Differential Geometry on Matrix Groups

by

Luyining Gan

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
May 2, 2020

Keywords: geodesic, curvature, matrix groups, differential geometry

Approved by

Tin-Yau Tam, Chair, Professor of Mathematics and Statistics
Ming Liao, Co-chair, Professor of Mathematics and Statistics
Huajun Huang, Professor of Mathematics and Statistics
Georg Hetzer, Professor of Mathematics and Statistics

Abstract

In this dissertation, we study the differential geometry of the matrix groups. In literatures, many authors proposed to compute matrix means using the geodesic distance defined by an invariant Riemannian metrics on the matrix groups, and this approach has become predominant; see references [20, 19, 5]. Most of papers in this field deal with either the space of positive definite matrices or some special matrix group, and their results have a quite similar form. The geometric means of the unitary group $\mathrm{U}(n)$ [15] was proposed by Mello in 1990. Later, Moakher gave the geodesic means on special orthogonal group $\mathrm{SO}(3)$ [17] and symmetric positive definite matrix space $\mathrm{SPD}(n)$ [18].

In Chapter 1, the explicit geodesic and gradient forms of special matrix groups are given, which would provide theoretical basis of computing the geometric mean. We have also presented the results in a more unified form than those that have appeared in the current literature.

Then, we study the curvatures of the matrix groups. Because the curvature provides important information about the geometric structure of a Riemannian manifold. For example, it is related to the rate at which two geodesics emitting from the same point move away from each other: the lower the curvature is, the faster they move apart (see Theorem IX.5.1 in [2, Chapter IX.5]). Many important geometric and topological properties are implied by suitable curvature conditions. In Chapter 2, we give a simple formula for sectional curvatures on the general linear group, which is also valid for many other matrix groups. This formula appears to be new in literature and is extended to more general reductive Lie groups. Additionally, we also discuss the relation between commuting matrices and zero sectional curvature for $\mathrm{GL}(n, \mathbf{R})$.

Finally we would like to point out that the main results in this dissertation appear in Gan, Liao and Tam [6].

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor Professor Tin-Yau Tam and co-advisor Professor Ming Liao for the continuous support of my Ph.D study and related research, for their patience, motivation, and immense knowledge. Thanks for their insightful comments and encouragement, and for the hard questions which have widen my research from various perspectives. Their guidance has helped me in all the time of research and writing of this dissertation.

I also thank Professor Huajun Huang, Professor Georg Hetzer, my advisory committee members, and Professor Jianjun Dong, my university reader, who have provided good suggestions.

My sincere thank also go to Wei Gao and Jie Han. I have got many helpful discussions and great encouragements from them. We had some joint works and so many wonderful conversations during my visit in the Department of Mathematics at the University of Rhode Island.

At last, I really appreciate Ms. Kitty Siu and Professor Tin-Yau Tam who have helped me learn the culture of a new environment. With their encouragement and guidance, I have had a good time during my graduate period. I want to thank everyone who has helped me throughout my graduate life, especially my parents, my dear roommate Shuwen Xue, my best friends Xu Han and Yanan Wang. This dissertation and any big or small achievements in my graduate life would not be accomplished without their continuous encouragement and endless support.

Table of Contents

Abstract ii

Acknowledgments iv

List of Figures vii

1 Geodesics on Matrix Groups 1

1.1 Introduction 1

1.2 Preliminary 2

1.3 Riemannian metric on matrix spaces 6

1.3.1 Riemannian metric of group $\mathrm{GL}(n, \mathbf{C})$ 6

1.3.2 Riemannian metric of closed subgroups of $\mathrm{GL}(n, \mathbf{C})$ 7

1.3.3 Riemannian metric on positive definite matrices space 8

1.4 Geodesics 9

1.4.1 Geodesics in $\mathrm{GL}(n, \mathbf{C})$ 9

1.4.2 Geodesics in subgroups of $\mathrm{GL}(n, \mathbf{C})$ 13

1.4.3 Geodesics in $\mathbb{P}_n$ the space of positive definite matrices 13

1.4.4 Riemannian distance function 14

1.5 Gradients 15

1.5.1 Definition of Riemannian gradient 15

1.5.2 Gradient on matrix spaces 15

1.5.3 Gradient of criterion function 18

1.5.4 An alternative proof of Moakher's Theorem 19

1.5.5 Gradient on Unitary Group $\mathrm{U}(n)$ 22

1.5.6 Gradient on the Orthogonal Group $\mathrm{O}(n)$ 24

1.5.7 Gradient on the space of positive definite matrices $\mathbb{P}_n$ 26

2 Curvature of matrix and reductive Lie groups 28
2.1 Introduction 28
2.2 Covariant derivative 31
2.3 Curvature Tensor 34
2.4 Sectional Curvature 36
2.5 Zero curvature and commutative property 47
2.6 Curvature of subgroups 50
2.7 Reductive Lie group 52
Bibliography 56

List of Figures

1.1 Tangent Space 2

1.2 Tangent space $\mathbb{H}_n$ of $\mathbb{P}_n$ 8

2.1 Curvature of the sphere 28

2.2 Curvature of the plane 29

2.3 Curvature of the hyperbola 29

2.4 Tangent space $\mathfrak{gl}(n, \mathbf{R})$ of $\mathrm{GL}(n, \mathbf{R})$ 34

2.5 Totally geodesic sub-manifolds of the torus 51

Chapter 1

Geodesics on Matrix Groups

1.1 Introduction

Traditionally, arithmetic mean has been widely used to average numerical data to obtain a mean value, which minimizes the sum of the square distances to the given point. More recently, averaging the matrix data has become increasingly important in many applications, especially when those matrices form a compact group or taken from the space of positive definite matrices. As matrices may be regarded as vectors in a Euclidean space, they may be averaged using the Euclidean distance. However, the matrix data are often taken from a special matrix group, if they are averaged using the Euclidean distance, the mean value may lie outside the group. Moreover, the matrix data are usually naturally associated to matrix multiplication, rather than to matrix addition. Therefore, it is more suitable to use a distance that is invariant with respect to the matrix multiplication to compute the matrix mean, instead of the Euclidean distance which is invariant with respect to matrix addition.

In literature, to address the above concerns, many authors proposed to compute matrix means using the geodesic distance defined by an invariant Riemannian metric on the matrix group, and this approach has become predominant; see references [20, 19, 5]. Most of papers in this field deal with either the space of positive definite matrices or some special matrix group, and their results have a quite similar form. The geometric means of the unitary group $\mathrm{U}(n)$ [15] was proposed by Mello in 1990. Later, Moakher gave the geodesic means on special orthogonal group $\mathrm{SO}(3)$ [17] and symmetric positive definite matrix space $\mathrm{SPD}(n)$ [18]. The purpose of this dissertation is to study some differential geometric aspects of matrix groups, related to averaging matrix data. We will present these results in a more unified form than those that have appeared in literature.

1.2 Preliminary

Definition 1.2.1. *Let V be a vector space and K be a group acting on V linearly. An inner product $\langle \cdot, \cdot \rangle$ on V is called K-invariant if for any $k \in K$ and any $u, v \in V$, $\langle ku, kv \rangle = \langle u, v \rangle$.*

Let M be a manifold and G be a Lie group acting on M. Any $g \in G$ is viewed as a map $g \colon M \to M$. It also induces a linear map $Dg : T_x M \to T_{gx} M$ for any $x \in M$, called the differential map of g at the point x, where $T_x M$ denotes the tangent space over M at x. See Figure 1.1 for illustration.

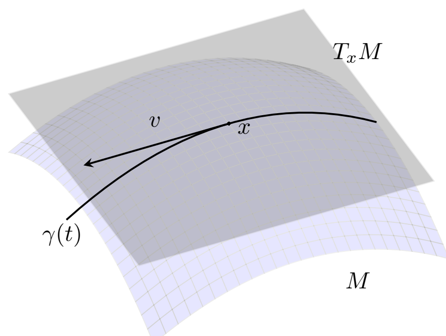

Figure 1.1: Tangent Space

A Riemannian metric on a manifold M is a family of inner products $\langle \cdot, \cdot \rangle_x$ on tangent spaces $T_x M$, $x \in M$ such that for each pair of vector fields X, Y on M, $x \mapsto \langle X, Y \rangle_x$ is a smooth function. A manifold equipped with a Riemannian metric is called a Riemannian manifold.

Definition 1.2.2. *A Riemannian metric $\{\langle \cdot, \cdot \rangle \colon x \in M\}$ is called G-invariant if for any $x \in M, u, v \in T_x M$ and $g \in G$,*

$$\langle Dg(u), Dg(v) \rangle_{gx} = \langle u, v \rangle_x. \tag{1.1}$$

Definition 1.2.3. *Let M be a manifold and G be a Lie group acting on G. The G-action is called transitive if for all $x, y \in M$, there exists $g \in G$ such that $gx = y$.*

Now, fix a point o in M. The group $K := \{g \in G \colon go = o\}$ is a closed subgroup of G, called the isotropy subgroup of G at o. Note that K acts on T_oM linearly through differentiable map. For simplicity, denote Dk by k for $k \in K$ in the following.

Proposition 1.2.4. *[10, p.200] Fix an inner product $\langle \cdot, \cdot \rangle_o$ at T_oM and denote Dg by g here for simplicity. A necessary and sufficient condition for the existence of a G-invariant Riemannian metric $\{\langle \cdot, \cdot \rangle_x \colon x \in M\}$ on M such that $\langle \cdot, \cdot \rangle_x = \langle \cdot, \cdot \rangle_o$ is that $\langle \cdot, \cdot \rangle_o$ is K-invariant on T_oM. Moreover, such a Riemannian metric is unique and is given by*

$$\langle X, Y \rangle_x = \langle g^{-1}X, g^{-1}Y \rangle_o, \tag{1.2}$$

for any $x \in M$, $X, Y \in T_xM$ and where $g \in G$ is chosen to satisfy $go = x$.

Consider a special case first. If $K = \{e\}$, where e is the identity element of G, then any inner product $\langle \cdot, \cdot \rangle$ on T_oM induces a G-invariant Riemannian metric on M.

Definition 1.2.5. *In particular, because a Lie group G acts on itself by left translation $(g, x) \mapsto gx$, so any inner product on $T_eG = \mathfrak{g}$ (Lie algebra) induces a unique Riemannian metric on G, that is invariant under left translation, called a left invariant Riemannian metric on G.*

Let M be an n-dimensional Riemannian manifold with Riemannian metric $\{\langle \cdot, \cdot \rangle_x; x \in M\}$. On a coordinate neighborhood U with local coordinates $x_1, \ldots, x_n$, let

$$g_{jk}(x) = \langle \frac{\partial}{\partial x_j}, \frac{\partial}{\partial x_k} \rangle_x, \quad x \in U, \tag{1.3}$$

and let $\{g^{jk}(x)\}$ be the inverse matrix of $\{g_{jk}(x)\}$. The Christoffel symbols are defined by

$$\Gamma^i_{jk} = \frac{1}{2} \sum_p g^{ip} (\frac{\partial}{\partial x_j} g_{pk} + \frac{\partial}{\partial x_k} g_{jp} - \frac{\partial}{\partial x_p} g_{jk}). \tag{1.4}$$

For any two vector fields X and Y on M, we may write

$$X(x) = \sum_i a_i(x) \frac{\partial}{\partial x_i} \quad \text{and} \quad Y(x) = \sum_i b_i(x) \frac{\partial}{\partial x_i} \tag{1.5}$$

for $x \in U$, where $a_i(x)$ and $b_i(x)$ are some smooth functions on U. Define

$$D_X Y = \sum_i (X b_i) \frac{\partial}{\partial x_i}, \tag{1.6}$$

and

$$\Gamma(X, Y) = \sum_{i,j,k} \Gamma^i_{jk} a_j b_k \frac{\partial}{\partial x_i}. \tag{1.7}$$

Then $D_X Y$ and $\Gamma(X, Y)$ are vector fields on U, but their values depend on the choice of local coordinates and so cannot be regarded as vector fields on M. However,

$$\nabla_X Y = D_X Y + \Gamma(X, Y) \tag{1.8}$$

does not depend on local coordinates [3, p.314], and so is a vector field on M, called the covariant derivative of Y with respect to X, which preserves the Riemannian metric in the sense that

$$X\langle Y, Z\rangle. = \langle \nabla_X Y, Z\rangle. + \langle Y, \nabla_X Z\rangle. \tag{1.9}$$

Let γ: $[a, b] \to M$ be a smooth path. Its derivative $\dot\gamma$ is a vector field along γ. Extend it to a smooth vector field on M. Then the vector field $\nabla_{\dot\gamma}\dot\gamma$ along γ does not depend on the choice of extension. The path is called a geodesic if

$$\nabla_{\dot\gamma}\dot\gamma = 0. \tag{1.10}$$

Thus, a geodesic is defined to be a smooth path γ such that the covariant derivative of $\dot{\gamma}$ with respect to itself is zero. In local coordinates, this may be written as

$$\ddot{\gamma} + \Gamma(\dot{\gamma}, \dot{\gamma}) = 0. \tag{1.11}$$

Let $\gamma\colon t \to \gamma(t),\ (t \in [a, b])$ be a smooth path in M with $\gamma(a) = A$ and $\gamma(b) = B$. Its length is defined as

$$L(\gamma) = \int_a^b \|\dot{\gamma}(t)\| \ dt, \tag{1.12}$$

where

$$\|\dot{\gamma}(t)\| = \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}}.$$

The Riemannian distance between A and B is defined to be the infimum of the length functional $L(\gamma)$ in (1.12) taken among all smooth paths γ with $\gamma(a) = A$ and $\gamma(b) = B$. A smooth path γ from A to B is a geodesic if and only if it or any of its reparametrization is an extremal point of $L(\gamma)$. The geodesic γ may also be characterized as an extremal point of the following energy functional:

$$E(\gamma) = \int_a^b \|\dot{\gamma}\|^2 \, dt, \tag{1.13}$$

because the associated Euler-Lagrange differential equation is just the geodesic Equation ((1.10)).

It is well known that the length of a geodesic $\gamma(t)$ may not realize the Riemannian distance between its end points, but it is locally minimizing in the sense that the small enough geodesic segment realizes the distance between the end points. Because the covariant derivative preserves the Riemannian metric, by (1.10), for any geodesic $\gamma(t)$, $t \in [a,\, b]$, $\|\dot{\gamma}(t)\|$ is a constant, and hence its length is given by

$$L(\gamma) = (b - a) \, \|\dot{\gamma}\| \, . \tag{1.14}$$

A submanifold H of M is called a sub-Riemannian manifold of M if H is equipped with a Riemannian metric $\{\langle\cdot,\cdot\rangle_x^H; x \in H\}$ induced by that of M, that is, for $x \in H$, $\langle\cdot,\cdot\rangle_x^H$ is the restrition of $\langle\cdot,\cdot\rangle_x$ on T_xH. The following result provides a relation between geodesics in M and those in H, see section I.14 in [7] for more details.

Lemma 1.2.6. *If γ is a curve in the submanifold H, and suppose that the curve γ is a geodesic in M, then this curve γ is also a geodesic in H.*

Proof. Let p and q be any two points on γ, where $p = \gamma(r_0)$ and $q = \gamma(r)$. Let N_p be a spherical normal neighborhood of p in G. Then the geodesic segment can be defined

$$\gamma_{pq} : t \to \gamma(t) \quad \text{when} \quad |t - r_0| \leq |r - r_0|$$

is contained in N_p if r is sufficiently close to r_0. Then the length of γ_{pq} satisfies

$$L(\gamma_{pq}) = d_G(p,q) \leq d_H(p,q) \leq L(\gamma_{pq}).$$

Thus, $L(\gamma_{pq}) = d_H(p,q)$. So, γ_{pq} is a curve of the shortest length in H joining p and q, that is, an geodesic on submanifold H. $\square$

A geodesic in the Riemannian submanifold H may not be a geodesic in M. If all geodesics in H are geodesics in M, then H is called a total geodesic submanifold of M.

1.3 Riemannian metric on matrix spaces

1.3.1 Riemannian metric of group $\mathrm{GL}(n, \mathbf{C})$

The general linear group $\mathrm{GL}(n, \mathbf{C})$ is the group of $n \times n$ nonsingular matrices and it is a Lie group. Let $G = \mathrm{GL}(n, \mathbf{C})$. The identity element e is the identity matrix I in G. Then $T_IG = \mathbf{R}^{2n^2} = \mathbf{C}^{n^2}$, where $\mathbf{C}^{n^2}$ is the space of $n \times n$ complex matrices, which may be identified with the Lie algebra $\mathfrak{gl}(n, \mathbf{C})$ as vector spaces of $\mathrm{GL}(n, \mathbf{C})$. For $X, Y \in \mathbf{C}^{n^2}$, define an inner product on $\mathbf{C}^{n^2}$ by

$$\langle X, Y \rangle = \mathrm{Re}\,\mathrm{tr}\,(X^* Y), \tag{1.15}$$

where Re denotes the real part.

If $X = \{a_{jk} + ib_{jk}\}$ and $Y = \{\alpha_{jk} + i\beta_{jk}\}$, where $i = \sqrt{-1}$, then

$$\langle X, Y \rangle = \sum_{j,k} a_{jk}\alpha_{jk} + \sum_{j,k} b_{jk}\beta_{jk}.$$

So it is the usual Euclidean inner product, where X and Y are regarded as vectors in $\mathbf{R}^{2n^2}$. So, according to Proposition 1.2.4, the inner product $\langle \cdot, \cdot \rangle$ on $\mathbf{C}^{n^2}$ induces a unique left invariant Riemannian metric on $\mathrm{GL}(n, \mathbf{C})$, given by

$$\langle\, X, Y \,\rangle_A = \langle\, A^{-1}X, A^{-1}Y \,\rangle, \tag{1.16}$$

for any $A \in \mathrm{GL}(n, \mathbf{C})$ and $X, Y \in \mathbf{C}^{n^2}$.

1.3.2 Riemannian metric of closed subgroups of $\mathrm{GL}(n, \mathbf{C})$

A closed subgroup of $\mathrm{GL}(n, \mathbf{C})$ is a Lie subgroup. If G is a closed subgroup of $\mathrm{GL}(n, \mathbf{C})$, then the left invariant Riemannian metric on $\mathrm{GL}(n, \mathbf{C})$ induces a left invariant Riemannian metric on G by restriction, that is,

$$\langle X, Y \rangle_A = \langle A^{-1}X, A^{-1}Y \rangle, \quad \text{for any } A \in G, \quad X, Y \in \mathfrak{g},$$

where $\mathfrak{g}$ is Lie algebra of G. For example, since $\mathrm{GL}(n, \mathbf{R})$ is a closed subgroup of $\mathrm{GL}(n, \mathbf{C})$, then the induced left invariant Riemannian metric on $\mathrm{GL}(n, \mathbf{R})$ is given by

$$\langle\, X, Y \,\rangle_A = \langle\, A^{-1}X, A^{-1}Y \,\rangle = \mathrm{tr}\,[(A^{-1}X)^T (A^{-1}Y)]. \tag{1.17}$$

1.3.3 Riemannian metric on positive definite matrices space

Let $\mathbb{P}_n$ be the space of all positive definite matrices in $\mathrm{GL}(n, \mathbf{C})$. Since $\mathbb{P}_n$ is a closed submanifold of $\mathrm{GL}(n, \mathbf{C})$, it may equipped with the induced Riemannian metric from $\mathrm{GL}(n, \mathbf{C})$ and becomes a sub-Riemannian manifold of $\mathrm{GL}(n, \mathbf{C})$. In the literature, a different Riemmanian metric is usually used that is invariant under the following congruence transformation on $\mathbb{P}_n$:

$$(g, p) \mapsto gpg^*, \tag{1.18}$$

for any $g \in G = \mathrm{GL}(n, \mathbf{C})$ and $p \in \mathbb{P}_n$. Note that the isotropy subgroup K at the identity matrix $I \in \mathbb{P}_n$ is the unitary group $\mathrm{U}(n)$, the group of all $n \times n$ unitary matrices. It is easy to see that the inner product $\langle \cdot, \cdot \rangle$ in (1.15) is $\mathrm{U}(n)$-invariant, so by Proposition (1.2.4), it induces a unique Riemannian metric on $\mathbb{P}_n$, that is invariant under the G-action on $\mathbb{P}_n$ defined by (1.18). However, in literature [1], people often use a different Riemannian metric on $\mathbb{P}(n)$ that is congruence invariant given by

$$\langle X, Y \rangle_p = \langle p^{-1/2} X p^{-1/2}, p^{-1/2} Y p^{-1/2} \rangle, \tag{1.19}$$

for $X, Y \in T_p\mathbb{P}_n$ and $p \in \mathbb{P}_n$.

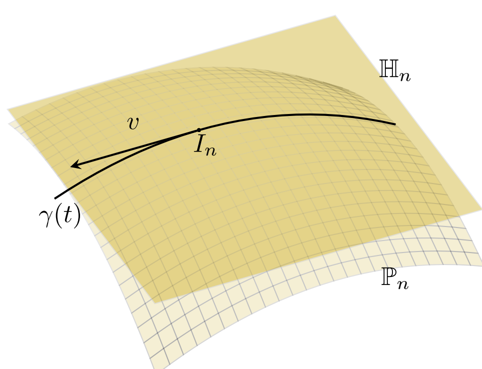

Figure 1.2: Tangent space $\mathbb{H}_n$ of $\mathbb{P}_n$

Because $T_p\mathbb{P}_n$ is the space of $n \times n$ Hermitian matrices $\mathbb{H}_n$, which is illustrated in Figure 1.2, by (1.15) and the properties of the trace,

$$\langle X, Y \rangle_p = \mathrm{Re}\,\mathrm{tr}\,(p^{-1}Xp^{-1}Y). \tag{1.20}$$

We remark that $\mathbb{H}_n$, as a tangent space of $\mathbb{P}_n$, contains $\mathbb{P}_n$. So the figure is somewhat misleading as the geometry is more intricate.

1.4 Geodesics

1.4.1 Geodesics in $\mathrm{GL}(n, \mathbf{C})$

Lemma 1.4.1. *For any $H(t), h(t) \in \mathbf{C}^{n\times n}$,*

$$\langle [H(t), h(t)], H(t) \rangle^E = \langle h(t), [H^*(t), H(t)] \rangle^E, \tag{1.21}$$

where $\langle \cdot, \cdot \rangle^E$ denotes the Euclidean metric.

Proof.

$$\begin{aligned}
& \langle [H(t), h(t)], H(t) \rangle^E \\
= \; & \mathrm{tr}[(H(t)h(t) - h(t)H(t))^* H(t)] \\
= \; & \mathrm{tr}[h^*(t)H^*(t)H(t) - H^*(t)h^*(t)H(t)] \\
= \; & \mathrm{tr}[h^*(t)H^*(t)H(t) - h^*(t)H(t)H^*(t)] \\
= \; & \mathrm{tr}[h^*(t)(H^*(t)H(t) - H(t)H^*(t))] \\
= \; & \langle h(t), [H^*(t), H(t)] \rangle^E.
\end{aligned}$$

□

Theorem 1.4.2. *Let $\gamma\colon [0,1] \to \mathrm{GL}(n,\mathbf{C})$ be geodesic connecting $A, B \in \mathrm{GL}(n,\mathbf{C})$. By minimizing the energy function*

$$\int_0^1 \langle\, H(t), H(t)\,\rangle^E \, dt, \tag{1.22}$$

where $H(t) = \gamma^{-1}(t)\dot{\gamma}(t)$, we have

$$\dot{H}(t) = H^*(t)H(t) - H(t)H^*(t). \tag{1.23}$$

Proof. For any real number ε, the pertubation of the curve $\gamma(t)$ is given by $\gamma(t,\varepsilon) = \gamma(t)e^{\varepsilon h(t)}$, where $h(t) \in \mathbf{C}^{n\times n}$. Let $\gamma(0,\varepsilon) = A$, $\gamma(1,\varepsilon) = B$ for any ε. Then, we have

$$H(t,\varepsilon) = \gamma^{-1}(t,\varepsilon)\dot{\gamma}(t,\varepsilon) \in \mathrm{GL}(n,\mathbf{C})$$

with $H(t,0) = H(t)$. So

$$\dot{\gamma}(t,\varepsilon) = \frac{d\gamma(t,\varepsilon)}{dt} = \dot{\gamma}(t)e^{\varepsilon h(t)} + \varepsilon\gamma(t)\dot{h}(t)e^{\varepsilon h(t)}.$$

So, we have

$$\begin{aligned}
&\frac{dH(t,\varepsilon)}{d\varepsilon}\\
= \;& \gamma^{-1}(t,\varepsilon)\frac{d\dot{\gamma}(t,\varepsilon)}{d\varepsilon} + \frac{d\gamma^{-1}(t,\varepsilon)}{d\varepsilon}\dot{\gamma}(t,\varepsilon)\\
= \;& \gamma^{-1}(t,\varepsilon)\frac{d(\dot{\gamma}(t)e^{\varepsilon h(t)} + \varepsilon\gamma(t)\dot{h}(t)e^{\varepsilon h(t)})}{d\varepsilon} - \gamma^{-1}(t,\varepsilon)\frac{d\gamma(t,\varepsilon)}{d\varepsilon}\gamma^{-1}(t,\varepsilon)\dot{\gamma}(t,\varepsilon)\\
= \;& \gamma^{-1}(t,\varepsilon)[\dot{\gamma}(t)h(t) + \gamma(t)\dot{h}(t) + \varepsilon\gamma(t)\dot{h}(t)]e^{\varepsilon h(t)} - \gamma^{-1}(t,\varepsilon)\gamma(t)h(t)e^{\varepsilon h(t)}H(t,\varepsilon).
\end{aligned}$$

Setting $\varepsilon = 0$, we have

$$\begin{aligned}
\frac{dH(t,\varepsilon)}{d\varepsilon}\Big|_{\varepsilon=0} &= \gamma^{-1}(t,0)(\dot{\gamma}h(t) + \gamma(t)\dot{h}(t)) - \gamma^{-1}(t,0)\gamma(t)h(t)H(t,0) \\
&= \gamma^{-1}(t,0)(\dot{\gamma}(t)h(t) + \gamma(t)\dot{h}(t) - \gamma(t)h(t)H(t)) \\
&= \dot{h}(t) + H(t)h(t) - h(t)H(t).
\end{aligned}$$

By minimizing the energy function $\int_0^1 \langle H(t), H(t)\rangle^E \, dt$, we have

$$\frac{d}{d\varepsilon}\int_0^1 \langle\, H(t,\varepsilon), H(t,\varepsilon)\,\rangle^E \, dt = 2\int_0^1 \langle\, \frac{dH(t,\varepsilon)}{d\varepsilon}, H(t,\varepsilon)\,\rangle^E \, dt.$$

Setting $\varepsilon = 0$, we have

$$\begin{aligned}
& \int_0^1 \langle\, \frac{dH(t,\varepsilon)}{d\varepsilon}, H(t,\varepsilon)\,\rangle^E \, dt|_{\varepsilon=0} \\
=\ & \int_0^1 \langle\, \dot{h}(t) + H(t)h(t) - h(t)H(t), H(t)\,\rangle^E \, dt \\
=\ & \int_0^1 \langle\, \dot{h}(t), H(t)\,\rangle^E + \langle\, H(t)h(t) - h(t)H(t), H(t)\,\rangle^E dt \\
=\ & \langle\, h(t), H(t)\,\rangle^E|_0^1 + \int_0^1 (-\langle h(t), \dot{H}(t)\,\rangle^E + \langle\, h(t), H^*(t)H(t) - H(t)H^*(t)\,\rangle^E)dt \\
=\ & \int_0^1 \langle\, h(t), -\dot{H}(t) + H^*(t)H(t) - H(t)H^*(t)\,\rangle^E \, dt = 0.
\end{aligned}$$

Since $h(t) \in \mathbf{C}^{n\times n}$ is arbitrary, we get

$$\dot{H}(t) = H^*(t)H(t) - H(t)H^*(t).$$

□

Theorem 1.4.3. *Let $\gamma\colon [0,1] \to \mathrm{GL}(n,\mathbf{C})$ be geodesic connecting $A, B \in \mathrm{GL}(n,\mathbf{C})$. The geodesic starting from $\gamma(0) = A$ with $\dot{\gamma}(0) = X$ is given by*

$$\gamma_{A,X}(t) = Ae^{t(A^{-1}X)^*}e^{t[A^{-1}X-(A^{-1}X)^*]}. \tag{1.24}$$

Proof. Since the solution of Equation (1.23) is the geodesic on the group $\mathrm{GL}(n, \mathbf{C})$, we need to verify the geodesic Equation (1.24) satisfying the Equation (1.23). Since

$$H(t) = \gamma^{-1}(t)\dot{\gamma}(t)$$

and

$$\dot{\gamma}(t) = Ae^{t(A^{-1}X)^*}(A^{-1}X)e^{t[A^{-1}X-(A^{-1}X)^*]},$$

we have

$$\begin{aligned} H(t) &= e^{t[(A^{-1}X)^*-A^{-1}X]}e^{-t(A^{-1}X)^*}A^{-1}Ae^{t(A^{-1}X)^*}(A^{-1}X)e^{t[A^{-1}X-(A^{-1}X)^*]} \\ &= e^{t[(A^{-1}X)^*-A^{-1}X]}(A^{-1}X)e^{t[A^{-1}X-(A^{-1}X)^*]}. \end{aligned}$$

Noting that

$$\dot{H}(t) = e^{t[(A^{-1}X)^*-A^{-1}X]}[(A^{-1}X)^*(A^{-1}X) - (A^{-1}X)(A^{-1}X)^*]e^{t[A^{-1}X-(A^{-1}X)^*]},$$

we compute

$$\begin{aligned} & H^*(t)H(t) - H(t)H^*(t) \\ = \; & e^{t[(A^{-1}X)^*-A^{-1}X]}(A^{-1}X)^*(A^{-1}X)e^{t[A^{-1}X-(A^{-1}X)^*]} \\ & -e^{t[(A^{-1}X)^*-A^{-1}X]}(A^{-1}X)(A^{-1}X)^*)e^{t[A^{-1}X-(A^{-1}X)^*]} \\ = \; & e^{t[(A^{-1}X)^*-A^{-1}X]}[(A^{-1}X)^*(A^{-1}X) - (A^{-1}X)(A^{-1}X)^*]e^{t[A^{-1}X-(A^{-1}X)^*]} \\ = \; & \dot{H}(t). \end{aligned}$$

□

1.4.2 Geodesics in subgroups of $\mathrm{GL}(n, \mathbf{C})$

Let G be a closed subgroup of $\mathrm{GL}(n, \mathbf{C})$. The $*$-invariant Riemmanian metric on G has been defined similar to equation (1.16). So, for $A, B \in G$, the geodesic starting from A with $\dot{\gamma}(0) = X$ has the same form

$$\gamma_{A,X}(t) = Ae^{t(A^{-1}X)^*}e^{t[A^{-1}X-(A^{-1}X)^*]}. \tag{1.25}$$

For $A, B \in \mathrm{U}(n)$ or $\mathrm{O}(n)$, this geodesic connecting A and B takes a simple form as

$$\gamma_{A,X}(t) = Ae^{t(A^{-1}X)}. \tag{1.26}$$

For example, since $\mathrm{GL}(n, \mathbf{R})$ is a closed subgroup of $\mathrm{GL}(n, \mathbf{C})$. The $*$-invariant Riemmanian metric on $\mathrm{GL}(n, \mathbf{R})$ is left invariant Riemmanian metric. Let $\gamma \colon [0, 1] \to \mathrm{GL}(n, \mathbf{R})$ be geodesic connecting $A, B \in \mathrm{GL}(n, \mathbf{R})$ with $\gamma(0) = A$ and $\gamma(1) = B$. The geodesic starting from A and $\dot{\gamma}(0) = X$ is

$$\gamma_{A,X}(t) = Ae^{t(A^{-1}X)^T}e^{t[A^{-1}X-(A^{-1}X)^T]}. \tag{1.27}$$

An explicit formula for geodesics in $\mathrm{GL}(n, \mathbf{R})$ under a more general left invariant metric can be found in Martin-Neff [14].

1.4.3 Geodesics in $\mathbb{P}_n$ the space of positive definite matrices

Theorem 1.4.4. *The geodesic starting from $P(0) = P$ in the space of positive definite matrices with $\dot{P}(0) = X \in T_P P$ under the congruence invariant metric is*

$$\gamma_{P,X}(t) = P^{1/2}e^{t(P^{-1/2}XP^{-1/2})}P^{1/2}. \tag{1.28}$$

Proof. The geodesics on $\mathbb{P}_n$ can be found as on $\mathrm{GL}(n, \mathbf{C})$ by a variation method, essentially repeating the computation in the proofs of Lemma 1.4.1, Theorem 1.4.2 and Theorem 1.4.3, but using the inner product $\langle X, Y \rangle = \mathrm{Re}\,\mathrm{tr}\,(XY)$ instead of $\langle X, Y \rangle = \mathrm{Re}\,\mathrm{tr}\,(X^*Y)$. The computation is simpler now because various matrices are now Hermitian. In particular, we now have $\dot{H}(t) = 0$, where $H(t) = \gamma(t)^{-1}\dot{\gamma}(t)$ as before. It is easy to verify $\gamma(t) = \exp(tX)$ is the solution of the above differential equation with $\gamma(0) = I$ and $\dot{\gamma}(0) = X$, so it is a geodesic starting at I. The general geodesic expression in (1.28) can be obtained by a congruence transformation. $\square$

1.4.4 Riemannian distance function

Let G be a Riemannian manifold and let $\gamma\colon [a, b] \to G$ with $\gamma(a) = A$ and $\gamma(b) = B$. The length of γ is defined by

$$L(\gamma) = \int_a^b \|\dot{\gamma}(t)\|\; dt. \tag{1.29}$$

The distance $d_G(A, B)$ between A and B on G is defined as the infimum of the length function $L(\gamma)$.

On $\mathbf{R}^n$ with the Euclidean metric, any straight line segment is the shortest piecewise smooth curve segment between its endpoints. Thus, the distance function is equal to the Euclidean distance $d_{\mathbf{R}^n}(A, B) = \|A - B\|$.

On other spaces, it is not as easy as $\mathbf{R}^n$. Matrix logarithm would be introduced here to define the Riemannian distance. When all the eigenvalues of matrix A lies in the nonnegative real line, called principal logarithm, denoted by $\log A$. Moreover, for any given matrix norm $\|\cdot\|$, if $\|A - I\| < 1$, then $\log A$ is defined as

$$\log A = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(A - I)^k}{k}.$$

With Remannian metric, for $\mathrm{U}(n)$, $\mathrm{O}(n)$ and $\mathbb{P}_n$, the Riemannian distance function on G is defined by

$$d_G(A, B) = \left\| \log B^{-1} A \right\|, \tag{1.30}$$

when $\|B^{-1}A - I\| < 1$.

1.5 Gradients

1.5.1 Definition of Riemannian gradient

In order to average data, geodesic and gradient are two important tools. Since the gradient gives the rate and direction of fastest increase, it is widely used in numerical computation, for example, the geodesic-based RSDA [5].

Let $\partial_A f$ denote the Euclidean gradient of function f with Euclidean metric and $\nabla_A f$ be the Riemannian gradient of function f with Riemannian metric.

Definition 1.5.1. *Let $f\colon \mathrm{GL}(n, \mathbf{C}) \to \mathbf{R}$ be a differentiable function and $\nabla_A f \in \mathrm{T}_A\mathrm{GL}(n, \mathbf{C})$ denote the Riemannian gradient of function f with respect to the metric (1.16), which is defined by the following condition*

$$\langle \nabla_A f, X \rangle_A = \langle X, \partial f(A) \rangle^E. \tag{1.31}$$

1.5.2 Gradient on matrix spaces

Theorem 1.5.2. *The Riemannian gradient of a sufficiently regular function $f\colon \mathrm{GL}(n, \mathbf{C}) \to \mathbf{C}$ associated to the Riemannian metric (1.16) satisfies*

$$\nabla_A f = AA^* \partial f(A). \tag{1.32}$$

Proof. According to Equation (1.31), for every $X \in \mathrm{T}_A\mathrm{GL}(n, \mathbf{C})$, we have

$$
\begin{aligned}
\langle X, \nabla f \rangle_A &= \langle A^{-1}X, A^{-1}\nabla f(A) \rangle \\
&= \mathrm{tr}\,[(A^{-1}X)^*(A^{-1}\nabla f(A))] \\
&= \mathrm{tr}\,[X^*(A^{-1})^*A^{-1}\nabla f(A)] \\
&= \langle X, \partial f(A) \rangle_E \\
&= \mathrm{tr}\,[X^*\partial f(A)].
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
(A^{-1})^*A^{-1}\nabla f(A) &= \partial f(A) \\
\nabla f(A) &= AA^*\partial f(A).
\end{aligned}
$$

□

Since X is arbitrary,

$$\mathrm{tr}\,[X^*((A^{-1})^*A^{-1}\nabla f(A) - \partial f(A))] = 0,$$

that is, for any $X \in \mathrm{T}_A\mathrm{GL}(n, \mathbf{R})$,

$$\langle X, (A^{-1})^*A^{-1}\nabla f(A) - \partial f(A) \rangle = 0.$$

Let G be a closed subgroup of $\mathrm{GL}(n, \mathbf{C})$. The left invariant Riemmanian metric on G have been defined similar to Equation (1.16). According to Proposition 1.2.4, the Riemannian gradient of the regular function f still satisfies $\nabla_A f = AA^*\partial f(A)$. For example, since $\mathrm{GL}(n, \mathbf{R})$ is a closed subgroup of $\mathrm{GL}(n, \mathbf{C})$, its Riemannian gradient is $\nabla_A f = AA^T\partial f(A)$.

Theorem 1.5.3. *The Riemannian gradient of a sufficiently regular function* $f\colon \mathrm{GL}(n, \mathbf{R}) \to \mathbf{R}$ *associated to the Riemannian metric (1.17) satisfies*

$$\nabla_A f = AA^T \partial f(A). \tag{1.33}$$

Proof. According to Equation (1.31), for every $X \in \mathrm{T}_A\mathrm{GL}(n, \mathbf{R})$, we have

$$\begin{aligned}
\langle\, X, \nabla f \,\rangle_A &= \langle\, A^{-1}X, A^{-1}\nabla f(A) \,\rangle \\
&= \mathrm{tr}\,[(A^{-1}X)^T(A^{-1}\nabla f(A))] \\
&= \mathrm{tr}\,[X^T A^{-T} A^{-1} \nabla f(A)] \\
&= \langle\, X, \partial f(A) \,\rangle_E \\
&= \mathrm{tr}\,[X^T \partial f(A)].
\end{aligned}$$

Thus, we have

$$\begin{aligned}
A^{-T}A^{-1}\nabla f(A) &= \partial f(A) \\
\nabla f(A) &= AA^T \partial f(A).
\end{aligned}$$

□

Since X is arbitrary,

$$\mathrm{tr}\,[X^T[A^{-T}A^{-1}\nabla f(A) - \partial f(A)]] = 0,$$

that is, for every $X \in \mathrm{T}_A\mathrm{GL}(n, \mathbf{R})$,

$$\langle\, X, A^{-T}A^{-1}\nabla f(A) - \partial f(A) \,\rangle = 0.$$

Theorem 1.5.4. *The Riemannian gradient of a sufficiently regular function $f\colon P(n) \to \mathbf{R}$ associated to the Riemannian metric (1.17) satisfies*

$$\nabla_P f = P \frac{\partial f(P) + \partial^* f(P)}{2} P. \tag{1.34}$$

Proof. According to Equation (1.31), for every $X \in \mathbb{H}_n$, we have

$$\begin{aligned} \langle X, \nabla f \rangle_p &= \langle P^{-1/2} X P^{-1/2}, P^{-1/2} \nabla f(P) P^{-1/2} \rangle \\ &= \mathrm{Re}\,\mathrm{tr}\,[(P^{-1}X)(P^{-1}\nabla f(P))] \\ &= \mathrm{Re}\,\mathrm{tr}\,[X(P^{-1}\nabla f(P) P^{-1})] \\ &= \langle X, \partial f(P) \rangle_E \\ &= \mathrm{Re}\,\mathrm{tr}\,[X \partial f(P)]. \end{aligned}$$

Thus, we have

$$\begin{aligned} X(P^{-1}\nabla f(P) P^{-1}) &= X \frac{\partial f(P) + \partial^* f(P)}{2} \\ \nabla f(P) &= P \frac{\partial f(P) + \partial^* f(P)}{2} P. \end{aligned}$$

□

1.5.3 Gradient of criterion function

Let $f\colon G \to \mathbf{R}$ be a criterion function as the Riemannian distances between sample points from the collection $\{B_1, B_2, \cdots, B_N\}$ and a fixed point A as below,

$$f(A) := \frac{1}{N} \sum_{k=1}^{N} \left\| \log(B_k^{-1} A) \right\|^2 . \tag{1.35}$$

where $A, B_1, B_2, \cdots, B_N \in G$.

By Karcher's theorem [8], if the data to be averaged is sufficiently close to each other, then the criterion function possesses exactly one minimizer. Thus the average matrix value of $\{B_1, B_2, \cdots, B_N\}$ is defined as

$$\bar{A} = \min_{A \in G} f(A).$$

1.5.4 An alternative proof of Moakher's Theorem

The following formula plays a key roll in computing the Euclidean gradient ∂f of the criterion function f on $\mathrm{U}(n)$, $\mathrm{O}(n)$ and $\mathbb{P}_n$. Moakher's proof [18] is based on an integral formula of matrix log. We have found a more direct and simpler proof.

Theorem 1.5.5. *[18] Let $A(t)$ be a matrix-valued function of the real variable t. Assume that for all t in its domain, $A(t)$ is an invertible matrix which does not have eigenvalues on the closed negative real line. Then*

$$\frac{d}{dt}\mathrm{tr}\,\left[(\log A(t))^2\right] = 2\mathrm{tr}\,\left[\log A(t) A^{-1}(t) \frac{d}{dt} A(t)\right]. \tag{1.36}$$

Proof. Here are some usual formula that are used in the following proof.

1. $\log A(t) = \sum\limits_{n=1}^{\infty} (-1)^{n-1} \frac{(A(t)-I)^n}{n}$ for $\|A(t) - I\| < 1$.
2. $\frac{d}{dt} \log A(t) = \sum\limits_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n} \left[\sum\limits_{k=0}^{n-1} (A(t) - I)^k \frac{d}{dt} A(t) (A(t) - I)^{n-k-1}\right]$.
3. $\mathrm{tr}\,(AB) = \mathrm{tr}\,(BA)$.
4. Any two analytic functions of the same matrix variable can commute.
5. $\sum\limits_{n=1}^{\infty} A^n = (I - A)^{-1}$ for $\|A\| < 1$.

For $\|A(t)-I\|<1$, we have

$$\begin{aligned}
\frac{d}{dt}\mathrm{tr}\,[(\log A(t))^2] &= 2\,\mathrm{tr}\,\left[\log A(t)\frac{d}{dt}\log A(t)\right]\\
&= 2\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}\mathrm{tr}\,\left[\log A(t)\sum_{k=0}^{n-1}(A(t)-I)^k\frac{d}{dt}A(t)(A(t)-I)^{n-k-1}\right]\\
&= 2\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}\mathrm{tr}\,\left[\sum_{k=0}^{n-1}(A(t)-I)^{n-k-1}\log A(t)(A(t)-I)^k\frac{d}{dt}A(t)\right]\\
&= 2\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}\mathrm{tr}\,\left[\sum_{k=0}^{n-1}\log A(t)(A(t)-I)^{n-k-1}(A(t)-I)^k\frac{d}{dt}A(t)\right]\\
&= 2\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}\mathrm{tr}\,\left[\sum_{k=0}^{n-1}\log A(t)(A(t)-I)^{n-1}\frac{d}{dt}A(t)\right]\\
&= 2\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}\mathrm{tr}\,\left[n\log A(t)(A(t)-I)^{n-1}\frac{d}{dt}A(t)\right]\\
&= 2\,\mathrm{tr}\,\left[\log A(t)\sum_{n=1}^{\infty}(-1)^{n-1}(A(t)-I)^{n-1}\frac{d}{dt}A(t)\right]\\
&= 2\,\mathrm{tr}\,\left[\log A(t)[I-(I-A(t))]^{-1}\frac{d}{dt}A(t)\right]\\
&= 2\,\mathrm{tr}\,\left[\log A(t)A^{-1}(t)\frac{d}{dt}A(t)\right].
\end{aligned}$$

Since $A(t)$ is analytic function and two analytic functions of the same matrix variable are always equivalent if they are equivalent in the open set. So, this formula fits all the analytic functions. Since

$$\frac{d}{dt}A(t)=\sum_{i,j}\frac{\partial}{\partial a_{ij}}A\frac{da_{ij}}{dt},$$

where $\frac{\partial A}{\partial a_{ij}}$ is an analytic function, the differentiable functions also satisfy this formula. $\square$

We remark that [20, Theorem 2] and [19, Equation (10)] are false since Moakher [18] computed $\frac{d}{dt}\mathrm{tr}\,(\log A(t))^2$, not $\frac{d}{dt}\mathrm{tr}\,(\log^T A(t)\log A(t))$. Also, a key step has the same mistake (see the penultimate equation [4, p.257]). Actually, both in the proof of [18, Proposition 2.1] and our alternative proof of Moakher's Theorem, we use the fact that two analytic functions of the same matrix variable commute. Since taking transpose is not an analytic function,

$\log^T A(t)$ and $\log A(t)$ do not commute in general in both Lorentz group and symplectic group. So, Moakher's Theorem cannot be applied in Lorentz group and symplectic group to compute $\frac{d}{dt}\mathrm{tr}\left[\log^T A(t)\log A(t)\right]$. This subtle oversight led to the incorrect conclusions on Riemannian gradients in [19] and [20].

Example 1.5.6. *According to [20, Equation (10)], the Euclidean gradient of the criterion function* $f(A) = \frac{1}{N}\sum_{k=1}^{N}\left\|\log(B_k^{-1}A)\right\|^2$*, where* $A, B_k \in \mathrm{Sp}(2\mathrm{n}, \mathbf{R})$ *for* $k = 1, 2, \ldots, N$*, is*

$$\partial_A f = \frac{2}{N}\sum_{k=1}^{N} A^{-T}\log(B_k^{-1}A).$$

Assume that $\partial_A f = 0$*. Then the above* A *is a critical point of the criterion function. We generate an example of* $\mathrm{Sp}(2, \mathbf{R})$ *when* $N = 2$*. Choose*

$$B_1 = \begin{bmatrix} -0.8012 & 0.3917 \\ -0.3882 & -1.0583 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 1.3109 & 0.2694 \\ -1.8177 & 0.3892 \end{bmatrix} \in \mathrm{Sp}(2, \mathbf{R}).$$

Since $\partial_A f = 0$*, we have* $\log(B_1^{-1}A) + \log(B_2^{-1}A) = 0$*. Taking matrix exponential, we get* $B_1^{-1}A = (B_2^{-1}A)^{-1}$*. By multiplying by* $B_1^{-1}A$*, we obtain* A *by computing* $B_1(B_1^{-1}B_2)^{\frac{1}{2}}$*, that is,*

$$A = \begin{bmatrix} 0.4822 & 0.6255 \\ -2.0868 & -0.6329 \end{bmatrix}.$$

However, we get $f(e^{0.99\log A}) - f(A) \approx -0.0522 < 0$ *and* $f(e^{1.01\log A}) - f(A) \approx -0.0555 > 0$*, which shows that* A *is not a critical point of the criterion function.*

Example 1.5.7. *According to [19, Equation (18)], the Euclidean gradient of the criterion function* $f(A) = \frac{1}{N}\sum_{q=1}^{N}\left\|\log(B_q^{-1}A)\right\|^2$*, where* $A, B_q \in \mathrm{O}(\mathrm{n}, \mathrm{k})$ *for* $q = 1, 2, \ldots, N$*, is*

$$\partial_A f = \frac{2}{N}\sum_{q=1}^{N} A^{-T}\log(B_q^{-1}A).$$

Assume that $\partial_A f = 0$. Then the above A is a critical point of the criterion function. We generate an example of $\mathrm{O}(1,1)$ *for $N = 2$. Choose*

$$B_1 = \begin{bmatrix} 2.8318 & -0.4587 \\ -0.4587 & 0.4274 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 7.7282 & -0.9171 \\ -0.9171 & 0.2382 \end{bmatrix} \in \mathrm{O}(1,1).$$

Similar to Example 1.5.6, we obtain A by computing $B_1(B_1^{-1}B_2)^{\frac{1}{2}}$, that is,

$$A = \begin{bmatrix} 4.6594 & -0.6070 \\ -0.6070 & 0.2937 \end{bmatrix}.$$

However, we get $f(e^{0.99 \log A}) - f(A) \approx -0.0103 < 0$ and $f(e^{1.01 \log A}) - f(A) \approx 0.0111 > 0$, which shows that A is not a critical point of the criterion function.

1.5.5 Gradient on Unitary Group $\mathrm{U}(n)$

The Lie group of unitary matrices is donoted by

$$\mathrm{U}(n) = \{Q \in \mathrm{GL}(n, \mathbf{C})\colon QQ^* = I\}.$$

The corresponding Lie algebra is denoted by

$$\mathfrak{u}(n) = \{A \in \mathfrak{gl}(n, \mathbf{C})\colon A^* = -A\}.$$

Theorem 1.5.8. *The Riemannian gradient of the criterion function (1.35) on* $\mathrm{U}(n)$ *is*

$$\nabla_A f = \frac{2}{N}\sum_{k=1}^{N} A \log(B_k^{-1}A). \tag{1.37}$$

Proof. Since $A, B_k \in \mathrm{U}(n)$, then $B_k^{-1}A \in \mathrm{U}(n)$. Then $\log(B_k^{-1}A) \in \mathfrak{u}(n)$, so $\log^*(B_k^{-1}A) = -\log(B_k^{-1}A)$. By Moakher's paper [18], we have

$$\begin{aligned}
df(A) &= \frac{1}{N}\sum_{k=1}^{N} d\left\|\log(B_k^{-1}A)\right\|^2 \\
&= \frac{1}{N}\sum_{k=1}^{N} d\left[\mathrm{tr}\left(\log^*(B_k^{-1}A)\log(B_k^{-1}A)\right]\right. \\
&= -\frac{1}{N}\sum_{k=1}^{N} d\left[\mathrm{tr}\left(\log^2(B_k^{-1}A)\right]\right. \\
&= -\frac{2}{N}\sum_{k=1}^{N} \mathrm{tr}\ \left[\log(B_k^{-1}A)(B_k^{-1}A)^{-1}d((B_k^{-1}A)\right] \\
&= -\frac{2}{N}\sum_{k=1}^{N} \mathrm{tr}\ \left[\log(B_k^{-1}A)A^{-1}dA\right] \\
&= \langle\, (-\frac{2}{N}\sum_{k=1}^{N}(\log(B_k^{-1}A)A^{-1})^*, dA\,\rangle^E.
\end{aligned}$$

Then, the Eculidean gradient is

$$\partial_A f = -\frac{2}{N}\sum_{k=1}^{N}\left[A^{-*}\log^*(B_k^{-1}A)\right] = \frac{2}{N}\sum_{k=1}^{N}\left[A\log(B_k^{-1}A)\right].$$

For any $X \in \mathrm{T}_A U(n)$, we have

$$\begin{aligned}
\langle X, \nabla_A f\rangle_A &= \langle A^{-1}X, A^{-1}\nabla_A f\rangle \\
&= \mathrm{Re}\ \mathrm{tr}\,[(A^{-1}X)^*(A^{-1}\nabla_A f)] \\
&= \langle X, \partial_A f\rangle^E \\
&= \mathrm{Re}\ \mathrm{tr}\,[(A^{-1}X)^* A^{-1}\partial_A f].
\end{aligned}$$

Then, we have

$$\mathrm{Re}\ \mathrm{tr}\,[(A^{-1}X)^*(A^{-1}\nabla_A f - A^{-1}\partial_A f)] = 0.$$

For $X \in \mathrm{T}_A U(n)$, there is

$$\langle (A^{-1}X)^*, (A^{-1}\nabla_A f - A^{-1}\partial_A f)\rangle = 0.$$

Note that $A^{-1}\nabla_A f$ is the projection of $A^{-1}\partial_A f$, that is,

$$\begin{aligned} A^{-1}\nabla_A f &= \tfrac{1}{2}(A^{-1}\partial_A f - \partial_A^* f A^{-*}) \\ &= \tfrac{1}{2}(A^{-1}\partial_A f - \partial_A^* f A). \end{aligned}$$

Thus, from the Riemannian gradient expression (1.33)of $\mathrm{GL}(n, \mathbf{R})$, the Riemannian gradient of this criterion function on $\mathrm{U}(n)$ is

$$\nabla_A f = \frac{2}{N}\sum_{k=1}^{N} A \log(B_k^{-1}A).$$

□

1.5.6 Gradient on the Orthogonal Group $\mathrm{O}(n)$

The Lie group of orthogonal matrices is denoted by

$$\mathrm{O}(n) = \{Q \in \mathrm{GL}(n, \mathbf{R})\colon QQ^T = I\}.$$

Then the corresponding Lie algebra is denoted by

$$\mathfrak{o}(n) = \{A \in \mathfrak{gl}(n, \mathbf{R})\colon A^T = -A\}.$$

Theorem 1.5.9. *The Riemannian gradient of the criterion function (1.35) on* $\mathrm{O}(n)$ *is*

$$\nabla_A f = \frac{2}{N}\sum_{k=1}^{N} A \log(B_k^{-1}A) \tag{1.38}$$

Proof. Since $A, B_k \in \mathrm{O}(n)$, then $B_k^{-1}A \in \mathrm{O}(n)$. Then $\log(B_k^{-1}A) \in \mathfrak{o}(n)$, so $\log^T(B_k^{-1}A) = -\log(B_k^{-1}A)$. By Moakher's result [18], we have

$$
\begin{aligned}
df(A) &= \frac{1}{N}\sum_{k=1}^{N} d\left\|\log(B_k^{-1}A)\right\|^2 \\
&= \frac{1}{N}\sum_{k=1}^{N} d\left[\mathrm{tr}\left(\log^T(B_k^{-1}A)\log(B_k^{-1}A)\right]\right. \\
&= -\frac{1}{N}\sum_{k=1}^{N} d\left[\mathrm{tr}\left(\log^2(B_k^{-1}A)\right]\right. \\
&= -\frac{2}{N}\sum_{k=1}^{N} \mathrm{tr}\ \left[\log(B_k^{-1}A)(B_k^{-1}A)^{-1}d((B_k^{-1}A)\right] \\
&= -\frac{2}{N}\sum_{k=1}^{N} \mathrm{tr}\ \left[\log(B_k^{-1}A)A^{-1}dA\right] \\
&= \langle(-\frac{2}{N}\sum_{k=1}^{N}(\log(B_k^{-1}A)A^{-1})^T, dA\rangle^E.
\end{aligned}
$$

Then, the Eculidean gradient is

$$
\partial_A f = -\frac{2}{N}\sum_{k=1}^{N} A^{-T}\log^T(B_k^{-1}A) = \frac{2}{N}\sum_{k=1}^{N} A\log(B_k^{-1}A).
$$

For any $X \in \mathrm{T}_A\mathrm{O}(n)$, we have

$$
\begin{aligned}
\langle\, X, \nabla_A f\,\rangle_A &= \langle\, A^{-1}X, A^{-1}\nabla_A f\,\rangle \\
&= \mathrm{Re}\ \mathrm{tr}\,[(A^{-1}X)^T(A^{-1}\nabla_A f)] \\
&= \langle\, X, \partial_A f\,\rangle^E \\
&= \mathrm{Re}\ \mathrm{tr}\,[(A^{-1}X)^T A^{-1}\partial_A f].
\end{aligned}
$$

Then, we have

$$
\mathrm{Re}\ \mathrm{tr}\,[(A^{-1}X)^T A^{-1}\nabla_A f - A^{-1}\partial_A f)] = 0.
$$

For $X \in \mathrm{T}_A U(n)$, there is

$$\langle\, (A^{-1}X)^T, (A^{-1}\nabla_A f - A^{-1}\partial_A f)\,\rangle = 0.$$

Note that $A^{-1}\nabla_A f$ is the projection of $A^{-1}\partial_A f$, that is,

$$\begin{aligned} A^{-1}\nabla_A f &= \frac{1}{2}(A^{-1}\partial_A f - \partial_A^T f A^{-T}) \\ &= \frac{1}{2}(A^{-1}\partial_A f - \partial_A^T f A). \end{aligned}$$

Thus, from the Riemannian gradient expression (1.33)of $\mathrm{GL}(n, \mathbf{R})$, the Riemannian gradient of this criterion function on $\mathrm{O}(n)$ is

$$\nabla_A f = \frac{2}{N}\sum_{k=1}^{N} A \log(B_k^{-1}A).$$

$\square$

1.5.7 Gradient on the space of positive definite matrices $\mathbb{P}_n$

Theorem 1.5.10. *The Riemannian gradient of the criterion function (1.35) on $\mathbb{P}_n$ is*

$$\nabla_A f = \frac{1}{N}\sum_{k=1}^{N} A \log(B_k^{-1}A) + \frac{1}{N}\sum_{k=1}^{N} \log^*(B_k^{-1}A)A. \tag{1.39}$$

Proof. Since the Riemannian distance of $A, B \in \mathbb{P}_n$ is

$$d(A,B) = \left\|\log(B^{-1}A)\right\| = \left[\sum_{i=1}^{n} \log^2 \lambda_i\right]^{1/2},$$

where $\lambda_i, i = 1, \cdots, n$ are the eigenvalues of $B^{-1}A$. We note that $B^{-1}A$ is similar to the symmetric matrix $B^{-1/2}AB^{-1/2}$. Then by Moakher's result [18], we have

$$
\begin{aligned}
df(A) &= \frac{1}{N}\sum_{k=1}^{N} d\left\|\log(B_k^{-1}A)\right\|^2 \\
&= \frac{1}{N}\sum_{k=1}^{N} d\left[\mathrm{tr}\,(\log(B_k^{-1/2}AB_k^{-1/2}))^2\right] \\
&= \frac{2}{N}\sum_{k=1}^{N}\mathrm{tr}\,\left[\log(B_k^{-1/2}AB_k^{-1/2})(B_k^{-1/2}AB_k^{-1/2})^{-1}d(B_k^{-1/2}AB_k^{-1/2})\right] \\
&= \frac{2}{N}\sum_{k=1}^{N}\mathrm{tr}\,\left[\log(B_k^{-1/2}AB_k^{-1/2})B_k^{1/2}A^{-1}B_k^{1/2}B_k^{-1/2}(dA)B_k^{-1/2}\right] \\
&= \frac{2}{N}\sum_{k=1}^{N}\mathrm{tr}\,\left[B_k^{-1/2}\log(B_k^{-1/2}AB_k^{-1/2})B_k^{1/2}A^{-1}dA\right] \\
&= \frac{2}{N}\sum_{k=1}^{N}\mathrm{tr}\,\left[\log(B_k^{-1}A)A^{-1}dA\right] \\
&= \langle\,(\frac{2}{N}\sum_{k=1}^{N}(\log(B_k^{-1}A)A^{-1})^*, dA\,\rangle^E.
\end{aligned}
$$

Then, the Eculidean gradient is

$$
\partial_P f = \frac{2}{N}\sum_{k=1}^{N}(A^{-1}\log(B_k^{-1}A)).
$$

Thus, from the Riemannian gradient expression (1.33) of $\mathrm{GL}(n, \mathbf{R})$, the Riemannian gradient of this criterion function on $\mathbb{P}_n$ is

$$
\begin{aligned}
\nabla_A f &= A\frac{\partial_A f + \partial_A^* f}{2}A \\
&= \frac{1}{2}A\left[\frac{2}{N}\sum_k(A^{-1}\log(B_k^{-1}A)) + (\frac{2}{N}\sum_k(A^{-1}\log(B_k^{-1}A)))^*\right]A \\
&= \frac{1}{N}\sum_{k=1}^{N}A\log(B_k^{-1}A) + \frac{1}{N}\sum_{k=1}^{N}\log^*(B_k^{-1}A)A.
\end{aligned}
$$

$\square$

Chapter 2

Curvature of matrix and reductive Lie groups

2.1 Introduction

The curvature provides important information about the geometric structure of a Riemannian manifold. For example, it is related to the rate at which two geodesics emitting from the same point move away from each other: the lower the curvature is, the faster they move apart (see Theorem IX.5.1 in [2, Chapter IX.5]). Many important geometric and topological properties are implied by suitable curvature conditions.

For example, for the three geodesic triangles in Figure 2.1, Figure 2.2 and Figure 2.3, the geodesics x_0x_1 and x_0x_2 start from the same point x_0 and move away from each other. The second graph in each of Figure 2.1, Figure 2.2 and Figure 2.3 shows the projection of these triangles. The curvature of the sphere, is positive. The curvature of the plane is zero. The hyperbola has negative curvature. These illustrate the fact: the lower the curvature is, the faster they move apart.

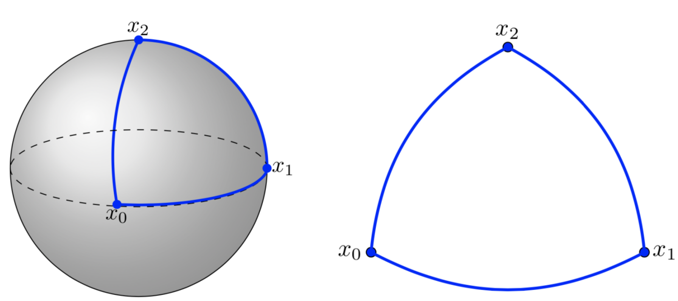

Figure 2.1: Curvature of the sphere

The classical geometric significance of the curvature tensor for a Riemannian manifold can be found in Helgason [7]. Let M be a Riemannian manifold of dimension 2 and let p be

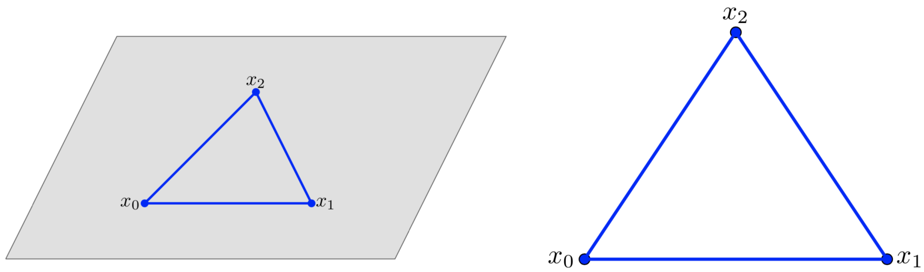

Figure 2.2: Curvature of the plane

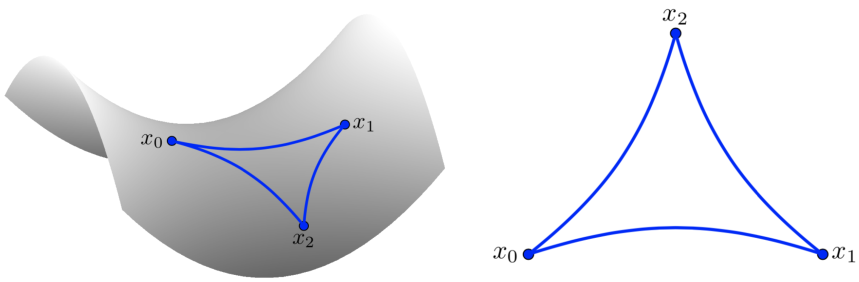

Figure 2.3: Curvature of the hyperbola

a point in M. Let $V_r(0)$ denote the open ball in the tangent place T_pM with center 0 and radius r. Suppose r is so small that Exp_p is a diffeomorphism of $V_r(0)$ onto the open ball $B_r(p)$. Let $A_0(r)$ and $A(r)$ denote the area of $V_r(0)$ and $B_r(p)$, respectively. The curvature of the 2-dimensional Riemannian manifold M of at the point $p \in M$ is defined as the limit

$$\kappa = 12 \lim_{r \to 0} \frac{A_0(r) - A(r)}{r^2 A_0(r)}.$$

For a general Riemannian manifold M of any dimension, let N_0 be a normal neighborhood of 0 in T_pM and let $N_p = \mathrm{Exp} N_0$. Let S be a two-dimensional vector subspace of T_pM. Then $\mathrm{Exp}(N_0 \cap S)$ is a connected submanifold of M of dimension 2 and has Riemannian structure induced by that of M. The curvature of $\mathrm{Exp}(N_0 \cap S)$ at p is called the sectional curvature of M at p along the plane section S.

In [16], Dr. Milnor gave an explicit formula for the curvature on a general Lie group equipped with a left invariant Riemannian metric. The curvature of a Riemannian manifold at a point can be described most easily by the bi-quadratic curvature function $\kappa(x,y) = \langle R_{xy}(x), y\rangle$. A given function $\kappa(x,y)$ can occur as curvature function for some Riemannian metric if and only if it is symmetric and bi-quadratic as a function of x and y, and vanishes whenever $x = y$. The real number $K = \kappa(u,v)$ is called the sectional curvature of the tangential 2-plane spanned by u and v if u and v are orthogonal unit vectors. The Lie algebra structure can then be described by $n \times n \times n$ array of structure constant

$$\alpha_{ijk} = \langle [e_i, e_j], e_k\rangle,$$

where $e_1, e_2, \ldots, e_n$ is an orthonormal basis of the Lie algebra. Then the explicit sectional curvature $\kappa(e_1, e_2)$ is given by the formula

$$\begin{aligned}\kappa(e_1,e_2) &= \sum_k (\frac{1}{2}\alpha_{12k}(-\alpha_{12k} + \alpha_{2k1} + \alpha_{k12}) \\ &\quad -\frac{1}{4}(\alpha_{12k} - \alpha_{2k1} + \alpha_{k12})(\alpha_{12k} + \alpha_{2k1} - \alpha_{k12}) - \alpha_{k11}\alpha_{k22}),\end{aligned}$$

which shows that the curvature can be computed completely from the information about the Lie algebra, together with its metric.

However, the curvature is usually not easy to compute explicitly. In the case of a Lie group equipped with a left invariant Riemannian metric, Milnor [16] obtained an explicit formula for sectional curvatures, but it is still quite complicated. To use it to compute a sectional curvature, one has to embed the spanning vectors of the section in an orthonormal frame and to compute the structure constants of the frame. Although it simplifies in many special cases, we have not seen a simple formula for the sectional curvature on the general linear group of matrices that is valid for all sections.

Our main result of this chapter is a simple and direct formula for the sectional curvature on the general linear group equipped with the left invariant Riemannian metric induced by the Frobenius norm. This formula also holds on any matrix group that is invariant under transposition, such as the orthogonal group and the Lorentz group, as they are totally geodesic submanifolds of the general linear group. Indeed, similar formula is valid for the more general reductive Lie groups. More details will be given later.

This chapter is organized as follows. After the preliminary material is introduced, we establish our main result, Theorem 2.4.8 on the sectional curvature for the general linear group in Section 2.4. In Section 2.5 we study the sectional curvature on $\mathrm{GL}(n, \mathbf{R})$ when the two tangent vectors are commuting matrices in $\mathfrak{gl}(n, \mathbf{R})$. In Section 2.6, we discuss the curvature of the subgroups of $\mathrm{GL}(n, \mathbf{R})$. In Section 2.7, our formula in Theorem 2.4.8 is extended to the reductive Lie groups.

2.2 Covariant derivative

Let G be $\mathrm{GL}(n, \mathbf{R})$. We may use the matrix elements g_{jk} of $g \in G$ as the local coordinates on G. Then any vector field X on G may be written as

$$X(g) = \sum_{j,k} a_{jk}(g) \frac{\partial}{\partial g_{jk}},$$

for some smooth functions $a_{jk}(g)$ on G. Assume X and Y are left-invariant vector fields on G, then

$$X(g) = \sum_{j,k} (gu)_{jk} \frac{\partial}{\partial g_{jk}}, \quad Y(g) = \sum_{i,j} (gv)_{ij} \frac{\partial}{\partial g_{ij}},$$

for $u, v \in \mathfrak{g}$ and $g \in G$.

For simplicity, we denote $X(g) = gu$ and $Y(g) = gv$. Then

$$\begin{aligned} D_X Y(g) &= X(gv) \\ &= \lim_{t\to\infty} \frac{(g+tX)v - gv}{t} \\ &= \lim_{t\to\infty} \frac{(g+tgu)v - gv}{t} \\ &= \lim_{t\to\infty} \frac{g(I+tu-I)v}{t} \\ &= guv. \end{aligned}$$

Lemma 2.2.1. *For $A \in \mathrm{GL}(n, \mathbf{R})$ and $X \in T_A\mathrm{GL}(n, \mathbf{R})$,*

$$\Gamma_A(X, X) = XX^*A^{-*} - AX^*A^{-*}A^{-1}X - XA^{-1}X, \tag{2.1}$$

where Γ denotes the Christoffel symbol.

Proof. Let a smooth path $\gamma : [0, 1] \to \mathrm{GL}(n, \mathbf{C})$ be a geodesic with $A \in \mathrm{GL}(n, \mathbf{C})$ and $X \in T_A\mathrm{GL}(n, \mathbf{C})$. According to Theorem 1.4.2, we know that the geodesic γ must satisfy the equation

$$\dot{H}(t) = H^*(t)H(t) - H(t)H^*(t),$$

where $H(t) = \gamma^{-1}(t)\dot{\gamma}(t)$. Then, the geodesic satisfies the following form.

$$-\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma} + \gamma^{-1}\ddot{\gamma} - \dot{\gamma}^*\gamma^{-*}\gamma^{-1}\dot{\gamma} + \gamma^{-1}\dot{\gamma}\dot{\gamma}^*\gamma^{-*} = 0$$

The geodesic equation, $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$, can be written in the form

$$\ddot{\gamma} + \Gamma_\gamma(\dot{\gamma}, \dot{\gamma}) = 0.$$

Thus,

$$\Gamma_\gamma(\dot{\gamma}, \dot{\gamma}) = \dot{\gamma}\dot{\gamma}^*\gamma^{-*} - \gamma\dot{\gamma}^*\gamma^{-*}\gamma^{-1}\dot{\gamma} - \dot{\gamma}\gamma^{-1}\dot{\gamma}.$$

Let $\gamma = A \in \mathrm{GL}(n, \mathbf{C})$ and $\dot{\gamma} = X \in T_A\mathrm{GL}(n, \mathbf{C})$. Then, we have

$$\Gamma_A(X, X) = XX^*A^{-*} - AX^*A^{-*}A^{-1}X - XA^{-1}X.$$

$\square$

Assume that the smooth path $\gamma : [0, 1] \to \mathrm{GL}(n, \mathbf{R})$ is the geodesic with $A \in \mathrm{GL}(n, \mathbf{R})$ and $X \in T_A\mathrm{GL}(n, \mathbf{R})$. Then, we have

$$\Gamma_A(X, X) = XX^TA^{-T} - AX^TA^{-T}A^{-1}X - XA^{-1}X. \tag{2.2}$$

For $A = g \in \mathrm{GL}(n, \mathbf{R})$,

$$\begin{aligned}\Gamma_g(X, X) &= gu(gu)^Tg^{-T} - g(gu)^Tg^{-T}g^{-1}gu - gug^{-1}gu \\ &= guu^T - gu^Tu - guu.\end{aligned}$$

$$\begin{aligned}\Gamma_g(X + Y, X + Y) &= g(u + v)(u + v)^T - g(u + v)^T(u + v) - g(u + v)(u + v) \\ &= g(uv^T + vu^T - u^Tv - v^Tu - uv - vu),\end{aligned}$$

$$\begin{aligned}\Gamma_g(X - Y, X - Y) &= g(u - v)(u - v)^T - g(u - v)^T(u - v) - g(u - v)(u - v) \\ &= g(-uv^T - vu^T + u^Tv + v^Tu + uv + vu).\end{aligned}$$

It is well known that Γ has the following properties:

1. $\Gamma(X, Y) = \Gamma(Y, X)$.
2. $\Gamma(\alpha_1X_1 + \alpha_2X_2, Y) = \alpha_1\Gamma(X_1, Y) + \alpha_2\Gamma(X_2, Y)$ and
   $\Gamma(X, \alpha_1Y_1 + \alpha_2Y_2) = \alpha_1\Gamma(X, Y_1) + \alpha_2\Gamma(X, Y_2)$.

Then, we have

$$\begin{aligned} 4\Gamma_g(X,Y) &= \Gamma_g(X+Y, X+Y) - \Gamma_g(X-Y, X-Y) \\ &= 2g(uv^T + vu^T - u^Tv - v^Tu - uv - vu). \end{aligned}$$

So, the covariant derivative can be expressed as

$$\begin{aligned} \nabla_X Y &= D_X Y + \Gamma(X,Y) \\ &= guv + \frac{1}{2}g(uv^T + vu^T - u^Tv - v^Tu - uv - vu) \\ &= \frac{1}{2}g(uv^T + vu^T - u^Tv - v^Tu + uv - vu) \\ &= \frac{1}{2}g([u, v^T] + [v, u^T] + [u, v]). \end{aligned}$$

2.3 Curvature Tensor

As stated in the previous section, let $G = \mathrm{GL}(n, \mathbf{R})$ be equipped with the left invariant Riemannian metric determined by the Frobenius inner product on the Lie algebra $\mathfrak{g} = \mathfrak{gl}(n, \mathbf{R})$ of G. Figure 2.4 shows that $\mathfrak{gl}(n, \mathbf{R})$ is the tangent space of $\mathrm{GL}(n, \mathbf{R})$ at the identity element I_n. Similar to $\mathbb{H}_n$ and $\mathbb{P}_n$, Figure 2.4 is a little bit misleading since $\mathfrak{gl}(n, \mathbf{R})$ contains $\mathrm{GL}(n, \mathbf{R})$ but not shown in this figure.

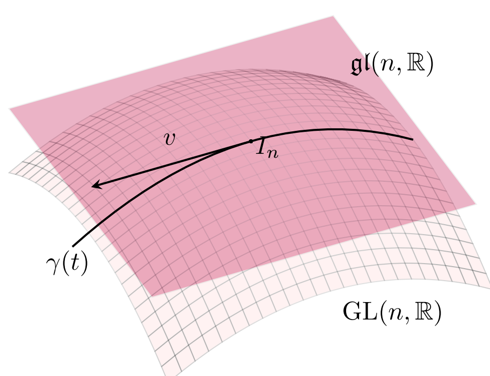

Figure 2.4: Tangent space $\mathfrak{gl}(n, \mathbf{R})$ of $\mathrm{GL}(n, \mathbf{R})$

The curvature tensor of the Riemannian connection is given by [7, p.43]

$$R(X,Y)Y = \nabla_X(\nabla_Y Y) - \nabla_Y(\nabla_X Y) - \nabla_{[X,Y]}Y, \tag{2.3}$$

where $[X,Y]$ is the Lie bracket of vector fields. The curvature tensor is to measure the intrinsic bending of G and the bending at each point is measured by the failure of mixed partial derivatives to commute. Let $X = gu$ and $Y = gv$ for $g \in \mathrm{GL}(n, \mathbf{R})$ and $u, v \in \mathbf{R}^{n\times n}$, the Lie bracket of vector fields is given by

$$\begin{aligned}
[X,Y] &= XY - YX \\
&= \sum_{j,k,p,q} (gu)_{jk}(gv)_{pq}\frac{\partial}{\partial g_{jk}}\frac{\partial}{\partial g_{pq}} + \sum_{j,k,p,q}(gu)_{jk}\frac{\partial}{\partial g_{jk}}(gv)_{pq}\frac{\partial}{\partial g_{pq}} \\
&\quad - \sum_{j,k,p,q} (gu)_{jk}(gv)_{pq}\frac{\partial}{\partial g_{jk}}\frac{\partial}{\partial g_{pq}} - \sum_{j,k,p,q}(gv)_{pq}\frac{\partial}{\partial g_{pq}}(gu)_{jk}\frac{\partial}{\partial g_{jk}} \\
&= \sum_{j,k,p,q}\left[(gu)_{jk}\frac{\partial}{\partial g_{jk}}(gv)_{pq}\frac{\partial}{\partial g_{pq}} - \sum_{j,k,p,q}(gv)_{pq}\frac{\partial}{\partial g_{pq}}(gu)_{jk}\frac{\partial}{\partial g_{jk}}\right] \\
&= \sum_{j,k,p,q}\left[(gu)_{jk}\left[\frac{\partial}{\partial g_{jk}}(\sum_r g_{pr}v_{rq})\right]\frac{\partial}{\partial g_{pq}} - (gv)_{pq}\left[\frac{\partial}{\partial g_{pq}}(\sum_r g_{jr}u_{rk})\right]\frac{\partial}{\partial g_{jk}}\right] \\
&= \sum_{p,q}\sum_{j,k}(gu)_{jk}\delta_{pj}v_{kq}\frac{\partial}{\partial g_{pq}} - \sum_{j,k}\sum_{p,q}(gv)_{pq}\delta_{pj}u_{qk}\frac{\partial}{\partial g_{jk}} \\
&= \sum_{p,q}\sum_{k}(gu)_{pk}v_{kq}\frac{\partial}{\partial g_{pq}} - \sum_{j,k}\sum_{q}(gv)_{jq}u_{qk}\frac{\partial}{\partial g_{jk}} \\
&= \sum_{p,q}(guv)_{pq}\frac{\partial}{\partial g_{pq}} - \sum_{j,k}(gvu)_{jk}\frac{\partial}{\partial g_{jk}} \\
&= guv - gvu \\
&= g[u,v].
\end{aligned}$$

Then

$$\nabla_{[X,Y]}Y = \tfrac{1}{2}g([[u,v],v^T] + [v,[u,v]^T] + [[u,v],v]).$$

Since

$$\begin{aligned}\nabla_X(\nabla_Y Y) &= \nabla_X(\tfrac{1}{2}g([v,v^T]+[v,v^T]+[v,v])) \\ &= \nabla_X(g[v,v^T]) \\ &= \tfrac{1}{2}g([u,[v,v^T]]+[[v,v^T],u^T]+[u,[v,v^T]]) \\ &= g[u,[v,v^T]]+\tfrac{1}{2}g[[v,v^T],u^T].\end{aligned}$$

and

$$\begin{aligned} & \nabla_Y(\nabla_X Y) \\ = \;& \nabla_Y(\tfrac{1}{2}g([u,v^T]+[v,u^T]+[u,v])) \\ = \;& \tfrac{1}{4}g\{[v,(2[u,v^T]+2[v,u^T]+[u,v]+[u,v]^T)]+[([u,v^T]+[v,u^T]+[u,v]),v^T]\} \\ = \;& \tfrac{1}{2}g[v,([u,v^T]+[v,u^T])]+\tfrac{1}{4}g[v,([u,v]+[u,v]^T)]+\tfrac{1}{4}g[([u,v^T]+[v,u^T]+[u,v]),v^T].\end{aligned}$$

Thus, the curvature tensor is

$$\begin{aligned}R(X,Y)Y =&\nabla_X(\nabla_Y Y)-\nabla_Y(\nabla_X Y)-\nabla_{[X,Y]}Y \\ =&g[u,[v,v^T]]+\tfrac{1}{2}g[[v,v^T],u^T]-\tfrac{1}{2}g[v,([u,v^T]+[v,u^T])]+\tfrac{1}{4}g[v,[u,v]] \\ &-\tfrac{3}{4}g[v,[u,v]^T]-\tfrac{1}{4}g[([u,v^T]+[v,u^T]),v^T]-\tfrac{3}{4}g[[u,v],v^T].\end{aligned} \tag{2.4}$$

2.4 Sectional Curvature

The sectional curvature of the section spanned by linearly independent u and v in $\mathfrak{g}$ is

$$S(u,v)=\frac{\langle R(u,v)v,u\rangle}{\langle u,u\rangle\langle v,v\rangle-\langle u,v\rangle^2}. \tag{2.5}$$

Note that the denominator is always positive. Indeed it is the area $|u\wedge v|$ of the parallelogram determined by u and v, so the sign of $S(u,v)$ is the same as that of $\langle R(u,v)v,u\rangle$. Moreover, when u and v are orthonormal, $S(u,v)=\langle R(u,v)v,u\rangle$.

When u and v are regarded as left invariant vector fields, the sectional curvature $S(u,v)$ is in general a function on G, but by the left invariance of the metric, it is a constant on G.

We will compute $\langle R(u,v)v,u\rangle$ for any $u,v \in \mathfrak{g}$, starting with some special cases. Recall that $\mathcal{S}$ and $\mathcal{A}$ are the spaces of symmetric and skew-symmetric matrices in $\mathfrak{g}$, respectively.

Because the Riemannian metric is left invariant, for any two left invariant vector fields $X = gu$ and $Y = gv$, the sectional curvature between X and Y does not depend on the point g, so may be computed at identity matrix I. Assume that $u \in \mathfrak{g}$ and $v \in \mathfrak{g}$ are orthonormal. The sectional curvature between $X = gu$ and $Y = gv$ is

$$\begin{aligned}
&\langle R(X,Y)Y,X\rangle \\
= \;& \operatorname{tr}[(R(X,Y)Y)X^T] \\
= \;& \operatorname{tr}\{[u,[v,v^T]]u^T - \tfrac{1}{2}[u^T,[v,v^T]]u^T - \tfrac{1}{2}[v,([u,v^T]+[v,u^T])]u^T + \tfrac{1}{4}[v,[u,v]]u^T \\
& \quad -\tfrac{3}{4}[v,[u,v]^T]u^T + \tfrac{1}{4}[v^T,([u,v^T]+[v,u^T])]u^T + \tfrac{3}{4}[v^T,[u,v]]u^T\}. \qquad (2.6)
\end{aligned}$$

If X,Y are orthonormal, then the sectional curvature between $X = gu$ and $Y = gv$ is $\langle R(X,Y)Y,X\rangle$ listed above. If X,Y are not orthonormal, then the sign of the sectional curvature between X and Y is the same as the sign of $\langle R(X,Y)Y,X\rangle$. Thus, we want to find $\langle R(X,Y)Y,X\rangle$ and focus on its sign, where u is arbitrary and v is symmetric or skew-symmetric.

To prove the theorem 2.4.1, we prove the following lemma.

Theorem 2.4.1. *Assume that $u \in \mathfrak{g}$ and $v \in \mathfrak{g}$ are symmetric. Then $\langle R(u,v)v,u\rangle$ is nonpositive. Additionally, it equals to zero if and only if u and v commute.*

Proof. Since u, v are symmetric,

$$\begin{aligned}
\langle R(u,v)v,u\rangle &= \operatorname{tr}\{\tfrac{1}{4}[v,[u,v]]u - \tfrac{3}{4}[v,[v,u]]u + \tfrac{3}{4}[v,[u,v]]u\} \\
&= \operatorname{tr}\{[v,[u,v]]u + \tfrac{3}{4}[v,[u,v]]u\} \\
&= \tfrac{7}{4}\operatorname{tr}\{[v,[u,v]]u\} \\
&= -\tfrac{7}{4}\langle [u,v],[u,v]\rangle \\
&= -\tfrac{7}{4}\,\|[u,v]\|^2 \\
&\leq 0,
\end{aligned}$$

with equality holds if and only if u, v commute. $\square$

Theorem 2.4.2. *Assume that $u \in \mathfrak{g}$ and $v \in \mathfrak{g}$ are skew-symmetric. Then $\langle R(u,v)v,u\rangle$ is nonnegative. Additionally, it is zero if and only if u and v commute.*

Proof.

$$\begin{aligned}
\langle R(u,v)v,u\rangle &= \operatorname{tr}\{-\tfrac{1}{4}[v,[u,v]]u + \tfrac{3}{4}[v,[v,u]]u + \tfrac{3}{4}[v,[u,v]]u\} \\
&= \operatorname{tr}\{\tfrac{1}{2}[v,[u,v]]u - \tfrac{3}{4}[v,[u,v]]u\} \\
&= -\tfrac{1}{4}tr\{[v,[u,v]]u\} \\
&= \tfrac{1}{4}\langle [u,v],[u,v]\rangle \\
&= \tfrac{1}{4}\,\|[u,v]\|^2 \\
&\geq 0,
\end{aligned}$$

with equality holds if and only if u, v commute. $\square$

Theorem 2.4.3. *Assume that u and v are in $\mathfrak{g}$, where u is symmetric and v are skew-symmetric. Then $\langle R(u,v)v,u\rangle$ is nonnegative. Additionally, it is zero if and only if u and v commute.*

Proof.

$$
\begin{aligned}
& \langle R(u,v)v,u\rangle \\
= \;& \operatorname{tr}\left\{[v,[u,v]]u + \tfrac{1}{4}[v,[u,v]]u - \tfrac{3}{4}[v,[u,v]]u + \tfrac{1}{2}[v,[u,v]]u - \tfrac{3}{4}[v,[u,v]]u\right\} \\
= \;& \tfrac{1}{4}\operatorname{tr}\left\{[v,[u,v]]u\right\} \\
= \;& \tfrac{1}{4}\langle [u,v],[u,v]\rangle \\
= \;& \tfrac{1}{4}\left\|[u,v]\right\|^2 \geq 0,
\end{aligned}
$$

with equality holds if and only if u, v commute. □

Let $\mathcal{S}$ be the space of symmetric matrices in $\mathfrak{g}$ and $\mathcal{A}$ be the space of skew-symmetric matrices in $\mathfrak{g}$, where $\mathfrak{g}$ is the Lie algebra of $\mathrm{GL}(n,\mathbf{R})$.

Theorem 2.4.4. *Let $X = gu$ and $Y = gv$ for $g \in \mathrm{GL}(n,\mathbf{R})$, $u \in \mathfrak{g}$ and $v \in \mathcal{A}$. Then*

$$\langle R(X,Y)Y,X\rangle = \tfrac{1}{4}\left\|[X,Y]\right\|^2,$$

which is nonnegative. Additionally, it is zero if and only if X and Y commute.

Proof. If $u \in \mathcal{S}$ or $u \in \mathcal{A}$, then it has a simple expression in Theorem 2.4.2 and Theorem 2.4.3. For these two cases, the sectional curvature is $\frac{1}{4}\left\|[X,Y]\right\|^2$ whose nonnegative with zero if and only if u, v are commute, that is, X and Y commute.

For any $u \in \mathfrak{g}$, we can decompose u uniquely as

$$u = \frac{u+u^T}{2} + \frac{u-u^T}{2}.$$

For simplicity, denote $(u+u^T)/2$ by u_1 and $(u-u^T)/2$ by u_2. It is trivial to know that u_1 is symmetric and u_2 is skew-symmetric. To find $\langle R(X,Y)Y,X\rangle$, we need to prove Claim 2.4.5

listed after this proof. Since the inner product has bilinear property, we have

$$
\begin{aligned}
& \langle R(X,Y)Y,X\rangle \\
= \; & \langle R(X_1+X_2,Y)Y,X_1+X_2\rangle \\
= \; & \langle R(X_1,Y)Y,X_1\rangle + \langle R(X_2,Y)Y,X_2\rangle + \langle R(X_1,Y)Y,X_2\rangle + \langle R(X_2,Y)Y,X_1\rangle \\
= \; & \tfrac{1}{4}\left\|[u_1,v]\right\|^2 + \tfrac{1}{4}\left\|[u_2,v]\right\|^2 \\
= \; & \tfrac{1}{4}\left\|[X,Y]\right\|^2 \\
\geq \; & 0.
\end{aligned}
$$

The equality holds if and only if $[X,Y]=0$, that is, X and Y commute. □

Claim 2.4.5. *Let $X_1 = gu_1$, $X_2 = gu_2$ and $Y = gv$ for $g \in \mathrm{GL}(n,\mathbf{R})$, $u_1 \in \mathcal{S}$ and $u_2, v \in \mathcal{A}$. Then we have $\langle R(X_1,Y)Y,X_2\rangle = 0$ and $\langle R(X_2,Y)Y,X_1\rangle = 0$.*

Proof. Because $[u_1,v]$ is symmetric and $[u_2,v]$ is skew-symmetric, $[u_1,v]$ and $[u_2,v]$ are orthogonal, that is, $\langle [u_1,v],[u_2,v]\rangle = 0$. Then, we have

$$
\begin{aligned}
& \langle R(X_1,Y)Y,X_2\rangle \\
= \; & \mathrm{tr}\,\{[u_1,[v,v^T]]u_2^T - \tfrac{1}{2}[u_1^T,[v,v^T]]u_2^T - \tfrac{1}{2}[v,([u_1,v^T]+[v,u_1^T])]u_2^T \\
& +\tfrac{1}{4}[v,[u_1,v]]u_2^T - \tfrac{3}{4}[v,[u_1,v]^T]u_2^T + \tfrac{1}{4}[v^T,([u_1,v^T]+[v,u_1^T])]u_2^T + \tfrac{3}{4}[v^T,[u_1,v]]u_2^T\} \\
= \; & -\tfrac{1}{4}\mathrm{tr}\,\{[v,[u_1,v]]u_2^T\} \\
= \; & \tfrac{1}{4}\mathrm{tr}\,\{[v,[u_1,v]]u_2\} \\
= \; & \tfrac{1}{4}\langle [u_1,v],[u_2,v]\rangle = 0,
\end{aligned}
$$

and

$$
\begin{aligned}
& \langle R(X_2,Y)Y,X_1\rangle \\
= \; & \operatorname{tr}\{[u_2,[v,v^T]]u_1^T - \tfrac{1}{2}[u_2^T,[v,v^T]]u_1^T - \tfrac{1}{2}[v,([u_2,v^T]+[v,u_2^T])]u_1^T + \tfrac{1}{4}[v,[u_2,v]]u_1^T \\
& -\tfrac{3}{4}[v,[u_2,v]^T]u_1^T + \tfrac{1}{4}[v^T,([u_2,v^T]+[v,u_2^T])]u_1^T + \tfrac{3}{4}[v^T,[u_2,v]]u_1^T\} \\
= \; & \tfrac{1}{4}\operatorname{tr}\{[v,[u_2,v]]u_1\} \\
= \; & \tfrac{1}{4}\langle [u_2,v],[u_1,v]\rangle = 0.
\end{aligned}
$$

□

Claim 2.4.6. *Let $X_1 = gu_1$, $Y = gv$ and $X_2 = gu_2 \in \mathcal{A}$ for $g \in \mathrm{GL}(n,\mathbf{R})$, $u_1, v \in \mathcal{S}$ and $u_2 \in \mathcal{A}$. Then we have $\langle R(X_1,Y)Y,X_2\rangle = 0$ and $\langle R(X_2,Y)Y,X_1\rangle = 0$.*

Proof. The proof is similar to Claim 2.4.5. □

Theorem 2.4.7. *Let $X = gu$ and $Y = gv$ for $g \in \mathrm{GL}(n,\mathbf{R})$, $u \in \mathfrak{g}$ and $v \in \mathcal{S}$. Let $X_1 = (X + X^T)/2$ and $X_2 = (X - X^T)/2$. Then*

$$
\langle R(X,Y)Y,X\rangle = -\tfrac{7}{4}\left\|[X_1,Y]\right\|^2 + \tfrac{1}{4}\left\|[X_2,Y]\right\|^2,
$$

which can be zero with nonzero $[X,Y]$.

Proof. If $u \in \mathcal{S}$, then it has a simple expression in Theorem 2.4.1. For this case, the sectional curvature is $-\frac{7}{4}\left\|[u,v]\right\|^2$ whose sign is nonpositive with zero if and only if u, v commute, that is, X and Y commute.

If $u \in \mathcal{A}$, then it has a simple expression in Theorem 2.4.3. For this case, the sectional curvature is $\frac{1}{4}\left\|[u,v]\right\|^2$ whose sign is nonnegative with zero if and only if u, v commute, that is, X and Y commute.

For any $u \in \mathfrak{g}$, we can decompose u as the form in the proof of Theorem 2.4.4. The following proof uses Claim 2.4.6.

$$\begin{aligned}
& \langle R(X,Y)Y,X\rangle \\
= \;& \langle R(X_1+X_2,Y)Y,X_1+X_2\rangle \\
= \;& \langle R(X_1,Y)Y,X_1\rangle + \langle R(X_2,Y)Y,X_2\rangle + \langle R(X_1,Y)Y,X_2\rangle + \langle R(X_2,Y)Y,X_1\rangle \\
= \;& -\tfrac{7}{4}\left\|[u_1,v]\right\|^2 + \tfrac{1}{4}\left\|[u_2,v]\right\|^2 \\
= \;& -\tfrac{7}{4}\left\|[X_1,Y]\right\|^2 + \tfrac{1}{4}\left\|[X_2,Y]\right\|^2.
\end{aligned}$$

Thus, given a symmetric Y, we can choose X to have $\langle R(X,Y)Y,X\rangle = 0$ with a nonzero $[X,Y]$. $\square$

Theorem 2.4.8. *Let $X = gu, Y = gv \in \mathrm{GL}(n,\mathbf{R})$ for $u,v \in \mathfrak{g}$. Then $\langle R(X,Y)Y,X\rangle$ is given by*

$$\langle R(X,Y)Y,X\rangle = -2\left\|[X_1,Y_1]\right\|^2 + \tfrac{1}{4}\left\|[X,Y]\right\|^2 + 2\langle [Y_1,X_1],[Y_2,X_2]\rangle. \tag{2.7}$$

Proof. For any $u,v \in \mathfrak{g}$, we can decompose u,v as the following form

$$u = \frac{u+u^T}{2} + \frac{u-u^T}{2},$$

and

$$v = \frac{v+v^T}{2} + \frac{v-v^T}{2}.$$

For simplicity, denote $\frac{1}{2}(u+u^T)$ by u_1, $\frac{1}{2}(u-u^T)$ by u_2, $\frac{1}{2}(v+v^T)$ by v_1 and $\frac{1}{2}(v-v^T)$ by v_2. It is trivial to know that u_1, v_1 are symmetric and u_2, v_2 are skew-symmetric.

Since the inner product has bilinear property and the property of curvature, $\langle R(X,Y)Z,W\rangle = \langle R(Z,W)X,Y\rangle$, the general sectional curvature between X and Y can be written as

$$\begin{aligned}
& \langle R(X,Y)Y,X\rangle \\
= & \langle R(X,Y_1+Y_2)(Y_1+Y_2),X\rangle \\
= & \langle R(X,Y_1)Y_1,X\rangle + \langle R(X,Y_2)Y_2,X\rangle + 2\langle R(X,Y_1)Y_2,X\rangle.
\end{aligned}$$

For the term $\langle R(X,Y_1)Y_1,X\rangle$, the sectional curvature between an arbitrary X and a symmetric Y_1 is $-\frac{7}{4}\left\|[X_1,Y_1]\right\|^2 + \frac{1}{4}\left\|[X_2,Y_1]\right\|^2$, which has been given in Theorem 2.4.7.

For the term $\langle R(X,Y_2)Y_2,X\rangle$, the sectional curvature between an arbitrary X and a skew-symmetric Y_2 is $\frac{1}{4}\left\|[X,Y_2]\right\|^2$, which has been given in Theorem 2.4.4.

For the term $\langle R(X,Y_1)Y_2,X\rangle$, it can be written as

$$\begin{aligned}
& \langle R(X,Y_1)Y_2,X\rangle \\
= & \langle R(X_1+X_2,Y_1)Y_2,X_1+X_2\rangle \\
= & \langle R(X_1,Y_1)Y_2,X_1\rangle + \langle R(X_2,Y_1)Y_2,X_2\rangle + \langle R(X_1,Y_1)Y_2,X_2\rangle + \langle R(X_2,Y_1)Y_2,X_1\rangle.
\end{aligned}$$

For the term $\langle R(X_2,Y_1)Y_2,X_2\rangle$ and $\langle R(X_2,Y_1)Y_2,X_1\rangle$, we can find the curvature tensor $R(X_2,Y_1)Y_2$ first.

$$\begin{aligned}
R(X_2,Y_1)Y_2 &= \nabla_{X_2}(\nabla_{Y_1}Y_2) - \nabla_{Y_1}(\nabla_{X_2}Y_2) - \nabla_{[X_2,Y_1]}Y_2 \\
&= -\tfrac{3}{4}g[u_2,[v_1,v_2]] - \tfrac{1}{4}g[v_1,[v_2,u_2]] + \tfrac{1}{2}g[v_2,[v_1,u_2]].
\end{aligned}$$

Since $[Y_1,X_2]$ and $[Y_2,X_2]$ are orthogonal, we have

$$\langle R(X_2,Y_1)Y_2,X_2\rangle = \tfrac{1}{4}\langle [Y_1,X_2],[Y_2,X_2]\rangle = 0.$$

The sectional curvature $\langle R(X_2, Y_1)Y_2, X_1\rangle$ will be the following form:

$$
\begin{aligned}
& \langle R(X_2, Y_1)Y_2, X_1\rangle \\
= \;& \operatorname{tr}\{(-\tfrac{3}{4}[u_2, [v_1, v_2]] - \tfrac{1}{4}[v_1, [v_2, u_2]] + \tfrac{1}{2}[v_2, [v_1, u_2]])u_1^T\} \\
= \;& -\tfrac{3}{4}\langle [u_2, [v_1, v_2]], u_1\rangle - \tfrac{1}{4}\langle [v_1, [v_2, u_2]], u_1\rangle + \tfrac{1}{2}\langle [v_2, [v_1, u_2]], u_1\rangle \\
= \;& -\tfrac{1}{4}\langle [u_2, [v_1, v_2]], u_1\rangle - (\tfrac{1}{2}\langle [u_2, [v_1, v_2]], u_1\rangle + \tfrac{1}{2}\langle [v_2, [u_2, v_1]], u_1\rangle) - \tfrac{1}{4}\langle [v_1, [v_2, u_2]], u_1\rangle \\
= \;& -\tfrac{1}{4}\langle [u_2, [v_1, v_2]], u_1\rangle + \tfrac{1}{2}\langle v_1, [v_2, u_2]], u_1\rangle - \tfrac{1}{4}\langle [v_1, [v_2, u_2]], u_1\rangle \\
= \;& -\tfrac{1}{4}\langle [v_1, v_2], [u_1, u_2]\rangle + \tfrac{1}{4}\langle [v_1, u_1], [v_2, u_2]\rangle \\
= \;& -\tfrac{1}{4}\langle [Y_1, Y_2], [X_1, X_2]\rangle + \tfrac{1}{4}\langle [Y_1, X_1], [Y_2, X_2]\rangle.
\end{aligned}
$$

For the term $\langle R(X_2, Y_1)Y_2, X_2\rangle$ and $\langle R(X_1, Y_1)Y_2, X_2\rangle$, we can find the curvature tensor $R(X_1, Y_1)Y_2$.

$$
\begin{aligned}
R(X_1, Y_1)Y_2 =& \nabla_{X_1}(\nabla_{Y_1}Y2) - \nabla_{Y_1}(\nabla_{X_1}Y_2) - \nabla_{[X_1, Y_1]}Y_2 \\
=& -\tfrac{1}{4}g[u_1, [v_1, v_2]] + \tfrac{1}{4}g[v_1, [u_1, v_2]] + \tfrac{1}{2}g[v_2, [u_1, v_1]].
\end{aligned}
$$

Since $[Y_1, X_1]$ and $[Y_2, X_1]$ are orthogonal, we have

$$\langle R(X_1, Y_1)Y_2, X_1\rangle = -\tfrac{3}{4}\langle [Y_1, X_1], [Y_2, X_1]\rangle = 0.$$

The sectional curvature $\langle R(X_1,Y_1)Y_2,X_2\rangle$ has the following form:

$$\begin{aligned}
&\langle R(X_1,Y_1)Y_2,X_2\rangle\\
=\;&\operatorname{tr}\{(-\tfrac{1}{4}[u_1,[v_1,v_2]]+\tfrac{1}{4}[v_1,[u_1,v_2]]+\tfrac{1}{2}[v_2,[u_1,v_1]])u_2^T\}\\
=\;&\tfrac{1}{4}\langle[u_1,[v_1,v_2]],u_2\rangle-\tfrac{1}{4}\langle[v_1,[u_1,v_2]],u_2\rangle-\tfrac{1}{2}\langle[v_2,[u_1,v_1]],u_2\rangle\\
=\;&-(\tfrac{1}{4}\langle[u_1,[v_2,v_1]],u_2\rangle+\tfrac{1}{4}\langle[v_1,[u_1,v_2]],u_2\rangle)-\tfrac{1}{2}\langle[v_2,[u_1,v_1]],u_2\rangle\\
=\;&\tfrac{1}{4}\langle[v_2,[v_1,u_1]],u_2\rangle-\tfrac{1}{2}\langle[v_2,[u_1,v_1]],u_2\rangle\\
=\;&\tfrac{1}{4}\langle[v_1,u_1],[v_2,u_2]\rangle+\tfrac{1}{2}\langle[v_1,u_1],[v_2,u_2]\rangle\\
=\;&\tfrac{3}{4}\langle[v_1,u_1],[v_2,u_2]\rangle\\
=\;&\tfrac{3}{4}\langle[Y_1,X_1],[Y_2,X_2]\rangle.
\end{aligned}$$

So,

$$\begin{aligned}
&\langle R(X,Y_1)Y_2,X\rangle\\
=\;&\langle R(X_1,Y_1)Y_2,X_1\rangle+\langle R(X_2,Y_1)Y_2,X_2\rangle+\langle R(X_1,Y_1)Y_2,X_2\rangle+\langle R(X_2,Y_1)Y_2,X_1\rangle\\
=\;&\tfrac{3}{4}\langle[Y_1,X_1],[Y_2,X_2]\rangle-\tfrac{1}{4}\langle[Y_1,Y_2],[X_1,X_2]\rangle+\tfrac{1}{4}\langle[Y_1,X_1],[Y_2,X_2]\rangle\\
=\;&\langle[Y_1,X_1],[Y_2,X_2]\rangle-\tfrac{1}{4}\langle[Y_1,Y_2],[X_1,X_2]\rangle.
\end{aligned}$$

Then, according to Claim 2.4.10, we have

$$\begin{aligned}
&\langle R(X,Y)Y,X\rangle\\
=\;&\langle R(X,Y_1)Y_1,X\rangle+\langle R(X,Y_2)Y_2,X\rangle+2\langle R(X,Y_1)Y_2,X\rangle\\
=\;&-\tfrac{7}{4}\left\|[X_1,Y_1]\right\|^2+\tfrac{1}{4}\left\|[X_2,Y_1]\right\|^2+\tfrac{1}{4}\left\|[X,Y_2]\right\|^2+2\langle[Y_1,X_1],[Y_2,X_2]\rangle-\tfrac{1}{2}\langle[Y_1,Y_2],[X_1,X_2]\rangle\\
=\;&-2\left\|[X_1,Y_1]\right\|^2+\tfrac{1}{4}\left\|[X,Y_1]\right\|^2+\tfrac{1}{4}\left\|[X,Y_2]\right\|^2+2\langle[Y_1,X_1],[Y_2,X_2]\rangle-\tfrac{1}{2}\langle[Y_1,Y_2],[X_1,X_2]\rangle\\
=\;&-2\left\|[X_1,Y_1]\right\|^2+\tfrac{1}{4}\{\left\|[X,Y_1]\right\|^2+\left\|[X,Y_2]\right\|^2-2\langle[X_1,X_2],[Y_1,Y_2]\rangle\}+2\langle[Y_1,X_1],[Y_2,X_2]\rangle\\
=\;&-2\left\|[X_1,Y_1]\right\|^2+\tfrac{1}{4}\left\|[X,Y]\right\|^2+2\langle[Y_1,X_1],[Y_2,X_2]\rangle.
\end{aligned}$$

$\square$

The sectional curvature Equation (2.7) is also valid on many matrix groups, such as Lorentz groups and symplectic group, as these are totally geodesic submanifolds of the general linear group.

Corollary 2.4.9. *The sectional curvature between any X and Y is nonpositive if X and Y commute.*

Proof. Since any X and Y is sum of a symmetric matrix and a skew-symmetric matrix. Let X_1, Y_1 be the symmetric matrices and X_2, Y_2 be the skew-symmetric matrices and $X = X_1 + X_2, Y = Y_1 + Y_2$. We have

$$[X, Y] = [X_1 + X_2, Y_1 + Y_2] = [X_1, Y_1] + [X_2, Y_2] + [X_1, Y_2] + [X_2, Y_1].$$

Since X and Y commute, $[X, Y] = 0$. Then the skew-symmetric $[X_1, Y_1] + [X_2, Y_2]$ and symmetric $[X_1, Y_2] + [X_2, Y_1]$ are zero, respectively. We have $[X_1, Y_1] = -[X_2, Y_2]$. Thus, the formula of the sectional curvature between X and Y can be simplified as

$$\begin{aligned}\langle R(X, Y)Y, X\rangle &= -2\left\|[X_1, Y_1]\right\|^2 + \tfrac{1}{4}\left\|[X, Y]\right\|^2 + 2\langle [Y_1, X_1], [Y_2, X_2]\rangle \\ &= -4\left\|[X_1, Y_1]\right\|^2 \leq 0.\end{aligned}$$

$\square$

Claim 2.4.10. *For any $X, Y \in \mathrm{GL}(n, \mathbf{R})$, the following equation*

$$\left\|[X, Y]\right\|^2 = \left\|[X, Y_1]\right\|^2 + \left\|[X, Y_2]\right\|^2 - 2\langle [Y_1, Y_2], [X_1, X_2]\rangle$$

always holds.

Proof. According to the Jacobi identity, we have

$$
\begin{aligned}
\|[X,Y]\|^2 &= \langle [X,Y],[X,Y]\rangle \\
&= \langle [X,Y_1+Y_2],[X,Y_1+Y_2]\rangle \\
&= \langle [X,Y_1],[X,Y_1]\rangle + \langle [X,Y_2],[X,Y_2]\rangle + 2\langle [X,Y_1],[X,Y_2]\rangle \\
&= \|[X,Y_1]\|^2 + \|[X,Y_2]\|^2 + 2\{\langle [X_1,Y_1],[X_2,Y_2]\rangle + \langle [X_2,Y_1],[X_1,Y_2]\rangle\} \\
&= \|[X,Y_1]\|^2 + \|[X,Y_2]\|^2 + 2\{\langle -[X_2,[X_1,Y_1]],Y_2\rangle + \langle [X_1,[X_2,Y_1]],Y_2\rangle\} \\
&= \|[X,Y_1]\|^2 + \|[X,Y_2]\|^2 + 2\{\langle [X_2,[Y_1,X_1]],Y_2\rangle + \langle [X_1,[X_2,Y_1]],Y_2\rangle\} \\
&= \|[X,Y_1]\|^2 + \|[X,Y_2]\|^2 + 2\langle -[Y_1,[X_1,X_2]],Y_2\rangle \\
&= \|[X,Y_1]\|^2 + \|[X,Y_2]\|^2 - 2\langle [X_1,X_2],[Y_1,Y_2]\rangle .
\end{aligned}
$$

□

For general linear group $\mathrm{GL}(n,\mathbf{R})$, their total geodesic submanifolds, for example, Lorentz group and symplectic group, have the same form of sectional curvature as the manifold $\mathrm{GL}(n,\mathbf{R})$.

2.5 Zero curvature and commutative property

Let u and v be two commuting matrices in $\mathfrak{g} = \mathfrak{gl}(n,\mathbf{R})$. They span a 2-dimensional abelian Lie sub-algebra of $\mathfrak{g}$. Let A be the associated abelian Lie subgroup of $G = \mathrm{GL}(n,\mathbf{R})$. The induced Riemannian metric on A from the ambient space G is invariant under translations on A, so locally A is a Euclidean space and it has zero curvature. It is interesting to know whether the 2-dimensional section of $\mathfrak{g}$ spanned by u and v also has a zero sectional curvature on G. It is also interesting to know, if the sectional curvature spanned by u and v is zero, whether u and v are commuting. We will see that the answers to these two questions are negative in general, but are positive under additional conditions.

Theorem 2.5.1. *The sectional curvature between any X and a skew-symmetric Y is zero if and only if they commute.*

Proof. This is the fact by the Theorem 2.4.4. $\langle R(X,Y)Y,X\rangle = \frac{1}{4}\left\|[X,Y]\right\|^2$ for any X and a skew-symmetric Y. Thus, $\langle R(X,Y)Y,X\rangle \geq 0$ with equality if and only if $[X,Y]=0$, that is, X and Y commute. □

Theorem 2.5.2. *The sectional curvature between any two symmetric X and Y is zero if and only if they commute.*

Proof. This is the fact by the Theorem 2.4.1. $\langle R(X,Y)Y,X\rangle = -\frac{7}{4}\left\|[X,Y]\right\|^2$ for any two symmetric X and Y. Thus, $\langle R(X,Y)Y,X\rangle \leq 0$ with equality if and only if $[X,Y]=0$, that is, X and Y commute. □

Example 2.5.3. *For $X, Y \in \mathrm{GL}(n,\mathbf{R})$, $\langle R(X,Y)Y,X\rangle$ can be zero if X and Y do not commute.*

Proof. By the Theorem 2.4.8, $\langle R(X,Y)Y,X\rangle$ is given by Equation (2.7), that is,

$$\langle R(X,Y)Y,X\rangle = -2\left\|[X_1,Y_1]\right\|^2 + \frac{1}{4}\left\|[X,Y]\right\|^2 + 2\langle [Y_1,X_1],[Y_2,X_2]\rangle.$$

For example,

$$X = \begin{bmatrix} 1 & 1+\sqrt{\frac{3}{5}} \\ -1+\sqrt{\frac{3}{5}} & 1 \end{bmatrix},\ Y = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}.$$

Then, X and Y can be decomposed as

$$X = X_1 + X_2 \text{ and } Y = Y_1 + Y_2,$$

where

$$X_1 = \begin{bmatrix} 1 & \sqrt{\frac{3}{5}} \\ \sqrt{\frac{3}{5}} & 1 \end{bmatrix},\ X_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

and

$$Y_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix},\ Y_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Thus, $\langle R(X,Y)Y, X\rangle = 0$ by Equation (2.7) with nonzero [X, Y]. $\square$

From Corollary 2.4.9, we know that

$$\langle R(X,Y)Y, X\rangle = -4\,\|[X_1, Y_1]\|^2$$

is nonpositive if X and Y commute. We know that if one of these two matrices is skew-symmetric, the $\langle R(X,Y)Y, X\rangle$ can be zero if and only if X and Y commute. Here gives an example to illustrate that $\langle R(X,Y)Y, X\rangle$ can be nonzero if X and Y commute for any X and Y, where none is skew-symmetric.

Example 2.5.4. *For $X, Y \in \mathrm{GL}(n, \mathbf{R})$, $\langle R(X,Y)Y, X\rangle$ can be nonzero if X and Y commute.*

Proof. If two matrices are simultaneously diagonalizable, then both matrices commute. For example, we choose X and Y by the following method,

$$\begin{aligned} X &= \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\frac{\pi}{6} & -\sin\frac{\pi}{6} & 0 \\ \sin\frac{\pi}{6} & \cos\frac{\pi}{6} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} -\frac{\sqrt{3}}{2}+2 & -\frac{1}{2} & \frac{\sqrt{3}}{2}-1 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\sqrt{3}+2 & -1 & \sqrt{3}-1 \end{bmatrix}, \end{aligned}$$

$$Y = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\frac{\pi}{3} & -\sin\frac{\pi}{3} & 0 \\ \sin\frac{\pi}{3} & \cos\frac{\pi}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} & -\frac{1}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & \frac{\sqrt{3}}{2} \\ 1 & -\sqrt{3} & 0 \end{bmatrix}.$$

Then

$$X_1 = \frac{X + X^T}{2} = \begin{bmatrix} -\frac{\sqrt{3}}{2} + 2 & -\frac{1}{2} & -\frac{\sqrt{3}}{4} + 1 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} & -\frac{1}{4} \\ -\frac{\sqrt{3}}{4} + 1 & -\frac{1}{4} & \sqrt{3} - 1 \end{bmatrix}$$

and

$$Y_1 = \frac{Y + Y^T}{2} = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} & \frac{1}{4} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & -\frac{\sqrt{3}}{4} \\ \frac{1}{4} & -\frac{\sqrt{3}}{4} & 0 \end{bmatrix}.$$

Then, we know that $[X_1, Y_1] = X_1Y_1 - Y_1X_1 \neq 0$, that is, $\|[X_1, Y_1]\|^2 \neq 0$. Thus, $\langle R(X,Y)Y, X\rangle$ can be nonzero (negative) for any two matrices X and Y, of which none is skew-symmetric, if X and Y commute. $\square$

2.6 Curvature of subgroups

Let H be a closed subgroup of $G = \mathrm{GL}(n, \mathbf{R})$. Then H is a Lie subgroup of G and its Lie algebra $\mathfrak{h}$ is a sub-Lie algebra of $\mathfrak{g}$. The left invariant Riemannian metric on G induces a left invariant Riemannian metric on H by restricting to the tangent spaces of H. Then H becomes a sub-Riemannian manifold of G. For any $u, v \in \mathfrak{h}$, we may compute the sectional curvature $S(u, v)$ on G as defined by (2.5), and we may also compute the sectional curvature $S_H(u, v)$ on H. In general, they are different, but if H is a totally geodesic sub-manifold of G, then $S(u, v) = S_H(u, v)$. By definition, H is a totally geodesic sub-manifold of G if all the geodesics in G, starting in H and tangent to H, are contained in H and so are also geodesics in H. In this case, it is well known that $S(u,v) = S_H(u,v)$; see for example Theorem 12.2 in [7, Chapter I]. For exmaple, both the bigger blue circle and the smaller red circle in Figure 2.5 are totally geodesic sub-manifolds of the torus.

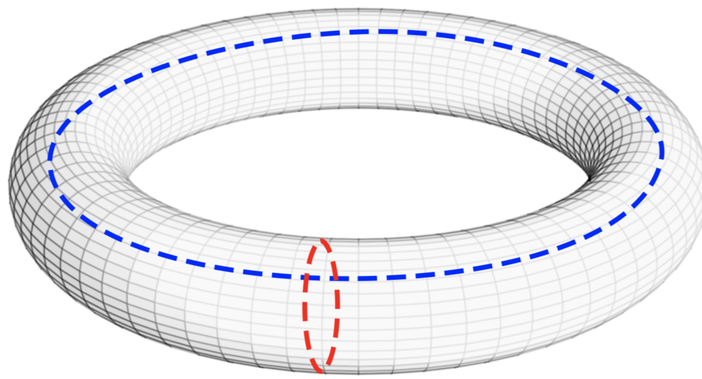

Figure 2.5: Totally geodesic sub-manifolds of the torus

By (3.9) in [14], the geodesic $\gamma(t)$ in $G = \mathrm{GL}(n, \mathbf{R})$ with $\gamma(0) = e$ and $\gamma'(0) = u \in \mathfrak{g}$ is given by

$$\gamma(t) = \exp(tu^\top)\exp(t(u - u^\top)).$$

Assume H is transpose-invariant, that is, for any $h \in H$, $h^\top \in H$. Then $\mathfrak{h}$ is also transpose-invariant. From the above geodesic expression, it is clear that any geodesic in G emitting from e and tangent to H is contained in H. Because the Riemannian metric is left invariant, this is true for the geodesic emitting from any point in H. It follows that H is a total geodesic sub-manifold of G. We have proved the following result.

Theorem 2.6.1. *Let H be a closed and transpose-invariant subgroup of $G = \mathrm{GL}(n, \mathbf{R})$, and let it be equipped with the left invariant Riemannian metric determined by the Frobenius inner product restricted to its Lie algebra $\mathfrak{h}$. Then Theorem 2.4.8 holds on H, that is, (2.13) holds for $u, v \in \mathfrak{h}$.*

2.7 Reductive Lie group

Let us recall the definition of reductive group [9, Chapter VII].

Definition 2.7.1. *The Harish-Chandra class $\mathcal{H}$ consists of 4-tuples (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 a nondegenerate, $\mathrm{Ad}(G)$-invariant, symmetric, bilinear form on $\mathfrak{g}$ such that*

1. *$\mathfrak{g}$ is reductive, that is, $\mathfrak{g} = \mathfrak{g}_1 + \mathfrak{z}$, where $\mathfrak{g}_1 = [\mathfrak{g}, \mathfrak{g}]$ and $\mathfrak{z}$ is the center of $\mathfrak{g}$.*

2. *$\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$ (called the Cartan decomposition), where $\mathfrak{k}$ is the $+1$-eigenspace and $\mathfrak{p}$ is the -1-eigenspace under θ.*

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 surjective diffeomorphism.*

5. *for every $g \in G$, the automorphism $\mathrm{Ad}(g)$ of $\mathfrak{g}$, extended to the complexification $\mathfrak{g}^{\mathbf{C}}$ of $\mathfrak{g}$ is contained in $\mathrm{Int}\mathfrak{g}^{\mathbf{C}}$.*

6. *the analytic subgroup G_1 of G with Lie algebra $\mathfrak{g}_1 = [\mathfrak{g}, \mathfrak{g}]$ has finite center.*

If $(G, K, \theta, B) \in \mathcal{H}$, then G is called a *reductive Lie group*.

The bilinear form $B(\cdot, \cdot) : \mathfrak{g} \times \mathfrak{g} \to \mathbf{R}$ induces an Euclidean inner product $B_\theta(\cdot, \cdot)$ on $\mathfrak{g}$ [9, p.448]:

$$\langle X, Y \rangle := B_\theta(X, Y) = -B(X, \theta Y).$$

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 B and thus under B_θ [9].

Example 2.7.2. *(1) $G = \mathrm{GL}(n, \mathbf{R})$ is reductive with $B(X, Y) = \mathrm{tr}\,(XY)$ and $\theta X = -X^\top$. Then $\langle X, Y \rangle = B_\theta(X, Y) = \mathrm{tr}\,(X^\top Y)$, where $X, Y \in \mathfrak{gl}(n, \mathbf{R})$.*

(2) $\mathrm{GL}(n, \mathbf{C})$ *is reductive with* $B(X,Y) = \mathrm{Re}\,[\mathrm{tr}\,(XY)]$*, where* Re *is the real part and* $\theta X = -X^*$*. Then* $\langle X, Y\rangle = B_\theta(X,Y) = \mathrm{Re}\,[\mathrm{tr}(X^*Y)]$*, where* $X, Y \in \mathfrak{gl}(n, \mathbf{C})$*.*

As in Section 3, the covariant derivative $\nabla_u v$ under a left invariant metric is [16, (5.3)]

$$\langle \nabla_u v, w\rangle = \frac{1}{2}(\langle [u,v], w\rangle - \langle [v,w], u\rangle - \langle [u,w], v\rangle). \tag{2.8}$$

In the rest of the paper, we will assume that G is equipped with the left invariant Riemannian metric determined by $\langle \cdot, \cdot\rangle = B_\theta(\cdot,\cdot)$ at $\mathfrak{g}$. Let $\|u\| = \langle u, u\rangle^{1/2}$ be the associated norm. It is easy to show

$$\langle [u,w], v\rangle = -\langle w, [\theta u, v]\rangle. \tag{2.9}$$

Proof.

$$\begin{aligned}\langle [u,w], v\rangle &= -B([u,w], \theta v) = -B(w, [\theta v, u]) = -B(w, \theta[v, \theta u]) \\ &= B_\theta(w, [v, \theta u]) = -\langle w, [\theta u, v]\rangle.\end{aligned}$$

□

By (2.8) and (2.9),

$$\langle \nabla_u v, w\rangle = \frac{1}{2}(\langle [u,v], w\rangle + \langle [\theta v, u], w\rangle + \langle [\theta u, v], w\rangle).$$

It follows that

$$\nabla_u v = \frac{1}{2}([u,v] - [u, \theta v] - [v, \theta u]). \tag{2.10}$$

It is easy to show that

$$[\mathfrak{k}, \mathfrak{k}] \subset \mathfrak{k}, \quad [\mathfrak{p}, \mathfrak{p}] \subset \mathfrak{k}, \quad [\mathfrak{k}, \mathfrak{p}] = [\mathfrak{p}, \mathfrak{k}] \subset \mathfrak{p}. \tag{2.11}$$

According to (2.10), we have

$$\nabla_u v = \begin{cases} \frac{1}{2}[u,v], & \text{if } u,v \in \mathfrak{p} \text{ or } u,v \in \mathfrak{k} \\ -\frac{1}{2}[u,v], & \text{if } u \in \mathfrak{p}, v \in \mathfrak{k} \\ \frac{3}{2}[u,v], & \text{if } u \in \mathfrak{k}, v \in \mathfrak{p}. \end{cases} \tag{2.12}$$

By (2.9), $\langle [w,u], v\rangle = -\langle u, [\theta w, v]\rangle$ and we have the following lemma.

Lemma 2.7.3. *Given $u, v, w \in \mathfrak{g}$, we have*

$$\langle [w,u], v\rangle = \begin{cases} \langle u, [w,v]\rangle, & \textit{if } w \in \mathfrak{p} \\ -\langle u, [w,v]\rangle, & \textit{if } w \in \mathfrak{k}. \end{cases}$$

The curvature tensor R and sectional curvature S are defined in the same ways as (2.3) and (2.5), respectively. We have the following result and we skip the proofs which are similar to those in Section 2.4.

Theorem 2.7.4. *Let G be a reductive Lie group. Let $u, v \in \mathfrak{g}$. Then*

$$\langle R(u,v)v, u\rangle = -2\left\|[u_1, v_1]\right\|^2 + \frac{1}{4}\left\|[u,v]\right\|^2 + 2\langle [u_1,v_1], [u_2,v_2]\rangle. \tag{2.13}$$

So

1. *Let $u, v \in \mathfrak{p}$. Then $\langle R(u,v)v,u\rangle = -\frac{7}{4}\left\|[u,v]\right\|^2 \leq 0$.*

2. *Let $u, v \in \mathfrak{k}$. Then $\langle R(u,v)v,u\rangle = \frac{1}{4}\left\|[u,v]\right\|^2 \geq 0$.*

3. *Let $u \in \mathfrak{p}$ and $v \in \mathfrak{k}$. Then $\langle R(u,v)v,u\rangle = \frac{1}{4}\left\|[u,v]\right\|^2 \geq 0$.*

4. *Let $u \in \mathfrak{g}$ and $v \in \mathfrak{k}$. Then $\langle R(u,v)v,u\rangle = \frac{1}{4}\left\|[u,v]\right\|^2 \geq 0$.*

5. Let $u \in \mathfrak{g}$ and $v \in \mathfrak{p}$. Then

$$\langle R(u,v)v, u\rangle = -\tfrac{7}{4}\left\|[u_1, v]\right\|^2 + \tfrac{1}{4}\left\|[u_2, v]\right\|^2.$$

Let H be a closed subgroup of $G = \mathrm{GL}(n, \mathbf{R})$ (or $G = \mathrm{GL}(n, \mathbf{C})$) that is invariant under (conjugate) transposition. It is known that [9, p.447] H is a reductive Lie group. By Theorem 2.7.4, we obtain an alternative proof of Theorem 2.6.1 and extend it to include complex matrix groups.

Corollary 2.7.5. *Let H be a closed subgroup of $\mathrm{GL}(n, \mathbf{R})$ or ($\mathrm{GL}(n, \mathbf{C})$) that is invariant under (conjugate) transposition, and let it be equipped with the left invariant Riemannian metric determined by the inner product in Example 2.7.2 restricted to its Lie algebra. Then Theorem 2.4.8 holds for H.*

Remark 2.7.6. Let G be a reductive Lie group with Lie algebra $\mathfrak{g}$. Let $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$ be a given Cartan decomposition corresponding to the Cartan involution θ. Let K be the analytic subgroup of G with Lie algebra $\mathfrak{k}$. Let $P = \exp \mathfrak{p}$. Note that $G = K \exp \mathfrak{p}$ and P is not a group, so Theorem 2.7.4 does not apply. When P is equipped with the symmetric space metric, it is a Riemannian manifold and the geodesic starting from $p \in P$ takes the form $p^{1/2}\exp(tu)p^{1/2}$, $u \in \mathfrak{p}$ [13]. It is related to the geometric means in the context of symmetric space of noncompact type and [13] evolves from the study of the matrix geometric means of two $n \times n$ positive definite matrices [1]. See [11, 12] for some recent interesting results and generalizations of matrix geometric means.

Bibliography

[1] R. Bhatia. *Positive Definite Matrices.* Princeton University Press, Princeton, NJ, 2007.

[2] I. Chavel. *Riemannian Geometry – A Modern Introduction. Vol. 108.* Cambridge University Press, Cambridge, 1993.

[3] L. Conlon. *Differentiable manifolds.* Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Boston, Inc., Boston, MA, second edition, 2001.

[4] S. Fiori. A Riemannian steepest descent approach over the inhomogeneous symplectic group: application to the averaging of linear optical systems. *Appl. Math. Comput.*, 283:251–264, 2016.

[5] D. Gabay. Minimizing a differentiable function over a differential manifold. *J. Optim. Theory Appl.*, 37(2):177–219, 1982.

[6] L. Gan, M. Liao, and T.-Y. Tam. Curvature of matrix and reductive lie groups. *J. Lie Theory*, 30(2):361–370, 2020.

[7] S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces. Vol. 34.* American Mathematical Society, Providence, RI, 2001. Corrected reprint of the 1978 original.

[8] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30(5):509–541, 1977.

[9] A. W. Knapp. *Lie groups beyond an introduction. Vol. 140.* Birkhäuser Boston, Inc., Boston, MA, second edition, 2002.

[10] S. Kobayashi and K. Nomizu. *Foundations of differential geometry. Vol. II.* Wiley Classics Library. John Wiley & Sons, Inc., New York, 1996. Reprint of the 1969 original, A Wiley-Interscience Publication.

[11] J. Lawson and Y. Lim. Weighted means and Karcher equations of positive operators. *Proc. Natl. Acad. Sci. USA*, 110(39):15626–15632, 2013.

[12] J. Lawson and Y. Lim. Karcher means and Karcher equations of positive definite operators. *Trans. Amer. Math. Soc. Ser. B*, 1:1–22, 2014.

[13] M. Liao, X. Liu, and T.-Y. Tam. A geometric mean for symmetric spaces of noncompact type. *J. Lie Theory*, 24(3):725–736, 2014.

[14] R. J. Martin and P. Neff. Minimal geodesics on $\mathrm{GL}(n)$ for left-invariant, right-$\mathrm{O}(n)$-invariant Riemannian metrics. *J. Geom. Mech.*, 8(3):323–357, 2016.

[15] P. A. Mello. Averages on the unitary group and applications to the problem of disordered conductors. *J. Phys. A*, 23(18):4061–4080, 1990.

[16] J. Milnor. Curvatures of left invariant metrics on Lie groups. *Advances in Math.*, 21(3):293–329, 1976.

[17] M. Moakher. Means and averaging in the group of rotations. *SIAM J. Matrix Anal. Appl.*, 24(1):1–16, 2002.

[18] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.*, 26(3):735–747, 2005.

[19] J. Wang, H. Sun, and S. Fiori. A Riemannian-steepest-descent approach for optimization on the real symplectic group. *Math. Methods Appl. Sci.*, 41(11):4273–4286, 2018.

[20] J. Wang, H. Sun, and D. Li. A geodesic-based Riemannian gradient approach to averaging on the Lorentz group. *Entropy*, 19(12):Paper No. 698, 13, 2017.