad is the adjoint representation $\mathrm{ad}:\mathfrak{g}\to\mathfrak{gl}(\mathfrak{g})$ of $\mathfrak{g}$ [29, p.80]. Consequently we have

$$\mathrm{Ad}\,(\exp X)=\exp(\mathrm{ad}\,X),\qquad \forall X\in\mathfrak{g}.$$

The group $\mathrm{Aut}\,\mathfrak{g}$ of all automorphisms of $\mathfrak{g}$ is a closed subgroup of $\mathrm{GL}(\mathfrak{g})$, hence is a Lie subgroup of $\mathrm{GL}(\mathfrak{g})$. The Lie algebra $\mathrm{Der}\,\mathfrak{g}$ of $\mathrm{Aut}\,\mathfrak{g}$ consists of all derivations of $\mathfrak{g}$ [23, p.127]. Since $\mathrm{ad}\,\mathfrak{g}$ is a subalgebra of $\mathrm{Der}\,\mathfrak{g}$, it corresponds to a connected subgroup $\mathrm{Int}\,\mathfrak{g}$ of $\mathrm{Aut}\,\mathfrak{g}$ [23, p.127], which is generated by $\exp(\mathrm{ad}\,\mathfrak{g})$ and called the *adjoint group* of $\mathfrak{g}$. Since $\exp(\mathrm{ad}\,X)=\mathrm{Ad}\,(\exp X)$ for all $X\in\mathfrak{g}$, we have $\mathrm{Int}\,\mathfrak{g}=\mathrm{Ad}\,G$ if G is connected. The Lie algebra $\mathfrak{g}$ is said to be *compact* if G is compact or, equivalently, the adjoint group $\mathrm{Int}\,\mathfrak{g}$ is compact.

The symmetric bilinear form B on $\mathfrak{g}$ defined by

$$B(X,Y)=\mathrm{tr}\,(\mathrm{ad}\,X\,\mathrm{ad}\,Y),\qquad \forall X,Y\in\mathfrak{g}$$

is called the *Killing form*, which is associative in the sense that

$$B([X,Y],Z)=B(X,[Y,Z]),\quad \forall X,Y,Z\in\mathfrak{g}.$$

If σ is an automorphism of $\mathfrak{g}$, then

$$\mathrm{ad}\,(\sigma X)=\sigma\circ\mathrm{ad}\,X\circ\sigma^{-1}$$

and thus $B(\sigma X,\sigma Y)=B(X,Y)$. In particular, B is $\mathrm{Ad}\,G$-invariant.

A Lie algebra $\mathfrak{g}$ is *Abelian* if $[\mathfrak{g},\mathfrak{g}]=0$; it is *simple* if it is not Abelian and has no nontrivial ideals; it is *solvable* if $D^k\mathfrak{g}=0$ for some k, where $D^0\mathfrak{g}=\mathfrak{g}$ and $D^{k+1}\mathfrak{g}=[D^k\mathfrak{g},D^k\mathfrak{g}]$; it is *nilpotent* if $C_k\mathfrak{g}=0$ for some k, where $C_0\mathfrak{g}=\mathfrak{g}$ and $C_{k+1}\mathfrak{g}=[C_k\mathfrak{g},\mathfrak{g}]$; it is *semisimple* if its (unique) maximal solvable ideal, called the *radical* of $\mathfrak{g}$ and denoted by $\operatorname{Rad}\mathfrak{g}$, is trivial (or, equivalently, its Killing form is nondegenerate); it is *reductive* if its center $\mathfrak{z}(\mathfrak{g}) = \operatorname{Rad}\mathfrak{g}$ (or, equivalently, $[\mathfrak{g}, \mathfrak{g}]$ is semisimple). A Lie algebra is semisimple if and only if it is isomorphic to a direct sum of simple algebras. A Lie group is called *semisimple (simple, reductive, solvable, nilpotent, Abelian)* if its Lie algebra is semisimple (simple, reductive, solvable, nilpotent, Abelian).

2.3 Complex Semisimple Lie Algebras

Let $\mathfrak{g}$ be a complex semisimple Lie algebra. An element $X \in \mathfrak{g}$ is called *nilpotent* if $\operatorname{ad} X$ is a nilpotent endomorphism; it is called *semisimple* if $\operatorname{ad} X$ is diagonalizable. Since $\mathfrak{g}$ is semisimple, it possesses nonzero subalgebras consisting of semisimple elements, which are Abelian and are called *toral subalgebras* of $\mathfrak{g}$ [27, p.35]. The normalizer of a subalgebra $\mathfrak{a}$ of $\mathfrak{g}$ is

$$N_{\mathfrak{g}}(\mathfrak{a}) = \{X \in \mathfrak{g} : \operatorname{ad} X(\mathfrak{a}) \subset \mathfrak{a}\};$$

it is the largest subalgebra of $\mathfrak{g}$ which contains $\mathfrak{a}$ and in which $\mathfrak{a}$ is an ideal. A subalgebra $\mathfrak{h}$ of $\mathfrak{g}$ is called a *Cartan subalgebra* of $\mathfrak{g}$ if it is self-normalizing, i.e., $\mathfrak{h} = N_{\mathfrak{g}}(\mathfrak{h})$, and nilpotent. The Cartan subalgebras of $\mathfrak{g}$ are exactly the maximal toral subalgebras of $\mathfrak{g}$ [27, p.80]. All Cartan subalgebras of $\mathfrak{g}$ are conjugate under the adjoint group $\operatorname{Int}\mathfrak{g}$ of inner automorphisms [27, p.82].

Let $\mathfrak{h}$ be a Cartan subalgebra of $\mathfrak{g}$. Since $\mathfrak{h}$ is Abelian, $\operatorname{ad}_{\mathfrak{g}}\mathfrak{h}$ is a commuting family of semisimple endomorphisms of $\mathfrak{g}$, which are thus simultaneously diagonalizable. In other words, $\mathfrak{g}$ is the direct sum of the subspaces

$$\mathfrak{g}_\alpha = \{X \in \mathfrak{g} : [H, X] = \alpha(H)X \text{ for all } H \in \mathfrak{h}\},$$

where α ranges over the dual $\mathfrak{h}^*$ of $\mathfrak{h}$. Note that $\mathfrak{g}_0 = \mathfrak{h}$ because $\mathfrak{h}$ is self-normalizing. A nonzero $\alpha \in \mathfrak{h}^*$ is called a *root* of $\mathfrak{g}$ relative to $\mathfrak{h}$ if $\mathfrak{g}_\alpha \neq 0$. The set of all roots, denoted by

Δ, is call the *root system* of $\mathfrak{g}$ relative to $\mathfrak{h}$. Thus we have the *root space decomposition*

$$\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{\alpha\in\Delta} \mathfrak{g}_\alpha.$$

The importance of root space decomposition lies on the fact that Δ characterizes $\mathfrak{g}$ completely.

The restriction of the Killing form on $\mathfrak{h}$ is nondegenerate and is given by

$$B(H, H') = \sum_{\alpha\in\Delta} \alpha(H)\alpha(H'), \qquad \forall H, H' \in \mathfrak{h}.$$

Consequently we can identify $\mathfrak{h}$ with $\mathfrak{h}^*$: each $\alpha \in \mathfrak{h}^*$ corresponds a unique $H_\alpha \in \mathfrak{h}$ with $\alpha(H) = B(H_\alpha, H)$ for all $H \in \mathfrak{h}$, and there is a nondegenerate bilinear form $\langle\cdot,\cdot\rangle$ defined on $\mathfrak{h}^*$ by $\langle\alpha, \beta\rangle = B(H_\alpha, H_\beta)$ for all $\alpha, \beta \in \mathfrak{h}^*$. The following is a collection of some properties of the root space decomposition [27, p.36–40]:

(1) Δ is finite and spans $\mathfrak{h}^*$.

(2) If $\alpha, \beta \in \Delta \cup \{0\}$ and $\alpha + \beta \neq 0$, then $B(\mathfrak{g}_\alpha, \mathfrak{g}_\beta) = 0$.

(3) If $\alpha \in \Delta$, then $-\alpha \in \Delta$, but no other scalar multiple of α is a root.

(4) If $\alpha \in \Delta$, then $[\mathfrak{g}_\alpha, \mathfrak{g}_{-\alpha}]$ is one dimensional, with basis H_α.

(5) If $\alpha \in \Delta$, then $\dim \mathfrak{g}_\alpha = 1$.

(6) If $\alpha, \beta \in \Delta$, then $\dfrac{2\langle\beta, \alpha\rangle}{\langle\alpha, \alpha\rangle} \in \mathbb{Z}$ and $\beta - \dfrac{2\langle\beta, \alpha\rangle}{\langle\alpha, \alpha\rangle}\alpha \in \Delta$.

2.4 Real Forms

Let $\mathfrak{g}$ be a complex Lie algebra. Then $\mathfrak{g}$ can be viewed as a real Lie algebra $\mathfrak{g}_\mathbb{R}$, called the *realification* of $\mathfrak{g}$. A *real form* of $\mathfrak{g}$ is a subalgebra $\mathfrak{g}_0$ of $\mathfrak{g}_\mathbb{R}$ such that $\mathfrak{g}_\mathbb{R} = \mathfrak{g}_0 \oplus i\mathfrak{g}_0$; in this case, $\mathfrak{g}$ is called the *complexification* of $\mathfrak{g}_0$. Let $\mathfrak{g}_0$ be a real form of $\mathfrak{g}$. Each $Z \in \mathfrak{g}$

can be uniquely written as $Z = X + iY$ with $X, Y \in \mathfrak{g}_0$. A map $\sigma : \mathfrak{g} \to \mathfrak{g}$ given by $X + iY \mapsto X - iY$ $(X, Y \in \mathfrak{g}_0)$ is called a *conjugation* of $\mathfrak{g}$ with respect to $\mathfrak{g}_0$. It is easy to see that

(1) $\sigma^2 = 1$,

(2) $\sigma(\alpha X) = \bar{\alpha}\sigma(X)$ for all $X \in \mathfrak{g}_0$ and $\alpha \in \mathbb{C}$,

(3) $\sigma(X + Y) = \sigma(X) + \sigma(Y)$ for all $X, Y \in \mathfrak{g}_0$, and

(4) $\sigma[X, Y] = [\sigma X, \sigma Y]$ for all $X, Y \in \mathfrak{g}_0$.

Thus σ is not an automorphism of $\mathfrak{g}$, but it is an automorphism of the real algebra $\mathfrak{g}_\mathbb{R}$. On the other hand, if $\sigma : \mathfrak{g} \to \mathfrak{g}$ satisfies the above properties, the set $\mathfrak{g}_0$ of fixed points of σ is a real form of $\mathfrak{g}$ and σ is the conjugation of $\mathfrak{g}$ with respect to $\mathfrak{g}_0$. Hence there is a one-to-one correspondence between real forms and conjugations of $\mathfrak{g}$.

Let B_0, B, and $B_\mathbb{R}$ denote the Killing forms of the Lie algebras $\mathfrak{g}_0$, $\mathfrak{g}$, and $\mathfrak{g}_\mathbb{R}$, respectively. Then [23, p.180]

$$B_0(X, Y) = B(X, Y), \qquad \forall X, Y \in \mathfrak{g}_0$$
$$B_\mathbb{R}(X, Y) = 2\mathrm{Re}\, B(X, Y), \qquad \forall X, Y \in \mathfrak{g}_\mathbb{R}.$$

Consequently $\mathfrak{g}_0$, $\mathfrak{g}$, and $\mathfrak{g}_\mathbb{R}$ are all semisimple if any of them is.

Every complex semisimple Lie algebra has a compact real form [23, p.181]. The compact real forms of complex simple Lie algebra are list in [23, p.516].

2.5 Cartan Decomposition

Let $\mathfrak{g}$ be a real semisimple Lie algebra, $\mathfrak{g}_\mathbb{C}$ its complexification, σ the conjugation of $\mathfrak{g}_\mathbb{C}$ with respect to $\mathfrak{g}$. A direct decomposition $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ of $\mathfrak{g}$ into a subalgebra $\mathfrak{k}$ and a vector subspace $\mathfrak{p}$ is called a *Cartan decomposition* if there exists a compact real form $\mathfrak{u}$ of

$\mathfrak{g}_{\mathbb{C}}$ such that $\sigma(\mathfrak{u}) \subset \mathfrak{u}$, $\mathfrak{k} = \mathfrak{g} \cap \mathfrak{u}$ and $\mathfrak{p} = \mathfrak{g} \cap i\mathfrak{u}$. If $\mathfrak{u}$ is any compact real form of $\mathfrak{g}_{\mathbb{C}}$ with a conjugation τ, then there exists an automorphism φ of $\mathfrak{g}_{\mathbb{C}}$ such that the compact real form $\varphi(\mathfrak{u})$ is invariant under σ, which guarantees the existence of a Cartan decomposition of $\mathfrak{g}$. In this case, the involutive automorphism $\theta = \sigma\tau$ is called a *Cartan involution* of $\mathfrak{g}$. The bilinear form B_θ of $\mathfrak{g}$ defined by

$$B_\theta(X, Y) = -B(X, \theta Y), \qquad \forall X, Y \in \mathfrak{g}$$

is symmetric and strictly positive definite. The following theorem establishes a one-to-one correspondence between Cartan decompositions of a real semisimple Lie algebra and its Cartan involutions [23, p.184] [39, p.144].

Theorem 2.2. *Let $\mathfrak{g}$ be a real semisimple Lie algebra with the direct sum of subspaces $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$. The following statements are equivalent.*

(1) $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ is a Cartan decomposition.

(2) The map $\theta : X + Y \mapsto X - Y$ $(X \in \mathfrak{k}, Y \in \mathfrak{p})$ is a Cartan involution of $\mathfrak{g}$.

(3) The Killing form is negative definite on $\mathfrak{k}$ and positive definite on $\mathfrak{p}$, and $[\mathfrak{k}, \mathfrak{k}] \subset \mathfrak{k}$, $[\mathfrak{p}, \mathfrak{p}] \subset \mathfrak{k}$, $[\mathfrak{k}, \mathfrak{p}] \subset \mathfrak{p}$.

Let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be a Cartan decomposition. It follows that $\mathfrak{k}$ and $\mathfrak{p}$ are the $+1$ and -1 eigenspaces of θ, respectively, and that $\mathfrak{k}$ is a maximal compactly embedded subalgebra of $\mathfrak{g}$. Moreover, $\mathfrak{k}$ and $\mathfrak{p}$ are orthogonal to each other with respect to both the Killing form B and the inner product B_θ.

In the special case of $\mathfrak{g}$ being a complex semisimple Lie algebra, if $\mathfrak{u}$ is a compact real form of $\mathfrak{g}$, then $\mathfrak{g}_{\mathbb{R}} = \mathfrak{u} \oplus i\mathfrak{u}$ is a Cartan decomposition [23, p.185].

The group level of Cartan decomposition is summarized below [23, p.252] [29, p.362].

Theorem 2.3. *Let G be a noncompact semisimple Lie group with Lie algebra $\mathfrak{g}$. Let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be the Cartan decomposition corresponding to a Cartan involution θ of $\mathfrak{g}$. Let K be the analytic subgroup of G with Lie algebra $\mathfrak{k}$. Then*

(1) K is connected, closed, and contains the center Z of G. Moreover, K is compact if and only if Z is finite. In this case, K is a maximal compact subgroup of G.

(2) There exists an involutive, analytic automorphism Θ of G whose fixed point set is K and whose differential at e is θ.

(3) The map $K \times \mathfrak{p} \to G$ given by $(k, X) \mapsto k \exp X$ is a diffeomorphism onto.

For any $k \in K$, $\mathrm{Ad}\, k$ leaves B invariant because $\mathrm{Ad}\, k \in \mathrm{Aut}\, \mathfrak{g}$; $\mathrm{Ad}\, k$ also leaves $\mathfrak{k}$ invariant because $\mathfrak{k}$ is the Lie algebra of K and hence $\mathrm{Ad}\, k$ leaves invariant the subspace of $\mathfrak{g}$ orthogonal to $\mathfrak{k}$, which is exactly $\mathfrak{p}$. If $X \in \mathfrak{g}$, write $X = X_{\mathfrak{k}} + X_{\mathfrak{p}}$ with $X_{\mathfrak{k}} \in \mathfrak{k}$ and $X_{\mathfrak{p}} \in \mathfrak{p}$ and we see that

$$\mathrm{Ad}\, k(\theta(X)) = \mathrm{Ad}\,(k)X_{\mathfrak{k}} - \mathrm{Ad}\,(k)X_{\mathfrak{p}} = \theta(\mathrm{Ad}\,(k)X_{\mathfrak{k}}) + \theta(\mathrm{Ad}\,(k)X_{\mathfrak{p}}) = \theta(\mathrm{Ad}\,(k)X),$$

i.e., $\mathrm{Ad}\, k$ commutes with θ. Hence $\mathrm{Ad}\, k$ leaves B_θ invariant as well.

2.6 Root Space Decomposition and Iwasawa Decomposition

Let $\mathfrak{g}$ be a real semisimple Lie algebra and $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ a Cartan decomposition with θ the corresponding Cartan involution. The bilinear form B_θ endows $\mathfrak{g}$ with the structure of a finite-dimensional inner product space. For any $X \in \mathfrak{g}$, with respect to B_θ, the adjoint of $\mathrm{ad}\, X$ is $-\mathrm{ad}\, \theta(X)$ [29, p.360]. If $X \in \mathfrak{p}$, then $\mathrm{ad}\, X$ is represented by a symmetric matrix with respect to an orthonormal basis of $\mathfrak{g}$. Thus the elements of $\mathfrak{p}$ are semisimple with real eigenvalues. Let $\mathfrak{a}$ be a maximal Abelian subspace of $\mathfrak{p}$. The commutative family $\mathrm{ad}\, \mathfrak{a}$ is simultaneously diagonalizable. For each real linear functional α on $\mathfrak{a}$, let

$$\mathfrak{g}_\alpha = \{\, X \in \mathfrak{g} : [H, X] = \alpha(H)X \text{ for all } H \in \mathfrak{a} \,\}.$$

It is easy to see that $\theta(\mathfrak{g}_\alpha) = \mathfrak{g}_{-\alpha}$ and $[\mathfrak{g}_\alpha, \mathfrak{g}_\beta] \subset \mathfrak{g}_{\alpha+\beta}$. If $\alpha \neq 0$ and $\mathfrak{g}_\alpha \neq \{0\}$, then α is called a *root* of $\mathfrak{g}$ with respect to $\mathfrak{a}$. Let Σ denote the set of all roots. The simultaneously diagonalization is expressed by $\mathfrak{g} = \mathfrak{g}_0 \oplus \bigoplus_{\alpha\in\Sigma} \mathfrak{g}_\alpha$, which is called the *root space decomposition* of $\mathfrak{g}$ with respect to $\mathfrak{a}$.

For each root α, the set

$$P_\alpha = \{X \in \mathfrak{a} : \alpha(X) = 0\}$$

is a subspace of $\mathfrak{a}$ of codimension 1. The subspaces $P_\alpha\,(\alpha \in \Sigma)$ divide $\mathfrak{a}$ into several open convex cones, called *Weyl chambers*. Fix a Weyl chamber $\mathfrak{a}_+$ and refer it as the *fundamental Weyl chamber*. A root α is *positive* if it is positive on $\mathfrak{a}_+$. Let Σ^+ denote the set of all positive roots. If $\alpha \in \Sigma^+$ and $X \in \mathfrak{g}_\alpha$, write $X = X_\mathfrak{k} + X_\mathfrak{p}$ with $X_\mathfrak{k} \in \mathfrak{k}$ and $X_\mathfrak{p} \in \mathfrak{p}$. Since $[\mathfrak{k}, \mathfrak{p}] \subset \mathfrak{p}$ and $[\mathfrak{p}, \mathfrak{p}] \subset \mathfrak{k}$, for any $H \in \mathfrak{a}$ we have $(\operatorname{ad} H)X_\mathfrak{k} = \alpha(H)X_\mathfrak{p}$ and $(\operatorname{ad} H)X_\mathfrak{p} = \alpha(H)X_\mathfrak{k}$, which imply $(\operatorname{ad} H)^2X_\mathfrak{k} = \alpha(H)^2X_\mathfrak{k}$, $(\operatorname{ad} H)^2X_\mathfrak{p} = \alpha(H)^2X_\mathfrak{p}$, and $\theta(X) = X_\mathfrak{k} - X_\mathfrak{p} \in \mathfrak{g}_{-\alpha}$. For $\alpha \in \Sigma^+$, define

$$\mathfrak{k}_\alpha = \{X \in \mathfrak{k} : (\operatorname{ad} H)^2X = \alpha(H)^2X \text{ for all } H \in \mathfrak{a}\},$$
$$\mathfrak{p}_\alpha = \{X \in \mathfrak{p} : (\operatorname{ad} H)^2X = \alpha(H)^2X \text{ for all } H \in \mathfrak{a}\}.$$

Let $\mathfrak{m}$ be the centralizer of $\mathfrak{a}$ in $\mathfrak{k}$, i.e.,

$$\mathfrak{m} = \{X \in \mathfrak{k} : \operatorname{ad}(X)H = 0 \text{ for all } H \in \mathfrak{a}\}.$$

The following result [36, p.107] is helpful in deriving the Hessian of a smooth function on K (see Lemma 4.5).

Lemma 2.4. *(1)* $\mathfrak{k} = \mathfrak{m} \oplus \sum_{\alpha\in\Sigma^+} \mathfrak{k}_\alpha$ *and* $\mathfrak{p} = \mathfrak{a} \oplus \sum_{\alpha\in\Sigma^+} \mathfrak{p}_\alpha$ *are direct sums whose components are mutually orthogonal under* B_θ*,*

(2) $\mathfrak{g}_\alpha \oplus \mathfrak{g}_{-\alpha} = \mathfrak{k}_\alpha \oplus \mathfrak{p}_\alpha$ *for all* $\alpha \in \Sigma^+$*, and*

(3) $\dim \mathfrak{g}_\alpha = \dim \mathfrak{k}_\alpha = \dim \mathfrak{p}_\alpha$ *for all* $\alpha \in \Sigma^+$*.*

The space $\mathfrak{n} = \bigoplus_{\alpha\in\Sigma^+} \mathfrak{g}_\alpha$ is a subalgebra of $\mathfrak{g}$. If $X \in \bigoplus_{\alpha\in\Sigma^+} \mathfrak{g}_{-\alpha}$, then

$$X = (X + \theta(X)) - \theta(X) \in \mathfrak{k} + \mathfrak{n}.$$

Since $\mathfrak{g}_0 = (\mathfrak{g}_0 \cap \mathfrak{k}) \oplus \mathfrak{a}$, we see that $\mathfrak{g} = \mathfrak{k} + \mathfrak{a} + \mathfrak{n}$. Applying θ we conclude that $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{a} \oplus \mathfrak{n}$, which is called *Iwasawa decomposition* of $\mathfrak{g}$ [23, p.263] [29, p.373].

The following theorem summarizes the group level of Iwasawa decomposition [29, p.374].

Theorem 2.5. *Let G be a semisimple Lie group with Lie algebra $\mathfrak{g}$. Let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{a} \oplus \mathfrak{n}$ be a Iwasawa decomposition. Let K, A, and N be the analytic subgroups of G with Lie algebras $\mathfrak{k}$, $\mathfrak{a}$, and $\mathfrak{n}$, respectively. Then $G = KAN$ and the map $(k, a, n) \mapsto kan$ is a diffeomorphism of $K \times A \times N$ onto G.*

2.7 Weyl Groups

Let the notations be as in Section 2.6. Let $\mathfrak{m}$ and M be the centralizers of $\mathfrak{a}$ in $\mathfrak{k}$ and in K, respectively, and M' the normalizer of $\mathfrak{a}$ in K, i.e.,

$$\mathfrak{m} = \{X \in \mathfrak{k} : \mathrm{ad}\,(X)H = 0 \text{ for all } H \in \mathfrak{a}\}$$
$$M = \{k \in K : \mathrm{Ad}\,(k)H = H \text{ for all } H \in \mathfrak{a}\}$$
$$M' = \{k \in K : \mathrm{Ad}\,(k)\mathfrak{a} \subset \mathfrak{a}\}.$$

Note that M and M' are also the centralizer and normalizer of A in K, respectively, and that they are closed Lie subgroups of K. Moreover, M is a normal subgroup of M', and the quotient group M'/M is finite because M and M' have the same Lie algebra $\mathfrak{m}$ [23, p.284]. The finite group $W(G,A) = M'/M$ is called the (analytically defined) *Weyl group* of G relative to A. For $w = m_w M \in W(G,A)$, the linear map $\mathrm{Ad}\,(m_w) : \mathfrak{a} \to \mathfrak{a}$ does not depend on the choice of $m_w \in M'$ representing w. It follows that $w \mapsto \mathrm{Ad}\,(m_w)$ is a faithful representation of $W(G,A)$ on $\mathfrak{a}$. Thus we may regard $w \in W(G,A)$ as the linear map $\mathrm{Ad}\,(m_w) : \mathfrak{a} \to \mathfrak{a}$ and $W(G,A)$ as a group of linear operators on $\mathfrak{a}$. The Weyl group $W(G,A)$ also acts on $\mathfrak{a}^*$ by $w \cdot \alpha = \alpha \circ w^{-1}$ for all $\alpha \in \mathfrak{a}^*$. The Killing form B is nondegenerate on $\mathfrak{a}$, and thus it induces an isomorphism of $\mathfrak{a}^*$ and $\mathfrak{a}$ by $\lambda \mapsto H_\lambda$ such that

$$\lambda(H) = B(H_\lambda, H), \qquad \forall \lambda \in \mathfrak{a}^*, \forall H \in \mathfrak{a}$$

This isomorphism induces an action of $W(G,A)$ on $\mathfrak{a}^*$ as follows. If we denote $H_{w\cdot\lambda} = w \cdot H_\lambda$ for all $\lambda \in \mathfrak{a}^*$, then for all $H \in \mathfrak{a}$

$$\begin{aligned}(w \cdot \lambda)(H) &= B(H_{w\cdot\lambda}, H) = B(w \cdot H_\lambda, H) = B(\mathrm{Ad}\,(m_w)H_\lambda, H)\\ &= B(H_\lambda, \mathrm{Ad}\,(m_w)^{-1}H) = \lambda(\mathrm{Ad}\,(m_w)^{-1}H)\\ &= (\lambda \circ \mathrm{Ad}\,(m_w)^{-1})(H)\end{aligned}$$

So the Weyl group $W(G,A)$ acts on $\mathfrak{a}^*$ by $w \cdot \lambda = \lambda \circ \mathrm{Ad}\,(m_w)^{-1} := \lambda \circ w^{-1}$.

For each root α, the reflection s_α about the hyperplane $P_\alpha = \{X \in \mathfrak{a} : \alpha(X) = 0\}$, with respect to the Killing form B, is a linear map on $\mathfrak{a}$ given by

$$s_\alpha(H) = H - \frac{2\alpha(H)}{\alpha(H_\alpha)} H_\alpha, \qquad \forall H \in \mathfrak{a},$$

where H_α is the element of $\mathfrak{a}$ representing α, i.e., $\alpha(H) = B(H_\alpha, H)$ for all $H \in \mathfrak{a}$. The group $W(\mathfrak{g}, \mathfrak{a})$ generated by $\{s_\alpha : \alpha \in \Sigma\}$ is called the (algebraically defined) *Weyl group* of $\mathfrak{g}$ relative to $\mathfrak{a}$. When viewed as groups of linear operators on $\mathfrak{a}$, the two Weyl groups $W(G,A)$ and $W(\mathfrak{g}, \mathfrak{a})$ coincide [29, p.383].

Chapter 3

Gradient Flows for the Minimum Distance to the Sum of Adjoint Orbits

This chapter introduces a famous problem on the sum of Hermitian matrices proposed by Weyl. Motivated by a recent paper of Li et al. [34], we study Weyl's problem in the context of semisimple as well as reductive Lie groups and give the gradient flow of a function corresponding to a generalized problem. This provides a unified approach to several results in [34].

3.1 Introduction

Given the eigenvalues of two $n \times n$ Hermitian matrices A and B, a famous problem of Weyl [52] is to give a complete description of the eigenvalues of $C = A + B$. The problem is completely solved and one may see [14, 17, 31] and their references for historical development. Its extension to compact Lie groups is given in [41, Theorem 9.3] and in particular the determination of singular values of the sum of two rectangular complex matrices is given in [41, p.447–450]. The solution of Weyl's problem and the result in [41] are not easy to be used as a checking tool for concrete matrices. For example, 10×10 Hermitian matrices yield too many inequalities [34], according to the Littlewood-Richardson rule. If we denote by $S(H)$ the unitary similarity orbit of a Hermitian matrix H, then Weyl's problem is to find necessary and sufficient conditions for

$$S(C) \subset S(A) + S(B) := \{UAU^{-1} + VBV^{-1} : U, V \in \mathrm{U}(n)\}$$

in terms of the eigenvalues of A, B and C. The set inclusion is equivalent to

$$\min_{U,V\in \mathrm{U}(n)} \|UAU^{-1} + VBV^{-1} - C\|_F = 0,$$

where $\|\cdot\|_F$ is the Frobenius norm on $\mathbb{C}_{n\times n}$.

Given complex matrices A_0, $A_1, \ldots, A_N$, Li et al. [34] studied the more general problem of finding the least squares approximation of A_0 by the sum of matrices from orbits $S(A_1), \ldots, S(A_N)$, i.e.,

$$\min\{\|X_1 + \cdots + X_N - A_0\|_F : (X_1, \ldots, X_N) \in S(A_1) \times \cdots \times S(A_N)\}, \tag{3.1}$$

where the orbits $S(A_i)$, $i = 1, \ldots, N$, are induced by some equivalence class on matrices such as unitary similarity, unitary equivalence, and unitary congruence. In each case, by Fréchet differentiation they derived the gradient flow of a corresponding smooth function, which was then used to design an algorithm to solve the respective optimization problem.

Motivated by the results in [34], we ask whether there might be a unified approach for studying these problems. The purpose of this chapter is to present such an approach in the context of semisimple as well as reductive Lie groups. Examples are given to illustrate our results and their relations to those in [34].

3.2 Formulation

Let G be a connected semisimple Lie group with Lie algebra $\mathfrak{g}$. Let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be the Cartan decomposition corresponding to a Cartan involution θ of $\mathfrak{g}$, where $\mathfrak{k}$ and $\mathfrak{p}$ are the $+1$ and -1 eigenspaces of θ, respectively. The Killing form B of $\mathfrak{g}$ induces a positive definite symmetric bilinear form B_θ on $\mathfrak{g}$ given by $B_\theta(X, Y) = -B(X, \theta Y)$ for all $X, Y \in \mathfrak{g}$. Note that $B_\theta|_{(\mathfrak{k}\times\mathfrak{k})} = -B$ and $B_\theta|_{(\mathfrak{p}\times\mathfrak{p})} = B$, and that $\mathfrak{k}$ and $\mathfrak{p}$ are orthogonal under both B and B_θ [29, p.359]. Given $X \in \mathfrak{g}$, we have the unique decomposition $X = X_\mathfrak{k} + X_\mathfrak{p}$ with $X_\mathfrak{k} \in \mathfrak{k}$ and $X_\mathfrak{p} \in \mathfrak{p}$. Let K be the analytic subgroup of G with Lie algebra $\mathfrak{k}$. Note that K is compact if and only if G has finite center [29, p.362]. We remark that simple classical groups have compact K [23, p.446–455]. Though K may not be compact in general, $\mathrm{Ad}_G K$ (or simply $\mathrm{Ad}\, K$) is compact in the adjoint group $\mathrm{Int}\, \mathfrak{g} = \mathrm{Ad}\, G$ and each $\mathrm{Ad}\, k \in \mathrm{Aut}\, \mathfrak{g}$ is orthogonal with respect to B_θ. Both $\mathfrak{k}$ and $\mathfrak{p}$ are $\mathrm{Ad}\, K$-invariant, i.e., $\mathrm{Ad}\,(k)\mathfrak{k} = \mathfrak{k}$ and $\mathrm{Ad}\,(k)\mathfrak{p} = \mathfrak{p}$ for all $k \in K$.

We cast Problem (3.1) in the context of semisimple Lie groups as follows. Given $A_0, A_1, \ldots, A_N \in \mathfrak{g}$ (not necessarily in $\mathfrak{p}$), find

$$\min_{k_i \in K} \left\| \sum_{i=1}^{N} \mathrm{Ad}\,(k_i) A_i - A_0 \right\|, \tag{3.2}$$

where the norm $\|\cdot\|$ is induced by B_θ, i.e., $\|X\|^2 = B_\theta(X, X)$ for all $X \in \mathfrak{g}$. In other words, we want to find the (minimum) distance between A_0 and the sum of the orbits $\mathrm{Ad}\,(K)A_i$, $i = 1, \ldots, N$. The sum $\mathrm{Ad}\,(K)A_1 + \cdots + \mathrm{Ad}\,(K)A_N$ is a union of orbits since it is invariant under $\mathrm{Ad}\, K$. The minimum is justified since $\mathrm{Ad}\, K$ is compact. Problem (3.2) is very difficult in view of the nontrivial solution to Weyl's problem, which is corresponding to $\mathfrak{sl}_n(\mathbb{C}) = \mathfrak{su}(n) \oplus i\mathfrak{su}(n)$ with $A, B, C \in \mathfrak{p}$ and $\mathfrak{p} = i\mathfrak{su}(n)$ consisting of Hermitian matrices. The solution of O'Shea and Sjmaar [41, Theorem 9.3] is very involved, which is essentially corresponding to a complex semisimple Lie algebra $\mathfrak{g} = \mathfrak{k} \oplus i\mathfrak{k}$, i.e., $\mathfrak{p} = i\mathfrak{k}$. Both are restricted to $A_0, A_1, \cdots A_N \in \mathfrak{p}$. Our goal is not to solve Problem (3.2). Instead we provide its gradient flow and related results. Since the general case does not differ much from the $N = 2$ case, our focus will be on this simpler case.

In Section 3.3, we derive the gradient flow of a smooth function associated to Problem (3.2). In Section 3.4, we show that several results in [34] can be recovered from our general results. In Section 3.5, we consider the special case when $N = 2$ and $A_1 \in \mathfrak{p}$ and $A_2 \in \mathfrak{k}$, for which the minimum and the maximum are given. Finally, some remarks are made for local and global extrema in Section 3.6.

3.3 Gradient Flow

Let the notations be as in Section 3.2. We first define the gradient flow of a smooth function on the analytic Lie subgroup K of G. It has a bi-invariant Riemannian structure [23, p.47] induced by the unique bi-invariant Riemannian structure Q on G such that $Q_e = B_\theta$ [23, p.148 #5]. More precisely, for each $k \in K$, the right translation $R_k : K \to K$ defined by $R_k(h) = hk$ is a diffeomorphism. Its derivative at the point $h \in K$ is denoted by $dR_k : T_h(K) \to T_{hk}(K)$, where $T_h(K)$ denotes the tangent space of K at h. The Riemannian structure on K is given by

$$\langle U, V\rangle_k := Q(U, V) = B_\theta(dR_k^{-1}(U), dR_k^{-1}(V)) = -B(dR_k^{-1}(U), dR_k^{-1}(V))$$

for all $U, V \in T_k(K)$. Note that this structure is bi-invariant since

$$\begin{aligned}
\langle U, V\rangle_k &= -B(\mathrm{Ad}\,(k^{-1})dR_k^{-1}(U), \mathrm{Ad}\,(k^{-1})dR_k^{-1}(V)) \\
&= -B(dL_k^{-1}(U), dL_k^{-1}(V)).
\end{aligned}$$

We simply write $\langle U, V\rangle$ for $\langle U, V\rangle_k$ if there is no danger of confusion. With respect to this Riemannian structure, if $\varphi : K \to \mathbb{R}$ is a smooth function, the gradient $\nabla\varphi_k \in T_k(K)$ of φ at k is given by

$$d\varphi_k(V) = \langle \nabla\varphi_k, V\rangle_k, \qquad V \in T_k(K). \tag{3.3}$$

Since $\langle\cdot,\cdot\rangle_k$ is nondegenerate, $\nabla\varphi_k$ is well defined and thus we have the gradient vector field $\nabla\varphi : K \to T_k(K)$. So the gradient flow of φ becomes

$$\frac{dk}{dt} = -\nabla\varphi_k. \tag{3.4}$$

Now we return to Problem (3.2) and focus on the simpler case when $N = 2$, i.e.,

$$\min_{k,h\in K} \|\mathrm{Ad}\,(k)A + \mathrm{Ad}\,(h)B - C\|, \qquad A, B, C \in \mathfrak{g}. \tag{3.5}$$

The general case is similar. Note that for any $k, h \in K$

$$\begin{aligned}
& \|\mathrm{Ad}\,(k)A + \mathrm{Ad}\,(h)B - C\|^2 \\
= \;& \|\mathrm{Ad}\,(k)A\|^2 + \|\mathrm{Ad}\,(h)B\|^2 + \|C\|^2 + 2B_\theta(\mathrm{Ad}\,(k)A, \mathrm{Ad}\,(h)B) \\
& \quad -2B_\theta(\mathrm{Ad}\,(k)A, C) - 2B_\theta(\mathrm{Ad}\,(h)B, C) \\
= \;& \|A\|^2 + \|B\|^2 + \|C\|^2 + 2B_\theta(\mathrm{Ad}\,(k)A, \mathrm{Ad}\,(h)B) \\
& \quad -2B_\theta(\mathrm{Ad}\,(k)A, C) - 2B_\theta(\mathrm{Ad}\,(h)B, C).
\end{aligned}$$

Thus Problem (3.5) is equivalent to the problem of finding

$$\min_{k,h\in K} f(k, h), \tag{3.6}$$

where the smooth function $f : K \times K \to \mathbb{R}$ is defined by

$$f(k, h) := B_\theta(\mathrm{Ad}\,(k)A, \mathrm{Ad}\,(h)B) - B_\theta(\mathrm{Ad}\,(k)A, C) - B_\theta(\mathrm{Ad}\,(h)B, C). \tag{3.7}$$

Apply the previous discussion on $K \times K$. The tangent space $T_{(k,h)}(K \times K)$ of $K \times K$ at the point (k, h) is endowed with a bi-invariant Riemannian structure given by

$$\langle U, V\rangle = -B_{(\mathfrak{g},\mathfrak{g})}(dR^{-1}_{(k,h)}U, dR^{-1}_{(k,h)}V), \quad U, V \in T_{(k,h)}(K \times K), \tag{3.8}$$

where the Killing form $B_{(\mathfrak{g},\mathfrak{g})}$ is naturally induced by B. With respect to this structure, the gradient $\nabla f_{(k,h)}$ of f at the point (k, h) is given by

$$df_{(k,h)}(V) = \langle\nabla f_{(k,h)}, V\rangle, \qquad V \in T_{(k,h)}(K \times K). \tag{3.9}$$

The gradient flow of (3.7) becomes

$$\frac{d(k,h)}{dt} = -\nabla f_{(k,h)}. \tag{3.10}$$

Theorem 3.1. *The gradient flow of f defined in* (3.7) *is given by*

$$\frac{d(k,h)}{dt} = (-dR_{k(t)}[\theta(C - \mathrm{Ad}\,(h)B), \mathrm{Ad}\,(k)A]_{\mathfrak{k}}, -dR_{h(t)}[\theta(C - \mathrm{Ad}\,(k)A), \mathrm{Ad}\,(h)B]_{\mathfrak{k}}), \tag{3.11}$$

or equivalently,

$$\frac{dk}{dt} = -dR_{k(t)}[\theta(C - \mathrm{Ad}\,(h)B), \mathrm{Ad}\,(k)A]_{\mathfrak{k}}, \tag{3.12}$$
$$\frac{dh}{dt} = -dR_{h(t)}[\theta(C - \mathrm{Ad}\,(k)A), \mathrm{Ad}\,(h)B]_{\mathfrak{k}}. \tag{3.13}$$

When $A, B, C \in \mathfrak{p}$, the gradient flow becomes

$$\frac{d(k,h)}{dt} = (dR_{k(t)}[C - \mathrm{Ad}\,(h)B, \mathrm{Ad}\,(k)A], dR_{h(t)}[C - \mathrm{Ad}\,(k)A, \mathrm{Ad}\,(h)B]). \tag{3.14}$$

When $A, B, C \in \mathfrak{k}$, the gradient flow becomes

$$\frac{d(k,h)}{dt} = (-dR_{k(t)}[C - \mathrm{Ad}\,(h)B, \mathrm{Ad}\,(k)A], -dR_{h(t)}[C - \mathrm{Ad}\,(k)A, \mathrm{Ad}\,(h)B]). \tag{3.15}$$

Proof. Each element $V \in T_{(k,h)}(K \times K)$ is of the form $V = dR_{(k,h)}(X, Y)$ for some unique $(X, Y) \in (\mathfrak{k}, \mathfrak{k})$. The curve $e^{t(X,Y)}(k, h)$ passes through (k, h) with tangent vector V. Note that $B([X, Y], Z) = B(X, [Y, Z])$ for all $X, Y, Z \in \mathfrak{g}$ [23, p.131], and that $\mathfrak{k}$ and $\mathfrak{p}$ are orthogonal under B. We have

$$
\begin{aligned}
df_{(k,h)}(V) &= \frac{d}{dt}\Big|_{t=0} f(e^{t(X,Y)}(k,h)) = \frac{d}{dt}\Big|_{t=0} f((e^{tX}k, e^{tY}h)) \\
&= \frac{d}{dt}\Big|_{t=0} \{B_\theta(\mathrm{Ad}\,(e^{tX}k)A, \mathrm{Ad}\,(e^{tY}h)B) - B_\theta(\mathrm{Ad}\,(e^{tX}k)A, C) \\
&\qquad - B_\theta(\mathrm{Ad}\,(e^{tY}h)B, C)\} \\
&= -B([X, \mathrm{Ad}\,(k)A], \theta\mathrm{Ad}\,(h)B) - B([Y, \mathrm{Ad}\,(h)B], \theta\mathrm{Ad}\,(k)A) \\
&\qquad + B([X, \mathrm{Ad}\,(k)A], \theta C) + B([Y, \mathrm{Ad}\,(h)B], \theta C) \\
&= -B([\mathrm{Ad}\,(k)A, \theta\mathrm{Ad}\,(h)B], X) - B([\mathrm{Ad}\,(h)B, \theta\mathrm{Ad}\,(k)A], Y) \\
&\qquad - B([\theta C, \mathrm{Ad}\,(k)A], X) - B([\theta C, \mathrm{Ad}\,(h)B], Y) \\
&= -B_{(\mathfrak{g},\mathfrak{g})}([\theta(C - \mathrm{Ad}\,(h)B), \mathrm{Ad}\,(k)A], [\theta(C - \mathrm{Ad}\,(k)A), \mathrm{Ad}\,(h)B], X \oplus Y) \\
&= \langle dR_{(k,h)}([\theta(C - \mathrm{Ad}\,(h)B), \mathrm{Ad}\,(k)A]_{\mathfrak{k}}, [\theta(C - \mathrm{Ad}\,(k)A), \mathrm{Ad}\,(h)B]_{\mathfrak{k}}), V\rangle \quad \text{by (3.8)}
\end{aligned}
$$

Thus

$$
\nabla f_{(k,h)} = (dR_k[\theta(C - \mathrm{Ad}\,(h)B), \mathrm{Ad}\,(k)A]_{\mathfrak{k}}, dR_h[\theta(C - \mathrm{Ad}\,(k)A), \mathrm{Ad}\,(h)B]_{\mathfrak{k}}) \tag{3.16}
$$

and the gradient flow takes the desired form (3.11). The results for $A, B, C \in \mathfrak{p}$ and $A, B, C \in \mathfrak{k}$ follow from the facts that $\mathfrak{p}$ and $\mathfrak{k}$ are $\mathrm{Ad}\,K$-invariant, and that $[\mathfrak{p}, \mathfrak{p}] \subset \mathfrak{k}$ and $[\mathfrak{k}, \mathfrak{k}] \subset \mathfrak{k}$ [29, p.359], and that $\mathfrak{k}$ and $\mathfrak{p}$ are the $+1$ and -1 eigenspaces of θ, respectively. □

Similarly we have the following result for the general case.

Theorem 3.2. *The gradient flow of the smooth function associated to Problem* (3.2) *is given by the following system of differential equations*

$$
\frac{dk_i}{dt} = -dR_{k_i(t)}[\theta(A_0 - \sum_{j \neq i} \mathrm{Ad}\,(k_j)A_j), \mathrm{Ad}\,(k_i)A_i]_{\mathfrak{k}}, \quad i = 1, \ldots, N. \tag{3.17}
$$

Remark 3.3. The above results in this section are also true for reductive Lie groups (proofs are skipped), which are members of the Harish-Chandra class [29, p.446]. More precisely, a *reductive Lie group* is a 4-tuple (G, K, θ, B), where G is a Lie group, K is a compact subgroup of G, θ is a Lie algebra involution of the Lie algebra $\mathfrak{g}$ of G, and B is an $\operatorname{Ad} G$-invariant, θ-invariant, nondegenerate symmetric bilinear form on $\mathfrak{g}$ such that

(1) $\mathfrak{g}$ is reductive, i.e., $\mathfrak{g} = \mathfrak{z} \oplus [\mathfrak{g}, \mathfrak{g}]$, where $\mathfrak{z}$ is the center of $\mathfrak{g}$ and $[\mathfrak{g}, \mathfrak{g}]$ is semisimple,

(2) $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$, where $\mathfrak{k}$ and $\mathfrak{p}$ are the $+1$ and -1 eigenspaces of θ, respectively, and $\mathfrak{k}$ is also the Lie algebra of K,

(3) $\mathfrak{k}$ and $\mathfrak{p}$ are orthogonal with respect to B, and B is negative definite on $\mathfrak{k}$ and positive definite on $\mathfrak{p}$,

(4) the map $K \times \exp \mathfrak{p} \to G$ given by multiplication is a diffeomorphism onto,

(5) for every $g \in G$, the automorphism $\operatorname{Ad} g$ of $\mathfrak{g}$, extended to the complexification $\mathfrak{g}^{\mathbb{C}}$ of $\mathfrak{g}$, is contained in $\operatorname{Int} \mathfrak{g}^{\mathbb{C}}$, and

(6) the semisimple connected subgroup G_{ss} of G with Lie algebra $[\mathfrak{g}, \mathfrak{g}]$ has finite center.

Example 3.4. Let G be a semisimple Lie group with finite center, let B be the Killing form on the Lie algebra $\mathfrak{g}$ of G, let $\theta : \mathfrak{g} \to \mathfrak{g}$ be a Cartan involution, let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be the Cartan decomposition with respect to θ, and let K be the analytic subgroup of G with Lie algebra $\mathfrak{k}$. Then (G, K, θ, B) is a reductive Lie group.

Remark 3.5. The gradient flow method has been used for other studies. For instance, a variety of algorithms in numerical analysis can be approached through gradient flows on adjoint orbits associated with semisimple Lie groups [9]. The recent review [43] gives a comprehensive account on the foundations of gradient flows on Riemannian manifolds including new applications to quantum control. See also [3, 8, 12, 24, 47]. See [53] for some study involving the sum of adjoint orbits associated with a compact connected Lie group.

3.4 Examples

In this section, we show by examples that several results in [34] can be recovered from our general results. Fréchet derivative is the main tool in [34] to derive various gradient flows. Our approach in Theorem 3.1 is intrinsic, i.e., no ambient space is required as in Fréchet differentiation.

Example 3.6. Consider the reductive group $G = \mathrm{GL}_n(\mathbb{C})$ with Lie algebra $\mathfrak{g} = \mathfrak{gl}_n(\mathbb{C})$ and $\theta(X) = -X^*$. Then $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ is just the Hermitian decomposition with $\mathfrak{k} = \mathfrak{u}(n)$ consisting of skew-Hermitian matrices and $\mathfrak{p} = i\mathfrak{u}(n)$ consisting of Hermitian matrices. Let the nondegenerate symmetric bilinear form be defined by $B(X,Y) = \mathrm{Re}\,\mathrm{tr}\, XY$ for all $X, Y \in \mathfrak{g}$. Now $B_\theta(X,Y) = \mathrm{Re}\,\mathrm{tr}\, XY^*$ and the norm induced by B_θ is the Frobenius norm. Problem (3.5) is then

$$\min_{U,V \in \mathrm{U}(n)} \|UAU^* + VBV^* - C\|_F. \tag{3.18}$$

By Remark 3.3 (we cannot directly apply Theorem 3.1 since $\mathfrak{gl}_n(\mathbb{C})$ is not semisimple) the associated gradient flow becomes

$$\begin{aligned} \frac{dU}{dt} &= -dR_U[\theta(C - \mathrm{Ad}\,(V)B), \mathrm{Ad}\,(U)A]_{\mathfrak{k}} = [(C^* - VB^*V^*), UAU^*]_{\mathfrak{k}} U, \\ \frac{dV}{dt} &= -dR_V[\theta(C - \mathrm{Ad}\,(U)A), \mathrm{Ad}\,(V)B]_{\mathfrak{k}} = [(C^* - UA^*U^*), VBV^*]_{\mathfrak{k}} V, \end{aligned}$$

which are exactly the formulae in [34, Section 2.4].

Example 3.7. Consider the group $G = \mathrm{U}_{p,q}$ [29, p.115], whose Lie algebra is

$$\mathfrak{u}_{p,q} = \left\{ \begin{pmatrix} X_1 & Y \\ Y^* & X_2 \end{pmatrix} : X_1 \in \mathfrak{u}(p), X_2 \in \mathfrak{u}(q), Y \in \mathbb{C}_{p\times q} \right\}.$$

Let $\theta(X) = I_{p,q} X I_{p,q}$ where $I_{p,q} = (-I_p) \oplus I_q$. Then we have

$$
\begin{aligned}
\mathfrak{k} &= \left\{ \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} : X_1 \in \mathfrak{u}(p), X_2 \in \mathfrak{u}(q) \right\}, \\
\mathfrak{p} &= \left\{ \begin{pmatrix} 0 & Y \\ Y^* & 0 \end{pmatrix} : Y \in \mathbb{C}_{p\times q} \right\}, \\
K &= \mathrm{U}(p) \times \mathrm{U}(q) = \left\{ \begin{pmatrix} U & 0 \\ 0 & V^* \end{pmatrix} : U \in \mathrm{U}(p), V^* \in \mathrm{U}(q) \right\}.
\end{aligned}
$$

The group action of K on $\mathfrak{p}$ is given by

$$
\hat{A} := \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} \mapsto \begin{pmatrix} U & 0 \\ 0 & V^* \end{pmatrix} \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} \begin{pmatrix} U^* & 0 \\ 0 & V \end{pmatrix} = \begin{pmatrix} 0 & UAV \\ (UAV)^* & 0 \end{pmatrix},
$$

under which the orbit of $\hat{A} \in \mathfrak{p}$ is the set

$$
\mathrm{Ad}\,(K)\hat{A} = \left\{ \begin{pmatrix} 0 & UAV \\ (UAV)^* & 0 \end{pmatrix} : U \in \mathrm{U}(p), V \in \mathrm{U}(q) \right\}.
$$

We set $B(X,Y) := \mathrm{Re}\,\mathrm{tr}\, XY$. For $\hat{A}, \hat{B}, \hat{C} \in \mathfrak{p}$, the minimization problem

$$
\min_{k,h\in K} \|\mathrm{Ad}\,(k)\hat{A} + \mathrm{Ad}\,(h)\hat{B} - \hat{C}\| \tag{3.19}
$$

is equivalent to

$$
\min_{U,X\in \mathrm{U}(p), V,Y\in \mathrm{U}(q)} \|UAV + XBY - C\|_F, \tag{3.20}
$$

which is studied in [34, Section 3.1]. So we want to minimize the function

$$
f(U,V,X,Y) := \mathrm{Re}\,\mathrm{tr}\,(UAV(XBY-C)^* - XBYC^*).
$$

Then from (3.14) the associated gradient flow to the minimization problem (3.19) becomes

$$\frac{dk}{dt} = dR_{k(t)}[\hat{C} - \mathrm{Ad}\,(h)\hat{B}, \mathrm{Ad}\,(k)\hat{A}], \tag{3.21}$$

$$\frac{dh}{dt} = dR_{h(t)}[\hat{C} - \mathrm{Ad}\,(k)\hat{A}, \mathrm{Ad}\,(h)\hat{B}]. \tag{3.22}$$

Set $k = U \oplus V^*, h = X \oplus Y^* \in K$. Then (3.21) and (3.22) become

$$\begin{aligned}
\frac{dU}{dt} &= -2\{(UAV)(C - XBY)^*\}_{\mathfrak{u}(p)}U, \\
\frac{dV^*}{dt} &= -2\{(UAV)^*(C - XBY)\}_{\mathfrak{u}(q)}V^*, \\
\frac{dX}{dt} &= -2\{(XBY)(C - UAV)^*\}_{\mathfrak{u}(p)}X, \\
\frac{dY^*}{dt} &= -2\{(XBY)^*(C - UAV)\}_{\mathfrak{u}(q)}Y^*,
\end{aligned}$$

which match the formulae in [34, Section 3.1].

Example 3.8. Consider the simple Lie algebra $\mathfrak{g} = \mathfrak{sp}_n(\mathbb{R})$ and let $\theta(A) = -A^\top$ for all $A \in \mathfrak{g}$. Then we have

$$\begin{aligned}
\mathfrak{sp}_n(\mathbb{R}) &= \left\{ \begin{pmatrix} A_1 & A_2 \\ A_3 & -A_1^\top \end{pmatrix} : A_2^\top = A_2,\ A_3^\top = A_3,\ A_1, A_2, A_3 \in \mathbb{R}_{n\times n} \right\} \\
K &= \left\{ \begin{pmatrix} U_1 & U_2 \\ -U_2 & U_1 \end{pmatrix} : U_1^\top U_1 + U_2^\top U_2 = I,\ U_1^\top U_2 = U_2^\top U_1,\ U_1, U_2 \in \mathbb{R}_{n\times n} \right\} \\
\mathfrak{k} &= \left\{ \begin{pmatrix} A_1 & A_2 \\ -A_2 & A_1 \end{pmatrix} : A_1^\top = -A_1,\ A_2^\top = A_2,\ A_1, A_2 \in \mathbb{R}_{n\times n} \right\} \\
\mathfrak{p} &= \left\{ \begin{pmatrix} A_1 & A_2 \\ A_2 & -A_1 \end{pmatrix} : A_1^\top = A_1,\ A_2^\top = A_2,\ A_1, A_2 \in \mathbb{R}_{n\times n} \right\}.
\end{aligned}$$

As in [44], we identify K with $\mathrm{U}(n)$ through the map $\gamma : K \to \mathrm{U}(n)$ defined by

$$\gamma \begin{pmatrix} U_1 & U_2 \\ -U_2 & U_1 \end{pmatrix} = U_1 + iU_2.$$

The map γ preserves matrix multiplication as well as addition. We identify $\mathfrak{k}$ with $\mathfrak{u}(n)$ in the same way. Let S denote the space of $n \times n$ complex symmetric matrices. We identify $\mathfrak{p}$ with S via the map $\delta : \mathfrak{p} \to S$ defined by

$$\delta \begin{pmatrix} A_1 & A_2 \\ A_2 & -A_1 \end{pmatrix} = A_2 + iA_1.$$

Note that $U^{-1} = U^\top$ for all $U \in K$ and that for all $A \in \mathfrak{p}$,

$$\begin{aligned} \delta(\mathrm{Ad}\,(U)A) &= \delta \left[\begin{pmatrix} U_1 & U_2 \\ -U_2 & U_1 \end{pmatrix} \begin{pmatrix} A_1 & A_2 \\ A_2 & -A_1 \end{pmatrix} \begin{pmatrix} U_1 & U_2 \\ -U_2 & U_1 \end{pmatrix}^{-1} \right] \\ &= (U_1 + iU_2)(A_2 + iA_1)(U_1 + iU_2)^\top. \end{aligned}$$

Hence with these identifications, the adjoint action of K on $\mathfrak{p}$ corresponds to the map $A \mapsto UAU^\top$ for $A \in S$ and $U \in \mathrm{U}(n)$.

For $A, B, C \in S$, the minimization problem [34, Section 4.1]

$$\min_{U,V \in \mathrm{U}(n)} \|UAU^\top + VBV^\top - C\|$$

corresponds to Problem (3.5) with $\mathfrak{g} = \mathfrak{sp}_n(\mathbb{R})$ and $A, B, C \in \mathfrak{p}$. The associated gradient flow is given by (3.14):

$$
\begin{aligned}
\frac{dU}{dt} &= -[\mathrm{Ad}\,(U)A, C - \mathrm{Ad}\,(V)B]U \\
&= -(UAU^\top \tilde{C} - \tilde{C}UAU^\top)U, \quad (\text{with } \tilde{C} = C - VBV^\top) \\
\frac{dV}{dt} &= -[\mathrm{Ad}\,(V)B, C - \mathrm{Ad}\,(U)A]V \\
&= -(VBV^\top \tilde{C} - \tilde{C}VBV^\top)V, \quad (\text{with } \tilde{C} = C - UAU^\top).
\end{aligned}
$$

Remark 3.9. The results in [34, Section 4.1] are for all $A, B, C \in \mathbb{C}_{n\times n}$. When $A, B, C \in S$, our results in Example 3.8 match [34].

3.5 Special Case: $A \in \mathfrak{p}$, $B \in \mathfrak{k}$

We now consider the special case of Problem (3.5) when $A \in \mathfrak{p}$ and $B \in \mathfrak{k}$, or vice versa. The consideration is mainly motivated by [34, Section 2.3] in which the reductive $\mathfrak{gl}_n(\mathbb{C})$ is studied. We first consider the semisimple case and then make a remark on the reductive case. Let $\mathfrak{a}$ be a maximal Abelian subspace of $\mathfrak{p}$. Let $W_\mathfrak{a} = M'/M$ be the Weyl group of G relative to $\mathfrak{a}$, where

$$
\begin{aligned}
M &= \{k \in K : \mathrm{Ad}\,(k)H = H \text{ for all } H \in \mathfrak{a}\}, \\
M' &= \{k \in K : \mathrm{Ad}\,(k)H \in \mathfrak{a} \text{ for all } H \in \mathfrak{a}\}.
\end{aligned}
$$

Since M' and M have the same Lie algebra [23, p.284], $W_\mathfrak{a}$ is a finite group. For each $w = k_w M \in W_\mathfrak{a}$, $\mathrm{Ad}\,k_w : \mathfrak{a} \to \mathfrak{a}$ is independent of the representative $k_w \in M'$ so that we may regard $W_\mathfrak{a}$ as a subgroup of $\mathrm{GL}(\mathfrak{a})$. The root space decomposition of $\mathfrak{g}$ with respect to $\mathfrak{a}$ is

$$
\mathfrak{g} = \mathfrak{g}_0 \oplus \sum_{\alpha \in \Sigma} \mathfrak{g}_\alpha,
$$

where Σ is the root system and

$$\mathfrak{g}_\alpha = \{X \in \mathfrak{g} : \operatorname{ad} H(X) = \alpha(H)X \text{ for all } H \in \mathfrak{a}\}$$

for $\alpha \in \mathfrak{a}^*$. The hyperplanes $P_\alpha = \{H \in \mathfrak{a} : \alpha(H) = 0\}$ for $\alpha \in \Sigma$ divide $\mathfrak{a}$ into finitely many open convex cones, which are called *Weyl chambers*. The Weyl group acts transitively on the Weyl chambers. Fix a Weyl chamber $\mathfrak{a}_+$ and refer it as the *fundamental Weyl chamber*, whose opposite Weyl chamber is then $-\mathfrak{a}_+$. A root α is called *positive* if it is positive on $\mathfrak{a}_+$, and a positive root is called *simple* if it is not the sum of two positive roots. The set of simple roots is denoted by Δ. For each root α the reflection s_α about the hyperplane P_α in $\mathfrak{a}$, with respect to B_θ, is a linear map given by

$$s_\alpha(H) = H - \frac{\alpha(H)}{\alpha(H_\alpha)} H_\alpha, \qquad \text{for all } H \in \mathfrak{a},$$

where $H_\alpha \in \mathfrak{a}$ is the element representing α, i.e., $\alpha(H) = B_\theta(H, H_\alpha)$ for all $H \in \mathfrak{a}$. It is known that the Weyl group $W_\mathfrak{a}$ is generated by the simple reflections s_α for $\alpha \in \Delta$ [29, Proposition 2.62, Theorem 6.57]. For each $w \in W_\mathfrak{a}$, define the *length* of w to be the smallest integer l such that w can be expressed as a product of l simple reflections. There is a unique element of maximal length [28, Section 1.8], which we denote by $\omega_\mathfrak{a}$ and call the *longest element* of $W_\mathfrak{a}$. This element is also uniquely characterized as the element in $W_\mathfrak{a}$ that sends the fundamental Weyl chamber $\mathfrak{a}_+$ to $-\mathfrak{a}_+$. So it is also known as the *opposition element* [42, p.88]. For example, when $\mathfrak{g} = \mathfrak{sl}_n(\mathbb{C})$ and $\mathfrak{a}$ consists of diagonal matrices in $\mathfrak{p} = i\mathfrak{su}(n)$, let $\mathfrak{a}_+$ be the set of diagonal matrices with non-increasing diagonal entries of zero trace. Then we have

$$\omega_\mathfrak{a}(\operatorname{diag}(a_1, a_2, \ldots, a_n)) = \operatorname{diag}(a_n, \ldots, a_2, a_1) \in -\mathfrak{a}_+.$$

There is another Weyl group associated with K that is different from the Weyl group $W_\mathfrak{a}$ of $(\mathfrak{g}, \mathfrak{a})$. By the proof of [30, Proposition 2.3], we have $K = K_0 Z$, where K_0 is compact semisimple and Z is the center of K. Let T_0 be any maximal Abelian subgroup of K_0 and $\mathfrak{t}_0$ its Lie algebra. The Weyl group of (K_0, T_0), defined to be the quotient of the normalizer of T_0 in K modulo T_0, acts by automorphisms of T_0, hence by invertible operators on $\mathfrak{t}_0$ and the maximal Abelian subalgebra $\mathfrak{t} = \mathfrak{t}_0 \oplus \mathfrak{z}$ of $\mathfrak{k}$, where $\mathfrak{z}$ is the Lie algebra of Z. We therefore define in this way the Weyl group $W_\mathfrak{t}$ for a fixed maximal Abelian subalgebra $\mathfrak{t}$ of $\mathfrak{k}$. A fundamental Weyl chamber $(\mathfrak{t}_0)_+$ in $\mathfrak{t}_0$ gives a corresponding fundamental Weyl chamber $\mathfrak{t}_+ = (\mathfrak{t}_0)_+ \oplus \mathfrak{z}$ in $\mathfrak{t}$. The longest element $w_\mathfrak{t}$ of $W_\mathfrak{t}$ is the one that sends the fundamental Weyl chamber $\mathfrak{t}_+$ to $-\mathfrak{t}_+$.

Let the notations be as above. Note that $\mathfrak{p} = \mathrm{Ad}\,(K)\mathfrak{a}$ [29, p.378] and the Weyl groups defined above can be viewed as subgroups of $\mathrm{Ad}\,K$. Since $A, C - \theta C \in \mathfrak{p}$, there exist $k', u \in K$ such that $\mathrm{Ad}\,(k')A \in \mathfrak{a}_+$ and $\mathrm{Ad}\,(u)(C - \theta C)/2 \in \mathfrak{a}_+$. Since $B, C + \theta C \in \mathfrak{k}$, there exist $h', v \in K$ such that $\mathrm{Ad}\,(h')B \in \mathfrak{t}_+$ and $\mathrm{Ad}\,(v)(C + \theta C)/2 \in \mathfrak{t}_+$.

Lemma 3.10. *Let $X, Y \in \mathfrak{a}$ (respectively, $\mathfrak{t}$). Suppose that $\omega_1 \cdot X$ and Y are in the same Weyl chamber and that $\omega_2 \cdot X$ and Y are in opposite Weyl chambers. Then*

$$\|Y - \omega_1 \cdot X\| \le \|Y - \omega \cdot X\| \le \|Y - \omega_2 \cdot X\|$$

for all $\omega \in W_\mathfrak{a}$ (respectively, $W_\mathfrak{t}$).

Proof. Note that $W_\mathfrak{a}$ and $W_\mathfrak{t}$ can be viewed as subgroups of $\mathrm{Ad}\,K$. The Killing form is $\mathrm{Ad}\,K$-invariant, so it is both $W_\mathfrak{a}$-invariant and $W_\mathfrak{t}$-invariant. Thus the B_θ-induced norm $\|\cdot\|$, which is convex, is also both $W_\mathfrak{a}$-invariant and $W_\mathfrak{t}$-invariant. The lemma then follows from [25, Corollary 3.10, Proposition 2.8]. □

For the special case when $A \in \mathfrak{p}$ and $B \in \mathfrak{k}$, the minimization problem (3.5) has a solution. The following result extends [34, Theorem 2.1].

Theorem 3.11. *Let $A \in \mathfrak{p}$, $B \in \mathfrak{k}$, and $C \in \mathfrak{g}$. Let $k', u \in K$ such that $\mathrm{Ad}\,(k')A \in \mathfrak{a}_+$ and $\mathrm{Ad}\,(u)(C - \theta C)/2 \in \mathfrak{a}_+$, and let $h', v \in K$ such that $\mathrm{Ad}\,(h')B \in \mathfrak{t}_+$ and $\mathrm{Ad}\,(v)(C + \theta C)/2 \in \mathfrak{t}_+$.*

Then

$$\begin{aligned}
&\min_{k,h\in K} \|\mathrm{Ad}\,(k)A + \mathrm{Ad}\,(h)B - C\|^2 \\
= \;& \|\mathrm{Ad}\,(u^{-1}k')A + \mathrm{Ad}\,(v^{-1}h')B - C\|^2 \qquad (3.23)\\
= \;& \|\mathrm{Ad}\,(u^{-1}k')A - (C-\theta C)/2\|^2 \\
&\quad + \|\mathrm{Ad}\,(v^{-1}h')B - (C+\theta C)/2\|^2
\end{aligned}$$

and

$$\begin{aligned}
&\max_{k,h\in K} \|\mathrm{Ad}\,(k)A + \mathrm{Ad}\,(h)B - C\|^2 \\
= \;& \|\mathrm{Ad}\,(u^{-1})\omega_{\mathfrak{a}}^{-1}\cdot \mathrm{Ad}\,(k')A + \mathrm{Ad}\,(v^{-1})\omega_{\mathfrak{t}}^{-1}\cdot \mathrm{Ad}\,(h')B - C\|^2 \qquad (3.24)\\
= \;& \|\mathrm{Ad}\,(u^{-1})\omega_{\mathfrak{a}}^{-1}\cdot \mathrm{Ad}\,(k')A - (C-\theta C)/2\|^2 + \\
&\quad \|\mathrm{Ad}\,(v^{-1})\omega_{\mathfrak{t}}^{-1}\cdot \mathrm{Ad}\,(h')B - (C+\theta C)/2\|^2,
\end{aligned}$$

where $\omega_{\mathfrak{a}}$ and $\omega_{\mathfrak{t}}$ are the longest elements of the Weyl groups $W_{\mathfrak{a}}$ and $W_{\mathfrak{t}}$, respectively.

Proof. Since $\mathfrak{k}$ and $\mathfrak{p}$ are orthogonal under B_θ and $\mathfrak{k}$ and $\mathfrak{p}$ are invariant under $\mathrm{Ad}\,K$, we have

$$\begin{aligned}
&\|\mathrm{Ad}\,(k)A + \mathrm{Ad}\,(h)B - C\|^2 \\
= \;& \|(\mathrm{Ad}\,(k)A - (C-\theta C)/2) + (\mathrm{Ad}\,(h)B - (C+\theta C)/2)\|^2 \\
= \;& \|\mathrm{Ad}\,(k)A - (C-\theta C)/2\|^2 + \|\mathrm{Ad}\,(h)B - (C+\theta C)/2\|^2.
\end{aligned}$$

So

$$\begin{aligned}
&\min_{k,h\in K} \|\mathrm{Ad}\,(k)A + \mathrm{Ad}\,(h)B - C\|^2 \\
= \;& \min_{k\in K} \|\mathrm{Ad}\,(k)A - (C-\theta C)/2\|^2 + \min_{h\in K} \|\mathrm{Ad}\,(h)B - (C+\theta C)/2\|^2
\end{aligned}$$

and

$$\begin{aligned}
& \max_{k,h\in K} \|\mathrm{Ad}\,(k)A + \mathrm{Ad}\,(h)B - C\|^2 \\
= & \max_{k\in K} \|\mathrm{Ad}\,(k)A - (C-\theta C)/2\|^2 + \max_{h\in K} \|\mathrm{Ad}\,(h)B - (C+\theta C)/2\|^2.
\end{aligned}$$

Noting that $W_{\mathfrak{a}} \subset \mathrm{Ad}\,K$ and $W_{\mathfrak{t}} \subset \mathrm{Ad}\,K$ and that the longest Weyl group element maps a Weyl chamber to its opposite chamber, the result then follows from Lemma 3.10. $\square$

Example 3.12. We would like to use $\mathfrak{sl}_n(\mathbb{R}) = \mathfrak{k} \oplus \mathfrak{p}$ with $n = 2m \geq 2$ to illustrate Theorem 3.11. Here $\mathfrak{k} = \mathfrak{so}(n)$ and $\mathfrak{p}$ is the space of $n \times n$ real symmetric matrices of zero trace. Thus $K = \mathrm{SO}(n)$. We may choose [7, p.219]

$$\begin{aligned}
\mathfrak{t} &= \left\{ \begin{pmatrix} 0 & \alpha_1 \\ -\alpha_1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & \alpha_m \\ -\alpha_m & 0 \end{pmatrix} : \alpha_1, \ldots, \alpha_m \in \mathbb{R} \right\} \\
\mathfrak{t}_+ &= \left\{ \begin{pmatrix} 0 & \alpha_1 \\ -\alpha_1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & \alpha_m \\ -\alpha_m & 0 \end{pmatrix} : \alpha_1 \geq \cdots \geq \alpha_{m-1} \geq |\alpha_m| \right\} \\
\mathfrak{a} &= \{\mathrm{diag}\,(\alpha_1, \ldots, \alpha_n) : \alpha_1, \ldots, \alpha_n \in \mathbb{R}, \sum_{i=1}^n a_i = 0\} \\
\mathfrak{a}_+ &= \{\mathrm{diag}\,(\alpha_1, \ldots, \alpha_n) : \alpha_1 \geq \cdots \geq \alpha_n, \sum_{i=1}^n a_i = 0\}.
\end{aligned}$$

The Weyl group $W_{\mathfrak{a}}$ acts as the symmetric group S_n on $\mathfrak{a}$, i.e.,

$$\mathrm{diag}\,(\alpha_1, \ldots, \alpha_n) \to \mathrm{diag}\,(\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(n)}), \quad \sigma \in S_n$$

and the Weyl group $W_{\mathfrak{t}}$ acts on $\mathfrak{t}$ in the following way

$$\begin{pmatrix} 0 & \alpha_1 \\ -\alpha_1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & \alpha_m \\ -\alpha_m & 0 \end{pmatrix} \to \begin{pmatrix} 0 & \pm\alpha_{\sigma(1)} \\ \mp\alpha_{\sigma(1)} & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & \pm\alpha_{\sigma(m)} \\ \mp\alpha_{\sigma(m)} & 0 \end{pmatrix}$$

in which the total number of sign changes on α's is even and $\sigma \in S_m$.

Suppose $C \in \mathfrak{sl}_n(\mathbb{R})$. According to the Hermitian decomposition and the spectral decomposition of real symmetric and skew symmetric matrices,

$$C = U\left[\begin{pmatrix} 0 & g_1 \\ -g_1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & g_m \\ -g_m & 0 \end{pmatrix}\right] U^{-1} + V \mathrm{diag}\,(f_1, \ldots, f_n) V^{-1} \tag{3.25}$$

for some $U, V \in \mathrm{SO}(n)$, $g_1 \geq \cdots \geq g_{m-1} \geq |g_m|$, and $f_1 \geq \cdots \geq f_n$.

Let $A \in \mathfrak{p}$ and $B \in \mathfrak{k}$. Then

$$A = Z \mathrm{diag}\,(a_1, \ldots, a_n) Z^{-1} \tag{3.26}$$

and

$$B = W\left[\begin{pmatrix} 0 & b_1 \\ -b_1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & b_m \\ -b_m & 0 \end{pmatrix}\right] W^{-1} \tag{3.27}$$

for some $Z, W \in \mathrm{SO}(n)$, $b_1 \geq \cdots \geq b_{m-1} \geq |b_m|$, and $a_1 \geq \cdots \geq a_n$. The Killing form of $\mathfrak{sl}_n(\mathbb{R})$ [23, p.180, p.186] is $B(X,Y) = 2n\mathrm{tr}\, XY$. Since $\theta(X) = -X^\top$,

$$B_\theta(X,Y) = -B(X, \theta Y) = 2n\mathrm{tr}\, XY^\top$$

and

$$\|X\|^2 = 2n\mathrm{tr}\, XX^\top,$$

a scalar multiple of the Frobenius norm. By Theorem 3.11

$$\min_{k,h \in \mathrm{SO}(n)} \|kAk^{-1} + hBh^{-1} - C\|^2 = 2n\left(\sum_{j=1}^{n} |f_j - a_j|^2 + 2\sum_{j=1}^{m} |g_j - b_j|^2\right). \tag{3.28}$$

Note that the scalar multiple $2n$ is not in [34, Theorem 2.1] because the Frobenius norm is used there.

The longest element of $W_{\mathfrak{a}}$ sends

$$\operatorname{diag}(\alpha_1, \dots, \alpha_n) \to \operatorname{diag}(\alpha_n, \dots, \alpha_1)$$

and the longest element [42, p.88] of $W_{\mathfrak{t}}$ sends

$$\begin{aligned}
&\begin{pmatrix} 0 & \alpha_1 \\ -\alpha_1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & \alpha_m \\ -\alpha_m & 0 \end{pmatrix} \\
\to &\begin{cases} \begin{pmatrix} 0 & -\alpha_1 \\ \alpha_1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & -\alpha_m \\ \alpha_m & 0 \end{pmatrix} & \text{if } m \text{ is even} \\ \begin{pmatrix} 0 & -\alpha_1 \\ \alpha_1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & -\alpha_{m-1} \\ \alpha_{m-1} & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & \alpha_m \\ -\alpha_m & 0 \end{pmatrix} & \text{if } m \text{ is odd} \end{cases}
\end{aligned}$$

So

$$\begin{aligned}
&\max_{k,h \in \mathrm{SO}(n)} \|kAk^{-1} + hBh^{-1} - C\|^2 \\
= &\begin{cases} 2n(\sum_{j=1}^{n} |f_j - a_{n-j+1}|^2 \\ \quad +2\sum_{j=1}^{m} |g_j + b_j|^2) & \text{if } m \text{ is even} \\ 2n(\sum_{j=1}^{n} |f_j - a_{n-j+1}|^2 \\ \quad +2(\sum_{j=1}^{m-1} |g_j + b_j|^2 + |g_m - b_m|^2)) & \text{if } m \text{ is odd} \end{cases}
\end{aligned} \tag{3.29}$$

However when we view $C \in \mathfrak{sl}_n(\mathbb{C}) = \mathfrak{su}(n) \oplus i\mathfrak{su}(n)$,

$$\begin{aligned}
C &= iU_1 \operatorname{diag}(g_1, \dots, g_{m-1}, |g_m|, -|g_m|, -g_{m-1}, \dots, -g_1) U_1^{-1} \\
&\quad + V \operatorname{diag}(f_1, \dots, f_n) V^{-1}
\end{aligned}$$

for some $U_1 \in \mathrm{SU}(n)$. Similarly view $A \in i\mathfrak{su}(n)$ and $B \in \mathfrak{su}(n)$ so that $A = Z\mathrm{diag}\,(a_1, \ldots, a_n)Z^{-1}$ as before and

$$B = iW_1\mathrm{diag}\,(b_1, \ldots, b_{m-1}, |b_m|, -|b_m|, -b_{m-1}, \ldots, -b_1)W_1^{-1}$$

for some $W_1 \in \mathrm{SU}(n)$. By Theorem 3.11 or [34, Theorem 2.1],

$$\begin{aligned} &\min_{k,h \in \mathrm{SU}(n)} \|kAk^{-1} + hBh^{-1} - C\|^2 \\ = \;& 2n\left(\sum_{j=1}^{n} |f_j - a_j|^2 + 2(\sum_{j=1}^{m-1} |g_j - b_j|^2 + |\,|g_m| - |b_m|\,|^2)\right). \end{aligned} \tag{3.30}$$

Clearly (3.30) is smaller than or equal to (3.28) because of the triangle inequality $|\,|g_m| - |b_m|\,| \leq |g_m - b_m|$. Indeed $\mathrm{SO}(n) \subset \mathrm{SU}(n)$ is the underlying reason.

The two Weyl groups $W_\mathfrak{a}$ and $W_\mathfrak{t}$ for $\mathfrak{sl}_n(\mathbb{C})$ are equal to S_n. So by Theorem 3.11 or [34, Theorem 2.1],

$$\begin{aligned} &\max_{k,h \in \mathrm{SU}(n)} \|kAk^{-1} + hBh^{-1} - C\|^2 \\ = \;& 2n(\sum_{j=1}^{n} |f_j - a_{n-j+1}|^2 + 2(\sum_{j=1}^{m-1} |g_j + b_j|^2 + (|g_m| + |b_m|)^2)) \end{aligned} \tag{3.31}$$

which is independent of the parity of m.

In conclusion, given $C \in \mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p} \subset \mathfrak{g}_\mathbb{C}$, $A \in \mathfrak{p}$ and $B \in \mathfrak{k}$, where $\mathfrak{g}_\mathbb{C} = \mathfrak{u} \oplus i\mathfrak{u}$ ($\mathfrak{u} = \mathfrak{k} \oplus i\mathfrak{p}$ so that $A \in i\mathfrak{u}$ and $B \in \mathfrak{u}$) is the complexification of $\mathfrak{g}$, it may not be true that the corresponding extrema given in Theorem 3.11 are the same for $\mathfrak{g}$ and $\mathfrak{g}_\mathbb{C}$.

Theorem 3.13. *Let $C \in \mathbb{R}_{n\times n}$, $A \in \mathbb{R}_{n\times n}$ be symmetric with eigenvalues $a_1 \geq \cdots \geq a_n$ and $B \in \mathbb{R}_{n\times n}$ be skew symmetric with eigenvalues $\pm ib_1, \ldots, \pm ib_m$ when $n = 2m$ and $\pm ib_1, \ldots, \pm ib_m, 0$ when $n = 2m+1$. Let A, B, C have decompositions* (3.26), (3.27) *and* (3.25) *respectively, with $f_1 \geq \cdots \geq f_n$ and*

1. *$b_1 \geq \cdots \geq b_{m-1} \geq |b_m|$ and $g_1 \geq \cdots \geq g_{m-1} \geq |g_m|$ when $n = 2m$*

2. $b_1 \geq \cdots \geq b_m \geq 0$ *and* $g_1 \geq \cdots \geq g_m \geq 0$ *when* $n = 2m+1$.

Let $\|\cdot\|$ *be the Frobenius norm on* $\mathbb{R}_{n\times n}$. *Then*

$$\{\|kAk^{-1} + hBh^{-1} - C\|^2 : k, h \in \mathrm{SO}(n)\} = [\ell_0, L_0].$$

1. If $n = 2m$, *then*

$$\ell_0 = \sum_{j=1}^{n} |f_j - a_j|^2 + 2\sum_{j=1}^{m} |g_j - b_j|^2 \tag{3.32}$$

and

$$L_0 = \begin{cases} \sum_{j=1}^{n} |f_j - a_{n-j+1}|^2 + 2\sum_{j=1}^{m} |g_j + b_j|^2 & \text{if } m \text{ even} \\ \sum_{j=1}^{n} |f_j - a_{n-j+1}|^2 + 2(\sum_{j=1}^{m-1} |g_j + b_j|^2 & \\ \qquad\qquad +|g_m - b_m|^2) & \text{if } m \text{ odd} \end{cases} \tag{3.33}$$

2. If $n = 2m+1$, *then*

$$\begin{aligned} \ell_0 &= \sum_{j=1}^{n} |f_j - a_j|^2 + 2\sum_{j=1}^{m} |g_j - b_j|^2 \\ L_0 &= \sum_{j=1}^{n} |f_j - a_{n-j+1}|^2 + 2\sum_{j=1}^{m} |g_j + b_j|^2. \end{aligned}$$

Proof. Note that the subspaces of symmetric matrices and skew symmetric matrices of $\mathbb{R}_{n\times n}$ are orthogonal with respect to the inner product $(X,Y) = \operatorname{tr} XY^\top$. So

$$\|kAk^{-1} + hBh^{-1} - C\|^2 = \|kAk^{-1} - (C + C^\top)/2\|^2 + \|hBh^{-1} - (C - C^\top)/2\|^2.$$

Hence

$$\ell_0 = \min_{k\in\mathrm{SO}(n)} \|kAk^{-1} - (C + C^\top)/2\|^2 + \min_{h\in\mathrm{SO}(n)} \|hBh^{-1} - (C - C^\top)/2\|^2.$$

Consider $n = 2m$. For the first term using (3.28) (dropping the scalar $2n$ and with $B = 0$ and $C - C^\top = 0$ in mind) and for the second term using [34, Theorem 2.1] (with $A = 0$ and $C + C^\top = 0$ in mind), we have ℓ_0. Similarly from (3.29) we have L_0. The odd case is simpler. $\square$

Remark 3.14. Theorem 3.11 is also true for reductive Lie groups with connected K (so the Weyl group $W_{\mathfrak{t}}$ is defined). For example, the reductive group $(\mathrm{GL}_n(\mathbb{C}), \mathrm{U}(n), \theta, B)$ with $\theta(X) = -X^*$ and $B(X,Y) = \operatorname{tr} XY$ yields [34, Theorem 2.1].

The reductive group $(\mathrm{GL}_n(\mathbb{R}), \mathrm{O}(n), \theta, B)$ with $\theta(X) = -X^\top$ and $B(X,Y) = \operatorname{tr} XY$ has non-connected $K = \mathrm{O}(n)$. Now $\mathfrak{k} = \mathfrak{so}(n)$ and $\mathfrak{p}$ is the space of $n \times n$ real symmetric matrices. Though $K = \mathrm{O}(n)$ is not connected, $\mathrm{O}(n) = \mathrm{SO}(n) \cup \hat{\mathrm{SO}}(n)$ where

$$\hat{\mathrm{SO}}(n) = \operatorname{diag}(-1, 1, \ldots, 1)\mathrm{SO}(n)$$

is the set of orthogonal matrices with determinant -1. So we cannot apply Theorem 3.11 directly. But the problem can be solved by two approaches. The first approach is to use $\mathrm{O}(n) = \mathrm{SO}(n) \cup \hat{\mathrm{SO}}(n)$ and apply Example 3.12. The second approach is to use $\mathrm{O}(n) \subset \mathrm{U}(n)$ and apply [34, Theorem 2.1] as follows.

We only consider even n since the odd case is similar. Let $A \in \mathfrak{p}$, $B \in \mathfrak{k}$ and $C \in \mathfrak{gl}_n(\mathbb{R})$. Then

$$C = U \left[\begin{pmatrix} 0 & g_1 \\ -g_1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & g_m \\ -g_m & 0 \end{pmatrix} \right] U^{-1} + V \operatorname{diag}(f_1, \ldots, f_n) V^{-1} \tag{3.34}$$

for some $U, V \in \mathrm{O}(n)$, $g_1 \geq \cdots \geq g_m \geq 0$, and $f_1 \geq \cdots \geq f_n$. Moreover

$$A = Z \operatorname{diag}(a_1, \ldots, a_n) Z^{-1} \tag{3.35}$$

and

$$B = W\left[\begin{pmatrix} 0 & b_1 \\ -b_1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & b_m \\ -b_m & 0 \end{pmatrix}\right] W^{-1} \tag{3.36}$$

for some $Z, W \in \mathrm{O}(n)$, $b_1 \geq \cdots \geq b_m \geq 0$, and $a_1 \geq \cdots \geq a_n$. Since $\mathrm{O}(n) \subset \mathrm{U}(n)$, by [34, Theorem 2.1]

$$\begin{aligned} \min_{k,h\in\mathrm{O}(n)} \|kAk^{-1} + hBh^{-1} - C\|^2 &\geq \min_{k,h\in\mathrm{U}(n)} \|kAk^{-1} + hBh^{-1} - C\|^2 \\ &= \sum_{j=1}^{n} |f_j - a_j|^2 + 2\sum_{j=1}^{m} |g_j - b_j|^2, \end{aligned}$$

where $\|\cdot\|$ denotes the Frobenius norm. Clearly the right side is attainable so we have

$$\min_{k,h\in\mathrm{O}(n)} \|kAk^{-1} + hBh^{-1} - C\|^2 = \sum_{j=1}^{n} |f_j - a_j|^2 + 2\sum_{j=1}^{m} |g_j - b_j|^2. \tag{3.37}$$

Similarly

$$\max_{k,h\in\mathrm{O}(n)} \|kAk^{-1} + hBh^{-1} - C\|^2 = \sum_{j=1}^{n} |f_j - a_{n-j+1}|^2 + 2\sum_{j=1}^{m} |g_j + b_j|^2. \tag{3.38}$$

The outcomes are identical to those [34, Theorem 2.1] if we view $A, B, C \in \mathfrak{gl}_n(\mathbb{C})$.

The following summaries the above discussion and asserts that the set $\{\|kAk^{-1} + hBh^{-1} - C\|^2 : k, h \in \mathrm{O}(n)\}$ is connected though $\mathrm{O}(n)$ is not.

Theorem 3.15. *Let $C \in \mathbb{R}_{n\times n}$, $A \in \mathbb{R}_{n\times n}$ be symmetric with eigenvalues $a_1 \geq \cdots \geq a_n$ and $B \in \mathbb{R}_{n\times n}$ be skew symmetric with eigenvalues $\pm ib_1, \ldots, \pm ib_m$ when $n = 2m$ and $\pm ib_1, \ldots, \pm ib_m, 0$ when $n = 2m+1$ ($b_1 \geq \cdots \geq b_m \geq 0$). Let A, B, C have decompositions* (3.35), (3.36) *and* (3.34) *respectively. Let $\|\cdot\|$ be the Frobenius norm on $\mathbb{R}_{n\times n}$. Then*

$$\{\|kAk^{-1} + hBh^{-1} - C\|^2 : k, h \in \mathrm{O}(n)\} = [\ell, L],$$

where

$$
\begin{aligned}
\ell &= \sum_{j=1}^{n} |f_j - a_j|^2 + 2\sum_{j=1}^{m} |g_j - b_j|^2 \\
L &= \sum_{j=1}^{n} |f_j - a_{n-j+1}|^2 + 2\sum_{j=1}^{m} |g_j + b_j|^2.
\end{aligned}
$$

Proof. We only deal with the even case $n = 2m$ (the odd case is simpler). We have already established (3.37) and (3.38). Let

$$S := \{\|kAk^{-1} + hBh^{-1} - C\|^2 : k, h \in \mathrm{O}(n)\}.$$

The set S contains

$$S_1 = \{\|kA_1k^{-1} + hB_1h^{-1} - C_1\|^2 : k, h \in \mathrm{SO}(n)\}$$

in which A_1, B_1, C_1 are A, B, C as in (3.35) (3.36), (3.34) respectively and in addition $U, Z, W \in \mathrm{SO}(n)$ (we can also assume $V \in \mathrm{SO}(n)$ otherwise replace V by $V\mathrm{diag}\,(-1, 1, \ldots, 1)$). Now S_1 is an interval since $\mathrm{SO}(n)$ is compact connected, i.e., $S_1 = [\ell_1, L_1]$. By Theorem 3.13

$$\ell_1 = \sum_{j=1}^{n} |f_j - a_j|^2 + 2\sum_{j=1}^{m} |g_j - b_j|^2 = \ell$$

and

$$L_1 = \begin{cases} \sum_{j=1}^{n} |f_j - a_{n-j+1}|^2 + 2\sum_{j=1}^{m} |g_j + b_j|^2 & \text{if } m \text{ is even} \\ \sum_{j=1}^{n} |f_j - a_{n-j+1}|^2 + 2(\sum_{j=1}^{m-1} |g_j + b_j|^2 + |g_m - b_m|^2) & \text{if } m \text{ is odd} \end{cases}$$

The set S also contains

$$S_2 = \{\|kA_2k^{-1} + hB_2h^{-1} - C_2\|^2 : k, h \in \mathrm{SO}(n)\} = [\ell_2, L_2]$$

in which A_2, B_2, C_2 are A, B, C as in (3.35) (3.36), (3.34) respectively except the last block B_2 is the negative of the last block of B and in addition $U, V, Z, W \in \mathrm{SO}(n)$ as before. Again by Theorem 3.13

$$\ell_2 = \sum_{j=1}^{n} |f_j - a_j|^2 + 2(\sum_{j=1}^{m-1} |g_j - b_j|^2 + |g_m + b_m|)$$

and

$$L_2 = \begin{cases} \sum_{j=1}^{n} |f_j - a_{n-j+1}|^2 + 2\sum_{j=1}^{m} |g_j + b_j|^2 & \text{if } m \text{ is odd} \\ \sum_{j=1}^{n} |f_j - a_{n-j+1}|^2 + 2(\sum_{j=1}^{m-1} |g_j + b_j|^2 + |g_m - b_m|^2) & \text{if } m \text{ is even} \end{cases}$$

Note that $L_1 \geq \ell_2 \geq \ell_1$ so S_1 and S_2 intersect. Moreover $L = L_1 \in S_1$ if m is even and $L = L_2 \in S_2$ if m is odd. We then have the desired result. $\square$

3.6 Global Extrema

The gradient flow we obtained in Section 3.3 could be used to design an algorithm to solve Problem (3.5) as the following coupled discretized gradient system alike [34]:

$$k_{m+1} = k_m \exp\{-\alpha_m[\theta(C - \mathrm{Ad}\,(h_m)B), \mathrm{Ad}\,(k_m)A]_{\mathfrak{k}}\}, \quad m = 1, 2, \ldots$$
$$h_{m+1} = h_m \exp\{-\beta_m[\theta(C - \mathrm{Ad}\,(k_m)A), \mathrm{Ad}\,(h_m)B]_{\mathfrak{k}}\}, \quad m = 1, 2, \ldots$$

where $\alpha_m, \beta_m > 0$ are the steps. But the gradient flow method has pitfalls of local minima.

We now consider a special case: $B = 0$ and $A, C \in \mathfrak{p}$. So the problem is to study the distance between C and the adjoint orbit $\mathrm{Ad}\,(K)A$ of A. Such problem would have a unique local minimum except for a measure zero set of A and C. Since $B = 0$ and $A, C \in \mathfrak{p}$, Problem (3.5) becomes

$$\min_{k \in K} \|\mathrm{Ad}\,(k)A - C\|, \tag{3.39}$$

which is equivalent to the problem

$$\max_{k \in K} B(\mathrm{Ad}\,(k)A, C). \tag{3.40}$$

Let $\mathfrak{a}$ be a maximal Abelian subspace of $\mathfrak{p}$. Since $\mathfrak{p} = \mathrm{Ad}\,(K)\mathfrak{a}$ and the Killing form B is $\mathrm{Ad}\,K$-invariant, we may assume that $A, C \in \mathfrak{a}$. Define a smooth function $f_{C,A} : K \to \mathbb{R}$ as

$$f_{C,A}(k) = B(\mathrm{Ad}\,(k)A, C).$$

This induces a function $\hat{f}_{C,A} : K/K_A \to \mathbb{R}$ since $K/K_A \approx \mathrm{Ad}\,(K)A$ [49, p.214], where $K_A := \{k \in K : \mathrm{Ad}\,(k)A = A\}$ is the centralizer of A in K.

The set of non-regular elements in $\mathfrak{a}$ is of measure zero, since it is the union of finite many hyperplanes in $\mathfrak{a}$. If C is regular, then $\hat{f}_{C,A}$ is a Morse function on K/K_A. If in addition A and C are in the same Weyl chamber, then $f_{C,A}$ has a unique local minimum and local maximum (which are global minimum and maximum) [49, Proposition 4.4].

As an example, consider $G = \mathrm{GL}_n(\mathbb{C})$. Without loss of generality we may assume that $A = \mathrm{diag}\,(a_1, \ldots, a_n)$ and $C = \mathrm{diag}\,(c_1, \ldots, c_n)$ with $a_1 \geq \cdots \geq a_n$ and $c_1 \geq \cdots \geq c_n$. Then (3.40) becomes

$$\max_{U \in \mathrm{U}(n)} \mathrm{tr}\, CUAU^*. \tag{3.41}$$

It is well known that for (3.41) the global maximum is $\sum_{i=1}^n a_i c_i$. The permutation matrix group $\{P_\sigma : \sigma \in S_n\} \subset \mathrm{U}(n)$ is part of the critical set of $f_{C,A}$. For each $\sigma \in S_n$, $\sum_{i=1}^n a_i c_{\sigma_i}$ is a critical value of $f_{C,A}$. Under the assumption that A and C are regular, i.e., $a_1 > \cdots > a_n$ and $c_1 > \cdots > c_n$, the optimization problem (3.39) has a unique local (global) minimum [12, Theorem 4.1]. A similar result is true for $\mathrm{GL}_n(\mathbb{R})$.

Chapter 4

Convexity of Generalized Numerical Ranges Associated with Lie Algebras

This chapter is essentially a brief survey on some generalized numerical ranges associated with Lie algebras. The classical numerical range of a complex square matrix is the image of the unit sphere under the quadratic form. One of the most beautiful properties is that the numerical range of a matrix is always convex. We give another proof of the convexity of a generalized numerical range associated with a compact Lie group via a connectedness result of Atiyah and a Hessian index result of Duistermaat, Kolk and Varadarajan.

4.1 Classical Numerical Range

Let $\mathbb{C}^n$ (resp., $\mathbb{R}^n$) be the vector space of all n-tuple complex (resp., real) numbers. Let $\mathbb{C}_{n\times n}$ denote the set of all complex $n \times n$ matrices. The *(classical) numerical range* of $A \in \mathbb{C}_{n\times n}$ is the set

$$W(A) = \{x^*Ax : x \in \mathbb{C}^n, x^*x = 1\} \subset \mathbb{C},$$

which is the image of the unit sphere in $\mathbb{C}^n$ under the quadratic map $x \mapsto x^*Ax$. Toeplitz-Hausdorff theorem [48, 22] asserts that $W(A)$ is convex for all $A \in \mathbb{C}_{n\times n}$, which perhaps is the most interesting property of numerical range. See [13] for an interesting geometric proof. The following is a collection of some other nice properties of numerical range, for which proofs and references can be found in [18, 26].

Proposition 4.1. The following statements hold for all $A \in \mathbb{C}_{n\times n}$.

(1) $W(A)$ is compact.

(2) $\sigma(A) \subset W(A)$, where $\sigma(A)$ is the spectrum of A.

(3) $W(A+\alpha I) = W(A) + \alpha$ for all $\alpha \in \mathbb{C}$, where $I \in \mathbb{C}_{n\times n}$ is the identity matrix.

(4) $W(\alpha A) = \alpha W(A)$ for all $\alpha \in \mathbb{C}$.

(5) $W(U^*AU) = W(A)$ for all $U \in \mathrm{U}(n)$, where $\mathrm{U}(n)$ is the unitary group.

(6) $W(A) = \operatorname{conv}\sigma(A)$ if A is normal, where $\operatorname{conv}\sigma(A)$ is the convex hull of $\sigma(A)$.

(7) $W(A \oplus B) = \operatorname{conv}(W(A) \cup W(B))$ for any $B \in \mathbb{C}_{k\times k}$ with $k \in \mathbb{N}$.

(8) $W(S) \subset W(A)$ for any principal submatrix S of A.

(9) If $A \in \mathbb{C}_{2\times 2}$ has eigenvalues λ_1 and λ_2, then $W(A)$ is an elliptical disk with λ_1 and λ_2 as foci, and minor axis of length $\sqrt{\operatorname{tr}(A^*A) - |\lambda_1|^2 - |\lambda_2|^2}$.

4.2 Generalized Numerical Ranges

There are many generalizations of the classical numerical range motivated by theories and applications in the last decades [20, 33]. Halmos [21] introduced the notion of *k-numerical range* of $A \in \mathbb{C}_{n\times n}$ for $1 \le k \le n$, which is defined by

$$W_k(A) = \left\{ \sum_{i=1}^{k} x_i^* A x_i : x_1, \ldots, x_k \text{ are orthonormal in } \mathbb{C}^n \right\}.$$

He conjectured and Berger [2] proved that $W_k(A)$ is always convex. Westwick [51] further generalized the k-numerical range to the *c-numerical range* of $A \in \mathbb{C}_{n\times n}$ for $c \in \mathbb{C}^n$ defined by

$$W_c(A) = \left\{ \sum_{i=1}^{n} c_i x_i^* A x_i : x_1, \ldots, x_n \text{ are orthonormal in } \mathbb{C}^n \right\}.$$

Westwick proved that $W_c(A)$ is always convex if $c \in \mathbb{R}^n$ and fails to be convex if $c \in \mathbb{C}^n$ in general. Let $c = (c_1, \ldots, c_n)^\top$ and $C = \operatorname{diag}(c_1, \ldots, c_n)$. Then one sees that $\mu \in W_c(A)$ if and only if $\mu \in \operatorname{tr}(CU^*AU)$ for some $U \in \mathrm{U}(n)$, where $\mathrm{U}(n)$ denotes the unitary group. This observation motivates the definition of *C-numerical range* of $A \in \mathbb{C}_{n\times n}$ for a general

$C \in \mathbb{C}_{n\times n}$ defined by

$$W_C(A) = \{\mathrm{tr}\,(CU^*AU) : U \in \mathrm{U}(n)\}.$$

This notion was first introduced by Goldberg and Straus in [19]. Note that $W_C(A)$ is the image of the unitary orbit

$$U(A) = \{U^*AU : U \in \mathrm{U}(n)\}.$$

under the linear functional on $\mathbb{C}_{n\times n}$ represented by C. Clearly $W_C(A) = W_A(C)$ and $W_C(A) = W_C(U^*AU)$ for all $U \in \mathrm{U}(n)$. Cheung and Tsing [11] proved that $W_C(A)$ is star-shaped.

4.3 Generalized Numerical Ranges Associated with Lie Algebras

4.3.1 Compact Case

Let $C \in \mathbb{C}_{n\times n}$ be Hermitian and let $A \in \mathbb{C}_{n\times n}$. Let $A = A_1 + iA_2$ be a Hermitian decomposition, where A_1 and A_2 are Hermitian. Then $W_C(A)$ can be identified with

$$W_C(A_1, A_2) := \{(\mathrm{tr}\, CU^*A_1U, \mathrm{tr}\, CU^*A_2U) : U \in \mathrm{U}(n)\} \subset \mathbb{R}^2.$$

Note that $\mathrm{U}(n)$ is a compact connected Lie group, whose Lie algebra $\mathfrak{u}(n)$ consists of all $n \times n$ skew Hermitian matrices. If $B \in \mathbb{C}_{n\times n}$ is Hermitian, then $iB, iC \in \mathfrak{u}(n)$ and

$$\mathrm{tr}\,(CU^*BU) = \mathrm{tr}\,(BUCU^*) = -\mathrm{tr}\,(iB)U(iC)U^*.$$

Thus one can assume that $A_1, A_2, C \in \mathfrak{u}(n)$ when concerning convexity of $W_C(A_1, A_2)$.

Westwick's proof uses the idea of Hausdorff's connectedness argument. He considered the function $f_B : \mathrm{U}(n)/D(n) \to \mathbb{R}$ given by $f_B([U]) = \mathrm{tr}\, CU^*BU$, where $B, C \in \mathbb{C}_{n\times n}$ are Hermitian, $D(n)$ is the subgroup of diagonal matrices in $\mathrm{U}(n)$, and $[U] = D(n)U$ for

$U \in \mathrm{U}(n)$. He showed that $f_B^{-1}(c)$ is connected for any $c \in \mathbb{R}$. Raïs [40] pointed out that there is a gap in Westwick's proof since the eigenvalues of C and B are assumed distinct.

Motivated by Westwick's paper, Raïs [40] considered a generalized numerical range associated with a compact Lie group. The following is Raïs [40] consideration. Let K be a compact connected Lie group with Lie algebra $\mathfrak{k}$. Let $\langle \cdot, \cdot \rangle$ be any $\mathrm{Ad}\, K$-invariant inner product on $\mathfrak{k}$, i.e.,

$$\langle \mathrm{Ad}\,(k)X, \mathrm{Ad}\,(k)Y \rangle = \langle X, Y \rangle, \qquad \forall X, Y \in \mathfrak{k}, \forall k \in K.$$

For any $X_1, X_2, C \in \mathfrak{k}$, the *$C$-numerical range* of the pair (X_1, X_2) is defined by

$$W_C(X_1, X_2) = \{(\langle X_1, \mathrm{Ad}\,(k)C \rangle, \langle X_2, \mathrm{Ad}\,(k)C \rangle) : k \in K\}.$$

Tam [46] proved that $W_C(X_1, X_2)$ is convex in $\mathbb{R}^2$. One may also consider the *joint C-numerical range* of $X_1, \ldots, X_p \in \mathfrak{k}$ defined by

$$W_C(X_1, \ldots, X_p) = \{(\langle X_1, \mathrm{Ad}\,(k)C \rangle, \ldots, \langle X_p, \mathrm{Ad}\,(k)C \rangle) : k \in K\}.$$

Tam's result is best possible in the sense that $W_C(X_1, \ldots, X_p)$ fails to be convex in general if $p \geq 3$ [6]. The main ideas in Tam's proof are applying a connectedness result of Atiyah [1] and using the symplectic structure of the co-adjoint orbit. Then the connectedness of the fibres of the map $\pi_C : \mathrm{Ad}\,(K)X \to \mathbb{R}$ defined by

$$\pi_C(Y) = \langle C, Y \rangle, \quad \forall Y \in \mathrm{Ad}\,(K)X$$

is established. The convexity of $W_C(X_1, X_2)$ then follows through rotation.

Very recently Markus and Tam [37] gave another proof of the convexity of $W_C(X_1, X_2)$. Without using symplectic technique, they proved the connectedness of the fibres of the map

$f_{C,X} : K \to \mathbb{R}$ for all $C, X \in \mathfrak{k}$ defined by

$$f_{C,X}(k) = \langle C, \mathrm{Ad}\,(k)X \rangle, \qquad \forall k \in K.$$

The fibre connectedness result in the compact group K of Markus and Tam is clearly stronger than the fibre connectedness result in the adjoint orbit $\mathrm{Ad}\,(K)X$:

$$\begin{tikzcd} K \arrow[rr, "\mathrm{Ad}\,(\cdot)X"] \arrow[dr, "f_{C,X}"'] & & \mathrm{Ad}\,(K)X \arrow[dl, "\pi_C"] \\ & \mathbb{R} & \end{tikzcd}$$

since the map $\mathrm{Ad}\,(\cdot)X : K \mapsto \mathrm{Ad}\,(K)X$ is continuous.

We shall give a third convexity proof (see Remark 4.10) via a connectedness result of Atiyah [1] and a Hessian index result of Duistermaat, Kolk and Varadarajan [16].

It is worthwhile to note that one may further assume that K is semisimple when concerning the convexity of $W_C(X_1, X_2)$. Since K is compact, we have $K = Z_0K_s = K_sZ_0$ [29, Prop 4.29] and $\mathfrak{k} = \mathfrak{k}_s \oplus \mathfrak{z}$, where K_s is semisimple with Lie algebra $\mathfrak{k}_s$, Z_0 is the identity component of the center Z of K, and $\mathfrak{z}_0$ is the Lie algebra of Z. Since $\mathrm{Ad}\,Z$ is trivial and $\mathrm{Ad}\,K$ acts trivially on $\mathfrak{z}$, for any $A = A_s + A_0$ with $A_s \in \mathfrak{k}_s$ and $A_0 \in \mathfrak{z}$ and for any $k = k_sz_0$ with $k_s \in K_s$ and $z_0 \in Z_0$, we have $\mathrm{Ad}\,(k)A = \mathrm{Ad}\,(k_s)A_s + A_0$. Thus

$$W_C(X_1, X_2) = \{(\langle X_{1s}, \mathrm{Ad}\,(k_s)C_s\rangle, \langle X_{2s}, \mathrm{Ad}\,(k_s)C_s\rangle) : k_s \in K_s\} + c,$$

where $c \in \mathbb{R}^2$ is a constant and $X_{1s}, X_{2s}, C_s \in \mathfrak{k}_s$.

4.3.2 Complex Semisimple Case

Tam [45] considered a generalized C-numerical range in the context of complex semisimple Lie algebras. Let $\mathfrak{g}$ be a complex semisimple Lie algebra and let $\mathfrak{k}$ be a compact real form of $\mathfrak{g}$. Then $\mathfrak{g} = \mathfrak{k} \oplus i\mathfrak{k}$ is a Cartan decomposition of $\mathfrak{g}$ with a corresponding Cartan involution

θ. The Killing form B induces an inner product B_θ on $\mathfrak{g}$ defined by

$$B_\theta(X, Y) = -B(X, \theta Y), \qquad \forall X, Y \in \mathfrak{g}.$$

Let G be a connected complex Lie group with Lie algebra $\mathfrak{g}$ and let K be the analytic subgroup of G with Lie algebra $\mathfrak{k}$. Given $X, C \in \mathfrak{g}$, the *C-numerical range* of X is defined by

$$W_C(X) = \{B_\theta(C, \mathrm{Ad}\,(k)X) : k \in K\}. \tag{4.1}$$

Note that the usual C-numerical range is for the reductive Lie algebra $\mathfrak{gl}_n(\mathbb{C})$ and that the compact case is essentially a special one with $X \in \mathfrak{k}$. Tam [45] conjectured that for any $X \in \mathfrak{g}$ and $f \in \mathfrak{g}^*$, the dual space of $\mathfrak{g}$, the set $f(\mathrm{Ad}\,(K)X)$ is star-shaped with respect to the origin.

The adjoint orbit $\mathrm{Ad}\,(K)X$ depends only on $\mathrm{Ad}_G K$, the analytic subgroup of the adjoint group $\mathrm{Int}\,\mathfrak{g}$ corresponding to $\mathrm{ad}_{\mathfrak{g}}\mathfrak{k}$, and thus $\mathrm{Ad}\,(K)X$ is independent of the choice of G. Let $\mathfrak{t}$ be a maximal abelian subalgebra of $\mathfrak{k}$. Then $\mathfrak{h} = \mathfrak{t} \oplus i\mathfrak{t}$ is a Cartan subalgebra of $\mathfrak{g}$. Let $\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{\alpha \in \Delta} \mathfrak{g}_\alpha$ be the root space decomposition of $\mathfrak{g}$ with respect to $\mathfrak{h}$, where

$$\begin{aligned} \mathfrak{g}_\alpha &= \{X \in \mathfrak{g} : [H, X] = \alpha(H)X \text{ for all } H \in \mathfrak{h}\}, \\ \Delta &= \{\alpha \in \mathfrak{h}^* : \alpha \neq 0 \text{ and } \dim \mathfrak{g}_\alpha \neq 0\}. \end{aligned}$$

Since $B(\mathfrak{g}_\alpha, \mathfrak{g}_\beta) = 0$ whenever $\alpha + \beta \neq 0$, we have the orthogonal projection $\pi : \mathfrak{g} \to \mathfrak{h}$. Cheung and Tam [10] proved that $\pi(\mathrm{Ad}\,(K)X)$ is star-shaped in $\mathfrak{h}$ with star center 0 for all $X \in \mathfrak{g}$. They further affirmed Tam's conjecture for the complex simple Lie algebras of type B [10]. The conjecture is valid for simple Lie algebras of type A [11], D, E_6 and E_7 [15]; it remains unknown for type C, E_8, F_4 and G_2.

4.3.3 Real Semisimple Case

Li and Tam [35] generalized C-numerical range in the context of real semisimple Lie algebras. Let $\mathfrak{g}$ be a real semisimple Lie algebra. Let G be a connected real semisimple Lie group with Lie algebra $\mathfrak{g}$. Let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be the Cartan decomposition of $\mathfrak{g}$ corresponding to a Cartan involution θ, where $\mathfrak{k}$ and $\mathfrak{p}$ are the $+1$ and -1 eigenspaces of θ, respectively. The Killing form B is positive definite on $\mathfrak{p}$ and negative definite on $\mathfrak{k}$. Let K be the analytic subgroup of G with Lie algebra $\mathfrak{k}$. For $C, X_1, \ldots, X_p \in \mathfrak{p}$, the *$C$-numerical range* of $(X_1, \ldots, X_p)$ is defined by

$$W_C(X_1, \ldots, X_p) = \{(B(C, \mathrm{Ad}\,(k)X_1), \ldots, B(C, \mathrm{Ad}\,(k)X_p)) : k \in K\}. \tag{4.2}$$

Since $\mathrm{Ad}\,(K)X$ is independent of the choice of connected G, so is (4.2). Li and Tam [35] proved that $W_C(X_1, X_2)$ is convex for all classical real simple Lie algebras except $\mathfrak{sl}_2(\mathbb{R})$. They also investigated $W_C(X_1, X_2, X_3)$ case by case for each classical real simple Lie algebra. It would be nice if we can show the convexity results of [35] in a unified way.

Remark 4.2. Reductive Lie algebras have similar structures with semisimple ones (see Remark 3.3). Thus the C-numerical range (4.2) is also well defined for reductive Lie algebras.

We begin with the notation of Morse function [38]. Let M be a manifold and $f : M \to \mathbb{R}$ a smooth function. A point $p \in M$ is called a *critical point* of f if the differential map $df_p : T_p(M) \to T_{f(p)}(\mathbb{R})$ is trivial. If p is a critical point of f, the *Hessian* H_p of f at p is a symmetric bilinear form on $T_p(M)$ defined by

$$H_p(v, w) = V_p(Wf), \qquad \forall v, w \in T_p(M),$$

where V and W are vector fields extended by v and w (i.e., $V_p = v, W_p = w$), respectively, and where $Wf \in C^\infty(M)$ is defined by $(Wf)(q) = W_q(f)$ for all $q \in M$. It is symmetric because

$$V_p(Wf) - W_p(Vf) = [V,W]_p(f) = 0$$

where $[V,W]$ is the Poisson bracket of V and W, and where $[V,W]_p(f) = 0$ since p is a critical point. It is well-defined because $V_p(Wf) = v(Wf)$ is independent of the extension V of v, while $W_p(Vf)$ is independent of W. If we choose a local chart about p, the Hessian can be represented by a real symmetric matrix. The *index* of H_p, referred to as the index of f at p, is the maximal dimension of a subspace of $T_p(M)$ on which H_p is negative definite, or equivalently the number of negative eigenvalues of the matrix associated with H_p. A smooth function on a manifold is called a *Morse function* if its Hessian is nondegenerate at every critical point. A *Morse-Bott function* [4, 5] is a smooth function on a manifold whose critical set is a closed submanifold and whose Hessian is nondegenerate in the normal direction. Equivalently, the kernel of the Hessian at a critical point equals the tangent space to the critical submanifold.

Now let $\mathfrak{a}$ be a maximal Abelian subspace of $\mathfrak{p}$ and let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{a} \oplus \mathfrak{n}$ be a Iwasawa decomposition of $\mathfrak{g}$. Let $G = KAN$ be the corresponding Iwasawa decomposition of G. Let $W = W(G,A) = M'/M$ be the Weyl group of G relative to A, where M' and M are the normalizer and centralizer of A in K, respectively. For $C, X \in \mathfrak{p}$, we consider the smooth function $f_{C,X} : K \to \mathbb{R}$ defined by

$$f_{C,X}(k) = B(C, \mathrm{Ad}\,(k)X). \tag{4.3}$$

The C-numerical range $W_C(X,Y)$ with $C, X, Y \in \mathfrak{p}$ is convex if every fibre $f_{C,X}^{-1}(c)$ with $c \in \mathbb{R}$ is connected (or empty) in K. Since $\mathfrak{p} = \cup_{k \in K} \mathrm{Ad}\,(k)\mathfrak{a}$ and B is $\mathrm{Ad}\,K$-invariant, we can assume that $C, X \in \mathfrak{a}$.

Example 4.3. Let $\mathfrak{g} = \mathfrak{sl}_2(\mathbb{R})$ and $G = \mathrm{SL}_2(\mathbb{R})$. Up to a multiple of 4, the Killing form is given by $B(X,Y) = \mathrm{tr}\,XY$ for all $X, Y \in \mathfrak{g}$. Let the Cartan involution θ be defined by

$\theta(X) = -X^\top$ for all $X \in \mathfrak{g}$. Then

$$\mathfrak{p} = \left\{ \begin{pmatrix} a & b \\ b & -a \end{pmatrix} : a, b \in \mathbb{R} \right\}$$

$$\mathfrak{k} = \mathfrak{so}(2) = \left\{ \begin{pmatrix} 0 & c \\ -c & 0 \end{pmatrix} : c \in \mathbb{R} \right\}$$

$$K = \mathrm{SO}(2) = \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbb{R} \right\}.$$

Let $\mathfrak{a} = \left\{ \begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix} : a \in \mathbb{R} \right\}$. Pick $C = X = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \in \mathfrak{a}$, $Y = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathfrak{p}$. For $k = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$, we have $f_{C,X}(k) = \mathrm{tr}\,(CkXk^{-1}) = 2\cos 2\theta$. Thus the fibre $f_{C,X}^{-1}(2) = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \right\}$ is clearly not connected in K. In fact, the C-numerical range of (X, Y) is not convex. More precisely, $W_C(X, Y) = \{(2\cos 2\theta, 2\sin 2\theta) : \theta \in \mathbb{R}\}$ is a circle on $\mathbb{R}^2$.

For each $X \in \mathfrak{a}$, let K_X and W_X denote the centralizers of X in K and in W, respectively. It is obviously that $M \subset K_X$, which guarantees that the notion $K_C w K_X$ makes sense for $w \in W$. The following two lemmas show that $f_{C,X}$ is a Morse-Bott function.

Lemma 4.4. ([16, p.314–316], [49, p.214]) The critical set of $f_{C,X}$ is

$$\begin{aligned} K_{C,X} &= \{k \in K : [C, \mathrm{Ad}\,(k)X] = 0\} \\ &= \bigcup_{w \in W} K_C w K_X \\ &= \bigcup_{w \in W_C \backslash W / W_X} K_C w K_X \end{aligned}$$

where the second union is disjoint and over a complete set of double coset representatives.

Lemma 4.5. ([16, p.317] [49, p.216]) Let $k = ux_w v$ with $u \in K_C$, $v \in K_X$, and x_w a representative of w in K. The Hessian H_k of $f_{C,X}$ at $k \in K$ is given by

$$\begin{aligned} & H_k(dL_k(Z), dL_k(Z)) \\ = & \left.\frac{d^2}{dt^2}\right|_{t=0} f_{C,X}(k \exp tZ) \\ = & -\sum_{\alpha\in\Sigma^+} \alpha(X)(w\cdot\alpha)(C)\|F_\alpha(\mathrm{Ad}\,(v)Z)\|^2, \quad \forall Z \in \mathfrak{k} \end{aligned} \tag{4.4}$$

where $dL_k : \mathfrak{k} \to T_k(K)$ denotes the differential at the identity of the left translation $L_k : K \to K$ given by $L_k(h) = kh$ and $F_\alpha : \mathfrak{k} \to \mathfrak{k}_\alpha$ is an orthogonal projection. In particular, $f_{C,X}$ is a Morse-Bott function and its index at k is

$$\sum_{\alpha\in\Sigma^+, \alpha(X)(w\cdot\alpha)(C)>0} \dim \mathfrak{g}_\alpha. \tag{4.5}$$

Remark 4.6. For each $\alpha \in \Sigma^+$, define $\mathfrak{k}_\alpha^v = \{Z \in \mathfrak{k} : \mathrm{Ad}\,(v)Z \in \mathfrak{k}_\alpha\}$. On each subspace $\mathfrak{k}_\alpha^v$ of $\mathfrak{k}$, depending on the value of $\alpha(X)(w\cdot\alpha)(C)$, exactly one of the following three cases happens for the Hessian: (1) positive definite, (2) negative definite, (3) trivial.

Noting that $\mathrm{Ad}\, v : \mathfrak{k} \to \mathfrak{k}$ is nonsingular and that $\mathfrak{k} = \mathfrak{m} \oplus \bigoplus_{\alpha\in\Sigma^+} \mathfrak{k}_\alpha$ (see Lemma 2.4), we see $\dim \mathfrak{k}_\alpha^v = \dim \mathfrak{k}_\alpha = \dim \mathfrak{g}_\alpha$ and $\mathfrak{k} = \mathrm{Ad}\,(v^{-1})\mathfrak{m} \oplus \bigoplus_{\alpha\in\Sigma^+} \mathfrak{k}_\alpha^v$. The index of $f_{C,X}$ is then

$$\sum_{\alpha\in\Sigma^+, \alpha(X)(w\cdot\alpha)(C)>0} \dim \mathfrak{k}_\alpha^v = \sum_{\alpha\in\Sigma^+, \alpha(X)(w\cdot\alpha)(C)>0} \dim \mathfrak{g}_\alpha.$$

The following example shows the explicit expression of the Hessian of $f_{C,X}$ for $\mathfrak{sl}_n(\mathbb{C})$.

Example 4.7. Let $\mathfrak{g} = \mathfrak{sl}_n(\mathbb{C})$ be viewed as a real semisimple Lie algebra. The Killing form of $\mathfrak{g}$ is given by $B(X, Y) = \mathrm{Re}\,\mathrm{tr}\, XY$ for all $X, Y \in \mathfrak{g}$ up to a scalar multiple of $4n$. Let the Cartan involution θ on $\mathfrak{g}$ be defined by $\theta(X) = -X^*$ for all $X \in \mathfrak{g}$. Then $\mathfrak{k} = \mathfrak{su}(n)$, $K = \mathrm{SU}(n)$, and $\mathfrak{p}$ consists of Hermitian matrices in $\mathfrak{g}$. Let $\mathfrak{a} \subset \mathfrak{p}$ be the subspace of (real) diagonal matrices. The root space decomposition of $\mathfrak{g}$ with respect to $\mathfrak{a}$ is

$$\mathfrak{g} = (\mathfrak{a} \oplus i\mathfrak{a}) \oplus \bigoplus_{i \neq j} \mathbb{C}E_{ij},$$

where E_{ij} is the matrix with 1 at the (i,j)-entry and 0 elsewhere. The root system is $\Sigma = \{e_i - e_j : 1 \leq i \neq j \leq n\}$, where $e_i \in \mathfrak{a}^*$ sends $A \in \mathfrak{a}$ to the i-th diagonal entry of A. The Weyl group W is isomorphic to the group P_n of permutation matrices. Let $\mathfrak{a}_+ \subset \mathfrak{a}$ be the fundamental Weyl chamber consisting of all diagonal matrices whose diagonal entries are in descending order. The set of positive roots is then $\Sigma^+ = \{e_i - e_j : 1 \leq i < j \leq n\}$. For each $\alpha = e_i - e_j \in \Sigma^+$, $\mathfrak{k}_\alpha = \{cE_{ij} - \bar{c}E_{ji} : c \in \mathbb{C}\}$. Pick $C = \mathrm{diag}\,(c_1, \ldots, c_n), X = \mathrm{diag}\,(x_1, \ldots, x_n) \in \mathfrak{a}$. The centralizers K_C (resp., K_X) of C (resp., X) in K consists of all matrices in $\mathrm{SU}(n)$ that commute with C (resp., X). The critical set of $f_{C,X}$ is thus $K_{C,X} = K_C P_n K_X$. For each $k = UPV$ with $U \in K_C$, $P \in P_n$, and $V \in K_X$, the Hessian of $f_{C,X}$ at k is

$$\begin{aligned} & -\sum_{\alpha \in \Sigma^+} (w \cdot \alpha)(C)\alpha(X)\|F_\alpha(\mathrm{Ad}\,(v)Z)\|^2, \qquad Z \in \mathfrak{su}(n) \\ = \quad & -8n \sum_{i<j} ((PCP^{-1})_{ii} - (PCP^{-1})_{jj})(x_i - x_j) \cdot |(VZV^{-1})_{ij}|^2. \end{aligned}$$

The index of $f_{C,X}$ at k is

$$\begin{aligned} & \sum_{i<j, ((PCP^{-1})_{ii} - (PCP^{-1})_{jj})(x_i - x_j) > 0} \dim_{\mathbb{R}}(\mathbb{C}E_{ij}) \\ = \quad & 2 \cdot |\{(i,j) : 1 \leq i < j \leq n \text{ and } ((PCP^{-1})_{ii} - (PCP^{-1})_{jj})(x_i - x_j) > 0\}| \end{aligned}$$

The following lemma of Atiyah is crucial.

Lemma 4.8. [1, p.4] Let $f : M \to \mathbb{R}$ be a Morse-Bott function on a compact connected manifold M. If neither f nor $-f$ has a critical manifold of index 1, then $f^{-1}(c)$ is connected (or empty) for every $c \in \mathbb{R}$.

The above lemmas enable one to focus on the computation of the index

$$\sum_{\alpha\in\Sigma^+,\alpha(X)(w\cdot\alpha)(C)>0} \dim \mathfrak{g}_\alpha$$

of $f_{C,X}$ in (4.5) and the index

$$\sum_{\alpha\in\Sigma^+,\alpha(X)(w\cdot\alpha)(C)<0} \dim \mathfrak{g}_\alpha$$

of $-f_{C,X}$. If neither of them is 1, the convexity of $W_C(X,Y)$ follows for real semisimple Lie groups G with finite center (in which case K is compact). As an application, we have the convexity of the C-numerical range for *complex* semisimple Lie groups.

Theorem 4.9. Let G be a complex semisimple Lie group viewed as a real Lie group, and let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ with $\mathfrak{p} = i\mathfrak{k}$ be a Cartan decomposition of the (real) Lie algebra $\mathfrak{g}$ of G. Then the C-numerical range $W_C(X,Y)$ defined in (4.2) is convex for all $C, X, Y \in \mathfrak{p}$.

Proof. Note that K is compact since $\mathfrak{k}$ is a compact real form of $\mathfrak{g}$. Since $\mathfrak{g}$ is complex semisimple, each $\mathfrak{g}_\alpha$ has even dimension over $\mathbb{R}$ [49, p.217] and thus the indices of $f_{C,X}$ and $-f_{C,X}$ are both even. By Atiyah's lemma $f_{C,X}^{-1}(c)$ is connected for all $c \in \mathbb{R}$. Rotating $W_C(X,Y)$ anti-clockwise by an angle $\theta \in \mathbb{R}$ yields $W_C(X',Y')$, where

$$(X',Y') = (\cos\theta X + \sin\theta Y, -\sin\theta X + \cos\theta Y) \in \mathfrak{p} \times \mathfrak{p}.$$

It follows that the intersection of $W_C(X,Y)$ with every straight line is connected, whence $W_C(X,Y)$ is convex. $\square$

Remark 4.10. Since $\mathfrak{p} = i\mathfrak{k}$ in Theorem 4.9. It is essentially the same as the compact case discussed in Section 4.3.1. This gives a third proof of Tam's result in [46].

The following example shows that the index condition is sufficient but not necessary for convexity of C-numerical range.

Example 4.11. Let $\mathfrak{g} = \mathfrak{sl}_3(\mathbb{R})$ and $G = \mathrm{SL}_3(\mathbb{R})$. Let the Cartan involution θ be defined by $\theta(X) = -X^\top$ for all $X \in \mathfrak{g}$. Then $\mathfrak{k} = \mathfrak{so}(3)$, $K = \mathrm{SO}(3)$, and $\mathfrak{p}$ is the space of all traceless symmetric matrices. Let $\mathfrak{a} \subset \mathfrak{p}$ be the subspace of diagonal matrices. The root space decomposition of $\mathfrak{g}$ relative to $\mathfrak{a}$ is

$$\mathfrak{g} = \mathfrak{a} \oplus \bigoplus_{i \neq j} \mathbb{R}E_{ij}.$$

The root system is $\Sigma = \{e_i - e_j : 1 \leq i \neq j \leq 3\}$. The centralizer M of $\mathfrak{a}$ in K consists of diagonal matrices in SO(3), and the normalizer of $\mathfrak{a}$ in K consists of generalized permutation matrices in SO(3) whose nonzero entries are either 1 or -1. Thus the Weyl group $W = M'/M$ is isomorphic to the group P_3 of permutation matrices. Let $\mathfrak{a}_+ \subset \mathfrak{a}$ be the fundamental Weyl chamber consisting of all diagonal matrices whose diagonal entries are in descending order. The set of positive roots is then $\Sigma^+ = \{e_1 - e_2, e_2 - e_3, e_1 - e_3\}$. Now pick $C, X \in \mathfrak{a}_+$ and consider the map $f_{C,X} : \mathrm{SO}(3) \to \mathbb{R}$ as defined in (4.3). Obviously $K_C = K_X = M$. By Lemma 4.4, the critical manifold of $f_{C,X}$ is $K_{C,X} = K_C W K_X = M'$. Because $\alpha(X) > 0$ and $\dim \mathfrak{g}_\alpha = 1$ for all $\alpha \in \Sigma^+$, the index given by (4.5) is equal to the number of positive roots sent to positive roots by the $w \in W$ under consideration. Since W_C and W_X are trivial, each Weyl group element can appear for some $k \in K_{C,X}$. Therefore we have the following six cases for the index of $f_{C,X}$ with the notations that $\alpha_1 := e_1 - e_2, \alpha_2 := e_2 - e_3, \alpha_3 := e_1 - e_3$.

Case 1: $w = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Since $w \cdot \alpha_1 = \alpha_1$, $w \cdot \alpha_2 = \alpha_2$, $w \cdot \alpha_3 = \alpha_3$, the index is 3.

Case 2: $w = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Since $w \cdot \alpha_1 = -\alpha_1$, $w \cdot \alpha_2 = \alpha_3$, $w \cdot \alpha_3 = \alpha_2$, the index is 2.

Case 3: $w = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$. Since $w \cdot \alpha_1 = \alpha_3$, $w \cdot \alpha_2 = -\alpha_2$, $w \cdot \alpha_3 = \alpha_1$, the index is 2.

Case 4: $w = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$. Since $w \cdot \alpha_1 = -\alpha_2$, $w \cdot \alpha_2 = -\alpha_1$, $w \cdot \alpha_3 = -\alpha_3$, the index is 0.

Case 5: $w = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. Since $w \cdot \alpha_1 = \alpha_2$, $w \cdot \alpha_2 = -\alpha_3$, $w \cdot \alpha_3 = -\alpha_1$, the index is 1.

Case 6: $w = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$. Since $w \cdot \alpha_1 = -\alpha_3$, $w \cdot \alpha_2 = \alpha_1$, $w \cdot \alpha_3 = -\alpha_2$, the index is 1.

Now $\dim K = \dim \mathrm{SO}(3) = 3$. This shows that for $\mathfrak{g} = \mathfrak{sl}_3(\mathbb{R})$ the condition of Lemma 4.8 is not satisfied, but $W_C(X, Y)$ is still convex [6].

Bibliography

[1] M. F. Atiyah, Convexity and commuting Hamiltonians, Bull. London Math. Soc., **308** (1982), 1–15.

[2] C. A. Berger, *Normal Dilations*, Ph.D. Thesis, Cornell University, 1963.

[3] A. Bloch, Ed., *Hamiltonian and Gradient Flows, Algorithms and Control*, Fields Institute Communications, Amer. Math. Soc., Providence, RI, 1994.

[4] R. Bott, Nondegenerate critical manifolds, Ann. of Math., **60** (1954), 248–261.

[5] R. Bott, Lectures on Morse theory, old and new, Bull. Amer. Math. Soc. (N. S.), **7** (1982) 331–358.

[6] L. Brickman, On the field of values of a matrix, Proc. Amer. Math. Soc., **12** (1961), 61–66.

[7] T. Bröcker and T. tom Dieck, *Representations of Compact Lie Groups*, Springer-Verlag, New York, 1985.

[8] R. W. Brockett, Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems, Linear Algebra Appl., **146** (1991), 79–91.

[9] R. W. Brockett, Differential geometry and the design of gradient algorithms, Proc. Symp. Pure Math., **54** (1993), 69–92.

[10] W. S. Cheung and T. Y. Tam, Star-shapedness and K-orbits in complex semisimple Lie algebras, Canad. Math. Bull., **54** (2011), 44–55.

[11] W. S. Cheung and N. K. Tsing, The C-numerical range of matrices is star-shaped, Linear Multilinear Algebra, *41* (1996), 245–250.

[12] M. T. Chu and K. Driessel, The projected gradient method for least squares matrix approximations with spectral constraints, SIAM J. Numer. Anal., **27** (1990), 1050–1060.

[13] C. Davis, The Toeplitz-Hausdorff theorem explained, Canad. Math. Bull., **14** (1971), 245–246.

[14] J. Day, W. So and R. C. Thompson, The spectrum of a Hermitian matrix sum, Linear Algebra Appl., **280** (1998), 289–332.

[15] D.Z. Djokovic and T.Y. Tam, Some questions about semisimple Lie groups originating in matrix theory, Canad. Math. Bull., **46** (2003), 332–343.

[16] J. J. Duistermaat, J. A. C. Kolk, and V. S. Varadarajan, Functions, flows and oscillatory integrals on flag manifolds and conjugacy classes in real semisimple Lie groups, Compositio Math., **49** (1983), 309–398.

[17] W. Fulton, Eigenvalues, invariant factors, highest weights, and Schubert calculus, Bull. Amer. Math. Soc., **37** (2000), 209–249.

[18] M. Goldberg, On certain finite-dimensional numerical ranges and numerical radii, Linear Multilinear Algebra, **7** (1979), 329–342.

[19] M. Goldberg and E. G. Straus, Elementary inclusion relations for generalized numerical ranges, Linear Algebra Appl., **18** (1977), 1–24.

[20] K. E. Gustafson and D. K. M. Rao, *Numerical Range: The Field of Values of Linear Operators and Matrices*, Springer, New York, 1997.

[21] P. R. Halmos, *A Hilbert Space Problem Book, 2ed.*, Springer-Verlag, New York, 1982.

[22] F. Hausdorff, Der Wertvorrat einer Bilinearform, Math Z., **3** (1919), 314–316.

[23] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press, New York, 1978.

[24] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*, Springer-Verlag, London, 1994.

[25] 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., (Series A) **66** (1999), 331–357.

[26] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[27] J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory*, Springer-Verlag, New York, 1972.

[28] J. E. Humphreys, *Reflection groups and Coxeter groups*, Cambridge University Press, New York, 1990.

[29] A. W. Knapp, *Lie Groups Beyond an Introduction*, 2ed., Birkhäuser, Boston, 2002.

[30] B. Kostant, On convexity, the Weyl group and the Iwasawa decomposition, Ann. Sci. École Norm. Sup., (4) **6** (1973), 413–455.

[31] A. Knutson and T. Tao, Honeycombs and sums of Hermitian matrices, Notices Amer. Math. Soc., **48** (2001), 175–186.

[32] J. Lee, *Introduction to Smooth Manifolds*, Springer-Verlag, New York, 2002.

[33] C. K. Li, C-numerical ranges and C-numerical radii, Linear Multilinear Algebra, **37** (1994), 51–82.

[34] C. K. Li, Y. T. Poon, and T. Schulte-Herbrüggen, Least-squares approximation by elements from matrix orbits achieved by gradient flows on compact Lie groups, Math. Comp., **80** (2011), 1601–1621.

[35] C. K. Li and T. Y. Tam, Numerical ranges arising from simple Lie algebras, Canad. J. Math., **52** (2000), 141–171.

[36] M. Liao, *Lévy Processes in Lie Groups*, Cambridge University Press, 2004.

[37] A. Markus and T. Y. Tam, Connectedness of some fibres on a compact connected Lie group, Linear Multilinear Algebra, **59** (2011), 1121–1126.

[38] J. Milnor, *Morse theory*, Princeton University Press, N.J. 1963.

[39] A. L. Onishchik and E. B. Vinberg, *Lie Groups and Lie Algebras III: Structure of Lie Groups and Lie Algebras*, Springer-Verlag, Berlin, 1994.

[40] M. Raïs, Remarques sur un theoreme de R. Westwick, unpublished manuscript.

[41] L. O'Shea and R. Sjamaar, Moment maps and Riemannian symmetric pairs, Math. Ann., **317** (2000), 415–457.

[42] H. Samelson, *Notes on Lie Algebras*, Springer-Verlag, New York, 1990.

[43] T. Schulte-Herbruggen, S. J. Glaser, G. Dirr, and U. Helmke, Gradient flows for optimization in quantum information and quantum dynamics: foundations and applications, Rev. Math. Phys., **22** (2010), 597–667.

[44] T. Y. Tam, A Lie-theoretic approach to Thompson's theorems on singular values-diagonal elements and some related results, J. London Math. Soc., (2) **60** (1999), 431–448.

[45] T. Y. Tam, On the shape of numerical ranges associated with Lie groups, Taiwanese J. Math., **5** (2001), 497–506.

[46] T. Y. Tam, Convexity of generalized numerical range associated with a compact Lie group, J. Austral. Math. Soc., **72** (2002), 57–66.

[47] T. Y. Tam, Gradient flows and double bracket equations, Differential Geom. Appl., **20** (2004), 209–224.

[48] O. Toeplitz, Das algebraishe Analogon zu einem Satze von Fejér, Math. Z., **2** (1918), 187–197.

[49] V. S. Varadarajan, The Method of Stationary Phase and Applications to Geometry and Analysis on Lie Groups, Book Chapter in *Algebraic and Analytic Methods in Representation Theory*, Academic Press, 1997, 167–242.

[50] F. Warner, *Foundations of Differentiable Manifolds and Lie Groups*, Springer-Verlag, New York, 1983.

[51] R. Westwick, A theorem on numerical range, Linear and Multilinear Algebra, **2** (1975), 311–315.

[52] H. Weyl, Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen, Math. Ann., **71** (1912), 441–479.

[53] A. Wright, Sums of adjoint orbits and L^2-singular dichotomy for $\mathrm{SU}(m)$, Adv. Math., **227** (2011), 253–266.