

FLORIDA STATE UNIVERSITY
COLLEGE OF ARTS AND SCIENCES

SMOOTHLY EVOLVING GEODESICS IN THE SPECIAL ORTHOGONAL GROUP:
DEFINITIONS, COMPUTATIONS AND APPLICATIONS

By

ZHIFENG DENG

A Dissertation submitted to the
Department of Mathematics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2024

Copyright © 2024 Zhifeng Deng. All Rights Reserved.

Zhifeng Deng defended this dissertation on November 13, 2024.

The members of the supervisory committee were:

Kyle A. Gallivan
Professor Co-Directing Dissertation

Wen Huang
Professor Co-Directing Dissertation

Pierre-Antoine Absil
Committee Member

Philip Bowers
Committee Member

Martin Bauer
Committee Member

Giray Ökten
Committee Member

Gordon Erlebacher
University Representative

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements.

To my dearest friend Franz and my parents, thank you for your constant support and encouragement. To the person who encouraged me to pursue an academic career, thank you.

ACKNOWLEDGMENTS

I extend my sincere thanks to my advisor, co-advisors, committee members, parents, friends, and the person who gave me the courage to pursue an academic career many years ago.

TABLE OF CONTENTS

List of Tables	viii
List of Figures	ix
Abstract	x
1 Introduction	1
1.1 Smoothly Evolving Geodesics on a Manifold	2
1.2 Basic Notions	3
1.2.1 Charts and Atlases	3
1.2.2 Tangent Space	4
1.2.3 Riemannian Metric, Geodesic and Exponential	5
1.2.4 Differentiation of Functions between Manifolds	7
1.2.5 Curve Length, Cut Locus and Conjugate Locus	7
1.2.6 Karcher Mean on a Manifold	9
1.3 Related Work	10
1.4 Research Overview and Dissertation Statement	12
2 Differentiating Matrix Exponential at Skew Symmetric Matrices	15
2.1 Introduction	15
2.2 Preliminaries	15
2.2.1 Matrix Logarithm	15
2.2.2 Real Schur Decompositions	17
2.2.3 Related Work	21
2.3 Problem Statement	22
2.4 Differential Formula	24
2.5 Inverse of the Differential Formula	27
2.6 Pseudoinverse of the Differential Formula	31
2.7 Routines and Implementations	34
2.7.1 Three-Stage Evaluation	35
2.7.2 Refinements on Existing Formulae	36
2.7.3 Pseudo Codes	37
2.8 Complexity Analyses and Numerical Results	37

2.8.1	Complexity	37
2.8.2	Experiments	39
3	Local Diffeomorphism in Skew Symmetric Matrices	41
3.1	Skew Symmetric Matrices with an Invertible Differential	41
3.2	Preimage of Exponential at Special Orthogonal Matrices	43
3.3	Diffeomorphism Structure in Skew Symmetric Matrices	50
3.3.1	Sufficient Condition of Constructing Diffeomorphism	51
3.3.2	Diffeomorphism on an Inscribed Ball	51
3.4	Nearby Logarithm	54
3.4.1	Comparison with the Original Definition	54
3.4.2	Algorithms	55
3.4.3	Visualizing Geodesics with Skew Symmetric Matrices	58
4	Smoothly Evolving Geodesic Problem on the Special Orthogonal Group	60
4.1	Introduction	60
4.2	Preliminaries	60
4.3	Problem Formulation	61
4.4	Solution Characterized by the Nearby Matrix Logarithm	62
4.5	Smoothly Evolving Geodesics of Endpoints Varying along Geodesic	65
4.5.1	Vector Fields and Geodesics	65
4.5.2	Co-Manifold Characterization	67
5	Velocity-Based Karcher Mean on the Special Orthogonal Group	71
5.1	Introduction	71
5.2	Example on a Circle	73
5.2.1	Objective Function with the Riemannian Distance	73
5.2.2	Objective Function with the Smoothly Evolving Arc Length	76
5.3	Velocity-Based Karcher Mean	77
5.3.1	Length-Based Karcher Mean Objective	77
5.3.2	Velocity-Based Karcher Mean on a Riemannian Manifold	79
5.3.3	Velocity-Based Karcher Mean on the Special Orthogonal Group	82
5.4	Numerical Experiments	84

6	The Endpoint Stiefel Geodesic Problem with the Canonical Metric	88
6.1	Introduction	88
6.2	Preliminaries	89
6.2.1	Riemannian Submersion	89
6.2.2	Stiefel Manifold with the Canonical Metric	91
6.2.3	Related Work	92
6.3	Problem Formulation	94
6.3.1	Matrix Equation	94
6.3.2	Preprocessing for Rank Reduction	95
6.3.3	Manifold Root-Finding Formulation	97
6.4	R-Newton Method of Solving a System on Manifold	98
6.4.1	Newton Direction	99
6.4.2	Algorithm	100
6.5	Numerical Experiments	101
7	Quotient Structure on the Fixed Rank Positive Semi-Definite Manifold	104
7.1	Introduction	104
7.2	Preliminaries	105
7.2.1	Geometric Interpretations in the FRPSD manifold	105
7.2.2	Objects in the Submersion	106
7.3	Riemannian Metric by Riemannian Submersion	107
7.3.1	Horizontal Lifting	107
7.3.2	Riemannian Metric Family Constructed by Designated Basis	108
7.3.3	Riemannian Geodesic with Motions in Subspaces and Ellipsoids	111
8	Conclusion and Future Research	116
	Bibliography	119
	Biographical Sketch	123

LIST OF TABLES

2.1	The feasible conditions of the root S in computing $\Delta_S \leftrightarrow \Delta_Q$	35
2.2	Complexity of the Directional Derivative of the Matrix Exponential	39
2.3	Complexity of the Directional Derivative of the Nearby Matrix Logarithm	39

LIST OF FIGURES

2.1	Illustration of $\sin(z)/z$ and $(\cos(z) - 1)/z$	28
2.2	Computation Time of $\Delta_S \mapsto \Delta_Q$ in $D \exp_S(\Delta) = Q\Delta_Q$	39
2.3	Computation Time of $\Delta_Q \mapsto \Delta_S$ in $D \exp_S(\Delta) = Q\Delta_Q$	40
3.1	Illustrations of the conjugate locus and the principal branch in Skew ₄	43
3.2	Illustration of Geodesics with Skew Symmetric Matrices.	59
5.1	Illustration of the Karcher Mean on an Unit Circle	74
5.2	Illustration of the Karcher Mean with Smoothly Evolving Arcs on an Unit Circle . . .	76
5.3	Karcher Mean with Evenly Spreading Spreading Data.	86
5.4	Karcher Mean with Randomly Spreading Spreading Data.	87
6.1	Performances on Solving the Stiefel Endpoint Geodesic Problem on St _{20,10}	103
7.1	Illustration of deformation $Q(t)$ and rotation $\Lambda(t)$ in ellipse.	105

ABSTRACT

This dissertation solves the problem of characterizing and computing a set of smoothly evolving geodesics emanating from the identity matrix that arrive at a smoothly varying endpoint on the special orthogonal group, namely the smoothly evolving geodesic problem. Since a set of smoothly evolving geodesics emanating from the identity matrix is equivalent to a set of smooth varying initial velocities in the tangent space at the identity matrix, the smooth evolving geodesic problem seeks a smooth curve of initial velocities $S(t)$ at the identity matrix that arrives at the given smooth curve $Q(t)$ on the special orthogonal group, such that $\exp(S(t)) = Q(t)$. Although the well-known matrix principal logarithm can find $\log(Q(t))$ that maps to Q under the matrix exponential, the resulting set of $\log(Q(t))$ is not always continuous with respect to the smoothly varying $Q(t)$. The smoothly evolving geodesic problem is solved by identifying the conjugate locus on the special orthogonal group and investigating features in the differential operator of the matrix exponential restricted to the set of skew symmetric matrices. Efficient and robust algorithms are further designed to compute such a smooth varying $S(t)$.

The smoothly evolving geodesic problem is motivated by an issue in the Karcher mean problem on a manifold that has a discontinuity in finding a shortest geodesic. This leads to the non-smooth objective function in the Karcher mean formulation with possibly multiple local minima and leads to the Karcher mean not being smoothly dependent on the given data set. Based on the computation of feasible smoothly evolving geodesics developed in the special orthogonal group, a novel Karcher mean generalization is proposed. The generalized Karcher mean considers the initial velocities emanating from the data points and splits the non-smooth objective function of the classic formulation into multiple smooth objective functions, which leads to a smooth optimization problem of the generalized formulation.

Futhermore, the smoothly evolving geodesic is applied to the quotient structures in the special orthogonal group. These quotient structures arise from the Riemannian submersion that defines a Riemannian structure on manifolds with the special orthogonal constraint. A root-finding formulation is then proposed to solve the endpoint geodesic problem in these manifolds. The endpoint geodesic problem, that seeks any geodesic between the given points, is a weaker form of the smoothly evolving geodesic problem. In the Stiefel manifold with the canonical metric, the proposed algorithm obtains more robust and efficient performance than the state of the art algorithm, especially

when the endpoints are well-separated. The speed up compared to the state of the art algorithm reaches to 10 in some cases. In the fixed rank positive semi-definite (FRPSD) matrix manifold, the geometric insights developed on the special orthogonal group is utilized to propose a new Riemannian metric on the FRPSD manifold which facilitates the development of novel and meaningful Riemannian geodesic interpretations. Although some of the basic notion remains open questions for this new metric, interesting features and propositions are discussed in this dissertation.

CHAPTER 1

INTRODUCTION

This dissertation investigates the inverse action of the Riemannian exponential on the special orthogonal group and identifies a set of smoothly evolving geodesics that arrive at a varying endpoint. The notion of smoothly evolving geodesics is crucial to many manifold applications as it maintains the differentiability of the objects at the varying endpoint, even when the endpoint goes beyond the cut locus of the emanating point. In particular, a smooth representation of the special orthogonal group in terms of rotating parameters is needed in signal processing [30], computer vision [10], and neural network [40]. In general, it is very hard to compute or identify such a set of geodesics on an arbitrary manifold, especially when the geodesic is non-minimal. One major contribution in this dissertation is to solve and compute such a set of geodesics in the special orthogonal group.

This dissertation also considers the geometric mean problem on a manifold as an important application of the smoothly evolving geodesics obtained in the special orthogonal group. The Karcher mean has been successfully applied in various researches, e.g., [5, 22, 10, 15]. However, the classic Karcher mean formulation in a Riemannian manifold may suffer from the non-differentiable distance function as pointed out in [20]. A novel generalization of the Karcher mean formulation is designed to address the non-smooth objective function issue in the classic Karcher mean formulation on a manifold. The solution to the new Karcher mean formulation is found and computed using the smoothly evolving geodesics in the special orthogonal group. Furthermore, the solution is shown to have smooth dependence with respect to the given data set. The tools are also used to solve geodesic problems on the Stiefel manifold and the set of fixed rank positive semi-definite (FRPSD) matrices. This work aims at solving the open question of the smoothly evolving geodesics on other important manifolds in the future. Furthermore, the technique and insights of the generalized Karcher mean on the special orthogonal group can be applied to the Stiefel manifold and the FRPSD manifold with corresponding smoothly evolving geodesics.

This dissertation is organized as follows. In **Chapter 1**, some basic notions in a Riemannian manifold are reviewed and they are followed by a brief history of the relevant topics and research. The chapter ends with a dissertation statement. Then, the following two chapters investigate two fundamental primitives to the smoothly evolving geodesics problem on the special orthogonal

group, namely the differential to matrix exponential studied in **Chapter 2** and the diffeomorphism studied in **Chapter 3**. With the necessary primitives developed, **Chapter 4** gives the more specific formulation of the smoothly evolving geodesic problem and presents a solution to it. **Chapter 5** through **7** apply the results to various applications. **Chapter 8** summarizes the major contributions of this dissertation.

1.1 Smoothly Evolving Geodesics on a Manifold

The Riemannian geodesic is one of the most fundamental objects in a manifold \mathcal{M} which is generalized from the notion of straight lines in an Euclidean setting. For a given fixed point $x \in \mathcal{M}$ in a manifold, the Riemannian geodesic $\gamma : [0, 1] \rightarrow \mathcal{M}$ emanating from x is characterized by the Riemannian exponential, which maps the tangent space $T_x\mathcal{M}$ of \mathcal{M} at x to the manifold itself

$$\begin{aligned} \text{Exp}_x : T_x\mathcal{M} &\rightarrow \mathcal{M} \\ v := \dot{\gamma}(0) &\mapsto \text{Exp}_x(v) := \gamma(1) \end{aligned} \tag{1.1}$$

where v is the velocity of γ at $t = 0$ and $\text{Exp}_x(v) = \gamma(1)$ is the endpoint at which γ arrives at $t = 1$. The entire geodesic is given by $\gamma(t) = \text{Exp}_x(t \cdot v), \forall t \in [0, 1]$. The problem of smoothly evolving geodesics on a manifold seeks a well-defined inverse $(\text{Exp}_x)^{-1}$ to the Riemannian exponential, such that for some smoothly evolving point $y(s) : [0, 1] \rightarrow \mathcal{M}$ in the manifold with the given initial $v(0) = (\text{Exp}_x)^{-1}(y(0))$, there is

$$\begin{aligned} (\text{Exp}_x)^{-1} : \mathcal{M} &\rightarrow T_x\mathcal{M} \\ y(s) &\mapsto v(s) \end{aligned} \tag{1.2}$$

where $\{v(s), s \in [0, 1]\}$ is smooth in $T_x\mathcal{M}$ that satisfies $\text{Exp}_x(v(s)) = y(s), \forall s \in [0, 1]$.

In the case where the differential of the Riemannian exponential at $v \in T_x\mathcal{M}$ is invertible at $\text{Exp}_x(v) = y$, the implicit function theorem on manifolds leads to the existence of two sufficiently small neighborhoods $\mathcal{U}_v \subset T_x\mathcal{M}$ and $\mathcal{U}_y \subset \mathcal{M}$ around $v \in T_x\mathcal{M}$ and $y \in \mathcal{M}$ respectively, on which $\text{Exp}_x : \mathcal{U}_v \rightarrow \mathcal{U}_y$ is an invertible smooth bijection. In other words, it concludes that a unique smooth $v(s), s \in [0, 1]$ solution to the smoothly evolving geodesics problem exists for any smooth $\{y(s), s \in [0, 1]\} \subset \mathcal{U}_y$ in the neighborhoods. However, this is usually not enough in practice. On one hand, the neighborhoods \mathcal{U}_v and \mathcal{U}_y may not be available explicitly. On the other hand, they may be too small for the varying $y(s)$. These considerations are addressed in the formulation of $y(s)$ and $v(s)$ as curves in (1.2). Rather than identifying an invertible smooth bijection $\text{Exp}_x : \mathcal{U}_v \rightarrow \mathcal{U}_y$ on open neighborhoods, (1.2) takes a weaker form that only requires an invertible bijection on curves.

Solving the smoothly evolving geodesic problem (1.2) and developing the respective computational routines yields the following beneficial results:

1. A smooth parameterization, centered at y , of the manifold \mathcal{M} that is realized in the tangent space $T_x\mathcal{M}$ where $x \neq y$. This is particularly useful for the special orthogonal group when x is taken to be the identity matrix.
2. Realizing (part of) the submanifold structure in \mathcal{M} including $y \in \mathcal{M}$ with respect to $x \in \mathcal{M}$ by restricting $y(s)$ in the submanifold. The submanifold identified in $T_x\mathcal{M}$ benefits from the ambient Euclidean setting in $T_x\mathcal{M}$.
3. The line search procedure along $y(s)$ in solving an optimization problem on \mathcal{M} can be converted to a line search along $v(s)$ in $T_x\mathcal{M}$.

1.2 Basic Notions

This section reviews some important results and propositions on a manifold that are closely related to this dissertation. They can be found in the classic textbooks on manifolds, such as [4] and [21].

1.2.1 Charts and Atlases

A local chart on an open subset \mathcal{U} of a d -dimensional manifold \mathcal{M} is an invertible map $\psi : \mathcal{U} \rightarrow \mathbb{R}^d$. The invertible map builds a one-to-one identification between the points in manifold and the points on the \mathbb{R}^d . An atlas on this manifold \mathcal{M} is a collection of charts $\{\psi_\alpha : \mathcal{U}_\alpha \rightarrow \mathbb{R}^d\}_\alpha$ that cover the \mathcal{M} as

$$\bigcup_{\alpha} \mathcal{U}_\alpha = \mathcal{M}.$$

The image $\psi_\alpha(x)$ of a point $x \in \mathcal{U}_\alpha \subset \mathcal{M}$ in \mathbb{R}^d is referred to as the coordinate of x under ψ_α .

The smoothness of a manifold is determined by the smoothness of the coordinate change map $\psi_\beta \circ \psi_\alpha^{-1}$ in $\mathcal{U}_\alpha \cap \mathcal{U}_\beta \neq \emptyset$, which converts the coordinates of the same point between different charts. For any curve $\tau(t), t \in [0, 1]$ sitting in $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$, it has two different coordinate forms under different charts as $\tau_\alpha(t) = (\psi_\alpha \circ \tau)(t)$ and $\tau_\beta(t) = (\psi_\beta \circ \tau)(t)$. These two curves represent the same τ and the smoothness in the conversion between them yields the smoothness of the τ they represent in \mathcal{M} . If the analysis on any curve τ can be carried from τ_α to τ_β up to p -times differentiations, then the manifold \mathcal{M} is said to be \mathcal{C}^p . The \mathcal{C}^∞ denotes the smooth manifold in which all analysis can be carried from one coordinate form to another under up to infinitely many differentiations.

The manifolds discussed in this dissertation are all C^∞ . Furthermore, the notion of a smooth object at a point $x \in \mathcal{M}$ means that the representation of the object composed with any chart containing x is smooth. For example, a function $f : \mathcal{M} \rightarrow \mathbb{R}$ is smooth at x if for any chart $\psi : \mathcal{U} \rightarrow \mathbb{R}^d$ containing x , there is a sufficiently small open neighborhood $\mathcal{U}_\delta \subset \mathcal{U}$ containing x , such that $f \circ \psi^{-1} : \{\psi(y) : y \in \mathcal{U}_\delta\} \rightarrow \mathbb{R}, \psi(y) \mapsto f(y)$ is smooth.

1.2.2 Tangent Space

In order to generalize the notion of straight lines to a manifold, one must define the velocity of curves in a manifold first. Consider a smooth curve $\{\tau(t) : t \in [-1, 1]\} \subset \mathbb{R}^n$ in the Euclidean space that emanates from $\tau(0) := x \in \mathcal{M}$, its velocity at $t = 0$ is given by

$$v_x := \left. \frac{d}{dt} \tau(t) \right|_{t=0} = \lim_{h \rightarrow 0} \frac{\tau(h) - \tau(0)}{h}.$$

In a manifold setting with $\{\tau(t) : t \in [-1, 1]\} \subset \mathcal{M}$, the subtraction that quantities difference between points $\tau(h)$ and $\tau(0)$ is no longer available in general. Since the velocity of $\tau(t)$ at $t = 0$ characterizes the infinitesimal motion of the curve at $t = 0$, one may consider expressing such an infinitesimal motion on a curve through the infinitesimal action it induces as an alternative. For arbitrary smooth function $f : \mathcal{M} \rightarrow \mathbb{R}$, the infinitesimal action at $t = 0$ the curve $\tau(t)$ induces is quantified as

$$\left. \frac{d}{dt} f(\tau(t)) \right|_{t=0} = \lim_{h \rightarrow 0} \frac{f(\tau(h)) - f(\tau(0))}{h}.$$

Following from this idea, the formal definition of the velocity of $\tau(t)$ at $x = \tau(0)$ is given by the collection of infinitesimal actions $\tau(t)$ acting on arbitrary function $f \in \mathcal{F}_x(\mathcal{M})$ as

$$\begin{aligned} \dot{\tau}(0) : \mathcal{F}_x(\mathcal{M}) &\rightarrow \mathbb{R} \\ f &\mapsto \dot{\tau}(0)f := \dot{\tau}(0)(f) \end{aligned} \tag{1.3}$$

where $\mathcal{F}_x(\mathcal{M})$ collects all function $f : \mathcal{M} \rightarrow \mathbb{R}$ that is smooth around x . In addition, when there are multiple smooth curves that obtain the same velocity, e.g., the $\dot{\gamma}(0) = \dot{\tau}(0)$ at $x = \gamma(0) = \tau(0)$, such a velocity is usually referred to as a tangent vector $v_x : \mathcal{F}_x(\mathcal{M}) \rightarrow \mathbb{R}$ at x that is independent with respect to the actual curves. The formal definition of tangent vectors follows.

Definition 1.2.1. A tangent vector v_x to a manifold \mathcal{M} at a point x is a mapping from $\mathcal{F}_x(\mathcal{M})$ to \mathbb{R} that is the velocity $\dot{\tau}(0)$ of some curve $\{\tau(t) : t \in [-1, 1]\}$ satisfying the following conditions

$$\begin{cases} \tau(0) = x \\ v_x(f) = \dot{\tau}(0)f, \forall f \in \mathcal{F}_x(\mathcal{M}) \end{cases} \tag{1.4}$$

Such a curve $\tau(t)$ is said to realize the tangent vector v_x . □

Note that a tangent vector can have different realizations of curve. In some literature, the notation for tangent vectors and curve velocities refers to the same notion. In this dissertation, the term “velocity” and the notation $\dot{\tau}(0)$ are used to emphasize some given realized curve $\tau(t)$ in the context while the term “tangent vector” and the notation v_x are used to emphasize the mapping nature in $\mathcal{F}_x(\mathcal{M}) \rightarrow \mathbb{R}$.

Finally, the collection of all tangent vectors v_x at $x \in \mathcal{M}$ forms the tangent space at x denoted as $T_x\mathcal{M}$. It collect all possible infinitesimal actions at x and has the same dimension with the manifold itself. More importantly, such a tangent space is a linear space that enjoys an ambient Euclidean setting.

1.2.3 Riemannian Metric, Geodesic and Exponential

A Riemannian structure of a manifold \mathcal{M} is built on an inner product operator g_x that maps from $T_x\mathcal{M} \times T_x\mathcal{M}$ to \mathbb{R} . The collection of all of these operators, smooth w.r.t. x , is denoted as the Riemannian metric g on \mathcal{M} . The inner product of $v_x, w_x \in T_x\mathcal{M}$ is given by

$$\langle v_x, w_x \rangle = g_x(v, w).$$

The norm $\sqrt{\langle v_x, v_x \rangle}$ induced by the inner product in $T_x\mathcal{M}$ is denoted as the g -norm, $\|v_x\|_g$.

Such a Riemannian metric fully characterizes a Riemannian manifold denoted as (\mathcal{M}, g) . It introduces the notions of distance, shortest curve and straight line to the manifold \mathcal{M} by the special curve known as the Riemannian geodesics. Recall that the length of a curve in an Euclidean setting is given by the integral of its velocity norm along the path. This idea applies to the Riemannian setting with the g -norm as

$$l_\tau := \int_0^1 \|\dot{\tau}(t)\|_g dt = \int_0^1 \sqrt{g_{\tau(t)}(\dot{\tau}(t), \dot{\tau}(t))} dt. \quad (1.5)$$

Taking the infimum among the lengths of all smooth curves connecting points $x, y \in \mathcal{M}$ yields the distance $d(x, y)$. Fortunately, the infimum can be obtained, i.e., a shortest curves connects $x, y \in \mathcal{M}$ with its length equals to the distance $d(x, y)$, if x and y are connected. Such a curve is denoted as a Riemannian geodesic.

In an Euclidean setting, the shortest curve coincides with the straight line parameterized as a curve $\{\gamma(t), t \in [0, 1]\} \in \mathbb{R}^n$ in constant velocity, i.e., $\dot{\gamma}(t) = v \in \mathbb{R}^n, \forall t \in [0, 1]$. Such a curve in an Euclidean space takes the unique form $\gamma(t) = \gamma(0) + t \cdot v$. In a manifold setting, one must define the notion of constant speed before defining a straight line. However, such a notion is not unique and

has to be path-dependent. In other words, depending on different curves $\tau(t)$ and $\gamma(t)$ connecting $x = \tau(0) = \gamma(0)$ and $y = \tau(1) = \gamma(1)$, a tangent vector $v_x \in T_x\mathcal{M}$ is considered “constant” with two different sets of tangent vectors $v_{\gamma,t} \in T_{\gamma(t)}\mathcal{M}$ and $v_{\tau,t} \in T_{\tau(t)}\mathcal{M}$ along the curves satisfying

$$\begin{cases} g_{\tau(t)}(\dot{\tau}(t), v_{\tau,t}) = g_x(\dot{\tau}(0), v_x) \\ g_{\gamma(t)}(\dot{\gamma}(t), v_{\gamma,t}) = g_x(\dot{\gamma}(0), v_x) \end{cases}, \forall t \in [0, 1] \quad (1.6)$$

where in general, $v_{\gamma,1} \neq v_{\tau,1} \in T_y\mathcal{M}$ if $\tau \neq \gamma$. The path- γ -dependent mapping between $v_x \in T_x\mathcal{M}$ and $v_{\gamma,t} \in T_{\gamma(t)}\mathcal{M}$ is found to be a linear operator $\mathcal{P}_{\gamma,0 \rightarrow t} : T_x\mathcal{M} \rightarrow T_{\gamma(t)}\mathcal{M}$ denoted as the parallel translation along γ . A curve $\gamma(t)$, $t \in [0, 1]$ is said to be a zero-accelerated curve or an affine geodesic to the parallel translation if it satisfies

$$\mathcal{P}_{\gamma,0 \rightarrow t}(\dot{\gamma}(0)) = \dot{\gamma}(t), \forall t \in [0, 1],$$

i.e., its velocities remains constant along itself.

Although there are infinitely many ways to define a parallel translation on \mathcal{M} , the fundamental theorem of Riemannian manifold picks out the unique parallel translation inducing an Riemannian geodesic as an affine geodesic, satisfying (1.6). This dissertation focuses on such a parallel translation with Riemannian geodesic but the study and the discussion presented in this work can certainly be generalized to arbitrary parallel translation in future work. Unless otherwise specified, the term “geodesic” in the rest of this dissertation refers to both the Riemannian geodesic in the context of “curve length” and the affine geodesic in the context of “constant speed” and “zero acceleration”.

Benefits from the notion of a zero-accelerated curve, a geodesic γ in \mathcal{M} that emanates from $\gamma(0) = x$ and arrives at $\gamma(1) = y$ can be uniquely characterized by its initial velocity $\dot{\gamma}(0)$, as one can integrate the ordinary differential equation $\mathcal{P}_{\gamma,0 \rightarrow t}(\dot{\gamma}(0)) = \dot{\gamma}(t)$ to find the unique solution $\gamma(t)$. The collection of such a map from the tangent vector $v = \dot{\gamma}(0) \in T_x\mathcal{M}$ to the arriving point $y = \gamma(1) \in \mathcal{M}$, forms a smooth mapping denoted as the Riemannian exponential

$$\begin{aligned} \text{Exp}_x : T_x\mathcal{M} &\rightarrow \mathcal{M} \\ v &\mapsto y \\ \text{s.t. } \begin{cases} v = \dot{\gamma}(0) \\ y = \gamma(1) \end{cases} &\text{ for geodesic } \gamma. \end{aligned}$$

In addition, this dissertation only considers the complete Riemannian manifold scenario stated in the Rinow-Hopf theorem, where the Riemannian exponential $\text{Exp}_x : T_x\mathcal{M} \rightarrow \mathcal{M}$ is well-defined and

smooth on the entire tangent space $T_x\mathcal{M}$. In other words, the geodesic $\gamma(t) = \text{Exp}_x(t \cdot v), t \in [0, 1]$ for any $v \in T_x\mathcal{M}$ can be extended to $\gamma(t) = \text{Exp}_x(t \cdot v), t \in [0, \infty)$. Note that the extended geodesic stays a zero accelerated curve but it is not necessary the shortest curve between the endpoints anymore. The lost shortest constraint is characterized by the notion of cut locus discussed later.

1.2.4 Differentiation of Functions between Manifolds

The idea of differentiating a smooth function $f : \mathcal{M} \rightarrow \mathbb{R}$ has been mentioned in the definition of tangent vectors on a manifold. Consider a smooth function $\varphi : \mathcal{N} \rightarrow \mathcal{M}$ between the manifolds \mathcal{N} and \mathcal{M} and let $\{x(t), t \in [0, 1]\} \subset \mathcal{N}$ be a smooth curve with $v_x = \dot{x}(0) \in T_x\mathcal{N}$ at $x = x(0)$. Let $y(t) := \varphi(x(t)), t \in [0, 1]$ be the image in \mathcal{M} , which is also a smooth curve with velocity $v_y := \dot{y}(0) \in T_y\mathcal{M}$ at $y = y(0)$. The tangent vector v_y is then the infinitesimal action of φ along v_x . The relationship between v_y and v_x forms the linear operator

$$\begin{aligned} D\varphi_x : T_x\mathcal{N} &\rightarrow T_y\mathcal{M} \\ \dot{x}(0) &\mapsto \dot{y}(0) := D\varphi_x[\dot{x}(0)] \end{aligned} \tag{1.7}$$

where $x(t), y(t) = \varphi(x(t)), \forall t \in [0, 1]$ are smooth curves on respective manifolds. This linear operator is denoted as the differential or the directional derivative of y and the notation $D\varphi_x[v]$ reads as “the differential of the function φ at the point x along the vector v ”.

Consider the Riemannian exponential $\text{Exp}_x : \mathcal{N} = T_x\mathcal{M} \rightarrow \mathcal{M}$ evaluated at v . Let $v(t)$ be a smooth curve emanating from $v = v(0) \in T_x\mathcal{M}$ with velocity $w \in T_v(T_x\mathcal{M}) = T_x\mathcal{M}$, e.g., the simple $v(t) = v + t \cdot w$. Let $y(t) = \text{Exp}_x(v(t))$ be the image under the Riemannian exponential, then the differential computes the infinitesimal motion $\dot{y}(t)$ at $t = 0$ as

$$D(\text{Exp}_x)_v[w] = \left. \frac{d}{dt} \text{Exp}_x(v(t)) \right|_{t=0} = \left. \frac{d}{dt} y(t) \right|_{t=0} \in T_y\mathcal{M}.$$

Note that this differential characterizes the perturbation to a set of geodesics parameterized by t as $\{\gamma_s(t) := \text{Exp}_x(s \cdot v(t)), s \in [0, 1]\}_t$ under the perturbation to the initial velocity $v(t)$ at $t = 0$. It is one of the main topics studied in this dissertation. Also note that this differential $D(\text{Exp}_x)_v : T_x\mathcal{M} \rightarrow T_y\mathcal{M}$ is a linear operator between two linear spaces with the same dimension d . If there are preferred bases on respective tangent spaces, the $D(\text{Exp}_x)_v$ can be expressed as a $d \times d$ matrix.

1.2.5 Curve Length, Cut Locus and Conjugate Locus

Inner product invariance is the essential property in the parallel translation that induces a Riemannian geodesic as an affine geodesic. It means that for any smooth curve $\{\tau(t), t \in [0, 1]\} \subset$

\mathcal{M} and any tangent vector $v_{\tau,0} \in T_{\tau(0)}\mathcal{M}$ with the parallel translated $v_{\tau,t} := \mathcal{P}_{\tau,0 \rightarrow t}(v_{\tau,0})$, the inner product given by the Riemannian metric g stays the same:

$$g_{\tau(0)}(v_{\tau,0}, \dot{\tau}(0)) = g_{\tau(t)}(v_{\tau,t}, \dot{\tau}(t)), \forall t \in [0, 1].$$

Recall that an affine geodesic $\gamma(t)$ has its velocity at any $\gamma(t)$ parallel translated from its initial velocity, i.e., $\dot{\gamma}(t) = \mathcal{P}_{\gamma,0 \rightarrow t}(\dot{\gamma}(0))$. It follows that the curve length of a Riemannian geodesic $\gamma(t), t \in [0, 1]$ is the norm of its velocity $\|\dot{\gamma}(s)\|_g = \sqrt{g_{\gamma(s)}(\dot{\gamma}(s), \dot{\gamma}(s))}$ at any $s \in [0, 1]$

$$\begin{aligned} l_\gamma &= \int_0^1 \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))} dt = \int_0^1 \sqrt{g_{\gamma(t)}(\mathcal{P}_{\gamma,0 \rightarrow t}(\dot{\gamma}(0)), \dot{\gamma}(t))} dt \\ &= \int_0^1 \sqrt{g_{\gamma(0)}(\dot{\gamma}(0), \dot{\gamma}(0))} dt = \sqrt{g_{\gamma(0)}(\dot{\gamma}(0), \dot{\gamma}(0))} = \sqrt{g_{\gamma(s)}(\dot{\gamma}(s), \dot{\gamma}(s))}, \forall s \in [0, 1] \end{aligned}$$

It further yields that the extended Riemannian geodesic $\gamma(t), t \in [0, T]$ has its length scaled linearly as $T \cdot \|\dot{\gamma}(t)\|_g$. For a complete Riemannian manifold (\mathcal{M}, g) , the Riemannian exponential can be extended infinitely, indicating the length of a Riemannian geodesic can be extended to infinity as well.

Such an infinitely extending length cannot always be the shortest length between two points on a bounded manifold. In other words, the statement of a Riemannian geodesic being the shortest curve must fail at some point. For example, the arcs of a great circle on a 2-sphere are geodesics and the extending arc fails to be shortest if it passes the opposite polar point. After passing the polar point, the shortest geodesic becomes the other arc that has not passed the polar point. This observation about the loss of shortest condition on an extending geodesic holds in general and the definition of the cut locus follows as the envelope of initial velocities in $T_x\mathcal{M}$ that emanates a shortest geodesic from $x \in \mathcal{M}$.

$$\text{Cut}_x := \{v \in T_x\mathcal{M} : \forall \sigma \in [0, 1], \{\text{Exp}_x(t\sigma v), t \in [0, 1]\} \text{ is the unique shortest geodesic}\}. \quad (1.8)$$

The $v \in \text{Cut}_x$ is denoted as a cut vector of x and the endpoint $y = \text{Exp}_x(v)$ it arrives is denoted as a cut point of x . In the arc example, the opposite polar point of x is a cut point, where there are two different cut vectors arriving at it with the same length.

Another possible cut vector scenario at v is to have $D(\text{Exp}_x)_v[w] = \mathbf{0} \in T_y\mathcal{M}$ for some $w \neq \mathbf{0} \in T_x\mathcal{M}$. Recall that $D(\text{Exp}_x)_v[w]$ describes an infinitesimal motion on the Riemannian geodesics emanating from x . Having a null direction $w \neq 0$ means one can produce infinitesimal change on

the geodesic along w while it costs no infinitesimal change on the arriving endpoint $y = \text{Exp}_x(v)$. Such a tangent vector defines a conjugate locus as follows.

$$\text{Conj}_x := \{v \in T_x\mathcal{M} : \exists w \neq \mathbf{0}, D(\text{Exp}_x)_v[w] = \mathbf{0}\}. \quad (1.9)$$

The $v \in \text{Conj}_x$ is referred to as a conjugate vector of x . Since $D(\text{Exp}_x)_v : T_x\mathcal{M} \rightarrow T_y\mathcal{M}$ is a linear operator between two linear spaces with the same dimension, having a null direction in $D(\text{Exp}_x)_v$ is equivalent to rank-deficiency in $D(\text{Exp}_x)_v$, i.e., $D(\text{Exp}_x)_v$ is being non-invertible. Therefore, a conjugate vector is a tangent vector at which the differential of the Riemannian exponential is non-invertible.

The two scenarios cover all possibilities of a cut vector. In summary, $v \in T_x\mathcal{M}$ is a cut vector if any or both of the following conditions hold:

1. The tangent vector v is the first conjugate vector along $\{s \cdot v : s \in [0, \infty)\} \subset T_x\mathcal{M}$.
2. There exists $w \neq v$ in $T_x\mathcal{M}$ satisfying $\|w\|_g = \|v\|_g$ and $\text{Exp}_x(v) = \text{Exp}_x(w)$.

It is important to distinguish the cut locus and the conjugate locus in this dissertation. The cut locus is considered a lot in the literature as it identifies a region where the Riemannian geodesics are the shortest curves. It not only relates the manifold structure with the metric space structure, but also provides a criterion of selecting a unique geodesic among the multiple geodesics between the given points in some manifolds. The classic Karcher mean formulation on a manifold is one of the many applications that is restricted within the cut locus. One of the major contributions in this work is to relax the cut locus and the shortest geodesic constraints to the smoothly evolving geodesic constraints, which is closely related to the conjugate locus.

1.2.6 Karcher Mean on a Manifold

The mean computation on a given data set has been an important analysis in various applications. The computed mean is usually considered the best representation in some measurements. When the data set $\{x_1, \dots, x_n\} \subset M$ lives on a metric space (M, d) , the Karcher mean formulation utilizes the distance function $d : M \times M \rightarrow \mathbb{R}$ to measure the objective as

$$\frac{1}{n} \sum_{i=1}^n d(x_i, y)^2, \forall y \in M$$

and defines the Karcher mean \bar{x} as a global minimum to the objective function

$$\bar{x} := \arg \min_{y \in M} \frac{\sum_{i=1}^n d(x_i, y)^2}{n}. \quad (1.10)$$

It derives the arithmetic mean $\bar{x} = \sum_{i=1}^n x_i/n$ of $\{x_1, \dots, x_n\} \subset \mathbb{R}^m, m \in \mathbb{Z}$ with the Euclidean space setting $(\mathbb{R}^m, (x, y) \mapsto |x - y|)$ and the geometric mean $\bar{x} = (\prod_{i=1}^n x_i)^{1/n}$ of positive numbers $\{x_1, \dots, x_n\} \subset \mathbb{R}_+ := \{x > 0 : x \in \mathbb{R}\}$ with the logarithmic metric space setting $(\mathbb{R}_+, (x, y) \mapsto |\log(x/y)|)$.

When it comes to a Riemannian manifold, it is natural to apply the classic Karcher mean formulation (1.10) directly to the metric space (\mathcal{M}, d) where d is the distance induced by the Riemannian metric. Recall that the distance between $x, y \in \mathcal{M}$ is the curve length of a shortest geodesic between them, the classic Karcher mean formulation on the manifold is then equivalent to

$$\bar{x} = \arg \min_{y \in \mathcal{M}} \frac{\sum_{i=1}^n g_{\gamma_i(0)}(\dot{\gamma}_i(0), \dot{\gamma}_i(0))}{n} \quad (1.11)$$

$$= \arg \min_{y \in \mathcal{M}} \frac{\sum_{i=1}^n g_{\gamma_i(1)}(\dot{\gamma}_i(1), \dot{\gamma}_i(1))}{n} \quad (1.12)$$

where $\gamma_i(t), t \in [0, 1]$ is a shortest geodesic between $\gamma_i(0) = x_i$ and $\gamma_i(1) = y$. Notice that the (1.11) has the g -norms evaluated at the initial velocities $\dot{\gamma}_i(0) \in T_{x_i}\mathcal{M}$, while the (1.12) has the g -norms evaluated at the arriving velocities $\dot{\gamma}_i(1) \in T_y\mathcal{M}$. By differentiating the objective (1.12) as a function of $y \in \mathcal{M}$, one obtains the set of critical points \mathbf{x}_* to (1.12) expressed in the set of solutions to a constraint on $T_y\mathcal{M}$:

$$\mathbf{x}_* = \arg \left\{ \sum_{i=1}^n \dot{\gamma}_i(1) = \mathbf{0} \right\} \quad (1.13)$$

where $\gamma_i(t), t \in [0, 1]$ is a shortest geodesic between x_i and y and there is $\bar{x} \in \mathbf{x}_*$.

Unfortunately, due to the possible multiple shortest geodesics and the non-smoothness in the Riemannian distance function, \mathbf{x}_* contains multiple critical points in general. Furthermore, the discontinuity of identifying a shortest geodesic around the cut locus yields that the set \mathbf{x}_* does not smoothly depend on the data set $\{x_1, \dots, x_i\}$ as they spread away. These features have been theoretical and computational issues to the Karcher mean problem on a manifold and, in some literature, the notion of the Karcher mean is relaxed to any critical point from the \mathbf{x}_* . One of the main contributions in this dissertation is to address these features on the special orthogonal group and to present a more appropriate generalization to (1.11) and (1.12) that is specific to a manifold setting.

1.3 Related Work

The concept of computing a set of smooth geodesics on a manifold is generalized from the idea of computing an inversion of a surjective function as a smooth multi-valued function. The complex

logarithm is one of the well-known examples that have the inversion of the complex exponential fully characterized. The first attempt to compute the smooth inversion of the Riemannian exponential on the special orthogonal group is given in [9]. Due to the lack of the inverse of the differential of the Riemannian exponential on the special orthogonal group, the inversion proposed in [9] is restricted within the cut locus of the identity matrix on the special orthogonal group.

The differential formula of an exponential map dates back to 1891, proven by Friedrich Schur, and it is sometimes also known as Duhamel’s formula. Duhamel’s formula on the matrix exponential of arbitrary square matrices yields the Baker–Campbell–Hausdorff formula and characterizes the conjugate locus on the general linear group. In 1995, Najfeld and Havel [28] derive the Duhamel’s formula on the matrix exponential of arbitrary diagonalized matrix that operates in complex arithmetics. This more restricted differential action has a simple inverse formula. Al-Mohy and Higham designs numerically stable and efficient algorithms to compute the Duhamel’s formula on the matrix exponential of any square matrix in 2009, [26], and to compute the differential to the matrix principal logarithm in 2013, [27]. However, the matrix principal logarithm has its range restricted within the principal branch of the matrix logarithm.

As the area of optimization on a manifold has steadily increased in interest to the optimization and application, more and more constrained optimizations and data sets are interpreted as unconstrained problems on a manifold. As one of the important milestones, Edelman et al. [11] identifies and discusses the Stiefel manifold with the canonical metric on the set of orthonormal bases as a Riemannian manifold inherited from the special orthogonal group as useful for computation and analysis of certain well-known situations in linear algebra. Although the Riemannian exponential map is given in that work, the inverse problem of finding a Riemannian geodesic between two given point is not solved until the recent work, e.g., [42], [29] and [33]. Unfortunately, these algorithms are only guaranteed for sufficiently close endpoints and degrade as the endpoints increase in separation.

The set of fixed rank positive semi-definite (FRPSD) matrices is a more complicated manifold related to the special orthogonal constraint and it arises in important applications like computer vision and statistical analysis. Vandereycken et al. propose a complete Riemannian structure in [36]. Although the Riemannian exponential on this manifold is derived, there is no way of computing its inverse to find a Riemannian geodesic between given points. The geometric interpretations of the resulting Riemannian geodesics are also not understood. These missing pieces limit its applications. In the spirit of having a simple and computationally tractable Riemannian geodesics, Massart and Absil propose a non-complete Riemannian structure in [23] that has been successfully

applied to various problems with the FRPSD constraint. In hopes of generalizing meaningful Riemannian geodesics on the FRPSD manifold, Bonnabel and Sepulchre [3] propose a Riemannian structure that has its geodesic infinitesimally approximated by a set of curves with special geometric interpretations.

The curves approximating the geodesic in [3] are then used to define a mean on the FRPSD manifold in [2]. The concept of this mean is generalized from the Karcher mean formulation on a metric space that is developed by Grove and Karcher [17] in 1970s. Unfortunately, the distance function on a manifold is usually not smooth globally and it depends on a shortest geodesic realizing the distance as its length, which may not be continuous globally as well. Both the mean proposed in [2] and the Karcher mean in the special orthogonal group, as reported in [20], have discussed the discontinuity in the distance function. As a result, the Karcher mean on a manifold is usually restricted to a “denser” data distribution such that the data set lies within the cut locus of the computed Karcher mean.

1.4 Research Overview and Dissertation Statement

As the Riemannian geometry becomes more and more developed, many classic Euclidean algorithms for solving unconstrained problems that has been adapted to their Riemannian generalizations for the respective problems on a manifold-constrained set, [1]. These Riemannian algorithms are built on the generalizations of geometric objects, e.g., the straight lines in the Euclidean settings are generalized to the Riemannian geodesics in the manifold setting. Unfortunately, such a generalization of straight lines is usually limited to a local scope bounded by the cut locus. It not only restricts the efficiency of the Riemannian algorithms built on these local generalizations, but may also introduce non-smoothness and discontinuity from the cut locus. Although there is no universal solution to address this locality issue or even to identify cut locus in arbitrary Riemannian manifolds, it is possible and worthwhile to develop a specialized solution for the special orthogonal group. On one hand, the special orthogonal group is well structured with rich properties and formulae explicitly available that helps the investigation. On the other hand, many other Riemannian manifolds with special orthogonal constraints are determined by the Riemannian structure on the special orthogonal group.

The first part of this dissertation investigates the behaviors of the Riemannian geodesics on the special orthogonal group and presents a novel characterization of the special orthogonal group

realized in the tangent space at the identity matrix, which is the set of skew symmetric matrices. The characterization identifies rich geometries of the Riemannian geodesics, possibly non-minimal, that emanate from the identity matrix in a smooth manner, i.e., the smooth geometry of the special orthogonal group beyond the scope of the cut locus is realized. Two efficient and reliable algorithms are designed to compute such a set of smoothly evolving (non-minimal) geodesics. These new insights and primitives are applied to the Karcher mean problem on the special orthogonal group and they overcome the issue of non-smooth objective function in the classic formulation.

The Stiefel manifold is a set of orthonormal bases and it is equipped with the Riemannian structure inherited from the special orthogonal group, known as the canonical metric. A point on the Stiefel manifold is a rectangular orthonormal matrix and it is identified with all of its special orthogonal completion as a submanifold in the special orthogonal group. Although the Riemannian exponential under the canonical metric is known, the endpoint geodesic problem that seeks a geodesic connecting two given endpoints is not solved until recent works in [42], [33] and etc. The second part of this dissertation investigates the recent algorithms on the endpoint geodesic problem on the Stiefel manifold and addresses their limitations with far-separated endpoints. Then, the insights and primitives developed on the special orthogonal group are applied to propose a novel Newton solver on a manifold root-finding formulation. Systematic numerical experiments further establish the dominant performance given by the proposed Newton solver.

The third part of this dissertation focuses on the manifold of fixed rank positive semi-definite matrices, namely the FRPSD manifold. This manifold usually emerges from the computer vision or the statistical analyses where the smooth varying geodesic with meaningful interpretations are essential to their application background. The first half of this part applies the tools developed on the special orthogonal group to solve the endpoint geodesic problem on the FRPSD manifold with an existing Riemannian structure proposed in [36]. Since it is still not understood how the Riemannian geodesic given in [36] is interpreted in practice, the second half of this part proposes a new Riemannian structure that is inspired by the attempt made in [3]. The properties of the new Riemannian structure are discussed and some interesting questions are left open.

Finally, the implementation of various algorithms and subroutines in the aforementioned topics plays an essential role in efficiency. Compared to the computations in Euclidean settings, the objects like the geodesic and the differential operator are a lot more expensive as they carry the non-trivial manifold constraints. Therefore, it is important to optimize all implementations from scratch and exploit the advantage in the manifold constraints as much as possible. For example, the

matrix exponential on skew symmetric matrices can be 80% times faster compared to the matrix exponential on real matrices. The last part of this dissertation collects these useful enhanced subroutines and primitives.

The following list highlights some of the most important contributions made in this dissertation.

1. Identify the conjugate locus on the special orthogonal group and derive efficient routines to compute the differential of the matrix exponential on the skew symmetric matrices and its inverse action.
2. Develop the notion of nearby matrix logarithm on the special orthogonal group that computes beyond the principal branch of the principal matrix logarithm. Reliable routines are designed to find a smooth skew symmetric $S(t), t \in [0, 1]$ for given $Q(t) = Q \exp(t \cdot \Delta)$ satisfying $\exp(S(t)) = Q(t), t \in [0, 1]$.
3. Based on the smoothly evolving geodesics in the special orthogonal group, a generalized Karcher mean is proposed, which is smoothly depending on the input data set.
4. Based on the quotient structure and the smoothly evolving geodesic in the special orthogonal group, a root-finding formulation of connecting given points with a geodesic is proposed on the Stiefel manifold with the canonical metric and the fixed rank positive semi-definite (FRPSD) matrix manifold with Vandereycken's metric.
5. A new Riemannian structure is proposed on the FRPSD manifold with meaningful Riemannian geodesic interpretations. Some of the earlier results in this new structure are derived and discussed.

CHAPTER 2

DIFFERENTIATING MATRIX EXPONENTIAL AT SKEW SYMMETRIC MATRICES

2.1 Introduction

The Riemannian structure on the special orthogonal group is closely related to the matrix exponential that maps from a skew symmetric matrix to a special orthogonal matrix. The differentiation of a matrix exponential has been studied in quantum theory, [37], and other literature, e.g., in economics and statistics [8, 7]. Various efficient algorithms for differentiating the matrix exponential have been proposed in [16, 28, 26]. While there have been much effort made on the matrix exponential map on general matrices, there is surprisingly less work specific to the set of skew symmetric matrices. This chapter investigates such a differential of the matrix exponential restricted to the set of skew symmetric matrices, which is an essential primitive in this dissertation. In particular, this chapter develops the explicit formulae in computing the restricted differential and its inverse. The set of skew symmetric matrices that have rank deficient differential are fully characterized and a pseudoinverse operator is designed in those rank deficient case.

This chapter is organized as follows. The introductory section briefly reviews the existing work on differentiating matrix exponential. Then, it introduces some necessary matrix factorizations and notations. The next section gives a detailed definition of the problem of interest, the formulae associated with differentiating the matrix exponential. The main body of this chapter derives these formulae for computing the restricted differential, its inverse and its pseudoinverse. Finally, this chapter presents the complexity analysis of the derived formulae along with some numerical results.

2.2 Preliminaries

2.2.1 Matrix Logarithm

Recall that the matrix exponential on a diagonalizable matrix X , i.e., there exists invertible complex matrix U and diagonal matrix Λ such that $X = U\Lambda U^{-1}$, is given by the entry-wise exponential on the diagonals $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ of D as

$$\exp(X) = P \exp(\Lambda) P^{-1} = P \operatorname{diag}(\exp(\lambda_1), \dots, \exp(\lambda_n)) P^{-1}$$

where $\text{diag}(A, B, \dots)$ denotes a (block) diagonal matrix with A, B, \dots on the diagonal. Notice that the exponential on complex numbers is periodic in $2\mathbf{i}\pi$ where \mathbf{i} is the imaginary unit from $\mathbf{i}^2 = -1$. It follows that for any matrix $X \in \mathbb{C}^{n \times n}$ and $\exp(X) = M$, the preimage of the matrix exponential at M , $\exp^{-1}(M) = \{\exp(Y) = M\}$, contains infinitely many solutions, including the given X . In most cases, the infinitely many solutions in $\exp^{-1}(M)$ form an isolated set of skew symmetric matrices. Similar to taking inverse of a periodic function on \mathbb{R} by returning the unique solution in a specified period, the principal logarithm of M , if it exists, is defined as the following unique point

$$\log(M) := \{X \in \mathbb{C}^{n \times n} : \exp(X) = M\} \cap \{X \in \mathbb{C}^{n \times n} : |\text{Im}(\lambda_i)| < \pi, \forall i = 1, \dots, n\}$$

where λ_i are the eigenvalues of X and $\text{Im} : (a + b \cdot \mathbf{i}) \mapsto b, \forall a, b \in \mathbb{R}$ takes the imaginary part of a complex number. The set $\{X \in \mathbb{C}^{n \times n} : |\text{Im}(\lambda_i)| < \pi, i = 1, \dots, n\}$ is known as the principal branch of the matrix exponential.

Note that not every matrix has its principal logarithm well defined. For example, take $Y \in \{X \in \mathbb{C}^{n \times n} : \exists 1 \leq i \leq n, \text{Im}(\lambda_i) = \pi\}$ on the boundary of the principal branch, then the preimage of $M = \exp(Y)$ has no intersection with the principal branch, as the imaginary parts in $\exp^{-1}(M)$ differ in multiples of $2\mathbf{i}\pi$. If there are $\lambda_j = a + \pi \cdot \mathbf{i}$ in Y , then any solution in the preimage $\tilde{Y} \in \exp^{-1}(M)$ has an eigenvalue in the form of $\tilde{\lambda}_j = a + (2k + 1)\pi \cdot \mathbf{i}$ where k is an integer. Notice that $|\text{Im}(\tilde{\lambda}_j)| = |(2k + 1)\pi| \geq \pi$, i.e., $\tilde{Y} \notin \{X \in \mathbb{C}^{n \times n} : |\text{Im}(\lambda_i)| < \pi, i = 1, \dots, n\}$.

The observations above apply to the more restricted case of the special orthogonal group. Further note that the eigenvalues of a skew symmetric matrices are either 0 or appear in purely-imaginary conjugate pairs $\pm\theta_i \cdot \mathbf{i}, i = 1, 2, \dots, m$ where $2m = n$ or $2m + 1 = n$. The statements specific to the special orthogonal group follow.

1. The matrix exponential $\exp : \mathbf{Skew}_n \rightarrow \mathbf{SO}_n$ is a smooth and surjective function.
2. The principal branch restricted on \mathbf{Skew}_n takes the form of

$$\mathbb{P} := \{S \in \mathbf{Skew}_n : \|S\|_2 < \pi\} \tag{2.1}$$

where $\|\cdot\|_2$ is the matrix 2-norm that returns the largest magnitude among the eigenvalues of the matrix.

3. The principal logarithm of a special orthogonal $Q \in \mathbf{SO}_n$, if it exists, is

$$\log(Q) = \{S \in \mathbf{Skew}_n : \|S\|_2 < \pi, \exp(S) = Q\}.$$

4. For a skew symmetric S with a pair of eigenvalues $\lambda_{\pm} = \pm\pi \cdot \mathbf{i}$, $Q = \exp(S)$ does not have its principal logarithm defined.

To see how the matrix 2-norm arises, notice the absolute value of the imaginary part of a purely-imaginary conjugate pair is the magnitude of the eigenvalues themselves, i.e., for $\lambda_{\pm} = \pm\theta_i \cdot \mathbf{i}$, $|\operatorname{Im}(\lambda_{\pm})| = |\theta_i|$.

2.2.2 Real Schur Decompositions

The real Schur decomposition plays an essential role in this dissertation and this section collects some important features and notations associated with it. Please refer to the textbook [16] for more details about the Schur decomposition.

Matrix Partitions. A real Schur decomposition converts any matrix into a block upper triangular matrix, in which the diagonal blocks are 2×2 or 1×1 , in the diagonal and zero entries below them. To simplify the expression in writing the action of such a block upper triangular matrix, the following notation in partitioning matrices into 2×2 blocks is introduced.

Definition 2.2.1. For any matrix M of size $n \times n$, denote

$$M_{[i,j]} := \begin{bmatrix} M_{2i-1,2j-1} & M_{2i-1,2j} \\ M_{2i,2j-1} & M_{2i,2j} \end{bmatrix}, \forall i, j \leq m$$

where $n = 2m$ or $n = 2m + 1$. For the odd $n = 2m + 1$, additionally denote $M_{[m+1,j]} := [M_{n,2j-1} \ M_{n,2j}]$, $\forall j \leq m$ in the leftover row, $M_{[i,m+1]} := \begin{bmatrix} M_{2i-1,n} \\ M_{2i,n} \end{bmatrix}$, $\forall i \leq m$ in the leftover column and $M_{[m+1,m+1]} := [M_{n,n}]$ in the leftover diagonal such that

$$M = \begin{bmatrix} M_{[1,1]} & \cdots & M_{[1,m]} \\ \vdots & \ddots & \vdots \\ M_{[m,1]} & \cdots & M_{[m,m]} \end{bmatrix}, n = 2m, \quad \text{or} \quad \begin{bmatrix} M_{[1,1]} & \cdots & M_{[1,m+1]} \\ \vdots & \ddots & \vdots \\ M_{[m+1,1]} & \cdots & M_{[m+1,m+1]} \end{bmatrix}, n = 2m + 1.$$

□

For a 5×5 matrix $M \in \mathbb{R}^{5 \times 5}$, the partition gives

$$\begin{bmatrix} M_{11} & M_{12} & M_{13} & M_{14} & M_{15} \\ M_{21} & M_{22} & M_{23} & M_{24} & M_{25} \\ M_{31} & M_{32} & M_{33} & M_{34} & M_{35} \\ M_{41} & M_{42} & M_{43} & M_{44} & M_{45} \\ M_{51} & M_{52} & M_{53} & M_{54} & M_{55} \end{bmatrix} = \begin{bmatrix} M_{[1,1]} & M_{[1,2]} & M_{[1,3]} \\ M_{[2,1]} & M_{[2,2]} & M_{[2,3]} \\ M_{[3,1]} & M_{[3,2]} & M_{[3,3]} \end{bmatrix}.$$

Schur Decomposition and Sepctral Decomposition. A *real Schur decomposition* of a real skew symmetric matrix $S \in \mathbf{Skew}_n$ is given by $S = RDR^T$ where R is an orthogonal matrix and D is a block diagonal matrix in the form of

$$D = \begin{cases} \text{diag}(D_{[1,1]}, D_{[2,2]}, \dots, D_{[m,m]}), & n = 2m \\ \text{diag}(D_{[1,1]}, D_{[2,2]}, \dots, D_{[m,m]}, D_{[m+1,m+1]}), & n = 2m + 1 \end{cases} \quad (2.2)$$

where $D_{[i,i]} = \begin{bmatrix} 0 & -\theta_i \\ \theta_i & 0 \end{bmatrix}$, $\theta_i \in \mathbb{R}$, $i = 1, 2, \dots, m$ and $D_{[m+1,m+1]} := 0$.

Note that the Schur decomposition on both a skew symmetric matrix and a special orthogonal matrix results in a block diagonal matrix with blocks in the form of $D_{[i,i]} = \begin{bmatrix} a_i & -b_i \\ b_i & a_i \end{bmatrix}$, rather than a block upper triangular matrix in general. Also notice that the real Schur decomposition of S is not unique but that all of them share the same block diagonal structure given in (2.2). The characterization of all Schur decompositions will be provided later in this section.

It follows immediately that a real Schur decomposition of any special orthogonal matrix $Q \in \mathbf{SO}_n$ is given by $Q = RER^T$ with $E = \exp(D)$ which consists of $E_{[i,i]} = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i) \end{bmatrix}$, $i = 1, \dots, m$ and an additional $E_{[m+1,m+1]} = 1$ when $n = 2m + 1$.

Another important implication of the structured Schur decomposition is that it reveals the spectral decomposition of S in the following simple way. Consider the unitary 2×2 matrix $U_2 := \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 \\ -\mathbf{i} & \mathbf{i} \end{bmatrix}$ where $\mathbf{i} = \sqrt{-1}$ is the imaginary unit, notice that it always consists of the eigenvectors of any D_i in (2.2) such that $D_{[i,i]} = U_2 \begin{bmatrix} -\theta_i \mathbf{i} & 0 \\ 0 & \theta_i \mathbf{i} \end{bmatrix} U_2^H$ where H denotes the conjugate transpose. Use the U_2 to form the $n \times n$ unitary matrix as

$$U := \begin{cases} \text{diag}(U_2, U_2, \dots, U_2), & n = 2m \\ \text{diag}(U_2, U_2, \dots, U_2, 1), & n = 2m + 1 \end{cases} \quad (2.3)$$

such that the unitary $V := RU$ consists of the eigenvectors of S , as

$$\begin{aligned} S &= RDR^T = R(UU^H)D(UU^H)R^T = V(U^H D U)V^H \\ &= \begin{cases} V \text{diag}(-\theta_1 \mathbf{i}, \theta_1 \mathbf{i}, \dots, -\theta_m \mathbf{i}, \theta_m \mathbf{i})V^H, & n = 2m \\ V \text{diag}(-\theta_1 \mathbf{i}, \theta_1 \mathbf{i}, \dots, -\theta_m \mathbf{i}, \theta_m \mathbf{i}, 0)V^H, & n = 2m + 1 \end{cases} \\ &= V\Lambda V^H. \end{aligned} \quad (2.4)$$

Characterization of Multiple Schur Decompositions. Recall that a spectral decomposition may not be unique if there exists a repeated eigenvalue. It gets more complicated in the Schur decomposition case, where matrices with distinct eigenvalues still have multiple Schur decompositions, possibly even with a different block diagonal matrix. This part presents an alternative way

of expressing a Schur decomposition and a characterization of all possible Schur decompositions on the given matrix based on the preferred Schur decomposition.

Definition 2.2.2 (Preferred Schur Decomposition). Let $X = RDR^T$ be a Schur decomposition of a skew symmetric matrix or a special orthogonal matrix X with block diagonal $X = \text{diag}(D_{[1,1]}, \dots)$ where $D_{[i,i]} = \begin{bmatrix} a_i & -b_i \\ b_i & a_i \end{bmatrix}$, $i = 1, \dots, m$ and the additional $D_{[m+1,m+1]} = 1$ or 0 in the $n = 2m + 1$ case. Then it is referred to as a *preferred Schur decomposition*, if the following conditions on the diagonal blocks are satisfied.

1. $b_i \geq 0, \forall i = 1, \dots, m$.
2. For a skew symmetric matrix, the diagonal blocks are placed in the following order

$$\begin{cases} b_i \geq b_{i+1}, & i = 1, \dots, m-1, \text{ if } X \in \mathbf{Skew}_n \\ a_i \leq a_{i+1}, & i = 1, \dots, m-1, \text{ if } X \in \mathbf{SO}_n \end{cases}$$

The resulting Schur decomposition is denoted as

$$\begin{aligned} X &= R \text{diag}(D_{a_j, b_j}, \dots, D_{a_r, b_r}) R^T \\ b_r &= 0 \text{ and } a_r = \begin{cases} 1, X \in \mathbf{SO}_n \\ 0, X \in \mathbf{Skew}_n \end{cases} \text{ for } n = 2m + 1 \end{aligned} \quad (2.5)$$

where $(a_i, b_i) \neq (a_j, b_j), \forall i \neq j$, $D_{a,b}$ are the block diagonal matrix that share the same (a, b) , $b_j \geq 0$ and $b_j > b_{j+1}$ for $X \in \mathbf{Skew}_n$ or $|a_j| \leq |a_j|$ for $X \in \mathbf{SO}_n$. \square

Proposition 2.2.3. Let X be a skew symmetric matrix or a special orthogonal matrix and let

$$X = RDR^T = R \text{diag}(D_{[1,1]}, \dots) R^T = R \text{diag}(D_{a_1, b_1}, \dots, D_{a_r, b_r}) R^T$$

be a preferred Schur decomposition. Let the dimension of $D_{a_i, b_i}, i = 1, \dots, r$ be $n_i = 2m_i + 1$ or $n_i = 2m_i$. Then, for any Schur decomposition

$$X = \tilde{R} \tilde{X} \tilde{R}^T = \tilde{R} \text{diag}(\tilde{X}_{[1,1]}, \dots) \tilde{R}^T,$$

there exist the orthogonal transformations P, G and Q , such that

$$\begin{cases} \tilde{R} & = RQGP \\ \tilde{D}_{[j,j]} & = D_{[i_j, i_j]}^*, \forall j = 1, \dots, m \end{cases}$$

where $\{i_j\}_{j=1}^m$ is a permutation of $\{1, 2, \dots, m\}$ and $D_{[i_j, i_j]}^*$ is either $D_{[i_j, i_j]}$ itself or its transpose, depending on the determinant of $G_{[i, i]}$. The orthogonal transformations are given as follows

1. *The Permutation:* $P = P_m \otimes I_2$ for $n = 2m$ or $P = \begin{bmatrix} P_m \otimes I_2 & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}$ for $n = 2m + 1$, where P_m permutes $1, 2, \dots, m$ into i_1, \dots, i_m and \otimes denotes the Kronecker product.

2. *The orthogonal transformations applied to vectors in groups of 2:*

$G = \text{diag}(G_1, \dots, G_m)$ for $n = 2m$ or $G = \text{diag}(G_1, \dots, G_m, 1)$ for $n = 2m + 1$ where G_i are any 2×2 orthogonal matrices.

3. *The orthogonal transformations applied to the vectors from repeated diagonal blocks:*

$Q = \text{diag}(Q_{a_1, b_1}, \dots, Q_{a_r, b_r})$ where Q_{a_i, b_i} are $n_i \times n_i$ orthogonal matrices as

$$\begin{cases} Q_{a_i, b_i} = Q_{m_i} \otimes I_2, \forall m_i \times m_i \text{ orthogonal } Q_{m_i} & \text{if } b_r \neq 0 \\ Q_{a_i, b_i} \text{ is any } n_i \times n_i \text{ orthogonal matrix} & \text{if } b_r = 0 \end{cases}$$

Angles in Skew Symmetric and Special Orthogonal Matrices. The structured real Schur decomposition on both the skew symmetric matrices and the special orthogonal matrices, especially the characterization of $\theta_1, \dots, \theta_m \in \mathbb{R}$ in the block diagonal matrix, is essential for the analyses derived in this work and the following notions are introduced.

Definition 2.2.4. Let $X = RDR^T$ be a preferred Schur decomposition of a skew symmetric matrix or a special orthogonal matrix X , with diagonal blocks $D_{[i,i]} = \begin{bmatrix} a_i & -b_i \\ b_i & a_i \end{bmatrix}$ with $b_i > 0$.

1. When X is skew symmetric, there are $a_i = 0$ and use $\theta_i = b_i$ as more geometric intuitive notation. The the set of θ_i are denoted as *the angles* of the X :

$$\Theta_X := \{\theta_i\}_{i=1}^m \in \mathbb{R}^m. \quad (2.6)$$

2. When Y is skew symmetric, there are $a_i^2 + b_i^2 = 1$ with $b_i \geq 0$. Then, there exists a unique $\theta_i \in [0, \pi]$, such that $a_i = \cos(\theta_i)$ and $b_i = \sin(\theta_i)$. Such a set of θ_i in range of $[0, \pi]$ are denoted as *the principal angles* of the X :

$$\Theta_X := \{\theta_i\}_{i=1}^m \in [0, \pi]^m. \quad (2.7)$$

On the other hand, given a set of angles $\Theta \in \mathbb{R}^m$ and the dimension $n = 2m$ or $n = 2m + 1$, the notation D^Θ and E^Θ are reserved for the block diagonal matrix in some Schur decompositions of a skew symmetric $S = RD^\Theta R^T$ and of a special orthogonal

$$Q = \exp(S) = R \exp(D^\Theta) R^T := RE^\Theta R^T$$

with diagonal blocks $D_{[i,i]}^\Theta = \begin{bmatrix} 0 & -\theta_i \\ \theta_i & 0 \end{bmatrix}$ and $E_{[i,i]}^\Theta = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i) \end{bmatrix}$ for $i = 1, \dots, m$. \square

2.2.3 Related Work

Theoretically speaking, the matrix exponential map $\exp : \mathbf{Skew}_n \rightarrow \mathbf{SO}_n$ is also the exponential map induced by the Lie group structure of \mathbf{SO}_n . This observation makes some general results in the Lie group context applicable to this dissertation. In particular, [31][Prop. 7, Sec. 1.2] gives the formula of the differential of the exponential map \exp induced by Lie group \mathcal{G}

$$\begin{aligned} D \exp_X : T_e \mathcal{G} &\rightarrow T_{\exp(X)} \mathcal{G} \\ Y &\mapsto \exp(X) \frac{1 - \exp(-\text{ad}_X)}{\text{ad}_X} Y \end{aligned}$$

where the $e \in \mathcal{G}$ is the identity element of the Lie group, the ad_X is the adjoint action on \mathcal{G} and the operator $\frac{1 - \exp(-\text{ad}_X)}{\text{ad}_X}$ is given by the power series

$$\frac{1 - \exp(-\text{ad}_X)}{\text{ad}_X} = \sum_{k=0}^{\infty} \frac{(-1)^k}{(k+1)!} (\text{ad}_X)^k.$$

Furthermore, the invertibility of $D \exp_X$ can be characterized by the eigenvalues of the adjoint action ad_X . Therefore, the work in this dissertation can be viewed as finding the respective explicit forms specific to the special orthogonal group. It is worth noting that such explicit forms for the general linear group have been studied, in which case the adjoint action is given by $\text{ad}_X(Y) = XY - YX$ with its eigenvalues being equal to the eigenvalues of X . Although the special orthogonal group can be viewed as a subgroup of the general linear group, some results cannot be blindly applied to the special orthogonal group as shown in this dissertation, which makes the derivation presented in this work necessary.

Computationally speaking, there have been many discussions and studies on the differential of the matrix exponential. In [28][**Algorithm 4.5**] Najfeld and Havel give a formula of the differential at diagonalizable foot. In [26][**Algorithm 7.4**] Al-mohy and Higham give the formula of the differential at general feet. In [27][**Algorithm 6.1**] Al-mohy and et al. give the formula of the inverse action at a foot within the principal branch. However, there is an absence of work specific to the skew symmetric matrices and the special orthogonal matrices. The invertibility condition of $D \exp_S : \mathbf{Skew}_n \rightarrow T_Q \mathbf{SO}_n$ is still not completely understood. Not to mention that there are many structures in the skew symmetry and special orthogonality unexploited which have important computational implications. The work in this dissertation is in hope of filling some of these gaps so that the theoretical geometric analysis on \mathbf{SO}_n may have stronger consequences in applications.

2.3 Problem Statement

Before giving any details of the formula $\text{D exp}_S : \mathbf{Skew}_n \rightarrow T_Q \mathbf{SO}_n$ at S with $Q = \exp(S)$, it is important to recall the characterization of $T_Q \mathbf{SO}_n$

$$T_Q \mathbf{SO}_n = \{Q\Omega : \Omega \in \mathbf{Skew}_n\}.$$

In most of the application scenarios, especially in the context of differentiable manifolds, it is the skew symmetric $\Omega \in \mathbf{Skew}_n$ in $\Delta_Q = Q\Omega \in T_Q \mathbf{SO}_n$ that participates in various computations and analyses rather than Δ_Q itself. For example, suppose a curve on \mathbf{SO}_n $\gamma(t) \in \mathbf{SO}_n, \forall t \in [0, 1]$ emanating from $\gamma(0) = Q$ along with $\frac{d}{dt}\gamma(t)|_{t=0} = Q\Omega$ is required. With $\Omega \in \mathbf{Skew}_n$, it is easy to construct $\gamma(t) := Q \exp(t\Omega)$. Although it is possible to construct another curve via $\Delta_Q = Q\Omega$ as $\gamma(t) := \text{Proj}_{\mathbf{SO}_n}(Q + t\Delta_Q)$ where $\text{Proj}_{\mathbf{SO}_n}$ is some (local) projector onto \mathbf{SO}_n , e.g., the polar projector $X \mapsto Q$ where $QR = X$ is a polar decomposition, such a curve does not share the rich geometry or convenient properties of $\exp(t\Omega)$. Therefore, the $\text{D exp}_S : \mathbf{Skew}_n \rightarrow T_Q \mathbf{SO}_n$ investigated in this work emphasizes the skew symmetric characterization Ω of $Q\Omega = \Delta_Q = \text{D exp}_S[\Delta_S] \in T_Q \mathbf{SO}_n$ as the \mathcal{L}_S defined below.

Definition 2.3.1. For any given $S \in \mathbf{Skew}_n$ and the respective $Q = \exp(S) \in \mathbf{SO}_n$, the linear map $\mathcal{L}_S : \mathbf{Skew}_n \rightarrow \mathbf{Skew}_n$ is defined as

$$\mathcal{L}_S(\Delta_S) := Q^T (\text{D exp}_S[\Delta_S]) \tag{2.8}$$

such that $\text{D exp}_S[\Delta_S] = Q\mathcal{L}_S(\Delta_S)$. □

This section derives the symbolic formulae for computing the \mathcal{L}_S and its (pseudo) inverse based on the existing formula for a more general case as reviewed in **Lemma 2.3.2**.

Lemma 2.3.2. [28][**Theorem 4.5**] For any diagonalizable $X = Z\Lambda Z^{-1} \in \mathbb{C}^{n \times n}$, the differential $\text{D exp}_X[\Delta]$ along $\Delta \in \mathbb{C}^{n \times n}$ is given by

$$\text{D exp}_X[\Delta] = Z ((Z^{-1}\Delta Z) \odot \Psi) Z^{-1}, \tag{2.9}$$

where \odot is the Hadamard product that performs entry-wise multiplication and the symmetric matrix Ψ has the entries

$$\psi_{ij} = \psi_{ji} = \begin{cases} \frac{e^{\lambda_j} - e^{\lambda_i}}{\lambda_j - \lambda_i} & \lambda_i \neq \lambda_j \\ e^{\lambda_i} & \lambda_i = \lambda_j \end{cases}. \tag{2.10}$$

Any skew symmetric matrix S is diagonalizable which makes (2.9) applicable. It follows immediately that the desired \mathcal{L}_S can be written for the diagonalizable foot $X = S \in \mathbf{Skew}_n$ as

$$\begin{aligned}
\mathcal{L}_X(\Delta) &= \exp(X)^{-1} D \exp_X[\Delta] \\
&= (Z \exp(-\Lambda) Z^{-1}) (Z ((Z^{-1} \Delta Z) \odot \Phi) Z^{-1}) \\
&= Z ((Z^{-1} \Delta Z) \odot (\exp(-\Lambda) \Psi)) Z^{-1} \\
&= Z ((Z^{-1} \Delta Z) \odot \Phi) Z^{-1}
\end{aligned} \tag{2.11}$$

where $\Phi = \exp(-\Lambda) \Psi$ has the following entries

$$\phi_{ij} = \begin{cases} \frac{e^{\lambda_j - \lambda_i} - 1}{\lambda_j - \lambda_i} & \lambda_i \neq \lambda_j \\ 1 & \lambda_i = \lambda_j \end{cases}.$$

Furthermore, the relationship between the spectral decomposition $P \Lambda P^H$ and the Schur decomposition $R D R^T$ of S in (2.4) yields \mathcal{L}_S in the form

$$\begin{aligned}
\mathcal{L}_S(\Delta_S) &= P ((P^H \Delta_S P) \odot \Phi) P^H \\
&= R U ((U^H (R \Delta_S R^T) U) \odot \Phi) U^H R^T
\end{aligned}$$

Notice that the Schur vectors R only act as a change of variables in the domain and range of \mathcal{L}_S in the expression above. For $\Delta_S, \Delta_Q \in \mathbf{Skew}_n$ with $\Delta_Q = \mathcal{L}_S(\Delta_S)$, let $M := R^T \Delta_S R$ and $N := R^T \Delta_Q R$ be the skew symmetric matrices computed by Δ_S, Δ_Q and R , such that

$$\begin{aligned}
\mathcal{L}_S(\Delta_S) &= R U ((U^H (R \Delta_S R^T) U) \odot \Phi) U^H R^T \\
\iff N &= U ((U^H M U) \odot \Phi) U^H.
\end{aligned}$$

Observe that the second linear map $M \mapsto N$ only depends on the eigenvalues of S which are completely determined by its angles. It is natural to further decompose the linear map \mathcal{L}_S into compositions of linear actions as follows.

Definition 2.3.3. For any $S \in \mathbf{Skew}_n$ with the Schur vectors R , the angles Θ and the $Q = \exp(S)$, denote the linear operator \mathcal{B}_R for the characterization under the base R as

$$\begin{aligned}
\mathcal{B}_R : \mathbf{Skew}_n &\rightarrow \mathbf{Skew}_n \\
\Delta &\mapsto R \Delta R^T
\end{aligned} \tag{2.12}$$

The respective core linear operator \mathcal{C}_Θ is given by

$$\begin{aligned}
\mathcal{C}_\Theta : \mathbf{Skew}_n &\rightarrow \mathbf{Skew}_n \\
M &\mapsto U ((U^H M U) \odot \Phi) U^H
\end{aligned} \tag{2.13}$$

such that

$$\mathcal{L}_S = \mathcal{B}_R \circ \mathcal{C}_\Theta \circ \mathcal{B}_R^{-1}.$$

Here, the eigenvalues Λ are determined by the angles Θ as in (2.4) and $\mathcal{B}_R^{-1}(\Delta) = R^T \Delta R$. \square

For any $S \in \mathbf{Skew}_n$ with the Schur vectors R , the angles Θ and the $Q = \exp(S)$, this section solves the following problems.

1. Derive an efficient formula for computing the linear action $\mathcal{L}_S : \mathbf{Skew}_n \rightarrow \mathbf{Skew}_n$.
2. Investigate the invertibility of $D \exp_S$, which is equivalent to the invertibility of \mathcal{C}_Θ .
3. Derive an efficient formula for computing the action of \mathcal{L}_S^{-1} in the invertible case.
4. Define a pseudoinverse action \mathcal{L}_S^\dagger with an efficient formula in the non-invertible case.

2.4 Differential Formula

The complexity of computing \mathcal{L}_S in the form of (2.11) is dominated by the 4 matrix multiplications with the complex P and P^H . For $\mathcal{L}_S = \mathcal{B}_R \circ \mathcal{C}_\Theta \circ \mathcal{B}_R^{-1}$, the 4 complex matrix multiplications are replaced by the 4 real matrix multiplications and the core map \mathcal{C}_Θ only involves complex matrix multiplications with the block diagonal U and the entry-wise Hadamard product. This means the complexity of $\mathcal{B}_R \circ \mathcal{C}_\Theta \circ \mathcal{B}_R^{-1}$ has already been reduced to 1/4 compared to (2.11). However, it remains beneficial to further exploit the formula of \mathcal{C}_Θ for faster computation as well as more insight into its invertibility.

Consider the $n = 5$ case as an example to investigate \mathcal{C}_Θ . Let $S \in \mathbf{Skew}_5$ be a skew symmetric 5×5 matrix with the Schur vectors R , angles $\Theta = \{\theta_1, \theta_2\}$, eigenvectors $P = RU$ and eigenvalues $\{-\theta_1 \mathbf{i}, \theta_1 \mathbf{i}, -\theta_2 \mathbf{i}, \theta_2 \mathbf{i}, 0\}$. Then, for any $M = \mathcal{B}_R^{-1}(\Delta_S) \in \mathbf{Skew}_n$, the block diagonal structure in U yields the $N = \mathcal{C}_\Theta(M)$ written as

$$\begin{aligned} N &= U \left((U^T M U) \odot \Phi \right) U^H \\ &= \begin{bmatrix} U_2 \left(U_2^H M_{[1,1]} U_2 \odot \Phi_{[1,1]} \right) U_2^H & U_2 \left(U_2^H M_{[1,2]} U_2 \odot \Phi_{[1,2]} \right) U_2^H & U_2 \left(U_2^H M_{[1,3]} \odot \Phi_{[1,3]} \right) \\ U_2 \left(U_2^H M_{[2,1]} U_2 \odot \Phi_{[2,1]} \right) U_2^H & U_2 \left(U_2^H M_{[2,2]} U_2 \odot \Phi_{[2,2]} \right) U_2^H & U_2 \left(U_2^H M_{[1,3]} \odot \Phi_{[2,3]} \right) \\ \left(M_{[3,1]} U_2 \odot \Phi_{[3,1]} \right) U_2^H & \left(M_{[3,2]} U_2 \odot \Phi_{[3,2]} \right) U_2^H & M_{[3,3]} \end{bmatrix} \end{aligned}$$

where the $M_{[i,j]}$ denotes blocks of sizes 2×2 , 1×2 and 2×1 as given in **Definition 2.2.1**.

In the expression above, there are only 2 nontrivial linear actions, the

$$U_2 \left(U_2^H M_{[i,j]} U_2 \odot \Phi_{[i,j]} \right) U_2^H, \forall j \leq i \leq m$$

acting on the $[i, j]$ -th 2×2 block and the

$$(M_{[m+1,j]}U_2 \odot \Phi_{[m+1,j]})U_2^H, \forall j \leq m$$

acting on the $[m+1, j]$ -th 1×2 block. These linear actions are independent of each other and they act in an in-place fashion, with $M_{[i,j]}$ computing $N_{[i,j]}$ respectively. Note that the skew symmetry in N and M saves the computation of the $[i, j]$ -th 2×2 blocks with $i < j$ and the $U_2(U_2^H M_{[i,m+1]} \odot \Phi_{[i,m+1]})$ with $\forall i \leq m$. These linear actions are given in the following lemma via some simple linear algebraic manipulations.

Lemma 2.4.1. For any $x = \theta_i \in \mathbb{R}$ that determines the $(2i-1)$ -th and the $(2i)$ -th eigenvalues $-\lambda_i \mathbf{i}$ and $\lambda_i \mathbf{i}$ and any $y = \theta_j \in \mathbb{R}$ that determines the $(2j-1)$ -th and the $(2j)$ -th eigenvalues $-\lambda_j \mathbf{i}$ and $\lambda_j \mathbf{i}$ of a skew symmetric matrix S , the linear action on $[i, j]$ -th 2×2 block

$$N_{[i,j]} = U_2 (U_2^H M_{[i,j]} U_2 \odot \Phi) U_2^H$$

can be written as a matrix $\mathcal{N}_{x,y} \in \mathbb{R}^{4 \times 4}$ in the forms of

$$\mathcal{N}_{x,y} := \frac{1}{2} \begin{bmatrix} a+c & -b-d & b-d & a-c \\ b+d & a+c & -a+c & b-d \\ -b+d & -a+c & a+c & -b-d \\ a-c & -b+d & b+d & a+c \end{bmatrix}, \quad \begin{cases} a = \frac{\sin(x-y)}{x-y}, & b = \frac{\cos(x-y)-1}{x-y} \\ c = \frac{\sin(x+y)}{x+y}, & d = \frac{\cos(x+y)-1}{x+y} \end{cases} \quad (2.14)$$

that acts on the vectorized system as $\mathcal{N}_{\theta_i, \theta_j} \cdot \text{vec}(M_{[i,j]}) = \text{vec}(N_{[i,j]})$, $\forall i, j \leq m$. In the limits when (1) $x = y \neq 0$, $a = 1$ and $b = 0$ or (2) $x = y = 0$, then $a = c = 1$ and $b = d = 0$.

Proof. For the 2×2 block $[i, j]$, $i, j \leq m$ associated with eigenvalues $-\theta_i \mathbf{i}$, $\theta_i \mathbf{i}$, $-\theta_j \mathbf{i}$ and $\theta_j \mathbf{i}$, there is

$$\Phi_{[i,j]} = \begin{bmatrix} \frac{e^{\mathbf{i}(-\theta_j+\theta_i)}-1}{\mathbf{i}(-\theta_j+\theta_i)} & \frac{e^{\mathbf{i}(\theta_j+\theta_i)}-1}{\mathbf{i}(\theta_j+\theta_i)} \\ \frac{e^{\mathbf{i}(-\theta_j-\theta_i)}-1}{\mathbf{i}(-\theta_j-\theta_i)} & \frac{e^{\mathbf{i}(\theta_j-\theta_i)}-1}{\mathbf{i}(\theta_j-\theta_i)} \end{bmatrix}. \quad \text{Denote } a, b, c \text{ and } d \text{ as}$$

$$\Phi_{[i,j]} := \begin{bmatrix} a - \mathbf{i}b & c - \mathbf{i}d \\ c + \mathbf{i}d & a + \mathbf{i}b \end{bmatrix} \quad \text{with} \quad \begin{cases} a = \frac{\sin(\theta_i - \theta_j)}{\theta_i - \theta_j} & b = \frac{\cos(\theta_i - \theta_j) - 1}{\theta_i - \theta_j} \\ c = \frac{\sin(\theta_i + \theta_j)}{\theta_i + \theta_j} & d = \frac{\cos(\theta_i + \theta_j) - 1}{\theta_i + \theta_j} \end{cases}.$$

Denote the real block $M_{[i,j]}$ as $\begin{bmatrix} \xi_1 & \xi_3 \\ \xi_2 & \xi_4 \end{bmatrix}$ and denote the complex block

$$U_2^H M_{[i,j]} U = \frac{1}{2} \begin{bmatrix} \eta_1 - \mathbf{i}\eta_4 & \eta_2 - \mathbf{i}\eta_3 \\ \eta_2 + \mathbf{i}\eta_3 & \eta_1 + \mathbf{i}\eta_4 \end{bmatrix}$$

to obtain the relation between ξ_i 's and η_i 's as

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix}.$$

Further writes $U_2 (U_2^H M_{[i,j]} U_2 \odot \Phi_{[i,j]}) U_2^H$ as expression of η_i 's as

$$\begin{aligned} G (G^H M_{[i,j]} G \odot \Phi_{[i,j]}) G^H &= \frac{1}{2} \begin{bmatrix} \eta_1 a + \eta_2 c - \eta_3 d - \eta_4 b & -\eta_1 b + \eta_2 d + \eta_3 c - \eta_4 a \\ \eta_1 b + \eta_2 d + \eta_3 c + \eta_4 a & \eta_1 a - \eta_2 c + \eta_3 d - \eta_4 b \end{bmatrix} \\ \Rightarrow \text{vec}(N_{[i,j]}) &= \frac{1}{2} \begin{bmatrix} a & c & -d & -b \\ b & d & c & a \\ -b & d & c & -a \\ a & -c & d & -b \end{bmatrix} \eta = \frac{1}{2} \begin{bmatrix} a+c & -b-d & b-d & a-c \\ b+d & a+c & -a+c & b-d \\ -b+d & -a+c & a+c & -b-d \\ a-c & -b+d & b+d & a+c \end{bmatrix} \xi. \end{aligned}$$

□

Lemma 2.4.2. For any $x = \theta_j \in \mathbb{R}$ that determines the $(2j-1)$ -th and the $(2j)$ -th eigenvalues $-\lambda_j \mathbf{i}$ and $\lambda_j \mathbf{i}$ of a skew symmetric matrix S , the linear action on $[m+1, j]$ -th 1×2 block

$$N_{[m+1,j]} = (M_{[m+1,j]} U_2 \odot \Phi_{[m+1,j]}) U_2^H, \forall j \leq m$$

can be written as a matrix $\mathcal{N}_x \in \mathbb{R}^{2 \times 2}$ in the forms of

$$\mathcal{L}_x := \begin{bmatrix} e & f \\ -f & e \end{bmatrix} \quad \text{with} \quad \begin{cases} e = \frac{\sin(x)}{x} \\ f = \frac{\cos(x) - 1}{x} \end{cases} \quad (2.15)$$

that acts on the vectorized system as $\mathcal{N}_{\theta_j} \cdot \text{vec}(M_{[m+1,j]}) = \text{vec}(N_{[m+1,j]})$, $\forall j \leq m$. In the limits when $x = 0$, then $e = 1$ and $f = 0$.

Proof. The proof is essentially the same with the proof of **Lemma 2.4.1** and therefore details are omitted. The important steps for the 1×2 case are

$$\begin{aligned} \Phi_{[3,j]} &= \begin{bmatrix} \frac{\exp(-\mathbf{i}\theta_j)-1}{-\mathbf{i}\theta_j} & \frac{\exp(\mathbf{i}\theta_j)-1}{\mathbf{i}\theta_j} \end{bmatrix} = [e + \mathbf{i}f \quad e - \mathbf{i}f] \\ [\xi_1 \quad \xi_2] U_2 &= \frac{\sqrt{2}}{2} [\xi_1 - \mathbf{i}\xi_2 \quad \xi_1 + \mathbf{i}\xi_2] \\ [\xi_1 \quad \xi_2] U_2 \odot \Phi_{[m+1,j]} &= \frac{\sqrt{2}}{2} [e\xi_1 + f\xi_2 + \mathbf{i}(f\xi_1 - e\xi_2) \quad e\xi_1 + f\xi_2 + \mathbf{i}(e\xi_2 - f\xi_1)] \\ ([\xi_1 \quad \xi_2] U_2 \odot \Phi_{[3,j]}) U_2^H &= [e\xi_1 + f\xi_2 \quad -f\xi_1 + e\xi_2] \end{aligned}$$

□

Theorem 2.4.3. For any skew symmetric $S \in \mathbf{Skew}_n$ with the Schur vectors R and the angles $\Theta = \{\theta_1, \dots, \theta_m\}$ where $n = 2m$ or $n = 2m + 1$, the core action $\mathcal{C}_\Theta(M) = N$ evaluated at $M = \mathcal{B}_R(\Delta_S), \forall \Delta_S \in \mathbf{Skew}_n$ assembles (2.14) on 2×2 blocks as

$$\begin{aligned} \begin{bmatrix} M_{[1,1]} & \cdots & M_{[1,m]} \\ \vdots & \ddots & \vdots \\ M_{[m,1]} & \cdots & M_{[m,m]} \end{bmatrix} &\mapsto \begin{bmatrix} N_{[1,1]} & \cdots & N_{[1,m]} \\ \vdots & \ddots & \vdots \\ N_{[m,1]} & \cdots & N_{[m,m]} \end{bmatrix} \\ \begin{bmatrix} \text{vec}(M_{[1,1]}) & \cdots & \text{vec}(M_{[1,m]}) \\ \vdots & \ddots & \vdots \\ \text{vec}(M_{[m,1]}) & \cdots & \text{vec}(M_{[m,m]}) \end{bmatrix} &\mapsto \begin{bmatrix} \mathcal{N}_{\theta_1, \theta_1} \cdot \text{vec}(M_{[1,1]}) & \cdots & \mathcal{N}_{\theta_1, \theta_m} \cdot \text{vec}(M_{[1,m]}) \\ \vdots & \ddots & \vdots \\ \mathcal{N}_{\theta_m, \theta_1} \cdot \text{vec}(M_{[m,1]}) & \cdots & \mathcal{N}_{\theta_m, \theta_m} \cdot \text{vec}(M_{[m,m]}) \end{bmatrix} \end{aligned}$$

and the additional (2.15) on 1×2 blocks when $n = 2m + 1$ as $\text{vec}(N_{[m+1,j]}) = \mathcal{N}_{\theta_j}(\text{vec}(M_{[m+1,j]}))$ and $N_{[j,m+1]} = -N_{[m+1,j]}^T, \forall j \leq m$.

Proof. The **Lemma 2.4.1** and **Lemma 2.4.2** give the constructive proof of **Theorem 2.4.3**. \square

When the matrix M and N are vectorized in the order of

$$\text{blk-vec}(X) := \text{vec}(\text{vec}(X_{[1,1]}), \text{vec}(X_{[2,1]}), \dots, \text{vec}(X_{[m,m]}), \dots),$$

the linear map $\mathcal{C}_\Theta : N \mapsto M$ can be written as the block diagonal matrix consists of $\mathcal{N}_{\theta_i, \theta_j}$ and the additional $\mathcal{N}_{\theta_j}, \forall i, j \leq m$ as $\mathcal{C}_\Theta := \text{diag}(\mathcal{N}_{\theta_1, \theta_1}, \dots, \mathcal{N}_{\theta_m, \theta_m}, \dots)$, such that $\mathcal{C}_\Theta \text{blk-vec}(M) = \text{blk-vec}(N)$.

Note that the a, b, c, d, e and f are well defined in all cases due to the nature of the function $\sin(z)/z$ and $(\cos(z) - 1)/z$ in the limit of $z \rightarrow 0$ as illustrated in **Figure 2.1**.

2.5 Inverse of the Differential Formula

This section investigates the invertibility of $\text{D exp}_S : \mathbf{Skew}_n \rightarrow T_Q \mathbf{SO}_n$ which is equivalent to the invertibility of $\mathcal{L}_S : \mathbf{Skew}_n \rightarrow \mathbf{Skew}_n$ as $\text{D exp}_S[\Delta_S] = Q \mathcal{L}_S(\Delta_S)$ where $Q \in \mathbf{SO}_n$ is always invertible. The invertibility of $\mathcal{L}_S : \mathbf{Skew}_n \rightarrow \mathbf{Skew}_n$ is equivalent to the invertibility of $\mathcal{C}_\Theta : \mathbf{Skew}_n \rightarrow \mathbf{Skew}_n$ as $\mathcal{L}_S(\Delta) = \mathcal{B}_R \circ \mathcal{C}_\Theta \circ \mathcal{B}_R^{-1}$ where $\mathcal{B}_R : \mathbf{Skew}_n \rightarrow \mathbf{Skew}_n$ is always invertible. According to **Theorem 2.4.3**, the core map \mathcal{C}_Θ is a collection of independent smaller linear systems $\mathcal{N}_{x,y} \in \mathbb{R}^{4 \times 4}$ and the additional $\mathcal{N}_x \in \mathbb{R}^{2 \times 2}$ where x, y take values from the angles Θ of S . Therefore, this section discusses the invertibilities of $\mathcal{N}_{x,y}$ and \mathcal{N}_x .

A natural approach to obtain an invertible \mathcal{C}_Θ is to ask if all $\mathcal{N}_{x,y}$, and the additional \mathcal{N}_x when $n = 2m + 1, \forall x, y \in \Theta$ are invertible. According to the symbolic form of these matrices given in (2.14) and (2.15), their symbolic inverse form are available as follows.

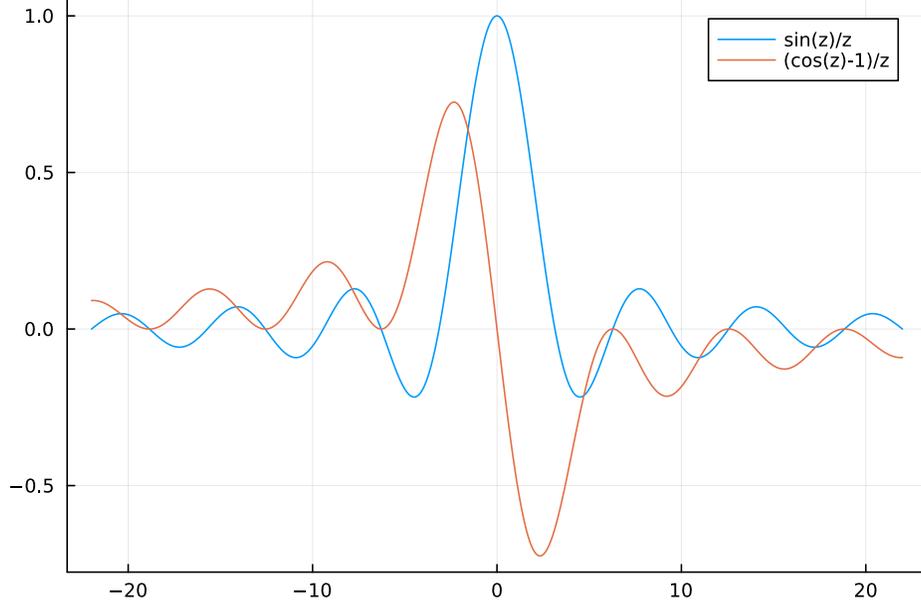


Figure 2.1: Illustration of $\sin(z)/z$ and $(\cos(z) - 1)/z$.

Proposition 2.5.1. *The symbolic inverse formulae, if exist, of the linear maps $\mathcal{N}_{x,y}$ defined in (2.16) and \mathcal{N}_x defined in (2.17) are in the form of*

$$\mathcal{N}_{x,y}^{-1} = \frac{1}{2} \begin{bmatrix} a' + c' & b' + d' & -b' + d' & a' - c' \\ -b' - d' & a' + c' & -a' + c' & -b' + d' \\ b' - d' & -a' + c' & a' + c' & b' + d' \\ a' - c' & b' - d' & -b' - d' & a' + c' \end{bmatrix} \text{ with } \begin{cases} a' = \frac{a}{a^2 + b^2}, & b' = \frac{b}{a^2 + b^2} \\ c' = \frac{c}{c^2 + d^2}, & d' = \frac{d}{c^2 + d^2} \end{cases} \quad (2.16)$$

when $a^2 + b^2 \neq 0$ and $c^2 + d^2 \neq 0$,

$$\mathcal{N}_x^{-1} = \begin{bmatrix} e' & -f' \\ f' & e' \end{bmatrix} \text{ with } \begin{cases} e' = \frac{e}{e^2 + f^2} \\ f' = \frac{f}{e^2 + f^2} \end{cases} \quad (2.17)$$

when $e^2 + f^2 \neq 0$.

Proof. Proof of verifying the symbolic inverse is obtained by direct computation of the matrix multiplications $\mathcal{N}_{x,y}\mathcal{N}_{x,y}^{-1} = I_4$ and $\mathcal{N}_x\mathcal{N}_x^{-1} = I_2$. The algebra is omitted. \square

Notice that the following trigonometric identities

$$\left\{ \begin{array}{l} \frac{\sin^2(z)}{z^2} + \frac{(\cos(z) - 1)^2}{z^2} = \frac{\sin^2(z) + \cos^2(z) + 2\cos(z) + 1}{z^2} = \frac{2 - 2\cos(z)}{z^2} \\ \frac{\sin(z)}{z} \Big/ \frac{2 - 2\cos(z)}{z^2} = \frac{z}{2} \cdot \frac{\sin(z)}{1 - \cos(z)} = \frac{z}{2} \cdot \frac{2\sin(z/2)\cos(z/2)}{2\sin^2(z/2)} = \frac{z}{2} \cdot \cot\left(\frac{z}{2}\right), \forall z \in \mathbb{R}. \\ \frac{1 - \cos(z)}{z} \Big/ \frac{2 - 2\cos(z)}{z^2} = \frac{z}{2} \end{array} \right.$$

further simplify the expressions of a', b', c', d', e' and f' as

$$\left\{ \begin{array}{l} a' = \frac{\theta_i - \theta_j}{2} \cdot \cot\left(\frac{\theta_i - \theta_j}{2}\right) \quad b' = \frac{\theta_i - \theta_j}{2} \\ c' = \frac{\theta_i + \theta_j}{2} \cdot \cot\left(\frac{\theta_i + \theta_j}{2}\right) \quad d' = \frac{\theta_i + \theta_j}{2}, \\ e' = \frac{\theta_i}{2} \cdot \cot\left(\frac{\theta_i}{2}\right) \quad f' = \frac{\theta_i}{2} \end{array} \right., \quad (2.18)$$

which introduces the following sufficient condition of the invertibility of \mathcal{C}_Θ .

Corollary 2.5.2. The core map $\mathcal{C}_\Theta : \mathbf{Skew}_n \rightarrow \mathbf{Skew}_n$ is invertible if for any two angles $\forall \theta_i, \theta_j \in \Theta$, including any angle repeated itself, there is

$$\begin{aligned} \theta_i \pm \theta_j \neq 2k\pi, \forall i, j \leq m, k = \pm 1, \pm 2, \dots \\ \theta_i \neq 2k\pi, \forall i \leq m, k = \pm 1, \pm 2, \dots \text{ additional for } n = 2m + 1. \end{aligned} \quad (2.19)$$

Proof. Under this condition (2.19), the simplified (2.18) are always well defined, which yields the invertibility of all $\mathcal{L}_{\theta_i, \theta_j}$ and the additional \mathcal{L}_{θ_i} . Note that the $\theta_i \pm \theta_j = 0$ case does not violate the condition, as $z \cot(z)$ is not define only at $z = k\pi, k = \pm 1, \dots$ \square

Note that **Corollary 2.5.2** is consistent with the classic result of differentiating the exponential map on the general linear group

$$\mathbf{GL}_n := \{X \in \mathbb{R}^{n \times n} : \det(X) \neq 0\},$$

c.f. [31][Prop. 7, Sec. 1.2], stating that $D \exp_A : \mathbb{R}^{n \times n} \rightarrow T_Q \mathbf{GL}_n = \mathbb{R}^{n \times n}$ is invertible if and only if for any 2 eigenvalues $\lambda_i, \lambda_j, \forall 1 \leq i, j \leq n$, there is $\lambda_i - \lambda_j \neq 2k\pi \mathbf{i}, k = \pm 1, \dots, m$ where $\mathbf{i} = \sqrt{-1}$. Indeed for a skew symmetric matrix A with the angles Θ , the eigenvalues are either $\pm \theta_i \mathbf{i}, 1 \leq i \leq m$ or 0. Then, $\theta_i \pm \theta_j = 2k\pi$ is equivalent to $\mathbf{i}\theta_i \pm \mathbf{i}\theta_j = 2k\pi \mathbf{i}$. For the differential

$$D \exp_S : \mathbb{R}^{n \times n} \rightarrow T_Q \mathbf{GL}_n, \Delta_S \mapsto Q \cdot \mathcal{B}_R \circ \mathcal{C}_\Theta \circ \mathcal{B}_R^{-1}(\Delta_S),$$

notice that $\mathcal{B}_R^{-1} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}, X \mapsto R^T X R$ is surjective, which make $\mathbb{R}^{n \times n}$ the domain of \mathcal{C}_Θ in $D \exp_S : \mathbb{R}^{n \times n} \rightarrow T_Q \mathbf{GL}_n$. Therefore, the maps $\mathcal{N}_{\theta_i, \theta_j}$ and the additional \mathcal{N}_{θ_i} are acting on free

variable $\text{vec}(M_{[i,j]})$ and the additional $\text{vec}(M_{[n,j]})$ and $\text{vec}(\text{vec}(M_{[i,n]}))$. Therefore, any of \mathcal{N} 's being rank deficient is equivalent to the action $\mathcal{C}_\Theta \circ \mathcal{B}_R^{-1} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ being rank deficient, which is equivalent to $\text{D exp}_S : \mathbb{R}^{n \times n} \rightarrow T_Q \mathbf{GL}_n$ not being invertible.

However, (2.19) is not necessary for $\text{D exp}_S : \mathbf{Skew}_n \rightarrow T_Q \mathbf{SO}_n$ to be invertible, as there are extra constraints on the domain of \mathcal{C}_Θ introduced by the skew symmetry, which is preserved under $\mathcal{B}_R^{-1} : \mathbf{Skew}_n \rightarrow \mathbf{Skew}_n$. For the upper or lower 2×2 blocks $M_{[i,j]}, i \neq j$ or the additional $M_{[i,m+1]}$ and $M_{[m+1,j]}$, skew symmetry relates $M_{[i,j]} = -M_{[j,i]}^T$, the entries within each of these blocks remain unconstrained in the respective small map and by the same argument given on $\text{D exp}_S : \mathbb{R}^{n \times n} \rightarrow T_Q \mathbf{GL}_n$, $\text{D exp}_S : \mathbf{Skew}_n \rightarrow T_Q \mathbf{SO}_n$ is not invertible if the following condition fails.

$$\theta_i \pm \theta_j \neq 2k\pi, \forall i \neq j \leq m, k = \pm 1, \pm 2, \dots$$

$$\theta_i \neq 2k\pi, \forall i \leq m, k = \pm 1, \pm 2, \dots \text{ additional for } n = 2m + 1.$$

For the diagonal block, the domain $M_{[i,i]}$ is no longer a free matrix in $\mathbb{R}^{2 \times 2}$ as the skew symmetry in $M_{[i,i]} = -M_{[i,i]}^T = \begin{bmatrix} 0 & -x \\ x & 0 \end{bmatrix}$ leaves only 1 degree of freedom and the repeated angles yield $a = 0$, $b = 1$ in (2.14) as $\theta_i - \theta_i = 0$. Together, the linear action of $\mathcal{L}_{\theta_i, \theta_i}$ acting on the diagonal block $M_{[i,i]}$ has the special form of

$$\mathcal{N}_{\theta_i, \theta_i}(\text{vec}(M_{[i,i]})) = \frac{1}{2} \begin{bmatrix} 1+c & -d & -d & 1-c \\ d & 1-c & -1-c & -d \\ d & -1+c & 1+c & -d \\ 1-c & d & d & 1+c \end{bmatrix} \begin{bmatrix} 0 \\ x \\ -x \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ x \\ -x \\ 0 \end{bmatrix}, \forall x, \theta_i \in \mathbb{R},$$

which is an identity map on $M_{[i,i]}$. Such an identity map is always invertible and therefore the $i = j$ case on the sufficient condition (2.19) can be dropped. The following **Theorem 2.5.3** summarizes the discussion above as a stronger necessary and sufficient condition of the invertibility in the more restricted linear map $\text{D exp}_S : \mathbf{Skew}_n \rightarrow T_Q \mathbf{SO}_n$.

Theorem 2.5.3. *The differential $\text{D exp}_S : \mathbf{Skew}_n \rightarrow T_Q \mathbf{SO}_n$ of the restricted matrix exponential $\exp : \mathbf{Skew}_n \rightarrow \mathbf{SO}_n$ at $S \in \mathbf{Skew}_n$, with the angles Θ and its exponential $Q = \exp(S)$, is invertible if and only if*

$$\begin{aligned} \theta_i + \theta_j &\neq 2k\pi, \forall i < j \leq m, k \in \mathbb{Z} \setminus \{0\} \\ \theta_i &\neq 2k\pi, \forall i \leq m, k \in \mathbb{Z} \setminus \{0\} \text{ additional for } n = 2m + 1. \end{aligned} \tag{2.20}$$

Proof. The previous argument completes the sufficiency of (2.20).

To see the necessity of (2.20), suppose $\theta_i + \theta_j = 2k\pi, i \neq j, k \neq 0$, there is $c = 0$ and $d = 0$ and a, b are undetermined, i.e.,

$$\mathcal{L}_{\theta_i, \theta_j} = \frac{1}{2} \begin{bmatrix} a & -b & b & a \\ b & a & -a & b \\ -b & -a & a & -b \\ a & -b & b & a \end{bmatrix}.$$

Let $M_{[i,j]} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, $M_{[j,i]} = -M_{[i,j]}^T$ and 0 elsewhere in M . There is

$$\text{vec}(N_{[i,j]}) = \mathcal{N}_{\theta_i, \theta_j} \text{vec}(M_{[i,j]}) = \mathbf{0}, \forall a, b \in \mathbb{R}.$$

□

Corollary 2.5.4. The conjugate locus of the identity matrix I_n in the tangent space $T_{I_n} \mathbf{SO}_n = \mathbf{Skew}_n$ is the set of skew symmetric matrices with the angles not satisfying (2.20), i.e.,

$$\text{Conj}_{I_n} := \left\{ S \in \mathbf{Skew}_n : \begin{array}{l} \exists i \neq j, k \neq 0, \text{ s.t. } \theta_i \pm \theta_j = 2k\pi \\ \text{or } \exists i, k \neq 0, \text{ s.t. } \theta_i = 2k\pi \text{ for } n = 2m + 1 \end{array} \right\} \quad (2.21)$$

2.6 Pseudoinverse of the Differential Formula

This section investigates the behavior of the rank-deficient differential at $S \in \mathcal{S}_-$ and it specifies a pseudoinverse of this rank-deficient linear map. Since the action of $\mathcal{N}_{\theta_i, \theta_i}$ on diagonal blocks is known to be the identity action, this section only considers $\mathcal{N}_{\theta_i, \theta_j}$ acting on 4 free variables from $M_{[i,j]}$ and the additional \mathcal{N}_{θ_i} acting on 2 free variables from $M_{[m+1,j]}$ with $i \neq j \leq m$.

Consider the most extreme case where $\theta_i + \theta_j = 2k\pi, \theta_i - \theta_j = 2l\pi$ for some integer $k, l \neq 0$. In this case, the a, b, c and d in (2.14) are all 0 and $\mathcal{N}_{\theta_i, \theta_j}$ degenerates to a trivial map $\mathbb{R}^4 \ni x \mapsto \mathbf{0}$ with rank 0. Such a trivial map has no meaningful pseudoinverse. Similarly, if $\theta_i = 2k\pi$ with some integer $k \neq 0$, \mathcal{N}_{θ_i} degenerates to a trivial map to $\mathbf{0}$ and it has no meaningful pseudoinverse either.

Then, for $\theta_i + \theta_j = 2k\pi$ for integer $k \neq 0$ and $\theta_i - \theta_j \neq 2l\pi, \forall l = \pm 1, \pm 2, \dots$, there is $c = 0$, $d = 0$ and $a^2 + b^2 \neq 0$. In this case, the action of $\mathcal{L}_{\theta_i, \theta_j}$ degenerates to

$$\mathcal{N}_{\theta_i, \theta_j} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a & -b & b & a \\ b & a & -a & b \\ -b & -a & a & -b \\ a & -b & b & a \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a(\xi_1 + \xi_4) - b(\xi_2 - \xi_3) \\ b(\xi_1 + \xi_4) + a(\xi_2 - \xi_3) \\ -b(\xi_1 + \xi_4) - a(\xi_2 - \xi_3) \\ a(\xi_1 + \xi_4) - b(\xi_2 - \xi_3) \end{bmatrix} := \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix}.$$

Notice that the following equations can be solved as

$$\begin{cases} \eta_1 = \eta_4, & \xi_1 + \xi_4 = 2 \cdot \frac{a\eta_1 + b\eta_2}{a^2 + b^2} \\ \eta_2 = -\eta_3, & \xi_2 - \xi_3 = 2 \cdot \frac{a\eta_2 - b\eta_1}{a^2 + b^2} \end{cases} \text{ with } \eta = \text{Proj}_{\mathcal{R}(\mathcal{N}_{\theta_i, \theta_j})} \left(\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \right) = \frac{1}{2} \begin{bmatrix} y_1 + y_4 \\ y_2 - y_3 \\ y_3 - y_2 \\ y_1 + y_4 \end{bmatrix}$$

where $y \in \mathbb{R}^4$ is an arbitrary vector, $\mathcal{R}(\mathcal{L}_{\theta_i, \theta_j}) \subset \mathbb{R}^4$ denotes the range of $\mathcal{N}_{\theta_i, \theta_j}$ and Proj is the orthogonal projector on \mathbb{R}^4 .

As expected, there are still 2 degrees of freedom available for ξ with $\mathcal{N}_{\theta_i, \theta_j} \xi = \eta$ for a given η . The extra constraints on ξ imposed by $\xi_1 = \xi_4$ and $\xi_2 = -\xi_3$ are then introduced to specify a particular pseudoinverse $\mathcal{N}_{\theta_i, \theta_j}^\dagger$ for $\eta \in \mathcal{R}(\mathcal{N}_{\theta_i, \theta_j})$ such that

$$\mathcal{N}_{\theta_i, \theta_j}^\dagger \begin{bmatrix} \eta_1 \\ \eta_2 \\ -\eta_2 \\ \eta_1 \end{bmatrix} = \frac{1}{a^2 + b^2} \begin{bmatrix} a\eta_1 + b\eta_2 \\ a\eta_2 - b\eta_1 \\ b\eta_1 - a\eta_2 \\ a\eta_1 + b\eta_2 \end{bmatrix}$$

Note that such a choice is not unique but it is chosen as the smallest solution in the preimage on \mathbb{R}^4 that satisfies $\mathcal{N}_{\theta_i, \theta_j} \xi = \eta$ as stated later in **Proposition 2.6.2**.

Similar analysis derives the formulae for $\theta_i - \theta_j = 2l\pi$ for some integer $l \neq 0$ and $\theta_i + \theta_j \neq 2k\pi, \forall k \in \pm 1, \dots$ as

$$\begin{cases} \eta_1 = -\eta_4, & \xi_1 - \xi_4 = 2 \cdot \frac{c\eta_1 + d\eta_2}{c^2 + d^2} \\ \eta_2 = \eta_3, & \xi_2 + \xi_3 = 2 \cdot \frac{c\eta_2 - d\eta_1}{c^2 + d^2} \end{cases} \text{ with } \eta = \text{Proj}_{\mathcal{R}(\mathcal{N}_{\theta_i, \theta_j})} \left(\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \right) = \frac{1}{2} \begin{bmatrix} y_1 - y_4 \\ y_2 + y_3 \\ y_2 + y_3 \\ y_4 - y_1 \end{bmatrix},$$

furthermore, under the extra constraints $\xi_1 = -\xi_4$ and $\xi_2 = \xi_3$ imposed on ξ , a pseudoinverse is given by

$$\mathcal{N}_{\theta_i, \theta_j}^\dagger \cdot \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_2 \\ -\eta_1 \end{bmatrix} := \frac{1}{c^2 + d^2} \begin{bmatrix} c\eta_1 + d\eta_2 \\ c\eta_2 - d\eta_1 \\ c\eta_2 - d\eta_1 \\ -c\eta_1 - d\eta_2 \end{bmatrix}.$$

The following **Definition 2.6.1** collects the formulae derived above and defines a pseudoinverse of a rank deficient \mathcal{C}_Θ .

Definition 2.6.1. For a rank deficient small system \mathcal{N}_{θ_i} with $\theta_i = 2k\pi$ for some integer $k \neq 0$, its trivial pseudoinverse is chosen as

$$\mathcal{N}_{\theta_i}^\dagger := \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

For a rank deficient small system $\mathcal{N}_{\theta_i, \theta_j}$, the $\mathcal{N}_{\theta_i, \theta_j}^\dagger, i \neq j$ denotes the pseudoinverse as the linear maps:

1. If \exists integers $k, l \neq 0$ s.t., $\theta_i + \theta_j = 2k\pi, \theta_i - \theta_j = 2l\pi$, then

$$\mathcal{N}_{\theta_i, \theta_j}^\dagger := \mathbf{0} \in \mathbb{R}^{4 \times 4}.$$

2. If \exists an integer $k \neq 0$ and \forall integers $l \neq 0$, s.t. $\theta_i + \theta_j = 2k\pi$, $\theta_i - \theta_j \neq 2l\pi$, then

$$\mathcal{N}_{\theta_i, \theta_j}^\dagger := \frac{1}{2} \begin{bmatrix} a' & b' & -b' & a' \\ -b' & a' & -a' & -b' \\ b' & -a' & a' & b' \\ a' & b' & -b' & a' \end{bmatrix}. \quad (2.22)$$

3. If \exists an integer $l \neq 0$ and \forall integers $k \neq 0$, s.t. $\theta_i + \theta_j \neq 2k\pi$, $\theta_i - \theta_j = 2l\pi$, then

$$\mathcal{N}_{\theta_i, \theta_j}^\dagger := \frac{1}{2} \begin{bmatrix} c' & d' & d' & -c' \\ -d' & c' & c' & d' \\ -d' & c' & c' & -d' \\ -c' & -d' & -d' & c' \end{bmatrix}. \quad (2.23)$$

The $\mathcal{C}_\Theta^\dagger : \mathbf{Skew}_n \rightarrow \mathbf{Skew}_n$ denotes the pseudoinverse of $\mathcal{C}_\Theta : \mathbf{Skew}_n \rightarrow \mathbf{Skew}_n$ that consists of $\mathcal{N}_{\theta_i, \theta_j}^\dagger$ acting on the vectorized off-diagonal blocks $N_{[i,j]}$, $i \neq j$, the identity map I_4 acting on the vectorized diagonal blocks $N_{[i,i]}$ and the additional $\mathcal{C}_\Theta^\dagger$, when $S \in \mathcal{S}_-$. The $\mathcal{C}_\Theta^\dagger$ degenerates to \mathcal{C}_Θ^{-1} when $S \in \mathcal{S}_+$, i.e., when \mathcal{C}_Θ is invertible. The pseudoinverse operators of $\mathcal{L}_S : \mathbf{Skew}_n \rightarrow \mathbf{Skew}_n$ and $\text{D exp}_S : \mathbf{Skew}_n \rightarrow T_Q \mathbf{SO}_n$ are given by $\mathcal{L}_S^\dagger := \mathcal{B}_R \circ \mathcal{C}_\Theta^\dagger \circ \mathcal{B}_R^{-1}$ and $(\text{D exp}_S)^\dagger [Q\Delta_Q] := \mathcal{L}_S^\dagger(\Delta_Q)$, respectively. \square

Notice that such a pseudoinverse operator is consistent with (2.16) by setting $a', b' = 0$ and/or $c', d' = 0$ and with (2.17) by setting $e', f' = 0$.

Also notice that the nontrivial small pseudoinverse consists of the projector onto the range space of the respective forward action and the explicit inverse formulae derived above. The verification of \mathcal{C}_Θ being a pseudoinverse by writing out $\mathcal{C}_\Theta \circ \mathcal{C}_\Theta^\dagger \circ \mathcal{C}_\Theta = \mathcal{C}_\Theta$ is omitted for simplicity. As mentioned in the construction, such a pseudoinverse is not unique and this particular $\mathcal{C}_\Theta^\dagger$ is chosen due to the following extra properties.

Proposition 2.6.2. *For the pseudoinverse $\mathcal{C}_\Theta^\dagger$ given in Definition 2.6.1, the following statements hold.*

1. *For any $N \in \mathbf{Skew}_n$ that may or may not be in the range space $\mathcal{R}(\mathcal{C}_\Theta)$, the pseudoinverse finds the nearest point to N on $\mathcal{R}(\mathcal{C}_\Theta)$ as*

$$\mathcal{C}_\Theta \circ \mathcal{C}_\Theta^\dagger(N) := N_* = \arg \min_{Y \in \mathcal{R}(\mathcal{C}_\Theta)} \|Y - N\|_{\mathbb{F}}^2.$$

2. *For any $N \in \mathcal{R}(\mathcal{C}_\Theta)$, the pseudoinverse finds the smallest solution as*

$$\mathcal{C}_\Theta^\dagger(N) := M_* = \arg \min_{\mathcal{C}_\Theta(X)=N} \|X\|_{\mathbb{F}}^2.$$

Proof. Since the small systems are independent of each other, it suffices to prove the similar statements about $\mathcal{N}_{\theta_i, \theta_j}^\dagger$ acting on $N_{[i,j]}$ and the additional $\mathcal{N}_{\theta_j}^\dagger$ acting on $N_{[m+1,j]}$. When the pseudoinverse is trivial, by construction there is $\mathcal{R}(\mathcal{N}_{\theta_i, \theta_j}) = \{\mathbf{0} \in \mathbb{R}^4\}$ or $\mathcal{R}(\mathcal{N}_{\theta_i}) = \{\mathbf{0} \in \mathbb{R}^2\}$, which yields the statements.

For the nontrivial pseudoinverse that only happens on the 4×4 system $\mathcal{N}_{\theta_i, \theta_j}^\dagger$ acting on $N_{[i,j]}$, $i \neq j$, without loss of generality, consider the $\theta_i + \theta_j = 2k\pi$ case. The $\theta_i - \theta_j = 2k\pi$ case can be similarly proved by interchanging the plus and minus between $\eta_1 + \eta_4$ and $\eta_2 - \eta_3$. The pseudoinverse by construction satisfies

$$\mathcal{N}_{\theta_i, \theta_j} \mathcal{N}_{\theta_i, \theta_j}^\dagger \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \eta_1 + \eta_4 \\ \eta_2 - \eta_3 \\ \eta_3 - \eta_2 \\ \eta_1 + \eta_4 \end{pmatrix},$$

which is an orthogonal projection from $y = \text{vec}(N_{[i,j]})$ to the range space characterized as

$$\mathcal{R}(\mathcal{N}_{\theta_i, \theta_j}) = \{\eta \in \mathbb{R}^4 : \eta_1 = \eta_4, \eta_2 = -\eta_3\}.$$

The orthogonal projection implies $\|\mathcal{N}_{\theta_i, \theta_j} \circ \mathcal{N}_{\theta_i, \theta_j}^\dagger (\text{vec}(N_{[i,j]})) - \text{vec}(N_{[i,j]})\|_2$ is minimized, where $\|\cdot\|_2$ denotes the vector 2 norm. It only remains to notice that the vector-2 norm is the same as the matrix Frobenius norm in the form of $\|\text{vec}(N_{[i,j]})\|_2 = \|N_{[i,j]}\|_F$. The proof of the first statement follows immediately.

For the second statement, without loss of generality, consider the $\theta_i + \theta_j = 2k\pi$ case. Notice that the preimage of $\mathcal{N}_{\theta_i, \theta_j}$ denoted as $\{\xi \in \mathbb{R}^4 : \mathcal{N}_{\theta_i, \theta_j} = \eta\}$ is given in the form of $\xi_1 + \xi_4 = 2C_1$ and $\xi_2 - \xi_3 = 2C_2$ where C_1 and C_2 are constants determined by c, d and y . The second statement follows from the classic results

$$\begin{aligned} \arg \min_{\xi_1 + \xi_4 = 2C_1} (\xi_1^2 + \xi_4^2) &= (C_1, C_1) \\ \arg \min_{\xi_2 - \xi_3 = 2C_2} (\xi_2^2 + \xi_3^2) &= (C_2, -C_2). \end{aligned}$$

□

2.7 Routines and Implementations

This section elaborates the algorithmic and implementation details of the derived formulae. Recall that the proposed new formulae are derived from the complex formulae (2.9) by exploiting the skew symmetry and avoiding complex arithmetic. For simplicity, the new formulae are denoted as “**real**” formulae while the formulae (2.9) are denoted as the “**complex**” formulae. For the

formulae proposed in [26] and [27], they both utilize the Padé series of the matrix exponential for the computation and therefore they are denoted as the “Padé” formulae in this section.

These formulae apply to different range of matrices S in $\text{Dexp}_S[\Delta_S] = Q\Delta_Q$ where $Q = \exp(S)$ and the following table presents a summary. Note that the row in “forward” means the computation $\Delta_S \mapsto \Delta_Q$ while the row in “backward” means the computation $\Delta_Q \mapsto \Delta_S$.

Table 2.1: The feasible conditions of the root S in computing $\Delta_S \leftrightarrow \Delta_Q$

Condition	Real	Complex	Padé
Forward	Skew Symmetric	Diagonalizable, [28]	All Complex Matrix , [26]
Backward	Skew Symmetric	Diagonalizable, [28]	Within Principal Branch, [27]

2.7.1 Three-Stage Evaluation

First of all, the computation of the formulae derived in this chapter all share three similar stages. They all require some intermediate results typically computed during the matrix exponential $S \mapsto \exp(S)$. These computed objects are then used to further compute the necessary parameters in the formulae. The final stage is to apply the formulae to an input variable. This 3-stage procedure applies to other existing formulae, not specific to the skew symmetric and special orthogonal matrices. These stages are denoted as “PreEval-Param-Action”. In the **PreEval** stage, the routine gets objects that are either available from previous computations or recomputed from the scratch. In the **Param** stage, all parameters in the formula other than the input variable are computed. The **Action** stage computes the action of the formula provided with an input Δ .

The stages like the **PreEval** and the **Param** are considered as a single “Preprocessing” stage in a typical algorithmic analysis. However, it is necessary in this case to distinguish them for the following reasons. Firstly, different formulae require different intermediate results that are computed in different ways of computing $S \mapsto \exp(S)$ and these objects do not convert easily from one to another. Secondly, the contribution of the **PreEval** stage to the overall complexity may vary from a very significant portion, if it is computed from the scratch, to nothing at all, if it reuses the previous computation appropriately. Finally, the contribution of the **Param** stage to the overall complexity is usually negligible. In conclusion, having **PreEval** and **Param** split provides important details in usage of these formulae.

In the **PreEval**, the formulae specific to skew symmetric and special orthogonal matrices require the Schur decomposition of the skew symmetric matrix, the formulae in [28] and [27] require the

spectral decomposition of the skew symmetric matrix and the formula in [26] requires the intermediate matrix products in the scaling and squaring method [18]. Note that the algorithmic details in various ways of computing $S \mapsto \exp(S)$ is beyond the scope of this dissertation. A brief discussion about this topic is given in later chapter on important primitives. Although the complexity of the matrix factorization significantly varies by the structure of the matrix being factored, this section no longer dive into details. Instead of elaborating all possible cases in the matrix factorization, this dissertation refers to the standard library of each formula that can handle most cases in a reasonable range with consistent complexity, which is discussed below.

2.7.2 Refinements on Existing Formulae

To demonstrate a comprehensive and reliable analysis of the complexity of the proposed new formulae specific to skew symmetric matrices, it is necessary to introduce appropriate refinements to the existing formulae that are designed for more general scenarios. These refinements should exploit part of the structure in the skew symmetry and accelerate the existing formulae so that the comparison is fair. Secondly, these refinements are expected to help narrowing down the vary range of complexity as the formulae are evaluated in a more targeted set of matrices.

The existing formulae proposed in [28], [26] and [27] are assumed to be evaluated on a normal matrix S , i.e., $SS^T - S^T S = \mathbf{0}$, i.e., S is unitarily diagonalizable with the spectral decomposition $Z\Lambda Z^H$ where Z^H is the Hermitian transpose of Z . It is clear that normal matrices include skew symmetric matrices S but not vice versa. With the additional normal structure, the formulae in [28] are free from computing the inverse of the eigenvectors of S as normal matrix has unitary eigenvectors and their inverse is given by the simple Hermitian transpose. For the formula in [27], the normal structure reduces the repeated solves on block upper triangular matrices into solves on block diagonal matrices. For the formula in [26], the differential of the matrix exponential is expressed as the linear combination of a series of dense matrix products, in which the normal structure cannot speed up the computation. Fortunately, this formula is already fast enough for the experiments.

Based on the discussion above, the `PreEval` stage of the formulae executes the following computations. The `real` formulae perform the `SYTRD`-like method proposed for the Schur decomposition of the skew symmetric matrices proposed in [24]. The `complex` formulae and `Padé` perform the `HETRD` in `CLAPACK` for the spectral decomposition of the Hermitian matrices. The `Padé` formula performs the dense matrix multiplications for the matrix products in the scaling and squaring method.

2.7.3 Pseudo Codes

Algorithm 1: Linear Operators of $D \exp_S$ and $D \exp_S^{-1}$

Input: Set of angles Θ
Output: Parameters for \mathcal{L}_Θ stored as vectors

```

1 for  $j = 1, \dots, m$  do
2    $L_j := [a_j \quad b_j]^T$ ; // Equation 2.16
3    $L_j^\dagger := [a'_j \quad b'_j]^T$ ; // Equation 2.17
4   for  $i = j + 1, \dots, m$  do
5      $L_{ij} := [a_{ij} + c_{ij} \quad a_{ij} - c_{ij} \quad b_{ij} + d_{ij} \quad b_{ij} - d_{ij}]^T$ ; // Equation 2.14
6      $L_{ij}^\dagger := [a'_{ij} + c'_{ij} \quad a'_{ij} - c'_{ij} \quad b'_{ij} + d'_{ij} \quad b'_{ij} - d'_{ij}]^T$ ; // Equation 2.15
7   Return  $L_{ij}, L_{ij}^\dagger, L_i, L_j^\dagger$  for  $i > j = 1, 2, \dots, m$ ;

```

Algorithm 2: Action of the Directional Derivative $D \exp_S$ or its Inverse.

Input: Real Schur vectors W of S with linear operators from **Algorithm 1** and Δ .
Output: Δ_Q from $D \exp_S[\Delta] = \exp(S)\Delta_Q$ or Δ_S from $D \exp_S[\Delta] = \exp(S)\Delta_Q$.

```

1  $X \leftarrow W^T \Delta W$ ;
2 for  $j = 1, \dots, m$  do
3   for  $i = j + 1, \dots, m$  do
4      $Y_{[i,j]} \leftarrow \mathcal{L}_{[i,j]}(X_{[i,j]})$  or  $Y_{[i,j]} \leftarrow \mathcal{L}_{[i,j]}^{-1}(X_{[i,j]})$ ; // Equation 2.14, 2.16
5      $Y_{[j,i]} \leftarrow -Y_{[i,j]}^T$ ; // Skew-symmetry
6    $Y_{[j,j]} \leftarrow X_{[j,j]}$ ;
7   if  $n = 2m + 1$  then
8      $Y_{[m,j]} \leftarrow \mathcal{L}_{[m,j]}(X_{[m,j]})$  or  $Y_{[m,j]} \leftarrow \mathcal{L}_{[m,j]}^{-1}(X_{[m,j]})$ ; // Equation 2.15, 2.17
9      $Y_{[j,m]} \leftarrow -Y_{[m,j]}^T$ ; // Skew-symmetry
10  Return  $WYW^T$ ;

```

2.8 Complexity Analyses and Numerical Results

This part presents the analysis of operation counts in the proposed new formulae and the existing formulae in [28], [26] and [27] and numerical experiments that validate the complexity analyses.

2.8.1 Complexity

Real formulae for skew symmetric S : A real Schur decomposition of S is required for $D \exp_S[\cdot]$ and $D(\exp_S)_Q^{-1}[\cdot]$ which takes approximately $25n^3$ floating point operations (FLOPs), [16]. All angles of S are known at no additional cost from the Schur decomposition. It remains to determine all the a, b, c, d from the angles by **Algorithm 1** so that all parameters in the direct formulae are obtained. This takes $8n^2$ FLOPs. Finally, the execution of **Algorithm 2** requires 4

matrix multiplications and $O(n^2) \approx 3n^2$ updates in 2×2 blocks, with a complexity approximately $8n^3 + 3n^2$ FLOPs.

Complex formulae for normal S : A spectral decomposition for an arbitrary normal matrix is obtained based on the real Schur decomposition [16] and it requires approximately $25n^3 + 9n^2$ FLOPs when S is skew symmetric. It remains to determine the $\exp(-\Lambda)\Psi$ matrix from

$$\begin{aligned}\Delta_Q &= \exp(S)^{-1} D \exp_S[\Delta_S] = Z \exp(-\Lambda) Z^{-1} (Z((Z^{-1} \Delta_S Z) \cdot \Psi) Z^{-1}) \\ &= Z(\exp(-\Lambda)((Z^{-1} \Delta_S Z) \cdot \Psi)) Z^{-1} \\ &= Z((Z^{-1} \Delta_S Z) \odot (\exp(-\Lambda)\Psi)) Z^{-1}\end{aligned}$$

where the last equal sign follows from the fact that $\exp(-\Lambda)$ is a diagonal matrix. These complex computations require approximately $8n^2$ FLOPs. Finally, the evaluation computes 4 complex matrix multiplications and 1 complex Hadamard product, which is $32n^3 + 3n^2$ FLOPs. For the inverse of the directional derivative, the Hadamard product with $\exp(-\Lambda)\Psi$ is replaced by the Hadamard division with $\exp(-\Lambda)\Psi$.

Padé formulae for normal S : Since the underlying algorithms for the matrix exponential and the matrix principal logarithm are fundamentally different, the two directional derivatives have different complexity. $D \exp_S[\cdot]$ does not require a decomposition and $D \log_Q[\cdot]$ requires a real Schur decomposition. The complexity of the implementations depends on the Padé order t of the matrix exponential and the rescaling process s on a large Δ_S and/or S . According to both [26] and [27], the recommended Padé order is $t = 13$ and no rescaling is considered in the complexity analysis, i.e., $s = 0$. Note that repeated evaluations with the same S or Q are required for these primitives and therefore those computations that only depend on S or Q are considered the preprocessing of the $D \exp_S$ or $D \log_Q$, which are reusable in repeated evaluations. The preprocessing for $D \exp_S$ requires 6 real matrix multiplications($12n^3$ FLOPs) and 1 LU decomposition($2/3n^3$ FLOPs), while the $D \log_S$ has no reusable terms for repeated evaluations. The evaluation of $D \exp_S$ requires 13 real matrix multiplications($26n^3$ FLOPs) and 2 LU solver on $n \times n$ matrices($4n^3 + 4n^2$ FLOPs). The evaluation of $D \log_S$ requires 17 real matrix multiplications for $34n^3$ FLOPs.

The following tables summarize the complexity discussion. Note that the decomposition computation is usually free from earlier computations in later chapters. The preprocessing computations can be reused in repeated evaluations. Also note that the preprocessing computation for $D(\exp_S)_{\exp(S)}^{-1}[\cdot]$ can reuse the preprocessing computation for $D \exp_S[\cdot]$ if it is available. In the case of reusing the preprocessing computation, the required computation is given in curly brackets.

Table 2.2: Complexity of the Directional Derivative of the Matrix Exponential

Algorithm	Decomposition	Preprocessing	Evaluation
Padé	0	$(12 + 2/3)n^3$	$30n^3 + 4n^2$
Complex	$25n^3 + 9n^2$	$8n^2$	$32n^3 + 3n^2$
Real	$25n^3$	$9n^2$	$8n^3 + 3n^2$

Table 2.3: Complexity of the Directional Derivative of the Nearby Matrix Logarithm

Algorithm	Decomposition	Preprocessing	Evaluation
Padé	$25n^3$	0	$34n^3$
Complex	$25n^3 + 9n^2$	$8n^2 \{0\}^*$	$32n^3 + 5n^2$
Real	$25n^3$	$5n^2 \{n^2\}^*$	$8n^3 + 3n^2$

* : Complexity that reuses the preprocessing from $D \exp_S[\cdot]$.

2.8.2 Experiments

For each formula, the following experiment is performed. Given a dimension n , two random skew symmetric matrices $S, \Delta \in \mathbf{Skew}_n$ are generated and $Q = \exp(S)$ is computed in advance. Then, the formula or its inverse is executed with the full 3-stages, i.e., the **PreEval** stage is executed from scratch. The computed times of the individual stage are recorded and the results are plotted against the dimension n , in **Figure 2.2** and **Figure 2.3**. Besides reporting time for computing the actions $\Delta_Q \mapsto \Delta_S$ and $\Delta_S \mapsto \Delta_Q$, the figures also present the combined time of **Param** and **Action** stages as the best scenario, where no **PreEval** is needed to be computed from scratch, and the combined time of all 3 stages as the worst scenario.

Note that the Padé Algorithm for the inverse action is not included as it is restricted within the principal branch and relies on a system solver that is significantly slower than the other 2 methods.

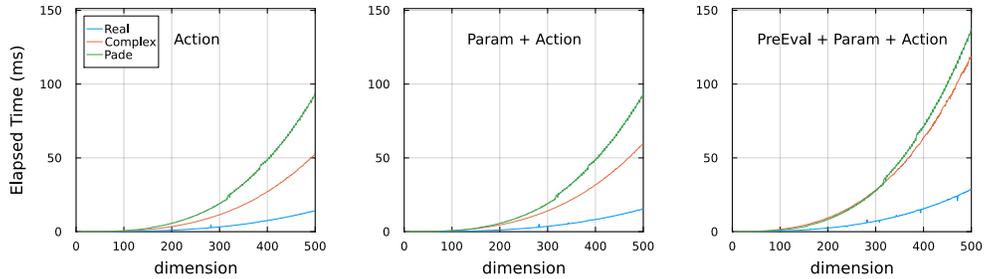


Figure 2.2: Computation Time of $\Delta_S \mapsto \Delta_Q$ in $D \exp_S(\Delta) = Q\Delta_Q$

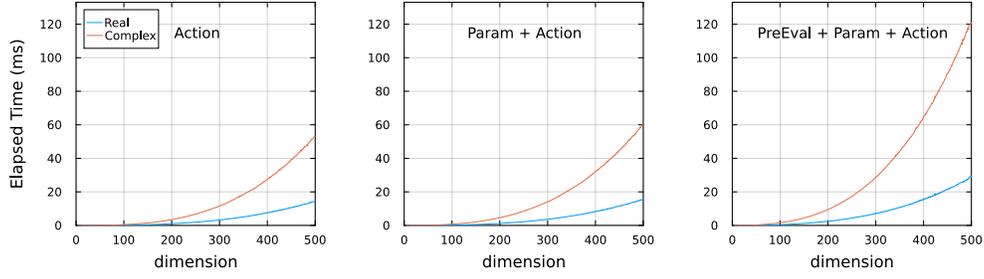


Figure 2.3: Computation Time of $\Delta_Q \mapsto \Delta_S$ in $D \exp_S(\Delta) = Q\Delta_Q$

Overall, the new formulae proposed in computing the linear map $\mathcal{L}_S(\Delta)$ and $\mathcal{L}_S^{-1}(\Delta)$ is 4 ~ 4.5 times as fast as the other formulae. The extra speedup beyond 4 is observed in large dimensions, in which case dense matrix products and complex arithmetics affects more in the real time performances.

CHAPTER 3

LOCAL DIFFEOMORPHISM IN SKEW SYMMETRIC MATRICES

Based on the differential of the matrix exponential on the skew symmetric matrices investigated in **Chapter 2**, this chapter further studies its implications between the special orthogonal group and the skew symmetric matrices via $\exp : \mathbf{Skew}_n \rightarrow \mathbf{SO}_n$ and presents a solution to the smoothly evolving geodesic problem (4.2) proposed in **Chapter 4**. In particular, this chapter first identifies a local diffeomorphism in $\exp : \mathbf{Skew}_n \rightarrow \mathbf{SO}_n$ in which the inverse of the matrix exponential is well-defined, unique and smooth. Then, some intriguing properties of the diffeomorphism are derived and discussed. The notion of the nearby logarithm proposed in [9] is re-interpreted under this diffeomorphism. Finally, two different algorithms are designed to compute the smoothly evolving geodesic problem (4.2).

3.1 Skew Symmetric Matrices with an Invertible Differential

In order to identify a diffeomorphism structure within $\exp : \mathbf{Skew}_n \rightarrow \mathbf{SO}_n$ around some skew symmetric matrix $S \in \mathbf{Skew}_n$, its differential at S restricted to the skew symmetric matrices, $D\exp_S : \mathbf{Skew}_n \rightarrow T_Q\mathbf{SO}_n$ where $Q = \exp(S)$, must be invertible. In other words, S is not on the conjugate locus Conj_{I_n} characterized in **Corollary 2.5.4**, denoted as

$$\mathcal{S} := \mathbf{Skew}_n \setminus \text{Conj}_{I_n} = \{S \in \mathbf{Skew}_n : S \text{ satisfies condition (2.20)}\}. \quad (3.1)$$

Recall that the conjugate locus Conj_{I_n} is described as the union of countable conditions indexed by the integers $k \neq 0$ and each one of them states that there exists a pair of angles $(\theta_i, \theta_j), i \neq j = 1, 2, \dots, m$ such that $\theta_i \pm \theta_j = 2k\pi$. Therefore, the skew symmetric S satisfying the k -indexed condition consists of the following two or three subsets

$$\begin{aligned} \mathcal{A}_{k,+} \cup \mathcal{A}_{k,-} &:= \{S \in \mathbf{Skew}_n : \exists \theta_i + \theta_j = 2k\pi\} \cup \{S \in \mathbf{Skew}_n : \exists \theta_i - \theta_j = 2k\pi\} \\ \mathcal{A}_{k,*} &:= \{S \in \mathbf{Skew}_n : \exists \theta_i = 2k\pi\} \text{ additional for } n = 2m + 1 \end{aligned} \quad (3.2)$$

such that

$$\begin{aligned} \text{Conj}_{I_n} &= \bigcup_{k=\pm 1, \pm 2, \dots} (\mathcal{A}_{k,+} \cup \mathcal{A}_{k,-}) \\ \text{or} &= \bigcup_{k=\pm 1, \pm 2, \dots} (\mathcal{A}_{k,+} \cup \mathcal{A}_{k,-} \cup \mathcal{A}_{k,\pi}) \text{ for } n = 2m + 1. \end{aligned}$$

It follows that $\mathcal{A}_{k,+}$, $\mathcal{A}_{k,-}$ and $\mathcal{A}_{k,*}$ are closed and connected subsets in \mathbf{Skew}_n . Therefore, the set \mathcal{S} is constructed by removing countably many closed subsets from \mathbf{Skew}_n . It immediately leads to the following conclusion on the subset structure of \mathcal{S} .

Proposition 3.1.1. *The set of skew symmetric matrices \mathcal{S} with an invertible differential to the matrix exponential is a collection of countable open and connected subsets denoted as \mathcal{S}_e , where $e \in \mathcal{E}$ belongs to a countable indices set. In other words,*

$$\mathcal{S} = \bigcup_{e \in \mathcal{E}} \mathcal{S}_e. \quad (3.3)$$

Proof. Simply notice that the closed subsets defined in (3.2) are closed subsets in the vector space \mathbf{Skew}_n . Removing countably many closed and connected subsets from a vector space results in countably many open and connected subsets. In this case, they are $\mathcal{S}_e, e \in \mathcal{E}$ where the countable index set \mathcal{E} is determined by the integer $k = \pm 1, \pm 2, \dots$ and the plus-minus-star signs that label the closed subsets (3.2) (but not themselves, i.e., $\mathcal{E} \neq (\{\pm 1, \dots\}, \{+, -, *\})$). \square

Definition 3.1.2. In particular, the $0 \in \mathcal{I}$ indexed subset \mathcal{S}_0 is reserved for the open subset that consists of the zero matrix $\mathbf{0} \in \mathbf{Skew}_n$, i.e.,

$$\mathcal{S}_0 := \{S \in \mathbf{Skew}_n : \forall i \neq j, \theta_i + \theta_j < 2\pi\}. \quad (3.4)$$

\square

For the case $n = 4$, there are only 2 angles of a skew symmetric matrix S , denote them as θ_1 and θ_2 . **Figure 3.1** illustrates the angles of the matrix at the conjugate locus in red. Note that each red dashed line corresponds to a closed subset in $\text{Conj}_{I_n} \subset \mathbf{Skew}_4$. The 4 dashed lines correspond to the 4 closed subsets $\mathcal{A}_{1,+}$, $\mathcal{A}_{1,-}$, $\mathcal{A}_{-1,+}$ and $\mathcal{A}_{-1,-}$. The interior red box in the middle is the 0-label subset \mathcal{S}_0 .

The skew symmetric matrix in the principal branch (2.1) has its angles bounded by $-\pi$ and π . The boundary of the principal branch is plotted as the black dashed line and the interior black box is the principal branch. Notice that the principal branch is fully contained in \mathcal{S}_0 and this observation applies to more general $n > 4$, as $\theta_i \pm \theta_j < 2\pi, \forall i \neq j = 1, \dots, m$ easily concludes $|\theta_i| < \pi, \forall i = 1, \dots, m$.

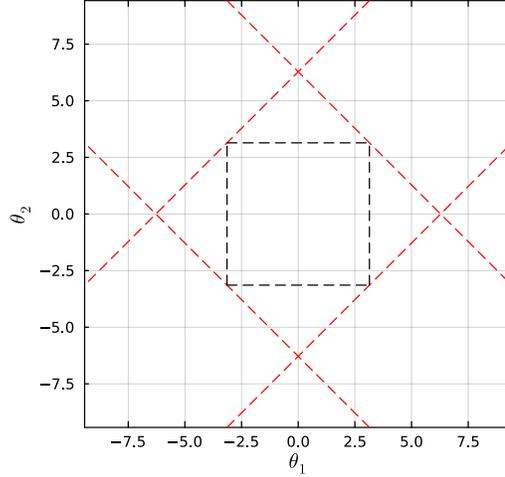


Figure 3.1: Illustrations of the conjugate locus and the principal branch in \mathbf{Skew}_4

3.2 Preimage of Exponential at Special Orthogonal Matrices

With the conjugate locus Conj_{J_n} fully characterized on \mathbf{Skew}_n , it leads to the next question characterizing the distribution on the preimage of the matrix exponential $\{X \in \mathbf{Skew}_n : \exp(X) = Q \in \mathbf{SO}_n\}$, especially for the $Q = \exp(S)$ with $S \in \text{Conj}_{J_n}$.

Consider a special orthogonal matrix Y that has a set of principal angles $\Theta \in [0, \pi]^m$ specified in the preferred Schur decomposition

$$Y = RER^T = R \text{diag}(E_{[1,1]}, \dots) R^T = R \text{diag}_{E_{a_1, b_1}, \dots, E_{a_r, b_r}}$$

as in (2.7). Let

$$X = RDR^T = R \text{diag}(D_{[1,i_1]}, \dots) R^T = R \text{diag}(D_{0, c_1}, \dots, D_{0, c_r}) R^T$$

be the skew symmetric matrix constructed from the angles Θ and Schur vectors R from Y . Recall that the notation $E_{[i,i]}$ stands for 2×2 blocks, and an additional 1 when $n = 2m + 1$, in the diagonal. The notation E_{a_r, b_r} stands for the repeated diagonal block $E_{a_j, b_j} = \begin{bmatrix} a_j & -b_j \\ b_j & a_j \end{bmatrix} \otimes I_{k_j}$ with multiplicity k_j , and an additional $E_{1,0} = \begin{bmatrix} I_2 \otimes I_{k_r-1} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}$ when $n = 2m + 1$. Similar structures follow for $D_{[i,i]}$ and D_{0, c_j} . The constructed skew symmetric $X = RDR^T$ serves as a reference skew symmetric matrix that characterizes the preimage

$$\mathbb{E}_Y^{-1} := \{X \in \mathbf{Skew}_n : \exp(X) = Y\}. \quad (3.5)$$

Lemma 3.2.1. For any $Y \in \mathbf{SO}_n$ with a preferred Schur decomposition $Y = RER^T$, with the principal angles $\Theta = \{\theta_1, \dots, \theta_m\} \in [0, \pi]$. Then the skew symmetric matrix $X = RDR^T$ with $D = \text{diag}\left(\begin{bmatrix} 0 & -\theta_1 \\ \theta_1 & 0 \end{bmatrix}, \dots\right)$ is a preferred Schur decomposition and it satisfies $\exp(X) = Y$.

Proof. By construction, there is

$$\exp(D_{[i,i]}) = \exp\left(\begin{bmatrix} 0 & -\theta_i \\ \theta_i & 0 \end{bmatrix}\right) = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i) \end{bmatrix} = E_{[i,i]}$$

such that $\exp(X) = R\exp(D)R^T = RER^T = Y$ follows immediately. To see that RDR^T is a preferred Schur decomposition, notice that $\theta_i \geq 0$ is satisfied as they are principal angles chosen to be nonnegative. For the ordering in cosine values, there is

$$\cos(\theta) \leq \cos(\omega) \iff \theta \geq \omega, \forall \theta, \omega \in [0, \pi],$$

as $\cos(x), x \in [0, \pi]$ is monotonically decreasing. Then, $\theta_i \geq \theta_{i+1}, i = 1, \dots, m-1$ follows. \square

Notice that if the principal angles Θ in Y are bounded by π , i.e., $0 < \theta_i < \pi, \forall i = 1, \dots, m$, the $X = RDR^T$ constructed in **Lemma 3.2.1** is the classic principal logarithm that is uniquely defined within the principal branch $\{X \in \mathbf{Skew}_n : \|X\|_2 < \pi\}$, as

$$\|X\|_2 = \|RDR^T\|_2 = \|D\|_2 = \max_{i=1, \dots, m} (|\theta_i|) < \pi.$$

Lemma 3.2.1 not only relates the special orthogonal $Y = RER^T$ with a skew symmetric $X = RDR^T$, but also implies that the repeated block diagonal structure in Q , if any, is preserved in the X constructed from the principal angles. This is not true for $S \in \mathbb{E}_Y^{-1}$ in general is discussed later in this section. The next **Lemma 3.2.2** states the implications of a shared repeated block diagonal structure in terms of the set of possible Schur vectors.

Lemma 3.2.2. For any two Y, Z skew symmetric or special orthogonal matrices with the corresponding preferred Schur decompositions $Y = RMR^T$ and $Z = RNR^T$, let M and N share the same block diagonal structures and repeated pattern given by

$$\begin{cases} N = \text{diag}(N_{a_1, b_1}, \dots, N_{a_r, b_r}), & N_{a_j, b_j} = \begin{bmatrix} a_j & -b_j \\ b_j & a_j \end{bmatrix} \otimes I_{k_j} \\ M = \text{diag}(M_{c_1, d_1}, \dots, M_{c_r, d_r}), & M_{c_j, d_j} = \begin{bmatrix} c_j & -d_j \\ d_j & c_j \end{bmatrix} \otimes I_{k_j} \end{cases}$$

where k_1, \dots, k_r is the multiplicity of the repeated diagonal block that satisfies $\sum_{j=1}^r k_j = 2m$. Then Y and Z share the same set of possible Schur vectors.

Proof. From **Proposition 2.2.3**, the set of all possible Schur vectors is $\{\tilde{R} = RQG\}$, generated by orthogonal transformations Q, G and P . The G term and the P term only depends on the dimension $n = 2m$ or $n = 2m + 1$. The Q term depends on the multiplicity and the position of repeated blocks, which are both identical in M and N . Therefore, the same sets of Q, G and P generates the same sets of Schur vectors of Y and Z . \square

According to **Lemma 3.2.2**, the $X = RDR^T$ constructed by the principal angles Θ of $Y = RER^T$ shares the same set of possible Schur vectors, i.e., for any given set of Schur vector \tilde{R} of Y , it is also a Schur vector of X , i.e., $\tilde{D} := \tilde{R}^T X \tilde{R}$ must be in forms of a block diagonal structure. This insight is further exploited in the following **Theorem** to characterize the preimage \mathbb{E}_Y^{-1} .

Theorem 3.2.3. *For any $Y \in \mathbf{SO}_n$ with a preferred Schur decomposition RER^T and the principal angles $\Theta \in [0, \pi]^m$, the Schur decomposition on the skew symmetric matrix $S = RD^\Theta R^T$ is a preferred Schur decomposition and S satisfies $\exp(S) = Y$. Furthermore, the preimage at Y is given by*

$$\{X \in \mathbf{Skew}_n : \exp(X) = Y\} = \left\{ \tilde{R} \tilde{D} \tilde{R}^T : \begin{array}{l} \tilde{R} = RQG, \tilde{D} = \text{diag}(\tilde{D}_{[1,1]}, \dots) \\ \tilde{D}_{[i,i]} = \begin{bmatrix} 0 & -\theta_i - 2k_i\pi \\ \theta_i + 2k_i\pi & 0 \end{bmatrix}, k_i \in \mathbb{Z} \end{array} \right\} \quad (3.6)$$

where $Q = \text{diag}(Q_{\cos(\theta_1), \sin(\theta_1)}, \dots, Q_{0,1})$ and $G = \text{diag}(G_{[1,1]}, \dots)$ with the additional condition $\det(G_{[i,i]}) = 1, i = 1, \dots, m$ are orthogonal transformations specified in **Proposition 2.2.3** applied to the preferred Schur decomposition $Y = RER^T$.

Proof. Let $X = RDR^T$ be a preferred Schur decomposition on the skew symmetric matrix X as constructed in **Lemma 3.2.1**. By **Lemma 3.2.2**, both X and Y share the same set of possible Schur vectors.

(RHS \implies LHS): Notice that $\det(G_{[i,i]}) = 1$, which yields $G_{[i,i]} \tilde{D}_{[i,i]} G_{[i,i]}^T = \tilde{D}_{[i,i]}, i = 1, \dots, m$. For $Q = \text{diag}(Q_{a_1, b_1}, \dots, Q_{a_r, b_r})$ where $Q_{a_j, b_j} = I_2 \otimes Q_{k_j}$ with any $k_j \times k_j$ orthogonal matrix acting on the j -th repeated diagonal block in E with the multiplicity of k_j . The j -th repeated diagonal block is expressed as $E_{a_j, b_j} = \begin{bmatrix} a_j & -b_j \\ b_j & a_j \end{bmatrix} \otimes I_{k_j}$. Simple algebra shows that $Q_{a_j, b_j} E_{a_j, b_j} Q_{a_j, b_j}^T = E_{a_j, b_j}$ as discussed in **Proposition 2.2.3**. Then, it can be concluded that ‘‘RHS \implies LHS’’

$$\begin{aligned} \exp(RQG\tilde{D}G^T Q^T R^T) &= \exp(RQ\tilde{D}Q^T R) = RQ \exp(\tilde{D}) Q^T R \\ &= RQ \text{diag}(\exp(\tilde{D}_{[1,1]}), \dots) Q^T R^T = RQ \text{diag}(E_{[1,1]}, \dots) Q^T R^T \\ &= RQE Q^T R^T = R \text{diag}(Q_{a_1, b_1} E_{a_1, b_1} Q_{a_1, b_1}^T, \dots) R^T \\ &= R \text{diag}(E_{a_1, b_1}, \dots) R^T = RER^T = Y \end{aligned}$$

(LHS \implies RHS): Let $S = R_S M R_S^T$ be a solution to $\exp(X) = Y$ with the Schur vectors R_S . Then R_S is also a Schur vector of Y as

$$Y = R_S \exp(M) R_S^T = R_S \operatorname{diag}(\exp(M_{[1,1]}), \dots) R_S^T.$$

By **Proposition 2.2.3** applied to Y , there exists orthogonal transformations $Q = I_n$, G and R such that

$$\tilde{R}_S (P^T G^T \operatorname{diag}(\exp(M_{[1,1]}), \dots) G P) \tilde{R}_S^T = \tilde{R}_S E \tilde{R}_S^T.$$

By **Proposition 2.2.3** applied to S , \tilde{R}_S remains a set of Schur vectors of S , i.e., $S = \tilde{R}_S \tilde{M} \tilde{R}_S$ is a Schur decomposition. Then, it remains to verify that the new Schur decomposition falls in (3.6).

Let $\tilde{E} = \operatorname{diag}(\tilde{E}_{[1,1]}, \dots)$ with $\tilde{E}_{[i,i]} := G_{[i,i]} \exp(M_{[i,i]}) G_{[i,i]}^T$, then the

$$P \tilde{E} P^T = \operatorname{diag}(\tilde{E}_{[i_1, i_1]}, \dots)$$

where P permutes i_1, \dots, i_m to $1, 2, \dots, m$. Since the solutions to

$$\tilde{E}_{[i_j, i_j]} = \exp \left(\begin{bmatrix} 0 & -\tilde{\theta}_{i_j} \\ \tilde{\theta}_{i_j} & 0 \end{bmatrix} \right) = \begin{bmatrix} \cos(\theta_j) & -\sin(\theta_j) \\ \sin(\theta_j) & \cos(\theta_j) \end{bmatrix}$$

is given by $\{\tilde{\theta}_{i_j} = \theta_j + 2k_j\pi : k_j \in \mathbb{Z}\}$, there is $M_{[i_j, i_j]} = \begin{bmatrix} 0 & -\theta_j - 2k_j\pi \\ \theta_j + 2k_j\pi & 0 \end{bmatrix}$. Then, it is ready to conclude ‘‘LHS \implies RHS’’ as

$$\begin{aligned} R_S M R_S^T &= R_S G M G^T R_S = R G P P^T M P P^T G^T R_S \\ &= \tilde{R}_S P^T \operatorname{diag}(M_{[1,1]}, \dots) P \tilde{R}_S = \tilde{R}_S \operatorname{diag}(M_{[i_1, i_1]}, \dots) \tilde{R}_S \\ &= \tilde{R}_S \begin{bmatrix} \begin{bmatrix} 0 & -\theta_1 - 2k_1\pi \\ \theta_1 + 2k_1\pi & 0 \end{bmatrix} & \mathbf{0} & \dots \\ \mathbf{0} & \begin{bmatrix} 0 & -\theta_2 - 2k_2\pi \\ \theta_2 + 2k_2\pi & 0 \end{bmatrix} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \tilde{R}_S^T \end{aligned}$$

where the first equality follows from the assumption $\det(G_{[i,i]}) = 1$ and the fourth equality follows from the fact that P^T permutes $1, \dots, m$ back to i_1, \dots, i_m by construction. \square

Computationally speaking, it is important to note that the extra condition $\det(G_{[i,i]}) = 1$ introduced in **Proposition 4.5.4** provides a convenient yet sufficient characterization based on the principal angles $\Theta \in [0, \pi]^m$, but it also suppresses the equivalent matrix expression of the Schur decompositions with $\tilde{R}_{[i]} = R_{[i]} G_{[i,i]}$, $k_i \in \mathbb{Z}$ where $\det(\tilde{G}_{[i,i]}) = -1$. Although the suppressed expression is equivalent to $G_{[i,i]} := \tilde{G}_{[i,i]} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $-k_i \in \mathbb{Z}$, i.e.,

$$\tilde{R}_{[i]} \begin{bmatrix} 0 & \theta_i - 2k_i\pi \\ -\theta_i + 2k_i\pi & 0 \end{bmatrix} \tilde{R}_{[i]}^T = \tilde{R}_{[i]} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -\theta_i + 2k_i\pi \\ \theta_i - 2k_i\pi & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tilde{R}_{[i]}^T,$$

the Schur decomposition on skew symmetric matrices computed from the standard library does not appear in the characterization chosen in **Theorem 3.2.3**. Therefore, extra considerations are needed to be taken care of if there are algorithmic demands on the form of Schur decompositions.

There are also two intriguing implications from the characterization (3.6) on particular skew symmetric matrices that sheds some light on the geometric insight of $\exp : \mathbf{Skew}_n \rightarrow \mathbf{SO}_n$. The first one justifies the cut locus Cut_{I_n} on \mathbf{SO}_n being $\{X \in \mathbf{Skew}_n : \|X\|_2 = \pi\}$ and the second one relates the skew symmetric matrices on the conjugate locus to skew symmetric matrices with repeated diagonal blocks structure.

Corollary 3.2.4. For any skew symmetric matrix $S \in \mathbf{Skew}_n$ on the cut locus $\text{Cut}_{I_n} = \{X \in \mathbf{Skew}_n : \|X\|_2 = \pi\}$, there exists a different $X \neq S$ also on the cut locus such that $\exp(S) = \exp(X)$ and $\|X\|_2 = \|S\|_2$.

Proof. Let $S = RDR^T$ be a preferred Schur decomposition with angles $\theta_i \geq \theta_{i+1}$. Since

$$\pi = \|S\|_2 = \|RDR^T\|_2 = \|D\|_2 = \max_{i=1, \dots, m} \left\| \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\|_2 = \max_{i=1, \dots, m} |\theta_i|,$$

there are $\theta_1 = \pi$ and $0 \geq \theta_i \leq \pi, i = 2, \dots, m$. Write S as the following

$$S = R_{[1]} \begin{bmatrix} 0 & -\pi \\ \pi & 0 \end{bmatrix} R_{[1]}^T + \sum_{i=2}^m R_{[i]} D_{[i,i]} R_{[i]}^T.$$

Let $Q = \exp(S) = RER^T$. By construction, Θ is the principal angles of Q . Therefore the preimage \mathbb{E}_Q^{-1} is characterized by (3.6) with S . Consider the X with -2π shift in the first block as

$$\mathbb{E}_Q^{-1} \ni X = R_{[1]} \begin{bmatrix} 0 & \pi \\ -\pi & 0 \end{bmatrix} R_{[1]}^T + \sum_{i=2}^m R_{[i]} D_{[i,i]} R_{[i]}^T$$

It is easy to see that $X \neq S$ as $S - X = R_{[1]} \begin{bmatrix} 0 & -2\pi \\ 2\pi & 0 \end{bmatrix} R_{[1]}^T \neq \mathbf{0}$. Then, there are two geodesics $\exp(t \cdot X)$ and $\exp(t \cdot S)$ arriving at Q from I_n with the same distances

$$\begin{cases} \|X\|_F = \|R_X D_X R_X^T\|_F = \|D_X\|_F = \sqrt{2(-\pi)^2 + 2 \sum_{i=2}^m \theta_i^2} \\ \|S\|_F = \|R_S D_S R_S^T\|_F = \|D_S\|_F = \sqrt{2\pi^2 + 2 \sum_{i=2}^m \theta_i^2} \end{cases}.$$

□

Corollary 3.2.5. A special orthogonal Q has repeated eigenvalues if and only if

$$\mathbb{E}_Q^{-1} \cap \text{Conj}_{I_n} \neq \emptyset,$$

i.e., there exists $S \in \text{Conj}_{I_n}$ such that $\exp(S) = Q$.

Furthermore, the repeated eigenvalues are characterized by a repeated block diagonal structure in its preferred Schur decomposition $Q = RER^T = R \text{diag}(E_{a_1, b_1}, \dots, E_{a_r, b_r})R^T$ where there is at least one diagonal block E_{a_j, b_j} larger than 2×2 . Such an E_{a_j, b_j} is characterized by the angles Ω in S . If there are k angles that violate (2.19) with $\omega = \omega_i$ as $i \neq i_1 \neq i_2 \neq \dots \neq i_k$ such that $\omega \pm \omega_{i_j} = 2k_j\pi$, then

$$E_{\cos(\omega), |\sin(\omega)|} = I_k \otimes \begin{bmatrix} \cos(\omega) & -|\sin(\omega)| \\ |\sin(\omega)| & \cos(\omega) \end{bmatrix}$$

or with $i_1 \neq i_2 \neq \dots \neq i_k$ such that $\omega_{i_j} = \pm 2k_j\pi$ and $n = 2m + 1$, then $E_{1,0} = I_{2k+1}$.

Proof. When $\omega_i = 2k\pi$ and $n = 2m + 1$, there is $\exp\left(\begin{bmatrix} 0 & -\omega_i \\ \omega_i & 0 \end{bmatrix}\right) = I_2$, which means a preferred Schur decomposition $RER^T = \exp(S)$ must have $E_{1,0}$ with dimension 3×3 at least. The corresponding canonical angle is 0 and the statement follows.

When $\omega_i \pm \omega_j = 2k\pi$, observe that $\omega_i \pm \omega_j = 2k\pi \Leftrightarrow \omega_i = (-\omega_j \pm 2k\pi)$. For $S = RD^\Omega R^T$ constructed by the angles ω_i , let $Q = \text{diag}\left(I_{2*(j-1)}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, I_{n-2j}\right)$ and $\mathcal{R} = RQ$ such that the j -th diagonal is flipped as

$$S = RD^\Omega R^T = \tilde{R} \text{diag}\left(\begin{bmatrix} 0 & -\omega_1 \\ \omega_1 & 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 & \omega_j \\ -\omega_j & 0 \end{bmatrix}, \dots\right) \tilde{R}^T.$$

Then, shift the j -th block with $\pm 2k\pi$ to get

$$\begin{aligned} X &= \tilde{R} \text{diag}\left(\dots, \begin{bmatrix} 0 & \omega_j \mp 2k\pi \\ -\omega_j \pm 2k\pi & 0 \end{bmatrix}, \dots\right) \tilde{R}^T \\ &= \tilde{R} \text{diag}\left(\dots, \begin{bmatrix} 0 & \omega_i \\ \omega_i & 0 \end{bmatrix}, \dots\right) \tilde{R}^T \end{aligned}$$

i.e., X has repeated diagonal blocks in i, j positions. Then, it is easy to verify $\exp(S) = \exp(X)$. Let the exponential of the repeated block be $E_{\cos(\omega_i), \sin(\omega_i)}$, convert it to a preferred Schur decomposition as $E_{\cos(\omega_i), |\sin(\omega_i)|}$ to find the corresponding (repeated) principal angles $\theta \in [0, \pi]$ satisfying $\cos(\theta) = \cos(\omega_i)$ and $\sin(\theta) = |\sin(\omega_i)|$.

The above discussion on constructing repeated diagonal blocks from the conjugate locus can be reversed to construct a skew symmetric matrix on the conjugate locus from a repeated angles. Therefore, the ‘‘if-and-only-if’’ statement follows. □

Corollary 3.2.5 reveals the relation between the set of special orthogonal matrices with repeated eigenvalues and the conjugate locus. Such a set is denoted as

$$\mathbb{Q} := \{\exp(S) : S \in \text{Conj}_{I_n}\} = \{Q \in \mathbf{SO}_n : Q \text{ with repeated eigenvalues}\}. \quad (3.7)$$

Note that having $Q \in \mathbb{Q}$ only implies there exists a $S \in \text{Conj}_{I_n}$ satisfying $\exp(S) = Q$, but not necessarily suggesting $\mathbb{E}_Q^{-1} \subset \text{Conj}_{I_n}$. In other words, there may be $X \in \mathbf{Skew}_n$ with $\exp(X) = Q \in \mathbb{Q}$ and an invertible $D \exp_X : \mathbf{Skew}_n \rightarrow T_Q \mathbf{SO}_n$. With the set \mathbb{Q} identified, the following theorem concludes an essential feature of the preimage $\mathbb{E}_Q^{-1}, \forall Q \notin \mathbb{Q}$.

Theorem 3.2.6. *For any $Y \in \mathbf{SO}_n \setminus \mathbb{Q}$, the preimage $\{X \in \mathbf{Skew}_n : \exp(X) = Y\}$ is an isolated set with points separated by a distance of at least 2π in matrix 2-norm, i.e.,*

$$\|A - B\|_2 > 2\pi, \forall A \neq B \in \mathbf{Skew}_n, \exp(A) = \exp(B) \quad (3.8)$$

Proof. Let $X = RDR^T$ be the reference skew symmetric as constructed in **Lemma 3.2.1** with the $Q = RER^T$ such that both A and B can be expressed as a shift from X in forms of (3.6) with $\xi, \eta \in \mathbb{Z}^m$, Q_A, G_A and Q_B, G_B such that

$$\begin{cases} A = \tilde{R}_A D^\xi \tilde{R}_A^T, & \tilde{R}_A = RQ_A G_A, D^\xi = \text{diag} \left(\begin{bmatrix} 0 & -\theta_1 - 2\xi_1\pi \\ \theta_1 + 2\xi_1\pi & 0 \end{bmatrix}, \dots \right) \\ B = \tilde{R}_B D^\eta \tilde{R}_B^T, & \tilde{R}_B = RQ_B G_B, D^\eta = \text{diag} \left(\begin{bmatrix} 0 & -\theta_1 - 2\eta_1\pi \\ \theta_1 + 2\eta_1\pi & 0 \end{bmatrix}, \dots \right) \end{cases}.$$

Since $Y \notin \mathbb{Q}$, there is no repeated diagonal block in E , the Q_A and Q_B are both the identity matrix I_n , which yields

$$\begin{aligned} \|A - B\|_2 &= \|RG_A D^\xi G_A^T R - RG_B D^\eta G_B^T R\|_2 = \|G_A D^\xi G_A^T - G_B D^\eta G_B^T\|_2 \\ &= \|D^\xi - D^\eta\|_2 = \max_{i=1, \dots, m} \|D_{[i,i]}^\xi - D_{[i,i]}^\eta\|_2 \\ &= \max_{i=1, \dots, m} |2(\xi_i - \eta_i)\pi| \geq 2\pi \end{aligned}$$

where the third equality follows from $\det(G_{A,[i,i]}) = 1$ and $\det(G_{B,[i,i]}) = 1$, the fourth equality follows from the 2-norm on block diagonal matrix and the fifth equality follows from the fact that $A \neq B \Leftrightarrow \xi \neq \eta$. \square

The condition $Y \notin \mathbb{Q}$ is necessary for **Theorem 3.2.6**. Otherwise, the following proposition finds a connected set of skew symmetric matrices that have the image under the exponential equals to the same Y with the repeated diagonal blocks.

Proposition 3.2.7. Consider $Y = RER^T \in \mathbb{Q}$ with principal angles Θ and $X = RDR^T \in \mathbf{Skew}_n$ be constructed by the principal angles, with $\theta_j = \theta_{j+1}$. Let $S = RD^\xi R^T \in \text{Conj}_{I_n}$ with shifts $\xi \in \mathbb{Z}^n$ in multiples of 2π , then for any $\omega \in [0, 2\pi]$ and $Q(\omega) := \begin{bmatrix} \cos(\omega) & -\sin(\omega) \\ \sin(\omega) & \cos(\omega) \end{bmatrix} \otimes I_2$ the set

$$\{S(\omega) : \omega \in [0, 2\pi]\} \subset \mathbb{E}_Y^{-1}$$

and $\|S(\omega) - S(0)\|_2$ ranges in $[0, 2|\xi_j - \xi_{j+1}|\pi]$ continuously where

$$S(\omega) := \sum_{i \neq j, j+1} R_{[i]} D_{[i,i]}^\xi R_{[i]}^T + \begin{bmatrix} R_{[j]} & R_{[j+1]} \end{bmatrix} Q(\omega) \begin{bmatrix} D_{[j,j]}^\xi & \mathbf{0} \\ \mathbf{0} & D_{[j+1,j+1]}^\xi \end{bmatrix} Q(\omega)^T \begin{bmatrix} R_{[j]}^T \\ R_{[j+1]}^T \end{bmatrix}.$$

Proof. This follows from the Gershgorin circle theorem. \square

3.3 Diffeomorphism Structure in Skew Symmetric Matrices

Based on the structures exploited on the connected subsets $\mathcal{S}_e, e \in \mathcal{E}$ and their boundaries Conj_{I_n} , it is now possible present the local diffeomorphism in the matrix exponential $\exp : \mathbf{Skew}_n \rightarrow \mathbf{SO}_n$ as follow.

Proposition 3.3.1. Given any $S \in \mathcal{S}_e, \forall e \in \mathcal{E}$, let $Q = \exp(S)$. There exists a small enough neighborhood of S denoted as $\mathcal{M}_S \subset \mathcal{S}_e$ with its image denoted as

$$\mathcal{N}_Q := \{Y = \exp(X) : X \in \mathcal{M}_S\},$$

such that the restricted matrix exponential $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q$ is a diffeomorphism.

Proof. Since the $\mathcal{S}_e, \forall e \in \mathcal{E}$ is an open and connected subset, it is always possible to build any small enough neighborhood $\mathcal{M}_S \subset \mathcal{S}_e$.

For such a \mathcal{M}_S , any $X \in \mathcal{M}_S$, including the $X = S$, has an invertible differential $D\exp_X : \mathbf{Skew}_n \rightarrow T_{\exp(X)}\mathbf{SO}_n$. By the inverse function theorem, there exists a small enough neighborhood of \mathcal{M}_S on which the matrix exponential is invertible.

Then, let that \mathcal{M}_S be this smaller subset, on which $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q$ is a bijection. Furthermore, the differential $D\exp_X : \mathbf{Skew}_n \rightarrow T_Y\mathbf{SO}_n, \forall X \in \mathcal{M}_S, Y = \exp(X) \in \mathcal{N}_Q$ is invertible, i.e., the inverse $\exp^{-1} : \mathcal{N}_Q \rightarrow \mathcal{M}_S$ is differentiable for any $Y \in \mathcal{N}_Q$. It concludes with the definition that states a bijection with an invertible differential is a diffeomorphism. \square

3.3.1 Sufficient Condition of Constructing Diffeomorphism

The locality of the diffeomorphism within $\mathcal{S}_e, \forall e \in \mathcal{E}$ cannot be extended to the entire \mathcal{S}_e , as the exponential $\exp : \mathcal{S}_e \rightarrow \mathbf{SO}_n$ may not be a bijection. To see that, simply notice that \mathcal{S}_0 contains $S \in \mathbf{Skew}_n$ with an angle $\theta_i \in (\pi, 2\pi)$. For example, in the illustration of \mathbf{Skew}_4 in **Figure 3.1**, the region within the red box but outside the black box contains such an angle greater than π . Replace θ_i in S with $\theta'_i = \theta_i - 2\pi \in (-\pi, 0)$ to obtain S' that still lies within \mathcal{S}_0 . In the illustration of \mathbf{Skew}_4 , S' lies within the black box. It is easy to see that $\exp(S) = \exp(S')$, i.e., $\exp : \mathcal{S}_0 \rightarrow \mathbf{SO}_n$ is not a bijection, which implies $\exp : \mathcal{S}_0 \rightarrow \mathbf{SO}_n$ is not a diffeomorphism.

Fortunately, the diffeomorphism within $\mathcal{S}_e, \forall e \in \mathcal{E}$ is lost solely due to the loss of bijection structure, as for any $X \in \mathcal{S}_e, \forall e \in \mathcal{E}$, the differential is always invertible. Combined with the fact that the preimage of the matrix exponential is a set of isolated points, a stronger statement on the diffeomorphism is proposed in below.

Proposition 3.3.2. *Let \mathcal{M}_S be an open neighborhood of $S \in \mathcal{S}_e$ for $e \in \mathcal{E}$ and let \mathcal{N}_Q be its image under the matrix exponential. Then the matrix exponential $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q$ is a diffeomorphism if for $\forall A, B \in \mathcal{M}_S, \|A - B\|_2 < 2\pi$.*

Proof. The matrix exponential can only fail its one-to-one nature in \mathcal{M}_S if there are at least 2 points $A \neq B$ from a preimage of some $Y \in \mathcal{N}_Q$ lie within \mathcal{M}_S . By construction, $Y \notin \mathbb{Q}$ which yields that the two points are separated in at least 2π distance, i.e., $\|A - B\|_2 \geq 2\pi$ as stated in **Theorem 3.2.6**. This statement contradicts the imposed condition. Therefore, the matrix exponential is a bijection on such a \mathcal{M}_S .

On the other hand, the $\mathcal{M}_S \subset \mathcal{S}_e$ by construction has all differential $D \exp_X, \forall X \in \mathcal{M}_S$ invertible, the diffeomorphism of $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q$ follows. \square

3.3.2 Diffeomorphism on an Inscribed Ball

The **Proposition 3.3.2** provides a sufficient condition to construct a diffeomorphism in $\mathcal{S}_e, \forall e \in \mathcal{E}$. The following proposition further shows that for any $S \in \mathcal{S}_e$, an inscribed ball of S , that is tangential to the boundary of \mathcal{S}_e , satisfies the proposed sufficient condition. This provides a practical way to identify a local diffeomorphism.

Proposition 3.3.3. *For any skew symmetric $S \in \mathcal{S}_e, \forall e \in \mathcal{E}$ with Schur vectors R and angles Θ , its distance under 2-norm to the conjugate locus, which is also its distance to the boundary of \mathcal{S}_e ,*

is fully determined by its angles Θ , denoted as δ_Θ . It is given by

$$2\delta_\Theta := \begin{cases} \min_{i \neq j=1, \dots, m, k \neq 0} \{|\theta_i + \theta_j - 2k\pi|, |\theta_i - \theta_j - 2k\pi|\}, & n = 2m \\ \min_{i \neq j=1, \dots, m, k \neq 0} \{|\theta_i + \theta_j - 2k\pi|, |\theta_i - \theta_j - 2k\pi|, |\theta_j - 2k\pi|\}, & n = 2m + 1 \end{cases}. \quad (3.9)$$

Furthermore, this distance is bounded from above as

$$\delta_\Theta \leq \begin{cases} \pi, & S \in \mathcal{S}_0 \\ \pi/2, & S \notin \mathcal{S}_0 \end{cases}. \quad (3.10)$$

Proof. Recall that a distance from a point to a subset is the infimum among the distance from that point to any point on the subset. In this case, the given point is $S \in \mathcal{S}_e$ and the subset is Conj_{J_n} . Since the Conj_{J_n} is a closed set that consists of countably many closed subsets, the distance from S to Conj_{J_n} becomes the infimum among the distances from S to all closed subsets. Then, one can characterize the distance from S to the closed subset $\mathcal{A}_{k,+}$

$$d_{k,+} := \text{dist}(S, \mathcal{A}_{k,+}).$$

Since $\mathcal{A}_{k,+}$ is a closed subset in \mathbf{Skew}_n , there must be a $X_{k,+} \in \mathcal{A}_{k,+}$ that realizes the distance, i.e., $\text{dist}(S, X_{k,+}) = \|S - X_{k,+}\|_2 = \text{dist}(S, \mathcal{A}_{k,+}) = d_{k,+}$. It is easy to find such a $X_{k,+}$ as $X_{k,+} = RD'R^T$ where R is the Schur vectors of S , D' consist of the same angles of S except the i, j -th angles that realize the minimum

$$\arg \min_{i \neq j=1, 2, \dots, m} |\theta_i + \theta_j - 2k\pi|.$$

Let ε be the difference $2k\pi - \theta_i - \theta_j$ that realized the minimum in magnitude and set the corresponding i, j -th angles as $\theta'_i = \theta_i + \varepsilon/2$ and $\theta'_j = \theta_j + \varepsilon/2$.

By construction, the resulting $X = RD'R^T$ has $\theta'_i + \theta'_j = 2k\pi$, i.e., it is on $\mathcal{A}_{k,\pm}$. Since $X = RD'R^T$ and $S = RDR^T$ share similar structures in the Schur decomposition, their distance under matrix 2-norm is given by

$$\begin{aligned} \|X - S\|_2 &= \|R(D' - D)R^T\|_2 = \|D' - D\|_2 \\ &= \left\| \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} & \cdots \\ \vdots & \ddots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \cdots & \begin{bmatrix} 0 & -\varepsilon/2 \\ \varepsilon/2 & 0 \end{bmatrix} & \cdots & \mathbf{0} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \cdots & \cdots \\ \mathbf{0} & \vdots & \mathbf{0} & \vdots & \begin{bmatrix} 0 & -\varepsilon/2 \\ \varepsilon/2 & 0 \end{bmatrix} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \right\|_2 = \frac{|\varepsilon|}{2}, \end{aligned}$$

i.e., $2d_{k,+} = |\varepsilon| = \min_{i \neq j=1,2,\dots,m} |\theta_i + \theta_j - 2k\pi|$.

Similarly, one can define $d_{k,-}$ and $d_{k,*}$ to the closed subset $\mathcal{A}_{k,-}$ and $\mathcal{A}_{k,*}$ realized by $X_{k,-}$ and $X_{k,*}$ respectively. Taking the minimum over $d_{k,+}$, $d_{k,-}$ and the optional $d_{k,*}$ for all $k = \pm 1, \dots$, yields (3.9).

Finally, to see the bound on the matrix 2-norm, consider $S \in \mathcal{S}_0$ first. In this case, the boundary of \mathcal{S}_0 is simply the 4 subsets $\mathcal{A}_{\pm 1, \pm}$. Let the distance be realized on $\mathcal{A}_{1,+}$ at θ_i, θ_j as $2\delta_\Theta = 2d_{1,+} = |\theta_i + \theta_j - 2\pi|$.

Suppose $2d_{1,+} > 2\pi$. Since, $S \in \mathcal{S}_0$, there is $\theta_i + \theta_j < 2\pi$, there is $-4\pi < \theta_i + \theta_j - 2\pi < -2\pi$. In this case,

$$\begin{aligned} 2d_{-1,+} &= \min_{i \neq j=1,2,\dots,m} |\theta_i + \theta_j + 2\pi| \\ &\leq |\theta_i + \theta_j + 2\pi| \\ &= |\theta_i + \theta_j - 2\pi + 4\pi| \\ &= 4\pi + (\theta_i + \theta_j - 2\pi) \leq 2\pi < 2d_{1,+}. \end{aligned}$$

This is a contradiction as $d_{1,+}$ realizes the shortest distance, therefore, $2d_{1,+} \leq 2\pi$. When $d_{1,-}$ realizes the distance, similar argument can be made by looking into $d_{-1,-}$.

When $S \in \mathcal{S}_e \neq \mathcal{S}_0$, the boundary of \mathcal{S}_e must include at least one of the following 3 pairs of subsets: (1) $\mathcal{A}_{k,+}$ and $\mathcal{A}_{k+1,+}$; (2) $\mathcal{A}_{k,-}$ and $\mathcal{A}_{k+1,-}$; (3) $\mathcal{A}_{k,*}$ and $\mathcal{A}_{k+1,*}$, where $k, k+1 \neq 0$ are integers. Apply the same argument above to this tighter pair of boundary yields $2\delta_\Theta \leq \pi$, which concludes the proof. Note that the missing closed subsets $\mathcal{A}_{0,+}$, $\mathcal{A}_{0,-}$ and $\mathcal{A}_{0,*}$ makes the \mathcal{S}_0 the only special case with $\delta_\Theta \leq \pi$. \square

Theorem 3.3.4. *For any $S \in \mathcal{S}_e, e \in \mathcal{E}$ with angles Θ , let*

$$\mathcal{M}_S := \{X \in \mathbf{Skew}_n : \|X - S\|_2 < \delta_\Theta\} \tag{3.11}$$

be the inscribed ball under the matrix 2-norm centered at S , where δ_Θ is the distance from S to Conj_{I_n} . Then $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q$ is a diffeomorphism where $\mathcal{N}_Q := \{\exp(X) : X \in \mathcal{M}_S\}$.

Proof. By construction, the inscribed ball \mathcal{M}_S lies within \mathcal{S}_e . Then, the triangle inequality of the matrix 2-norm yields $\forall A, B \in \mathcal{M}_S, \|A - B\|_2 \leq \|A - S\|_2 + \|B - S\|_2 < 2\pi$, i.e., the sufficient condition in **Proposition 3.3.2** is satisfied. \square

From this point, the \mathcal{M}_S in a diffeomorphism $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q$ is assumed to be the inscribed ball (3.11) unless otherwise specified. Note that the principal branch (2.1) happens to be the largest

inscribed ball centered at $\mathbf{0} \in \mathbf{Skew}_n$ with the radius $\delta_{\mathbf{0}} = \pi$. This result is also consistent with the classic result stating that the matrix exponential is invertible within the principal branch.

3.4 Nearby Logarithm

The immediate result following from the diffeomorphism $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q$ is the well defined inverse function $\exp^{-1} : \mathcal{N}_Q \rightarrow \mathcal{M}_S$, which is also a smooth bijection map with invertible differential. It leads to the definition of the nearby matrix logarithm, which is originally proposed in [9].

Definition 3.4.1. Given a skew symmetric $S \in \mathcal{S}$, Let $\mathcal{M}_S \subset \mathcal{S}$ be an open neighborhood consisting of S and let $\mathcal{N}_Q := \{\exp(X) : X \in \mathcal{M}_S\}$ be its image such that $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q$ is a diffeomorphism. Then, a nearby matrix logarithm is the inverse function of the diffeomorphism denoted as

$$\log_S : \mathcal{N}_Q \rightarrow \mathcal{M}_S. \quad (3.12)$$

□

Note that any diffeomorphism $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q$ meets the condition in **Definition 3.4.1**, i.e., the \mathcal{M}_S is not restricted to the inscribed ball in (3.11) or even the sufficient condition in **Proposition 3.3.2**. However, the inscribed ball is convenient in practice.

Also note that the skew symmetric S in \log_S serves more as a parameter that locally determines the nearby logarithm in a loose manner, rather than an input variable. Given a well-defined nearby matrix logarithm $\log_S : \mathcal{N}_Q \rightarrow \mathcal{M}_S, \forall X \in \mathcal{M}_S$, the same open subset \mathcal{M}_S can also be viewed as the neighborhood of X , such that $\log_X : \mathcal{N}_Q \rightarrow \mathcal{M}_S$ is the same nearby logarithm. One can still force S to be an input variable and define the nearby matrix logarithm around S as something like $\mathbf{Skew}_n \times \mathbf{SO}_n \rightarrow \mathbf{Skew}_n, (S, Y) \mapsto X$ where $\exp(X) = Y$, but such a function is not even continuous in the \mathbf{Skew}_n portion of its domain.

3.4.1 Comparison with the Original Definition

Compared to the vague description given in [9] stating that the nearby matrix logarithm around the given S seeks a solution $(X, \exp(X) = Y)$ nearest to $(S, \exp(S) = Q)$, **Definition 3.4.1** clarifies the condition of S around which the nearby logarithm is defined.

The description of the “nearest” solution of (X, Y) to (S, Q) is ambiguous, as the measurement to quantify the “nearest” notion can be made on \mathbf{Skew}_n or \mathbf{SO}_n , in different metrics. Furthermore,

the “nearest” notions on the two different metric space may not be equivalent. Therefore, **Definition 3.4.1** drops the “nearest” statement and replaces it with diffeomorphism. The impression of finding solution “near” to $\exp(S) = Q$ is kept naturally in the diffeomorphism description, as one may shrink \mathcal{M}_S to an arbitrarily small open neighborhood of S and the diffeomorphism structure stays on the shrunk neighborhood. For this reason, the term “nearby” is kept in the new definition even though the notion of “being the nearest solution” does not explicitly appear in it. In conclusion, the basically heuristic definition of the nearby matrix logarithm given in [9] is subsumed and accurately interpreted in the new definition given above.

3.4.2 Algorithms

This section develops two different approaches in computing the nearby logarithm $\log_S : \mathcal{N}_Q \rightarrow \mathcal{M}_S$, the Newton method **Algorithm 3** and the adaptive method **Algorithm 4**, that operate on **Skew_n** and **SO_n** respectively.

Newton Method. The Newton method in **Algorithm 3** is improved from the prototype algorithm proposed in [9], with the input $S \in \mathcal{S}_e$ and $Y \in \mathcal{N}_Q$ clarified and the N_i in line 6 is computed with the improved formula, rather than appealing to a brute-force linear solver on $\text{Dexp}_{X_i}[N_i] = Q_i M_i$ as a matrix free action, which is suggested in [9]. Note that this algorithm operates on **Skew_n** and generates a sequence of $\{X_i\} \subset \mathbf{Skew}_n$ that has convergence guaranteed by the classic Newton framework with a good initial guess S_0 .

Algorithm 3: Newton Method for Computing the Nearby Logarithm

Input: $S \in \mathcal{S}_e$ with diffeomorphism $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q$ and $Y \in \mathcal{N}_Q$ near $Q = \exp(S)$
Output: The unique inverse of $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q$ at Y as $X = \log_S(Y)$.

- 1 Initial guess $X_0 \leftarrow S$ if not provided;
- 2 $Y_0 \leftarrow Y$;
- 3 $i \leftarrow 0$;
- 4 **while** $\|Y_i - Y\| > \varepsilon$ **do**
- 5 $M_i \leftarrow \text{Skew}(Y_i^T Y - I_n)$; // Project difference to $T_{Y_i} \mathbf{SO}_n$
- 6 $N_i \leftarrow \mathcal{L}_{X_i}^{-1}(M_i)$; // **Algorithm 2**
- 7 $X_{i+1} \leftarrow X_i + \alpha \cdot N_i$; // Line search for step size α
- 8 $Y_{i+1} \leftarrow \exp(X_{i+1})$;
- 9 $i \leftarrow i + 1$;
- 10 **Return** X_i ;

There are three important algorithmic considerations to note in **Algorithm 3**. First of all, the M_i in line 5 is an estimation of the difference between Y_i and Y . Theoretically speaking, the $M_i = \log(Y_i^T Y)$ should be used. However, the computation cost in one iteration is dominated by the matrix exponential in line 8, including an extra matrix logarithm in line 5 would double

the execution time in one iteration. On the other hand, due to the close enough assumption of $Y \in \mathcal{N}_Q$, the estimation $\text{Skew}(Y_i^T Y - I_n)$ is good enough for the convergence. Therefore, the latter is used rather than the former. Secondly, the computation executed in line 6, $\mathcal{L}_{X_i}^{-1}(M_i) = N_i$, should reuse the information available in line 8 from the last iteration, $\exp(X_i) = Y_i$. It implies that the matrix exponential should be computed with the Schur decomposition approach $\exp(X_i) = \exp(R_i D_i R_i^T) = R_i^T \exp(D_i) R_i^T$ and the Schur vectors R_i and angles Θ_i in D_i should be kept for later computations in line 6. Finally, although the line search procedure is mentioned in line 7 for the sake of completeness, it is actually not recommended to perform complicated line search for the step size, as verifying the quality of a try on the step size requires a full matrix exponential. The current stable implementation of **Algorithm 3** uses the constant step size $\alpha = 1$. A more careful discussion of the linear search and convergence investigation is left as future work.

Recursive Algebraic Method. The recursive algebraic method takes a different approach by exploiting the fact that the smooth curve $\{Y(t) = \exp(S) \exp(t \cdot \Delta) : t \in [0, 1]\} \subset \mathcal{N}_Q$ where $Y(0) = \exp(S)$ and $Y(1) = Y$ produces a smooth curve $\{X(t) = \log_S(Y(t)) : t \in [0, 1]\} \in \mathbf{Skew}_n$ where $X(1) = \log_S(Y)$ is the desired solution.

Therefore, for any $\delta > 0$, there exists a sufficiently small mesh $0 = t_0 < t_1 < \dots < t_k = 1$, such that $\|X(t_{i-1}) - X(t_i)\|_2 < \delta, \forall i = 1, \dots, k$. Although computing the matrix 2-norm is almost as expensive as computing a matrix logarithm, it is not hard to estimate its lower bound as

$$\|M\|_2 \geq \max_{i,j=1,\dots,n} |M_{i,j}|$$

where $M_{i,j}$ are the entries in M . Then, for a preferred Schur decomposition $Y(t) = RER^T$ with $X = R\tilde{D}R^T$ where \tilde{D} are shifted from the principal Θ in Y , there is $\|S - X\|_2 = \|R^T S R - \tilde{D}\|_2$. Note that the freedom in the above difference is the 2π shifts on the diagonal. Let $M = R^T S R$, then

$$|M_{2i+2,2i+1} - \theta_i - 2\xi_i\pi| < \pi, \forall i = 1, \dots, m \quad (3.13)$$

is necessary to have $\|M - \tilde{D}\|_2 < \pi$, where $\xi_i \in \mathbb{Z}$. Such an integer vector ξ is unique and can be found from the integer part of $(M_{2i+2,2i+1} - \theta_i)/2\pi$. It leads to the following recursive algebraic

algorithm.

Algorithm 4: Algebraic Method for Computing the Nearby Logarithm

Input: $S \in \mathcal{S}_e, Q = \exp(S)$ with diffeomorphism $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q, \Delta \in \mathbf{Skew}_n$ where $\{\exp(S) \exp(t \cdot \Delta) : t \in [0, \alpha]\} \subset \mathcal{N}_Q$, threshold $\delta < \pi$ on matrix 2-norm, step size $\alpha \leq 1$ and shrinking parameter $\sigma < 1$.

Output: $X_\alpha = \log_S(Q(\alpha))$

- 1 If δ not provided, $\delta \leftarrow \delta_\Omega$, the distance from S to Conj_{I_n} ; // (3.10)
 - 2 $Q_\alpha \leftarrow Q \exp(\alpha \cdot \Delta_S)$;
 - 3 Factor Q_α for a preferred Schur decomposition $RE R^T$ and the principal angles Θ ;
 - 4 $M \leftarrow R^T S R$;
 - 5 Find the unique set of integers $\xi_i, i = 1, \dots, m$ that satisfies (3.13);
 - 6 $\tilde{M} \leftarrow M - \text{diag} \left(\begin{bmatrix} 0 & -\theta_1 - 2\xi_1\pi \\ \theta_1 + 2\xi_1\pi & 0 \end{bmatrix}, \dots \right)$;
 - 7 **if** $\|\tilde{M}\|_2 < \delta$ **then**
 - 8 $X_\alpha \leftarrow R \text{diag} \left(\begin{bmatrix} 0 & -\theta_1 - 2\xi_1\pi \\ \theta_1 + 2\xi_1\pi & 0 \end{bmatrix}, \dots \right) R^T$;
 - 9 **Return** (Q_α, X_α)
 - 10 **else**
 - 11 **Call** itself for $(Q_{\sigma\alpha}, X_{\sigma\alpha})$ from **Algorithm 4** with $(S, Q, \Delta, \delta, \sigma \cdot \alpha, \sigma)$;
 - 12 **Call** itself for (Q_α, X_α) from **Algorithm 4** with $(X_{\sigma\alpha}, Q_{\sigma\alpha}, \Delta, \delta, (1 - \sigma) \cdot \alpha, \sigma)$;
 - 13 **Return** (Q_α, X_α) ;
-

In this recursive algorithm, the divide-and-conquer idea is employed in lines 11 and 12, when the trial solution identified in line 5 does not satisfy the bound δ on the matrix 2-norm. The failure in line 5 suggest that the algebraic approach has no guarantee in computed solution. Then, a break point $Q_{\sigma\alpha} = Q \exp(\sigma\alpha \cdot \Delta)$ is inserted and then line 9 attempts to solve the easier problem $X_{\sigma\alpha} = \log_S(Q \exp(\sigma\alpha \cdot \Delta))$. When line 9 succeeds, the algorithm moves on the solve the nearby logarithm problem on the remaining $Q \exp(t \cdot \Delta), t \in [\sigma\alpha, \alpha]$. When the call in line 11 fails at its first attempt, another breakpoint is inserted at $t = \sigma^2\alpha$ and another two calls to **Algorithm 4** are made in line 11. This process continues until all calls to **Algorithm 4** succeed, and it thereby produces a mesh $\{t_0 = 0, \dots, t_k = 1\}$ with $X_i = \log_S(Q \exp(t_i \cdot \Delta))$ satisfying $\|X_{i+1} - X_i\|_2 < \delta$.

There are three important algorithmic observations in **Algorithm 4**. Firstly, the complexity within each call to **Algorithm 4** is dominated by three parts, the matrix exponential in line 2, the Schur factorization in line 3 and the matrix 2-norm evaluation in line 7. Improvements on these subroutines will help improve the overall performance considerably. Secondly, the divide-and-conquer strategy in line 11 and 12 can be adjusted so that they reuse the previous computed objects in the recursive calls as much as possible. For example, by setting $\sigma = 1/2$, the matrix exponential in line 2 in each call becomes $\exp((k/2^s)\Delta)$. By further restricting α being $1/2^s$, only $\exp((1/2^s)\Delta)$ are needed. These exponentials are available in one call to the scaling and squaring

algorithm [25] when computing $\exp(\Delta)$. Finally, the threshold $\delta < \pi$ significantly affects the total number of calls, it is currently using the distance from S to Conj_{I_n} as the inscribed ball guarantees a diffeomorphism. However, this condition is not necessary and it may be too restrictive in some cases. It is left as future work to determine a more efficient and precise bound δ .

3.4.3 Visualizing Geodesics with Skew Symmetric Matrices

The complexity of the **Algorithm 3** and **Algorithm 4** depends on various complicated factor as discussed above. They include the quality of the initial guess for **Algorithm 3** and the appropriate bound in (4.3) for **Algorithm 4**. Other than the algorithmic influences, the primitives that requires the real Schur decomposition on skew symmetric matrices and special orthogonal matrices also have significant influences to the overall complexity. This is an active research area with a lot of potential unexploited, see [24], so it is too earlier to make systematic comparison of the computing performances in the two presented algorithm. Therefore, this chapter performs a simple experiment to demonstrate that both algorithms have the correct functionality, which is to computed the nearby matrix logarithm around $S \in \mathbf{Skew}_n$ that is potentially beyond the principal branch.

Consider the geodesics $Q(t) = \exp(S) \exp(t \cdot \Delta), t \in [0, 1]$ used in the last experiment, instead of just getting $\log_S(Q(1))$, this part extends the geodesic with large enough $t \in [0, T]$ and inserts 5000 mesh points $0 = t_0 < t_1 < \dots < t_k = T$. Then, the nearby matrix logarithm is used step-by-step as $S_0 = S, S_i = \log_{S_{i-1}}(Q(t_i)), i = 1, \dots, k$. The returned S_i are expected to be a smooth curve on \mathbf{Skew}_n , which is illustrated in **Figure 3.2**. For a skew symmetric matrix $S \in \mathbf{Skew}_n$ with the angles θ_1 and θ_2 , **Figure 3.2** reports its angles to demonstrates the smoothness. Notice that the order in θ_1 and θ_2 as well as the signs in them can be arbitrarily flipped while the corresponding $S \in \mathbf{Skew}_n$ remains the same. Therefore, the eight possible combinations of angles in every $S \in \mathbf{Skew}_4$

$$\{(\theta_1, \theta_2), (\theta_2, \theta_1), (-\theta_1, \theta_2), (\theta_2, -\theta_1), (\theta_1, -\theta_2), (-\theta_2, \theta_1), (-\theta_1, -\theta_2), (-\theta_2, -\theta_1)\}$$

are checked and the most appropriate combination is selected to be reported in **Figure 3.2**.

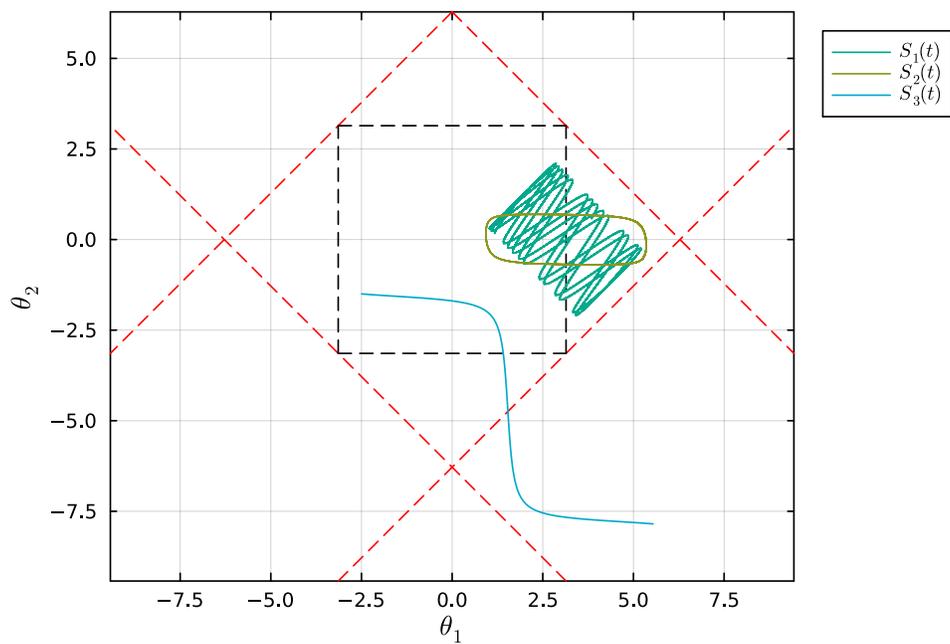


Figure 3.2: Illustration of Geodesics with Skew Symmetric Matrices.

Red Dashed: Conj_{I_n} ; Black Dashed: $\{S : \|S\|_2 = \pi\}$; Solids: Curves Computed by the Nearby Logarithm

CHAPTER 4

SMOOTHLY EVOLVING GEODESIC PROBLEM ON THE SPECIAL ORTHOGONAL GROUP

4.1 Introduction

The special orthogonal group \mathbf{SO}_n arises in many applications, [5, 22, 10, 15], and a special orthogonal matrix is usually interpreted as a rotation to set of independent variables. For example, a 3×3 special orthogonal matrix is usually used to represent a rotation of the 3 spacial coordinates and it can be extended to a 6×6 special orthogonal matrix if the rotations in the velocities are included. This information may be collected from a gyroscope and a GPS locator that describes an object's momentum in each time frame. Part of the anti-vibration algorithm in cameras takes the average of these rotation information of lens to adjust the graphic information at each time frame and to produce a picture with more consistent and stable quality. However, these applications suffer from the fact that the rotation information can only be smoothly represented in a small region and there is a lack of a smooth representation of all rotations in practice, see [40].

With the differential of matrix exponential restricted to the skew symmetric matrices and the nearby matrix logarithm on the special orthogonal matrices developed in the previous two chapters, it is possible to define and solve the smoothly evolving geodesic problem on \mathbf{SO}_n . The smoothly evolving geodesics computed in this chapter present a novel solutions to the above issues and they also yield important implications on other manifolds with special orthogonal constraints as discussed in later chapters.

4.2 Preliminaries

\mathbf{SO}_n is a set of orthogonal $n \times n$ matrices that have their determinant equal to 1, i.e.,

$$\mathbf{SO}_n := \{Q \in \mathbb{R}^{n \times n} : Q^T Q = I_n, \det Q = 1\}. \quad (4.1)$$

This is a Lie group and there is a natural Riemannian structure induces by the Lie structure, as developed in [11]. Some important notions are reviewed in this section.

As a Lie-group-induced Riemannian structure, \mathbf{SO}_n is a complete manifold, on which the Riemannian exponential is smoothly defined on the entire tangent space at any point. Furthermore, \mathbf{SO}_n is connected. It is then safe to conclude that the Riemannian exponential on \mathbf{SO}_n forms a smooth surjective map $\text{Exp}_Q : T_Q\mathbf{SO}_n \rightarrow \mathbf{SO}_n$ at any $Q \in \mathbf{SO}_n$. More details specific to \mathbf{SO}_n are given below.

The tangent space of \mathbf{SO}_n at special orthogonal matrix Q is characterized by the set of skew symmetric matrices as

$$T_Q\mathbf{SO}_n = \{QS : S \in \mathbf{Skew}_n\}$$

where $\mathbf{Skew}_n = \{S \in \mathbb{R}^{n \times n} : S + S^T = \mathbf{0}\}$ collects all skew symmetric $n \times n$ matrices. The Riemannian metric is given by the inner product between the skew symmetric matrices as

$$g_Q(QS_1, QS_2) = \frac{1}{2} \text{tr}(S_1^T S_2).$$

The corresponding Riemannian geodesic takes the simple form of

$$\text{Exp}_Q(QS) = Q \exp(S)$$

where $\exp(\cdot)$ is the matrix exponential.

Note that the identity matrix I_n lives in \mathbf{SO}_n and the cut locus of I_n in $T_{I_n}\mathbf{SO}_n$ takes the following simple form:

$$\text{Cut}_{I_n} = \{S : S \in \mathbf{Skew}_n, \|S\|_2 = \pi\}.$$

In other words, for any skew symmetric S with $\|S\|_2 < \pi$, the geodesic $Q(t) = \exp(t \cdot S), t \in [0, 1]$ is the unique shortest geodesic between I_n and $\exp(S)$. According to the Lie group structure, the characterization of Cut_{I_n} can be transported to any $Q \in \mathbf{SO}_n$ as follows

$$\text{Cut}_Q = \{Q \cdot S : S \in \mathbf{Skew}_n, \|S\|_2 = \pi\}.$$

4.3 Problem Formulation

With the notations introduced above, it is easy to write the smoothly evolving geodesic problem on \mathbf{SO}_n in the form of

$$\begin{cases} PS(0) = PS \\ P \exp(S(t)) = Q(t), \forall t \in [0, 1] \end{cases}$$

for substituting (1.2) with $P \in \mathbf{SO}_n$ for $x \in \mathcal{M}$, $PS \in T_P\mathbf{SO}_n$ for $v \in T_x\mathcal{M}$ and $P \exp(S(t))$ for $\text{Exp}_x(v(t))$. However, this formulation is equivalent to a simpler formulation that emanates from the identity matrix as in the following **Definition 4.3.1**.

Consider the general formulation of the smoothly evolving geodesic problem (1.2) on the special orthogonal group $\mathcal{M} = \mathbf{SO}_n$ with $x = I_n$, $T_x\mathcal{M} = \mathbf{Skew}_n$, $v = S$ and $\text{Exp}_x(v) = \exp(S)$. The smoothly evolving geodesic problem on \mathbf{SO}_n is defined as follows.

Definition 4.3.1 (Formulation at the Identity Matrix). For a smooth curve of special orthogonal matrices $\{Q(t), t \in [0, 1]\} \subset \mathbf{SO}_n$ with a skew-symmetric matrix S satisfying $\exp(S) = Q(0)$, the smoothly evolving geodesic problem on \mathbf{SO}_n seeks a smooth curve of skew-symmetric matrices $\{S(t), t \in [0, 1]\} \subset \mathbf{Skew}_n$ such that

$$\begin{cases} S(0) = S \\ \exp(S(t)) = Q(t), \forall t \in [0, 1] \end{cases} \quad (4.2)$$

□

It remains to show that the smoothly evolving geodesic problem at any point $P \in \mathbf{SO}_n$ can be solved under the formulation (4.2) at I_n , which follows from the transitive property of the geodesics on \mathbf{SO}_n . For an arbitrary smooth curve $\{Q(t), t \in [0, 1]\} \subset \mathbf{SO}_n$, a reference special orthogonal matrix $P \neq I_n$ and an initial velocity $PS \in T_P\mathbf{SO}_n$ such that $\text{Exp}_P(PS) = Q(0)$, there is

$$\begin{cases} PS(0) = PS \\ P \exp(S(t)) = Q(t), \forall t \in [0, 1] \end{cases} \iff \begin{cases} S(0) = S \\ \exp(S(t)) = P^T Q(t), \forall t \in [0, 1] \end{cases}$$

Let $\tilde{Q}(t) = P^T Q(t)$. Then, the smooth evolving geodesic problem of $Q(t)$ referenced at P is converted to (4.2) of $\tilde{Q}(t)$ referenced at I_n . The found solution $S(t)$ of the $\tilde{Q}(t)$ problem can be recovered as the solution $PS(t) \in T_P\mathbf{SO}_n$ of the $Q(t)$ problem.

The ability of shifting a geodesic emanating from P to a geodesic emanating from I_n and vice versa is known as the transitive property, which follows from the Lie group structure. On a more general manifold that may not enjoy similar features, one should expect more complications in solving the smoothly evolving geodesic problem.

4.4 Solution Characterized by the Nearby Matrix Logarithm

With the smoothly evolving geodesic problem on \mathbf{SO}_n formulated at the identity matrix as (4.2), it is simplified to the problem of finding a smooth inversion of the matrix exponential evaluated

in the skew symmetric matrices. This problem is locally solved by the nearby matrix logarithm developed in **Chapter 3**, as the local diffeomorphism $\exp : \mathcal{M}_S \rightarrow \mathcal{N}$ at a neighborhood of S uniquely specifies the inversion. This statement is summarized as follows.

Proposition 4.4.1. *Consider any skew symmetric $S \notin \text{Conj}_{I_n}$ with $Q = \exp(S)$, let $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q$ be a local diffeomorphism where \mathcal{M}_S is an open neighborhood that includes S . Then, for any smooth curve $\{Q(t) : t \in [0, 1]\} \subset \mathbf{SO}_n$ with $Q(0) = Q$, the solution to the smoothly evolving geodesic problem (4.2) with $Q(t)$ and $S(0) = S$ is uniquely given by the nearby matrix logarithm*

$$S(t) = \log_S(Q(t)).$$

Proof. The existence and the smoothness of the curve $S(t) = \log_S(Q(t))$ are guaranteed by the diffeomorphism. It is also easy to see that $\exp(S(t)) = Q(t)$ by construction. Therefore, $S(t)$ is a solution.

The uniqueness follows from characteristics of the isolated preimage. Since the neighborhood \mathcal{M}_S does not include any matrix in the conjugate locus by construction. **Theorem 3.2.6** suggests that the solution $S(t) = \log_S(Q(t))$ at any $t \in [0, 1]$ is an isolated point such that $\forall X \neq S(t)$ satisfying $\exp(X) = \exp(S(t)) = Q(t)$, there is $\|X - S(t)\|_2 \geq 2\pi$. If any other smooth curve $X(t)$ satisfying $\exp(X(t)) = Q(t)$ exists, then there is $\|X(t) - S(t)\|_2 > \pi, \forall t \in [0, 1]$. This is impossible, as $X(0) = S(0) = S$ is the shared initial point. \square

With **Theorem 3.3.4** constructing a practical diffeomorphism in an inscribed ball and the **Algorithm 3** and **Algorithm 4** computing the nearby logarithm reliably, it remains to partition the smooth curve $\{Q(t) : t \in [0, 1]\}$ into $0 = t_0 < t_1 < \dots < t_k = 1$, such that

$$\{Q(t) : t \in [t_{i-1}, t_i]\} \subset \mathcal{N}_{Q'}, \forall i = 1, \dots, k, \text{ for some diffeomorphism } \exp : \mathcal{M}_{S'} \rightarrow \mathcal{N}_{Q'}. \quad (4.3)$$

The divide-and-conquer idea is applicable to the partition process, which leads to the following algorithm that has a structure similar to the algebraic algorithm of the nearby matrix logarithm,

Algorithm 4.

Algorithm 5: Smoothly Evolving Geodesics Computed by the Nearby Matrix Logarithm

Data: Special orthogonal $\{Y(t), t \in [0, 1]\}$
Input: Section $[t, s] \in [0, 1]$, skew symmetric $X(t)$, shrinking parameter $\sigma < 1$
Output: $X(\alpha)$ from the solution $X(t)$ in (4.2)
1 if Condition (4.3) for $\{Y(t), t \in [t, s]\}$ is satisfied with $\exp : \mathcal{M}_{S'} \rightarrow \mathcal{N}_{Q'}$ **then**
2 | Return $\log_{S'}(Y(s))$; // **Algorithm 3 or Algorithm 4**
3 else
4 | $\delta \leftarrow (1 - \sigma)t + \sigma s$;
5 | Call itself for $X(\delta)$ from **Algorithm 5** with $([t, \delta], X(t), \sigma)$;
6 | Call itself for $X(s)$ from **Algorithm 5** with $([\delta, s], X(\delta), \sigma)$;
7 | Return $X(s)$;

There are two important algorithmic considerations in **Algorithm 5**. Firstly, the difficulties in verifying the condition in line 1, i.e., in (4.3), may vary from no computation to very expensive computations. It depends on the structure of $Q(t)$, e.g., it is difficult to locate a diffeomorphism covering $Q(t)$, if it exists, when $Q(t)$ gets arbitrarily close to the \mathbb{Q} in (3.7). On the other hand, when $Q(t)$ takes a simple form like $Q(t) = Q \exp(t \cdots \Delta)$, a sufficiently small section of $s - t < \epsilon$ guarantees (4.3). One of the practical criteria is to check

$$l_{Q(t) \rightarrow Q(s)} < L\delta_{\Theta}$$

where $l_{Q(t) \rightarrow Q(s)}$ is the (estimated) length of $Q(t)$ section, L is the (estimated) upper bound of operation norm \mathcal{L}_X^{-1} for $X \in \mathcal{M}_{S(t)}$ and δ_{Θ} is the radius of the inscribed ball $\mathcal{M}_{S(t)}$, which is the distance from $S(t)$ to Conj_{I_n} . Secondly, the appropriate choice in computing the nearby logarithm in line 2 also depends on the structure and the knowledge of $Q(t)$. When $Q(t) = Q \exp(t \cdot \Delta)$, the curve coincides with the intermediate curve used in **Algorithm 4**, which makes it an appropriate choice. In the case where users can generate a good initial guess for $\log_{S'}(Y(s))$, **Algorithm 3** becomes more feasible.

Unfortunately, it is not clear how to justify the global existence of the solution $X(t)$ to the smoothly geodesic problem (4.2) with arbitrary smooth curve $Y(t)$. A convenient sufficient condition is to assume $\{Y(t) : t \in [0, 1]\} \cap \mathbb{Q} = \emptyset$, i.e., there must exist an open cover $\{\mathcal{M}_{S_i} \subset \mathcal{S}_e\}_{i=1}^k$ for some $e \in \mathcal{E}$, such that $\{\exp(\mathcal{M})_{S_i}\}_{i=1}^k$ covers $\{Y(t) : t \in [0, 1]\}$ and the respective nearby logarithm of $\exp : \mathcal{M}_{S_i} \rightarrow \mathcal{N}_{Q_i}$ identifies the unique solution $\{X(t) : t \in [0, 1]\} \subset \mathcal{S}_e$. However, this condition is not necessary. For the $Y \in \mathbb{Q}$, there still exists an isolated solution $X \in \mathbb{E}_Y^{-1}$ with an invertible $D\exp_X$ as given in **Theorem 3.2.6**. For such $Y = \exp(X)$, let $S(t) = X$, there is still a local diffeomorphism around X that guarantee the existence of solution to (4.2) locally around X . Even

for the $X \in \text{Conj}_{I_n}$ that solves $\exp(X) = Y \in \mathbb{Q}$, the numerical experiment performed on the geodesic $Y(t) = Q \exp(t \cdot \Delta)$ in **Chapter 3** indicates that such a curve can be extended to infinity in general. This observation leads to the discussion in the next section about the $Y(t)$ being the geodesics on \mathbf{SO}_n rather than arbitrary smooth curves.

4.5 Smoothly Evolving Geodesics of Endpoints Varying along Geodesic

In the smoothly evolving geodesic problem (4.2), the smooth condition is the constraint on the varying endpoints $Q(t)$. Based on the local diffeomorphism $\exp : \mathcal{M}_S \rightarrow \mathcal{N}_Q$ established on $\mathcal{S}_e, e \in \mathcal{E}$, the smoothly evolving geodesic problem has a local solution in **Proposition 4.4.1**. This section considers the locality constraint in **Proposition 4.4.1** and further investigates the smoothly evolving geodesic problem (4.2) on the endpoints varying along geodesics emanating from Q as

$$Q(t, \Delta) := Q \exp(t \cdots \Delta) = \text{Exp}_Q(t \cdot Q\Delta).$$

The more restricted and structured varying endpoints yield a stronger and more global conclusion.

4.5.1 Vector Fields and Geodesics

First of all, recall that a Riemannian geodesic $\gamma(t)$ emanating from $\gamma(0)$ along $\dot{\gamma}(0)$ on the manifold \mathcal{M} is the unique solution to the ODE problem

$$\begin{cases} \dot{\gamma}(t) = \mathcal{P}_{\gamma, 0 \rightarrow t} \dot{\gamma}(0) \\ \gamma(0) = \gamma(0) \end{cases} \quad (4.4)$$

where $\mathcal{P}_{\gamma, 0 \rightarrow t} : T_{\gamma(0)}\mathcal{M} \rightarrow T_{\gamma(t)}\mathcal{M}$ is the parallel translation along γ . Then, the parallel translation on \mathbf{SO}_n along the geodesic $\gamma(t) = Q(t, \Delta)$ is given by the map

$$\mathcal{P}_{\gamma, 0 \rightarrow t}^{\mathbf{SO}_n}(QS) = Q \exp(t\Delta)S, \forall QS \in T_Q\mathcal{M} = \{QS : S \in \mathbf{Skew}_n\}.$$

In other words, any given geodesic $Q(t, \Delta)$ is equivalent with a vector field on \mathbf{SO}_n as

$$\mathfrak{V}_\Delta := \{Q\Delta : Q \in \mathbf{SO}_n\} \subset T\mathbf{SO}_n \quad (4.5)$$

where $T\mathbf{SO}_n = \bigcup_Q T_Q\mathbf{SO}_n$ is the tangent bundle of \mathbf{SO}_n . At any point Q , the vector field $\mathfrak{V}_\Delta(Q)$ assigns a tangent vector $Q\Delta \in T_Q\mathbf{SO}_n$. With this vector field given and the initial point $\gamma(0) = Q$ specified, the original geodesic $Q(t, \Delta)$ can be uniquely recovered.

Recall that the (pseudo) inverse operator at $X \in \mathbf{Skew}_n$ with $Y = \exp(X)$ is

$$\begin{aligned} D \exp_X^\dagger : T_{\exp(X)} \mathbf{SO}_n &\rightarrow \mathbf{Skew}_n \\ YS &\mapsto \mathcal{L}_X^\dagger(S), \forall S \in \mathbf{Skew}_n. \end{aligned}$$

Notice that the ‘‘prefix’’ Y is dropped and the vector field \mathfrak{V}_Δ , (4.5), to \mathbf{SO}_n naturally relates to the following vector field to the set of skew symmetric matrices

$$\mathfrak{S}_\Delta := \{\mathcal{L}_X^\dagger(\Delta) : X \in \mathbf{Skew}_n\}. \quad (4.6)$$

At any $X \in \mathbf{Skew}_n$, the vector field $\mathfrak{S}_\Delta(X)$ assigns the tangent vector $\mathcal{L}_X^\dagger(\Delta) \in T_X \mathbf{Skew}_n = \mathbf{Skew}_n$.

Proposition 4.5.1. *The vector field \mathfrak{S}_Δ constructed in (4.6) is locally smooth around X if $\Delta \in \mathcal{R}(\mathcal{L}_X)$, i.e., $Y\Delta$ is in the range space of $D \exp_X$, where $Y = \exp(X)$.*

Proof. When $X \notin \text{Conj}_{I_n}$, \mathcal{L}_X is invertible and the \mathcal{L}_S is smooth around X , according to the explicit formula derived in (2.8). Then, the smoothness of \mathfrak{S}_Δ around X follows.

When $X \in \text{Conj}_{I_n}$ and $\Delta \in \mathcal{R}(\mathcal{L}_X)$, by the nature of a pseudo inverse operator, there is $\mathcal{L}_X = \Delta$ and the designed projector onto $\mathcal{R}(\mathcal{L}_X)$ degenerates to the identical map. Without the non-smooth action given by the projector, for any $X(t) \in \mathbf{Skew}_n$ with $X(0) = X$ and $Y(t) = \exp(X(t))$. There is $Y(0) = \exp(X) \in \mathbb{Q}$. When $X(t)$ leaves Conj_{I_n} , i.e., $\dot{X}(0)$ is not tangential to Conj_{I_n} , $Y(t)$ leaves \mathbb{Q} and the restricted $\mathfrak{S}_\Delta(X(t)) = \mathcal{L}_X^\dagger(Y(t))$ is smooth. When $\dot{X}(0)$ is tangential to Conj_{I_n} , the designed pseudo inverse operator remains a smooth action, which follows from the explicit formula developed in **Definition ??**. \square

Note that for the $\Delta \notin \mathcal{R}(\mathcal{L}_X)$ case, the vector field \mathfrak{S}_Δ is no longer smooth around X , which only happens on the conjugate locus Conj_{I_n} . The almost every where smooth vector field yields the ODE problem with the unique solution that coincides with the solution of the smoothly evolving geodesic problem as stated in follows.

Theorem 4.5.2. *For any geodesic $Q(t, \Delta) = Q \exp(t \cdot \Delta)$ with $\exp(S) = Q$ and $\Delta \in \mathcal{R}(\mathcal{L}_S)$, there exists a unique (local) solution $S(t), t \in [0, \epsilon]$ to the ODE problem*

$$\begin{aligned} \dot{S}(t) &= \mathfrak{S}_\Delta(S(t)) \\ (S(0)) &= S \end{aligned} \quad (4.7)$$

This solution coincides with the smoothly evolving geodesic problem (4.2) with $Q(t, \Delta)$ and $S(0) = 0$.

Proof. The existence and uniqueness of the $S(t)$ as a solution to the ODE (4.7) is guaranteed by the smoothness that follows from **Proposition** 4.5.1.

To see that it is the solution to the smoothly evolving geodesic problem, notice that

$$\frac{d}{dt} \exp(S(t)) = D \exp_{S(t)} \left[\frac{d}{dt} S(t) \right] = D \exp_{S(t)} [\mathfrak{S}_\Delta(S(t))] = \exp(S(t)) \Delta.$$

By the existence and uniqueness of the solution to the ODE (4.4) with $\exp(S(0)) = Q$, the solution must be $\exp(S(t)) = Q \exp(t \cdots \Delta)$. Therefore, $\{S(t), t \in [0, \epsilon]\}$ satisfies the criteria of the solution to (4.2). \square

Note that the ODE formulation presented in this section is a more universal characterization of the smoothly evolving geodesic problem on an arbitrary Riemannian manifold. Therefore, it is expected to define the same curves stated in **Theorem** 4.5.2. Thanks to the rich geometry in the \mathbf{SO}_n , the solution to the smoothly evolving geodesic problem can be computed without introducing the ODE (4.7).

4.5.2 Co-Manifold Characterization

With the vector field \mathfrak{S}_Δ as a function of $\Delta \in \mathbf{Skew}_n$ that defines ODEs throughout \mathbf{Skew}_n , the following definition of a *co-manifold* characterization is introduced to establish a “copy” of the \mathbf{SO}_n around some $\exp(X) = Y$ realized in $T_{I_n} \mathbf{Skew}_n$.

Definition 4.5.3. For any $X \in \mathbf{Skew}_n$ and $Y = \exp(X)$, let \mathfrak{B} be a subspace of $\mathcal{R}(\mathcal{L}_X)$. Then a co-manifold characterization around X with \mathfrak{B} is defined as a map

$$\begin{aligned} \mathfrak{C}_{X, \mathfrak{B}} : \mathfrak{B} &\rightarrow \mathbf{Skew}_n \\ \Delta &\mapsto X(1) \end{aligned} \tag{4.8}$$

where Δ yields a solvable ODE (4.7) for $t \in [0, 1]$ and $X(1)$ is the solution $X(t)$ evaluated at $t = 1$. \square

Note that it is not completely understood if any Δ with arbitrary scale yields a solvable ODE, i.e., it is not clear if the local solution $\{X(t) : t \in [0, \epsilon]\}$ to ODE (4.7) can be extended to infinity. Therefore, $\mathfrak{C}_{X, \mathfrak{B}}$ may only be well defined within an envelope in \mathfrak{B} . Nevertheless, the well-defined part remains a smooth map.

Proposition 4.5.4. *The well defined co-manifold characterization $\mathfrak{C}_{X, \mathfrak{B}}$ is a smooth and invertible mapping, which further provides a smooth characterization of $\{Q \exp(\Delta) : \Delta \in \mathfrak{B}\}$ as $Q \exp(\Delta) \mapsto X(1)$.*

Proof. The map $Q \exp(\Delta) \mapsto X(1)$ is the solution to the smoothly evolving geodesics with endpoints restricted to vary along geodesic evaluated at the endpoint $Q(1) \mapsto X(1)$. However, this mapping is generally not invertible, as the solution is path-dependent. In this case with the path fully characterized by Δ , the mapping between $X(1)$ and Δ is indeed invertible.

When the existence and uniqueness of $\mathfrak{C}_{X,\mathfrak{B}}(\Delta)$ is guaranteed around Δ , the smoothness follows from the fact that it is locally realized by the nearby logarithm, which is a local diffeomorphism. \square

Note that the co-manifold characterization $\mathfrak{C}_{X,\mathfrak{B}}$ remains a local characterization. However, the set $\{Y \exp(S) : S \in \mathfrak{B}\}$ generated at $Y = \exp(X)$ with \mathfrak{B} has a very strong global manifold structure in \mathbf{SO}_n . It is reasonable to speculate that there is a manifold structure in the collection of the co-manifold characterization as defined in below.

Definition 4.5.5. Consider a submanifold \mathcal{G} in \mathbf{SO}_n generated at $Y \in \mathbf{SO}_n$ defined as

$$\mathcal{G} = \mathcal{G}(Y, \mathfrak{B}) := \{Y \exp(S) : S \in \mathfrak{B}\} \quad (4.9)$$

where \mathfrak{B} is a subspace. The co-manifold characterization of the submanifold \mathcal{G} is the union of the co-manifold characterization around X where $\exp(X) \in \mathcal{G}$ defined as

$$\mathfrak{C}_{\mathcal{G}} := \bigcup_{\{X:\exp(X) \in \mathcal{G}\}} \mathfrak{C}_{X,\mathfrak{B}}. \quad (4.10)$$

\square

The co-manifold characterization is a novel concept that attempts to translate and preserve the structure of a submanifold in \mathbf{SO}_n to \mathbf{Skew}_n which enjoys an embedded Euclidean structure. Although most of the geometry in the co-manifold is not understood globally, it provides a connected manifold-like feasible set that supports the retraction-like first order update according to the following observation.

Lemma 4.5.6. A geodesic triangle on \mathbf{SO}_n with the vertices $Q_A, Q_B, Q_C \in \mathbf{SO}_n$ and the edges $Q_A \exp(t \cdot X_{AB}), Q_A \exp(t \cdot X_{AC}), Q_B \exp(t \cdot Q_{BC})$ for $t \in [0, 1]$ and $X_{AB}, X_{BC}, X_{AC} \in \mathbf{Skew}_n$ is equivalent to a geodesic triangle with the vertices $I_n, Q_A^T Q_B, Q_A^T Q_C$ with the edges $\exp(t \cdot X_{AB}), \exp(t \cdot X_{AC})$ and $Q_A^T Q_B \exp(t \cdot Q_{BC})$.

The geodesic triangles on \mathbf{SO}_n can be arbitrarily shifted by any special orthogonal matrices. In **Lemma 4.5.6**, the triangle Q_A, Q_B, Q_C is shifted by the transpose of a vertex Q_A^T . Regardless of

how the triangle is shifted, the skew symmetric matrices X_{AB}, X_{BC} and X_{AC} that characterizes the edges remains unchanged. Notice that the triangle $I_n, Q_A^T Q_B, Q_A^T Q_C$ in \mathbf{SO}_n also yields a triangle in \mathbf{Skew}_n with vertices $\mathbf{0}, X_{AB}$ and X_{AC} with edges parameterized by $t \in [0, 1]$: $S_{AB}(t) = t \cdot X_{AB}$, $S_{AC}(t) = t \cdot X_{AC}$ and $S_{BC}(t)$ solved from the smoothly evolving geodesic problem

$$\begin{cases} \exp(S_{BC}(t)) = Q_A^T Q_B \exp(t \cdot X_{BC}) \\ S_{BC}(0) = X_{AB} \end{cases}.$$

The following proposition states a more general result.

Proposition 4.5.7. *Consider any $A \in \mathbf{Skew}_n$ with a solution $S_{AB}(t)$ that arrives at a $B \in \mathbf{Skew}_n$ that is characterized by $X_{AB} \in \mathbf{Skew}_n$ as follows,*

$$\begin{cases} \exp(S_{AB}(t)) = \exp(A) \exp(t \cdot X_{AB}) \\ S_{AB}(0) = A \end{cases}.$$

If there is another solution $S_{BC}(t)$ that emanates from B and arrives at a $C \in \mathbf{Skew}_n$ given by

$$\begin{cases} \exp(S_{BC}(t)) = \exp(B) \exp(t \cdot X_{BC}) \\ S_{BC}(0) = B \end{cases}.$$

Then, there exists a solution $S_{AC}(t)$ that emanates from A and arrives at a $C \in \mathbf{Skew}_n$ given by

$$\begin{cases} \exp(S_{AC}(t)) = \exp(A) \exp(t \cdot X_{AC}) \\ S_{AC}(0) = A \end{cases},$$

such that $S_{AB}(t), S_{BC}(t), S_{AC}(t)$ forms a triangle in \mathbf{Skew}_n and their exponential form a geodesic triangle in \mathbf{SO}_n .

Proof. When $A = \mathbf{0}$ which makes $\exp(A) = I_n$, the edge $S_{AB}(t)$ is fully characterized as $S_{AB}(t) = t \cdot B$. Then, the edge $S_{BC}(t)$ is given as an assumption. Let C be the arriving point in $S_{BC}(t)$ and the edge $S_{AC}(t) = t \cdot C$ naturally follows. For the $A \neq \mathbf{0}$ case, shift the geodesic triangle by $\exp(-A)$ to obtain the endpoint $Q'_A = \exp(-A)Q_A = I_n$, $Q'_B = \exp(-A)Q_B$ and $Q'_C = \exp(-A)Q_C$. Then, let $A' = \mathbf{0}$ so that B' and C' in the respective $\exp(B') = Q'_B$ and $\exp(C') = Q'_C$ can be solved by the nearby matrix logarithm, as indicated in the first scenario. Shift the geodesic triangle Q'_A, Q'_B, Q'_C and the corresponding skew symmetric triangle A', B', C' back to Q_A, Q_B, Q_C concludes the statement. \square

Corollary 4.5.8. For any sequential updates $S_i \rightarrow S_{i+1} \rightarrow \cdots \rightarrow S_{i+n}$ along the solutions to the smoothly geodesic problems in the form of

$$\begin{cases} \exp(S_{j \rightarrow j+1}(t)) = \exp(S_j) \exp(t \cdot X_{j \rightarrow j+1}) \\ S_{j \rightarrow j+1}(0) = S_j \\ S_{j \rightarrow j+1}(1) = S_{j+1} \end{cases}, j = 1, \dots, i + n - 1,$$

there exists a direct update $S_{i \rightarrow i+n}(t)$ in the form of the solution to

$$\begin{cases} \exp(S_{i \rightarrow i+n}(t)) = \exp(S_i) \exp(t \cdot X_{i \rightarrow i+n}) \\ S_{i \rightarrow i+n}(0) = S_i \\ S_{i \rightarrow i+n}(1) = S_{i+n} \end{cases}$$

such that it construct a polygon with $n + 1$ edges in \mathbf{Skew}_n .

Proof. Repeatedly apply **Proposition 4.5.7** to get the triangle with the edges $S_{i \rightarrow i+1}(t)$, $S_{i+1 \rightarrow i+2}(t)$ and $S_{i \rightarrow i+2}(t)$, and then the triangle with the edges $S_{i \rightarrow i+2}(t)$, $S_{i+2 \rightarrow i+3}(t)$ and $S_{i \rightarrow i+3}(t)$. It continues until the S_{i+n} is produced. \square

Corollary 4.5.8 is essential to the computations, especially in solving an optimization problem with iterative updates. In those computations, a series of data points $\{X_0, X_1, \dots\}$ are computed along the smoothly evolving geodesic and this corollary guarantees that there exists a smoothly evolving geodesic that connects the originated initial guess X_0 to any $X_i, i > 0$. Such a geodesic maintains connectivity and smoothness from the initial guess to any intermediate result as long as the update is constrained by the smoothly evolving geodesic problems in **Corollary 4.5.8**. Notice that the constrained update in **Corollary 4.5.8** is exactly the condition proposed on the co-manifold characterization, i.e., the sequential updates in the co-manifold characterization also maintains connectivity and smoothness inherited from a smoothly evolving geodesic. This is useful in many applications and analyses. For example, in the Stiefel manifold that consists of $n \times p$ orthonormal matrices, each matrix X is identified with the subset $\{Q \in \mathbf{SO}_n : \forall Q = [X \ X_\perp]\}$. This subset is a submanifold in \mathbf{SO}_n and can be fully characterized as $\left\{ Q \exp \left(\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Delta_{n-p} \end{bmatrix} \right) : \text{for some } Q = [X \ X_\perp], \forall \Delta_{n-p} \in \mathbf{Skew}_{n-p} \right\}$. The co-manifold characterization then further translates this submanifold characterized in Δ_{n-p} to $T_{I_n} \mathbf{SO}_n = \mathbf{Skew}_n$.

CHAPTER 5

VELOCITY-BASED KARCHER MEAN ON THE SPECIAL ORTHOGONAL GROUP

5.1 Introduction

The Karcher mean problem that finds an average of a given set of points arises in the literature and applications in various contexts. The Karcher mean on a Riemannian manifold has a significant amount of attention as the manifold structure provides an intrinsic representation of constrained objects. These objects are embedded in a parameterization space that has significantly larger dimension than the manifold in general. Furthermore, a Riemannian metric equips the manifold with a metric space that is inherently related to the motion on the constrained sets rather than the embedded space. Ever since the work by Edelman et al. in [11], the notion of Riemannian manifolds is proposed on various constrained objects and the Karcher mean problem on the respective Riemannian manifold has been studied extensively, see for example [20], [35], [2] and [39].

Recall that a Karcher mean on a metric space M with distance function $\text{dist} : M \times M \rightarrow \{x \geq 0\}$ is the global minimum to the following objective function

$$f(x) := \frac{1}{2k} \sum_{i=1}^k \text{dist}(x, y_i)^2, \forall x \in M$$

where $\{y_1, \dots, y_k\}$ are the given data points. In a Riemannian setting, the distance between 2 points is realized by the length of a shortest geodesic between them, which is measured by its velocity under the Riemannian metric. Then, the objective function is equivalent to

$$\begin{aligned} f(x) &= \frac{1}{2k} \sum_{i=1}^k g_x(\text{Log}_x(y_i), \text{Log}_x(y_i)) \\ &= \frac{1}{2k} \sum_{i=1}^k g_{y_i}(\text{Log}_{y_i}(x), \text{Log}_{y_i}(x)) \end{aligned} \tag{5.1}$$

where $\text{Log}_p : \mathcal{M} \rightarrow T_p\mathcal{M}$ for $p \in \mathcal{M}$ is the Riemannian logarithm. It returns an initial velocity at p that emanates a shortest geodesic arriving at the given $q \in \mathcal{M}$. This elegant and compact objective function provides a necessary and sufficient condition for $x \in \mathcal{M}$ being a critical point if

and only if

$$\sum_{i=1}^k \text{Log}_x(y_i) = \mathbf{0}. \quad (5.2)$$

This is obtained from differentiating the objective as $D f_x[\Delta] = \sum_{i=1}^k g_x(\text{Log}_x(y_i), \Delta)/k$. Unfortunately, there are severe theoretical concerns in utilizing the condition (5.2).

All concerns arise from the differentiability of $f(x)$, which depends on the distance function always being differentiable, i.e., the Riemannian logarithm always being differentiable. Unfortunately, this is impossible in general. Consider the \mathbf{SO}_n investigated in this dissertation as an example, there exist multiple shortest Riemannian geodesics for some pair of points. On these points, neither the distance function is differentiable, nor is the Riemannian logarithm well-defined or continuous. As a consequence, the objective function (5.1) may be globally non-smooth and non-convex, which may introduce multiple local minima to this objective function based on distances. Other than the theoretical difficulties in identifying the “unique” Karcher mean, the existence of multiple local minima implies the sensitivity and discontinuity within the computed Karcher mean with respect to the varying data set as well as the varying computing setup. In other words, when the given data points or the initial computing setup are perturbed (in a smooth manner), the computed Karcher mean may jump from one local minimum to another discontinuously.

Except for the rare cases where the Riemannian manifold is shown to be globally geodesically convex, e.g., the set of positive definite matrices in [39], the issues of differentiability persist. In most of the literature about Karcher mean on Riemannian manifolds, the global minimum requirement in the classic distance-based Karcher mean formulation is dropped and any local minimum characterized by (5.2) is accepted as a Karcher mean, e.g., [20, 38]. Some literature addresses and handles the uniqueness by restricting the data set to be close enough, e.g., [20]. When the Riemannian logarithm is not computationally available, the (5.2) with an alternative for Log_x is used to define a Karcher-mean-like mean, e.g., [2].

This chapter investigates the differentiability issues of the Karcher mean on \mathbf{SO}_n by looking into a \mathbf{SO}_2 example, i.e., the Karcher mean on a circle. Then, it applies the tools developed on \mathbf{SO}_n to propose a generalized Karcher mean that depends on velocity input on \mathbf{SO}_n , namely the velocity-based Karcher mean. The velocity-based Karcher mean is designed to maintain the differentiability in computations and numerical experiments are performed to demonstrate its potential in applications.

5.2 Example on a Circle

This section considers the simplest non-trivial Karcher mean problem on \mathbf{SO}_n , the Karcher mean problem on \mathbf{SO}_2 with 2 points Q_1 and Q_2 . Any 2×2 special orthogonal matrix takes the form of $Q = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$ where $c^2 + s^2 = 1$ and this characterization is equivalent with the unit circle $c^2 + s^2 = 1$ with the point $p = [c \ s]^T$ representing the given Q . Meanwhile, any 2×2 skew symmetric matrix takes the form of $\Delta = \Delta_{[0,\theta,0]} = \begin{bmatrix} 0 & -\theta \\ \theta & 0 \end{bmatrix}$ where $\theta \in \mathbb{R}$. In this very restricted setting of \mathbf{SO}_2 , the geodesic becomes

$$\text{Exp}_Q(t \cdot \Delta) = \begin{bmatrix} c \cos(t \cdot \theta) - s \sin(t \cdot \theta) & -c \sin(t \cdot \theta) - s \cos(t \cdot \theta) \\ c \sin(t \cdot \theta) + s \cos(t \cdot \theta) & c \cos(t \cdot \theta) - s \sin(t \cdot \theta) \end{bmatrix}, t \in [0, 1].$$

This geodesic is equivalent with the arc moving in the unit circle as

$$p(t) = (c \cos(t \cdot \theta) - s \sin(t \cdot \theta), c \sin(t \cdot \theta) + s \cos(t \cdot \theta)).$$

Note that the arc may overlap when $t\theta > 2\pi$. Furthermore, the length of the geodesic equals the length of the arc, which makes the Riemannian distance between two points realized by the shortest arc length between the corresponding points on the unit circle.

5.2.1 Objective Function with the Riemannian Distance

According to \mathbf{SO}_2 realized on the unit circle constructed above, consider the Karcher mean problem of the two points $Q_1, Q_2 \in \mathbf{SO}_n$ which are realized by y_1, y_2 as

$$y_1 = \begin{bmatrix} \cos(0.9\pi) \\ \sin(0.9\pi) \end{bmatrix}, \quad y_2 = \begin{bmatrix} \cos(0.6\pi) \\ \sin(0.6\pi) \end{bmatrix}$$

Consider the unit circle parameterized by

$$p_x := \begin{bmatrix} \cos(x) \\ \sin(x) \end{bmatrix}, x \in [-\pi, \pi].$$

Figure 5.1 illustrates the y_1, y_2 and x on a unit circle and plots the objective function $f(x) := f(p_x)$ against x . The shortest arcs that realize the Riemannian distances from x to y_1 and y_2 are also plotted in red.

Note that the middle point $[\cos(0.75\pi) \ \sin(0.75\pi)]^T$ with $x_1 = 0.75\pi$ of the arc between y_1 and y_2 is indeed the Karcher mean of y_1 and y_2 . This is easily verified from the plot of objective function on the right. However, the objective function $f(x)$ is not globally smooth as expected, it is not differentiable at $x = -0.1\pi$ and $x = -0.4\pi$. These break points partition the unit circle into

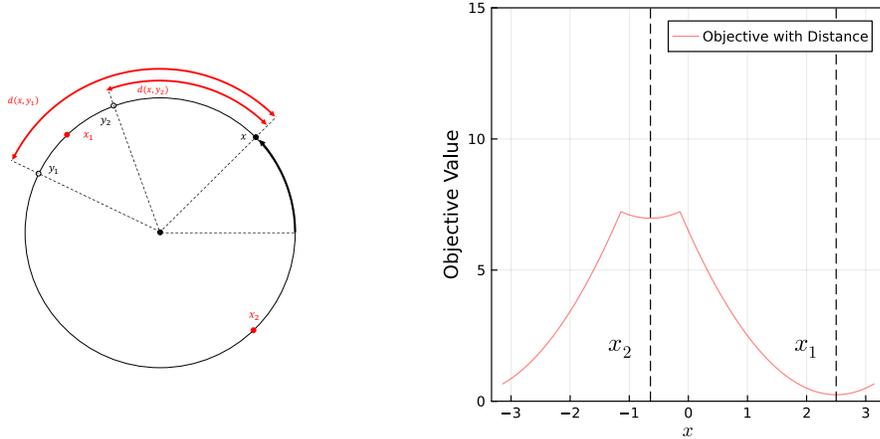


Figure 5.1: Illustration of the Karcher Mean on an Unit Circle

two regions, the region containing periodic segments $[-\pi, -0.4\pi]$ and $[-0.1\pi, \pi]$ that can be glued together at $-\pi$ and π and the $[-0.4\pi, -0.1\pi]$ region. Within both regions, the objective functions are not only smooth, but also convex, which grants them the respective unique global minimum, $x_1 = 0.75\pi$ in the first region and the $x_2 = -0.25\pi$ in the second region. Also note that the non-differentiable points have discontinuous Riemannian logarithms as expected. For example, the shortest arc from x around -0.1π to $y_1 = 0.9\pi$ flips from the arc in the upper half-circle to the arc in the lower half-circle or vice versa. This observation is reflected in the discontinuous Riemannian logarithm with x approaching -0.1π from above or below

$$\begin{cases} \lim_{x \rightarrow (-0.1\pi)_-} \text{Log}_{y_1} \left([\cos(x) \ \sin(x)]^T \right) = [-\pi \sin(0.9\pi) \ \pi \cos(0.9\pi)]^T \\ \lim_{x \rightarrow (-0.1\pi)_+} \text{Log}_{y_1} \left([\cos(x) \ \sin(x)]^T \right) = [\pi \sin(0.9\pi) \ -\pi \cos(0.9\pi)]^T \end{cases}$$

Even with this simple setup, this example has demonstrated the theoretical and computational subtleties in the Karcher mean formulation on \mathbf{SO}_n , or on a Riemannian manifold in general. First of all, the non-smooth structure from the discontinuous Riemannian logarithm introduces multiple local minima. The more data points it includes in the Karcher mean problem, the more discontinuous points are introduced. Eventually, it results in more complicated and non-convex structures in the objective function with more local minima. Secondly, there is no good way to tell if a local minimum is a global minimum. Even with just 2 points in the example, the found local minimum x_1 and x_2 are no different from each other as they both realize the local minimal objective value with the shortest geodesic connecting to the data points. In other words, one must

exhaust all local minima to find the true Karcher mean, which is impossible in practice. In an extreme case when the two points sit the opposite polar point with each other, i.e.,

$$\begin{cases} y_1 = [\cos(\theta) & \sin(\theta)]^T \\ y_2 = [\cos(\theta) & -\sin(\theta)]^T \end{cases},$$

the corresponding local minima x_1 and x_2 have the same objective value, i.e., there are two Karcher means. Last but not the least, some literature, e.g., [20], has relaxed the condition of a Karcher mean but it may still have computational concerns, as the computed mean is very sensitive with respect to the initial guess and the distribution of the data set. In other words, consider a set of relaxed Karcher means of given data set $\mathbf{y} = \{y_1, \dots, y_k\}$ as follows

$$\tilde{\mathbf{y}} := \{x \in \mathcal{M} : f(x) \text{ is a local minimum}\} \quad (5.3)$$

that consists of all local minima and let

$$\widehat{\text{Mean}}(x_0, \mathbf{y}) \in \tilde{\mathbf{y}} \quad (5.4)$$

be the computed relaxed Karcher mean from the initial guess x_0 . Then the computed mean $\widehat{\text{Mean}}(x_0, \mathbf{y})$ is not continuous with respect to both the initial guess x_0 and the data set \mathbf{y} , which includes the following scenarios.

1. When \mathbf{y} stays constant such that $\tilde{\mathbf{y}}$ stays constant, $\widehat{\text{Mean}}(x_0, \mathbf{y})$ may jump to a different local minimum in $\tilde{\mathbf{y}}$ as x_0 moves.
2. When x_0 stays fixed and \mathbf{y} varies in a smooth manner such that the set $\tilde{\mathbf{y}}$ also varies in a smooth manner, the computed mean $\widehat{\text{Mean}}(x_0, \mathbf{y})$ may jump from smoothly varying local minimum to another smoothly varying local minimum.
3. Some of the local minima may vanish as \mathbf{y} varies in a smooth manner, e.g., when y_1 and y_2 overlaps, x_2 becomes a kink in the objective function illustrated in **Figure 5.1**, i.e. it is no longer a local minimum.

These sensitive relationships are quite common in solving non-convex optimization problems in general, since the sets of local minima to those non-convex objective functions may not be smooth in the first place. However, one can expect more structure in the Karcher mean problem on a Riemannian manifold once the smoothness condition is re-introduced to the objective function.

5.2.2 Objective Function with the Smoothly Evolving Arc Length

The difficulties in the Karcher mean problem on \mathbf{SO}_n are consequences of the discontinuous Riemannian logarithm that is necessary for measuring the distance on a given point. Therefore, the notion of the smoothly evolving geodesic that comes with a smoothly evolving curve length is a perfect relaxed alternative to overcome the difficulties. In particular, this section applies the smoothly evolving geodesic to the same example given above.

For some $x \in [-\pi, \pi]$, in addition to the point $(\cos(x), \sin(x))$ it represents, let $\theta_1, \theta_2 \in \mathbb{R}$ determine two initial velocities given on y_1 and y_2 such that the arcs emanating from them along the respective velocities arrive at x as

$$\begin{cases} y_1 \exp(\Delta_{[0, \theta_1, 0]}) = p_x \\ y_2 \exp(\Delta_{[0, \theta_2, 0]}) = p_x \end{cases}.$$

Then, as $\{x(t) : t \in [0, 1]\}$ varies on the unit circle, there exist unique smooth functions $\theta_1(t), \theta_2(t)$ of t such that $\theta_i(0) = \theta_i$ and $y_i \exp(\Delta_{[0, \theta_i(t), 0]}) = p_{x(t)}$. These geodesics correspond to the arcs on the unit circle that may be greater than the half-circle. Finally, the distance in the objective function realized by the shortest geodesic is replaced by the curve length of the smoothly evolving geodesics/arcs. **Figure 5.2** illustrates the notion of smoothly evolving arcs in the unit circle and the corresponding objective function.

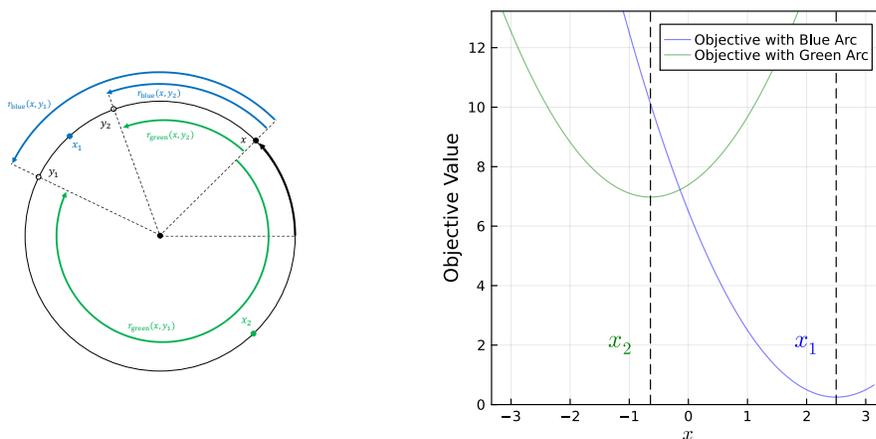


Figure 5.2: Illustration of the Karcher Mean with Smoothly Evolving Arcs on an Unit Circle

There are two different sets of initial θ_1, θ_2 presented in **Figure 5.2**, the blue set and the green set. For each set, the corresponding objective function degenerates to a globally smooth and convex

function that consists of the global minimum as the unique critical point. The green set finds the local minimum x_1 from the classic Karcher mean formulation and the blue set finds the local minimum x_2 . Even when the data point y_1 and y_2 overlaps, x_2 remains the global minimum to the green set. More importantly, the original non-smooth and non-convex objective function can be recovered from taking the minimum over the relaxed objective functions with all possible sets of initial velocity θ_1, θ_2 . This further validates the observation that distance realized by a shortest geodesic in the classic Karcher mean formulation is not appropriate.

In conclusion, the Karcher mean formulation in \mathbf{SO}_2 with the distance replaced by the curve length of a set of smoothly evolving geodesics has overcome the issues addressed in the previous section. In particular, the objective function with the velocities of the initial geodesic specified, namely the velocity-based objective function, is globally smooth and convex on the manifold and obtains a unique critical point as the global minimum. As long as the algorithmic objects vary in a smooth manner, the computation under any algorithm remains in the same framework of the velocity-based objective function and should converge to the corresponding global minimum. In other words, the computation of solving the optimization problem under this velocity-based objective function is reliable to the perturbation on the initial guess and the data set.

5.3 Velocity-Based Karcher Mean

In the \mathbf{SO}_2 example, utilizing the metric space nature in a Riemannian manifold in (5.1) yields theoretical and computational issues. While these issues cannot be handled by relaxing the global minimum condition to a local minimum condition, introducing the notion of the smoothly evolving arc/geodesic solves them naturally. This example suggests that the classic Karcher mean formulation (1.10) is not applicable for a Riemannian setting without some appropriate generalizations. Therefore, this section considers a more general statement that describes the Karcher mean in a length argument, which degenerates to the classic formulation in a Hilbert space. Based on that length argument, a novel Karcher mean formulation based on velocity argument is proposed as a special case specific to a Riemannian manifold setting.

5.3.1 Length-Based Karcher Mean Objective

Recall that the distance in a Hilbert space is realized by the length of the unique shortest curve between the given points. In this context, the classic Karcher mean with the objective (1.10) can be interpreted as a special case of the following statement.

Lemma 5.3.1. The Karcher mean of a set of data points $\{y_1, \dots, y_k\}$ in a metric space M finds a set of continuous curves, that connects each data point to itself, such that these curves realize the minimal sum of the squared lengths

$$\tilde{f}(\tilde{\tau}, x) := \frac{\sum_{i=1}^k l_{\tau_i}^2}{2k}, \tilde{\tau} = \{\tau_1, \dots, \tau_k\} \subset \Gamma(M) \quad (5.5)$$

where $\Gamma(M)$ denote the space of all continuous curves in M with well defined length, τ_i is a curve connecting $y_i = \tau_i(0)$ and $x = \tau_i(1)$ and l_{τ_i} is the length of $\tau_i(t), t \in [0, 1]$.

Consider a Hilbert space as the metric space M in which the shortest curve between given points is unique and realizes the distance with its length. Then, the minimal sum of the squared lengths is always achieved by the sum of the squared lengths of the shortest curves, i.e., the sum of the squared distances. The uniqueness of such shortest curves, given the data points and the Karcher mean, also makes it trivial to identify them. Eventually, the length-based argument given in **Lemma 5.3.1** degenerates to the classic formulation (1.10) in a Hilbert space, and it is referred to as the distance-based Karcher mean in this chapter.

When it comes to a Riemannian manifold \mathcal{M} where the points can be connected by multiple geodesics, the shortest curves between points are no longer globally continuous for arbitrary pair of points in \mathcal{M} . The \mathbf{SO}_2 example has already demonstrated the disadvantages in dropping the non-minimal geodesics for keeping the distance argument. Since the shortest curve constraint simplifies the length-based statement in **Lemma 5.3.1** only if it is uniquely and continuously defined over all pair of points on the metric space, one must resort back to the more general length-based statement to work with a Riemannian manifold. Furthermore, it necessary to further weaken the global constraint such that the mean associated to non-minimal geodesic can be included. This ideas lead to the following generalization of the length-based statement.

Definition 5.3.2. The generalized Karcher mean $(\tilde{\tau}_*, x_*)$ of a set of data points $\{y_1, \dots, y_k\}$ is a **local minimum** of the length-based objective (5.5) such that for any perturbation $(\tilde{\tau}, x) \neq (\tilde{\tau}_*, x_*)$, there is $\tilde{f}(\tilde{\tau}_*, x_*) \leq \tilde{f}(\tilde{\tau}, x)$ \square

Note that the perturbation includes the cases of $x_* = x$ while $\tilde{\tau}_* \neq \tilde{\tau}$ and the cases of $x_* \neq x$, i.e., it is both the curves and the shared endpoint x_* together that realize the local minimum. However, working on a space of curves $\Gamma(M)$ with infinite dimension is not desirable in many ways. The following section considers the generalized Karcher mean (the local minima) of the (5.5)

specific to a complete Riemannian manifold and derives a simplified version based on a velocity argument. This is similar to the process of the length-based arguments (5.5) degenerating to the classic distance-based Karcher mean (1.10) in a Hilbert space.

5.3.2 Velocity-Based Karcher Mean on a Riemannian Manifold

In a Riemannian manifold (\mathcal{M}, g) , consider a Riemannian geodesic $\gamma(t) = \text{Exp}_y(t \cdot v)$ with velocity $v \in T_y \mathcal{M}$ that connects arbitrary y and $x = \gamma(1)$. Although such a geodesic is no longer guaranteed to be a curve with the minimal length between x and y , it achieves the local minimal length in $\Gamma(\mathcal{M})$, the space of all continuous curves in \mathcal{M} with well defined length. In other words, for any other perturbed curve $\tau(t) \approx \text{Exp}_y(t \cdot v)$ that connects y and x , there is $l_\gamma \leq l_\tau$ where the length of a curve is obtained from the integration of $\sqrt{g_{\tau(t)}(\dot{\tau}(t), \dot{\tau}(t))}$. Note that in an inner product space like (\mathcal{M}, g) , the space of curves with well defined length can be further specified as follows. $\Gamma(\mathcal{M})$ is the space of continuous curves consists of the curves that can be partitioned into finite segments such that the aforementioned integral exists on each segment. But the characterization of $\Gamma(\mathcal{M})$ is not a primary focus since it is dropped in the simplified formulation.

On the other hand, the Hopf-Rinow theorem guarantees that the Riemannian exponential in a complete Riemannian manifold is smoothly defined on the entire tangent space at any point $y \in \mathcal{M}$. Suppose the manifold is also connected, which is a reasonable assumption for the Karcher mean problem and denote the set of initial velocities emanating from $y \in \mathcal{M}$

$$\mathfrak{V}(y, x) := \{v \in T_y \mathcal{M} : \text{Exp}_y(v) = x\} \subset T_y \mathcal{M}$$

that arrives at $x \in \mathcal{M}$ along the Riemannian geodesic $\gamma(t) := \text{Exp}_y(t_i \cdot v), t \in [0, 1]$. Then, $\mathfrak{V}(y, x)$ is well defined and nonempty for any $x, y \in \mathcal{M}$. Furthermore, the length-based argument given in **Definition 5.3.2** is simplified to the following velocity-based argument.

Proposition 5.3.3. *Let (\mathcal{M}, g) be a complete and connected Riemannian manifold and let $\mathbf{y} = \{y_1, \dots, y_k\} \subset \mathcal{M}$ be a set of data points. Then, any local minimum $(\tilde{\gamma}, x)$ of the length-based objective (5.5) is equivalent to a local minimum in the velocity-based objective*

$$f(\mathbf{v}, x) := \sum_{i=1}^k \frac{g_{y_i}(v_i, v_i)}{2k} = \sum_{i=1}^k \frac{g_x(w_i, w_i)}{2k}, \mathbf{v} = (v_1, \dots, v_k), v_i \in \mathfrak{V}(y_i, x), \quad (5.6)$$

with the Riemannian geodesics $\gamma_i(t) = \text{Exp}_{y_i}(t \cdot v_i)$ carrying velocities

$$\begin{cases} v_i := \dot{\gamma}_i(0) \in \mathfrak{V}(y_i, x) \subset T_{y_i} \mathcal{M} \\ w_i := \dot{\gamma}_i(1) \in T_x \mathcal{M} \end{cases}, \forall i = 1, \dots, k. \quad (5.7)$$

More specifically, for any local minimum (\mathbf{v}, x) of the velocity-based objective (5.6), $(\tilde{\gamma}, x)$ with $\gamma_i(t) = \text{Exp}_{y_i}(t \cdot v_i)$ is a local minimum of the length-based objective (5.5).

Proof. This statement follows from the fact that any curve in a complete Riemannian manifold has the local minimal length between the endpoints if and only if the curve is a Riemannian geodesic. Note that such a Riemannian geodesic may not be a shortest one.

Suppose there exists a local minimum $(\tilde{\tau}, x)$ to the length-based objective such that the $\tau_i(t)$ is not a Riemannian geodesic. Then, the length operator evaluated at $\tau_i \in \Gamma(\mathcal{M})$ is not a local minimum, i.e., there exists a perturbed $\gamma_i \neq \tau_i \in \Gamma(\mathcal{M})$ with the same endpoints, such that $l_{\gamma_i} < l_{\tau_i}$, i.e., $\tilde{f}(\{\tau_1, \dots, \gamma_i, \dots, \tau_k\}, x) < \tilde{f}(\{\tau_1, \dots, \tau_i, \dots, \tau_k\}, x)$. This is a contradiction to the assumption. Therefore, any local minimum $(\tilde{\gamma}, x)$ to (5.5) consists of Riemannian geodesics in the form of $\gamma_i(t) = \text{Exp}_{y_i}(t \cdot v_i), v_i \in \mathfrak{V}(y_i, x)$.

It remains to show that the corresponding (\mathbf{v}, x) is a local minimum of the velocity-based objective (5.6). Notice that for any $v_i \in T_{y_i}\mathcal{M}$, it uniquely defines a curve $\text{Exp}_{y_i}(t \cdot v_i) \in \Gamma(\mathcal{M})$ when \mathcal{M} is a complete Riemannian manifold and the length of such a curve is given by $\sqrt{g_{y_i}(v_i, v_i)}$. Therefore, the velocity-based objective (5.6) can be viewed as a restricted subproblem of the length objective as

$$\tilde{f}(\tilde{\tau}, x) := \frac{l_{\tilde{\tau}}^2}{k}, \tilde{\tau} = \{\tau_1, \dots, \tau_k\}, \tau_i(t) = \text{Exp}_{y_i}(t \cdot v_i).$$

Then, any local minimum to the full problem (5.5) is naturally a local minimum to the above subproblem, i.e., a local minimum to the velocity-based objective (5.6). \square

Note that in both the length-based objective and the velocity-based objective, the shared arriving point $x = \tau_i(1)$ in (5.5) and $x = \text{Exp}_{y_i}(v_i)$ in (5.6) for all $i = 1, \dots, k$ is redundant. However, keeping the shared x in the objective emphasizes that the perturbation to the objective should yield a shared motion at the point x as a crucial characteristic. In the length-based objective (5.5), if there is an infinitesimal change $(\tilde{\delta}_{\tilde{\tau}}, \Delta_x)$ applied to $(\tilde{\tau}, x)$ where $\Delta_x \in T_x\mathcal{M}$ and $\tilde{\delta}_{\tilde{\tau}} = \{\delta_i(t) \in T_{\tau_i(t)}\mathcal{M} : t \in [0, 1]\}_{i=1}^k$, then there must be

$$\delta_i(1)(t) = \Delta_x, \forall i = 1, \dots, k.$$

For the velocity-based objective (5.6), the redundant x becomes more informative in expressing the tangents to the feasible set as follows.

Proposition 5.3.4. *Consider the feasible set that is constrained by (5.7) and let (\mathbf{v}, x) be a point in the feasible set. Then, the tangent to the feasible set around (\mathbf{v}, x) takes the form of*

$$(\Delta_{\mathbf{v}}, \Delta_x) := \left(\left\{ (\text{D Exp}_{y_i})_{v_i}^{-1} [\Delta_x] \right\}_{i=1}^k, \Delta_x \right) \in T_{y_1} \mathcal{M} \times \cdots \times T_{y_k} \mathcal{M} \times T_x \mathcal{M}. \quad (5.8)$$

Proof. For each geodesic $\gamma_i(t) = \text{Exp}_{y_i}(t \cdot v_i)$ with the fixed emanating point y_i , the infinitesimal change $\Delta_x \in T_x \mathcal{M}$ on the arriving point x is fully characterized by the infinitesimal change $\Delta_{v_i} \in T_{v_i}(T_{y_i} \mathcal{M}) = T_{y_i} \mathcal{M}$ where $\Delta_{v_i} = (\text{D Exp}_{y_i})_{v_i}^{-1} [\Delta_x]$ have been carefully discussed in the smoothly evolving geodesic problem in previous chapters. \square

Note that special considerations are needed if there is an operator $\text{D Exp}_{y_i} : T_{y_i} \mathcal{M} \rightarrow T_x \mathcal{M}$ that is not invertible. In particular, if Δ_x does not fall into the range space of D Exp_{y_i} , such a motion to x within (\mathbf{v}, x) is not possible in the feasible set constrained by (5.7). It implies that the feasible set may not be a manifold globally. It remains future work to investigate this specific case. Nevertheless, the tangent to the feasible set characterized in (5.8) yields the following differential operator and gradient operator to the velocity-based objective (5.6).

Proposition 5.3.5. *The differential of the velocity-based objective (5.6) along $(\Delta_{\mathbf{v}}, \Delta_x)$ in the form of (5.8) is given by*

$$\begin{aligned} \text{D} f(\mathbf{v}, x)[(\Delta_{\mathbf{v}}, \Delta_x)] &= \sum_{i=1}^k \frac{g_x(w_i, \Delta_x)}{k} = g_x \left(\sum_{i=1}^k \frac{w_i}{k}, \Delta_x \right) \\ &= \sum_{i=1}^k \frac{g_{y_i}(v_i, (\text{D Exp}_{y_i})_x^{-1} [\Delta_x])}{k}. \end{aligned} \quad (5.9)$$

Proof. The first equation follows directly from differentiating the Riemannian metric with respect to the perturbation to the root along Δ_x . By doing so, the \mathbf{v} terms in the objective are independent and therefore dropped. For any $w \in T_x \mathcal{M}$, there is $\text{D} g_x(w, w)[\Delta_x] = 2g_x(w, \Delta_x)$. The second equality follows from the bi-linearity in the Riemannian metric. The third equality is obtained through the chain rule in differentiating the v terms as follows

$$\begin{aligned} \text{D} f_x(\mathbf{v})[(\Delta_{\mathbf{v}}, \Delta_x)] &:= \sum_{i=1}^k \frac{1}{2k} \text{D} (g_{y_i}(v_i, v_i))_{v_i} [(\Delta_{\mathbf{v}}, \Delta_x)] \\ &= \sum_{i=1}^k \frac{g_{y_i}(v_i, \text{D}(v_i)[(\Delta_{\mathbf{v}}, \Delta_x)])}{k} = \sum_{i=1}^k \frac{g_{y_i}(v_i, \Delta_{v_i})}{k} \end{aligned}$$

where $\Delta_{v_i} = (\text{D Exp})_{v_i}^{-1} [\Delta_x]$ as characterized in (5.8). \square

Corollary 5.3.6. The gradient descent direction

$$\text{grad } f_{(\mathbf{v}, x)} := (\text{grad } f_{v_1}, \dots, \text{grad } f_{v_k}, \text{grad } f_x)$$

of the velocity-based objective evaluated at (\mathbf{v}, x) is given by the $\text{grad } f_x$ as follows.

$$\begin{cases} \text{grad } f_x = \sum_{i=1}^k \frac{w_i}{k} \\ \text{grad } f_{v_i} = (\text{Exp}_{y_i})_{v_i}^{-1} [\text{grad } f_x] \end{cases}, \forall i = 1, \dots, k. \quad (5.10)$$

Furthermore, any (\mathbf{v}, x) is a critical local minimum if $\text{grad } f_{(\mathbf{v}, x)} = \mathbf{0}$, i.e., $\sum_{i=1}^k w_i = \mathbf{0}$.

5.3.3 Velocity-Based Karcher Mean on the Special Orthogonal Group

In the special orthogonal group, the tangent space $T_Q \mathbf{SO}_n$ at any $Q \in \mathbf{SO}_n$ is equivalent to \mathbf{Skew}_n as discussed in previous chapters. Applying the velocity-based Karcher mean notion to \mathbf{SO}_n yields the following specific formulation.

Proposition 5.3.7. *The objective of the velocity-based Karcher mean on \mathbf{SO}_n with the data set $\{Y_1, \dots, Y_k\} \subset \mathbf{SO}_n$ is given by*

$$f(\mathbf{S}, Q) = \sum_{i=1}^k \frac{\text{tr}(S_i^T S_i)}{4k}, \mathbf{S} = \{S_1, \dots, S_k\}, S_i \in \mathbb{E}_{Y_i^T Q}^{-1}, \quad (5.11)$$

with the gradient given by

$$\text{grad}(f_Q) \mathbf{s} = \frac{1}{2k} Q \sum_{i=1}^k S_i. \quad (5.12)$$

Proof. It follows from $g_Q(QA, QB) = \text{tr}(A^T B)/2$ and the geodesic $\text{Exp}_{Y_i}(tS_i)$, emanating from Y_i and arriving at Q , has $S_i \in \mathbb{E}_{Y_i^T Q}^{-1}$ such that $\exp(S_i) = Y_i^T Q$. \square

Apply the notion of the smoothly evolving geodesic to further impose a smooth structure between Q and \mathbf{S} as follows.

Definition 5.3.8. The velocity-based objective with the smooth constraint along $\{Q(t) : t \in [0, 1]\} \subset \mathbf{SO}_n$ is given by the solutions $\{X_i(t) : t \in [0, 1]\}$ to the smoothly evolving geodesic problem $Y_i^T Q(t)$ with initial condition $Q' \in \mathbf{SO}_n$ and $X_i' \in \mathbf{Skew}_n$, such that

$$\begin{cases} Q(0) = Q' \\ X_i(0) = X_i' \\ Y_i \exp(X_i(t)) = Q(t) \end{cases}, \forall i = 1, \dots, k.$$

The corresponding objective is $f(Q(t), \mathbf{X}(t)), t \in [0, 1]$ where $\mathbf{X}' = \{X'_1, \dots, X'_k\}$ is the initial condition and $\mathbf{X}(t) = \{X_1(t), \dots, X_k(t)\}$ are the solutions to the smoothly evolving geodesic problems to $Q(t) = Y_i \exp(X_i(t))$. \square

Recall that the skew symmetric matrix moving in a smooth manner that yields any geodesics in \mathbf{SO}_n constructs the smooth structure $\mathfrak{C}_{\mathbf{SO}_n} \subset \mathbf{Skew}_n$. This smooth structure can be imposed to the objective as follows.

Proposition 5.3.9. *The velocity-based objective (5.11) of a set of data points $\mathbf{Y} = \{Y_1, \dots, Y_k\} \subset \mathbf{SO}_n$ around every point (\mathbf{S}, Q) satisfying the constraint (5.7) is smooth.*

Proof. Firstly, the co-manifold characterization discussed in **Proposition 4.5.4** guarantees smoothness around every (S_i, Q) from the geodesic $\gamma_i(t) := Y_i \exp(t \cdot S_i)$ connecting $\gamma_i(0) = Y_i$ and $\gamma_i(1) = Q$. For any tangent $(\Delta_{\mathbf{S}}, \Delta_Q)$, satisfying (5.7), to the feasible set around (\mathbf{S}, Q) , it can be decomposed into the tangents (Δ_{S_i}, Δ_Q) on each (S_i, Q) . These tangents yields smoothly varying geodesics on γ_i as discussed in previous chapters. Therefore, the smoothly varying geodesics have smoothly varying lengths with respect to the tangents. Secondly, notice that the velocity-based objective is the sum of the squared lengths in these smoothly varying geodesics $\tilde{\gamma} = \{\gamma_1, \dots, \gamma_k\}$. As the length of these geodesics are smoothly varying along any tangents to the feasible set constrained by (5.7), the objective velocity-based objective (5.11) is smoothly varying on the feasible set. \square

While the feasible set constrained by (5.7) may not be a manifold, **Proposition 5.3.9** states that the velocity-based objective (5.11) where the point (\mathbf{S}, Q) in its feasible set moves in the smoothly evolving geodesic manner as specified in **Proposition 4.5.4** maintains smoothness. For the cases where there is a S_i on the conjugate locus Conj_{I_n} , the feasible set as well as the tangents to it are restricted due to the loss of invertibility in $D \exp_{S_i}$ but the velocity-based objective remains smooth in this more restricted case. While the gradient derived in (5.12) does not exists in this special case, it is still possible to move the (\mathbf{S}, Q) in a smooth manner such that the objective values decrease. The details of this specific technique is not finalized enough for this dissertation and it is therefore left as the future work. Suppose the special case with $S_i \in \text{Conj}_{I_n}$ is not encountered during the

computations, a prototype of a gradient descent method on the velocity-based Karcher mean with the smoothly evolving geodesic constraint is constructed as follows.

Algorithm 6: A Gradient Descent Method for the Velocity-Based Karcher Mean

Data: $Y_1, \dots, Y_k \in \mathbf{SO}_n$
Input: Initial condition $Q_{(0)}$ and $\Delta_{i,(0)} \in \mathbf{Skew}_n$ such that $Y_i \exp(\Delta_{i,(0)}) = Q_{(0)}$
Output: A velocity-based Karcher mean Q with $Y_i \exp(\Delta_i) = Q$ and $\sum_{i=1}^k \Delta_i = \mathbf{0}$

- 1 $j \leftarrow 0;$ // Iteration counter j
- 2 $S_{(j)} \leftarrow \sum_{i=1}^k \Delta_{i,(j)};$
- 3 **while** $\|S_{(j)}\|_F > \varepsilon$ **do**
- 4 Solve $\Delta_{i,(j)}(t)$ from the smoothly evolving geodesic problem $Y_i^T Q_{(j)} \exp(tS_{(j)})$ with
 initial condition $\Delta_{i,(j)}(0) = \Delta_{i,(j)};$
- 5 Line search on step size α along $Q_{(j)} \exp(tS_{(j)})$ and $\Delta_{i,(j)}(t);$
- 6 $Q_{(j+1)} \leftarrow Q_{(j)} \exp(\alpha S_{(j)});$
- 7 $\Delta_{i,(j+1)} \leftarrow \Delta_{i,(j)}(\alpha);$
- 8 $S_{(j+1)} \leftarrow \sum_{i=1}^k \Delta_{i,(j+1)};$
- 9 $i \leftarrow i + 1;$

10 **Return** $Q_{(j)}, \{\Delta_{1,(j)}, \dots, \Delta_{k,(j)}\};$

5.4 Numerical Experiments

To demonstrate the smoothness nature in the velocity-based Karcher mean computations, this section considers a set of data points $\{Y_i(t)\}_{i=1}^k \subset \mathbf{SO}_n$ that are smoothly varying with respect to $t \in [0, 1]$ by

$$Y_i(t) = Y_i \exp(t \cdot \Delta_i), S_i \in \mathbf{Skew}_n$$

where $Y_i(0) = Y_i, i = 1, \dots, k$ are denoted as the initial data point. For every specific $t \in [0, 1]$, the corresponding set of data points $\{Y_i(t)\}_{i=1}^k$ defines a Karcher mean problem labelled by t .

Then, let $(\mathbf{S}(t), Q(t)) = ((S_1(t), \dots, S_k(t)), Q(t))$ be a velocity-Karcher mean computed by some given condition such that they satisfy the following conditions for any t .

$$\begin{cases} Y_i(t) \exp(S_i(t)) = Q(t), i = 1, \dots, k \\ \sum_{i=1}^k S_i(t) = \mathbf{0}. \end{cases}$$

When the initial guess of **Algorithm 6** is set appropriately, the computed means $(\mathbf{S}(t), Q(t))$ that solves the t -labelled velocity-Karcher mean problem are expected to be smoothly dependent on the t . As a comparison, the classic Karcher mean of the t -labelled data $\{Y_i(t)\}_{i=1}^k$ is also computed by

the algorithm proposed in [20] with the iterative step

$$\begin{aligned}\Omega^{\text{iter}} &\leftarrow \frac{1}{2k} \sum_{i=1}^k \log(\exp(S^{\text{iter}})^T Y_i) \\ Q^{\text{iter}+1} &\leftarrow Q^{\text{iter}} \exp(\Omega^{\text{iter}}) \\ S^{\text{iter}+1} &\leftarrow \log(Y_i^T Q^{\text{iter}+1})\end{aligned}$$

where $Q_i \in \mathbf{SO}_n$ are the data point, $X^{\text{iter}} \in \mathbf{Skew}_n$ and $E^{\text{iter}} = \exp(X^{\text{iter}}) \in \mathbf{SO}_n$ is the current guess of the Karcher mean and $\log : \mathbf{SO}_n \rightarrow \mathbf{Skew}_n$ is the principal logarithm.

Two experiment setups are designed to demonstrate the feature discussed above. In the first setup, consider $(\mathbf{S}, Q) := (\mathbf{S}(0), Q(0))$ solves the Karcher mean problem with the initial data set $\{Y_1, \dots, Y_k\}$ such that

$$\begin{cases} Y_i \exp(S_i) = Q, i = 1, \dots, k \\ \sum_{i=1}^k S_i = \mathbf{0}. \end{cases}.$$

Let the data points $Y_i(t)$ leave the initial mean Q along $Y_i(t) = Y_i \exp(-tS_i)$. Notice that $Q = Y_i \exp(S_i)$, which means $Y_i(t) = Y_i \exp(-tS_i) = Q \exp(-(1+t)S_i)$. In this special setup, one can easily verify that $((1+t)\mathbf{S}, Q)$ is a solution to the t -labelled Karcher mean problem, i.e., the computed Q is expected to stay at where it is started. **Figure 5.3** reports two moments with $t = 2.8$ on the left and $t = 2.9$ on the right. The initial mean Q is denoted as the “true mean” and the crosses are the computed velocity-based Karcher mean using the nearby logarithm and the distance-based Karcher mean using the principal logarithm. The geodesics $\{Y_i(t) \exp(s \cdot (1+t)S_i) : s \in [0, 1]\}_{i=1}^k$ that connects the data point $Y_i(t)$ to the true mean Q are also plotted as the blue dashed line. In the experiment, for all $t \leq 2.8$, both the distance-based Karcher mean and the velocity-based Karcher mean converge to the true mean. As t increase to $t = 2.9$ and beyond, the velocity-based Karcher mean still converge to the true mean, while the distance-based Karcher mean suddenly jumps to a different converged point.

The second experiment has a more complicated motions in the varying endpoints. Instead of choosing $\Delta_i = -S_i$ to force a true mean Q staying at the same point, the skew symmetric Δ_i are now sample randomly. It causes a more complicated behavior of the computed velocity-based Karcher mean as the data points $Y_i(t) = Y_i \exp(t\Delta_i)$ varies with respect to t . In this experiment, $k = 10$ data points are used. As the $Y_i(t)$ varies smoothly with respect to t , it generates ten smooth trajectories, which are the black trajectories reported in **Figure 5.4**. For each t -labelled Karcher mean problem, the computed velocity-based Karcher mean and the computed distance-based Karcher mean are

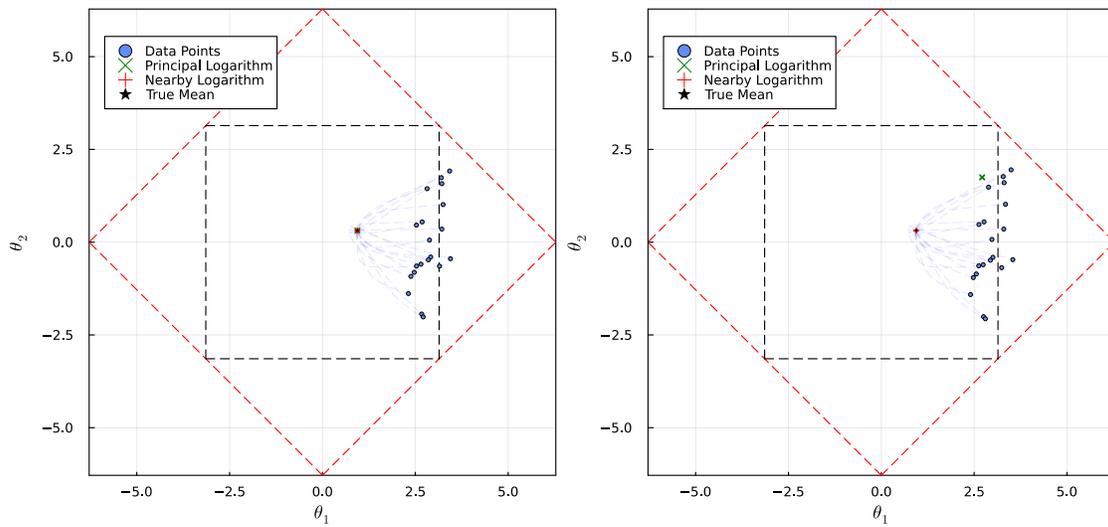


Figure 5.3: Karcher Mean with Evenly Spreading Spreading Data.

reported in green and blue respectively. The green and the blue trajectories behave as expected. The green trajectory appears to be smooth with respect to t , while the blue trajectory is broken into continuous segments. In other words, the experiment suggests that the computed velocity-based Karcher mean maintains the smooth dependence while the compute distance-based Karcher mean does not have the continuous feature.

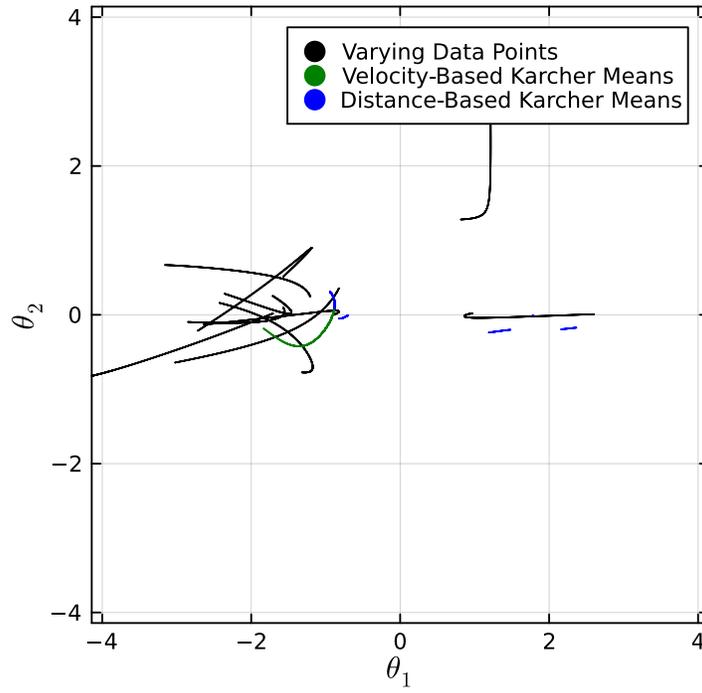


Figure 5.4: Karcher Mean with Randomly Spreading Spreading Data.

CHAPTER 6

THE ENDPOINT STIEFEL GEODESIC PROBLEM WITH THE CANONICAL METRIC

6.1 Introduction

The Stiefel manifold $\mathbf{St}_{n,p}$ consists of $n \times p$ orthogonal matrices

$$\mathbf{St}_{n,p} := \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\}$$

and it is one of the most important manifolds and arises in many applications. A point $U = [U_1 \ \cdots \ U_p]$ on the Stiefel manifold is usually considered as a preferred basis of the p -dimensional vector subspace $\text{col}(U)$ in \mathbb{R}^n so that for any point $x \in \text{col}(U) \subset \mathbb{R}^n$ on there is a unique \mathbb{R}^p representation $\alpha = (\alpha_1, \dots, \alpha_p)^T$ with $x = U\alpha = \sum_i^p \alpha_i U_i$. For example, the principal component analysis (PCA) on a distribution on n components that keeps p primary factors is represented by a rank p covariance matrix C with a spectral decomposition

$$C = U\Lambda U^T$$

where $U \in \mathbf{St}_{n,p}$ and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$. Each coefficient of the unique combination in a basis vector U_i is a principal component since it is not correlated with other coefficients.

The Riemannian structure of the Stiefel manifold is studied in [11] and two Riemannian metrics are presented, namely the embedded metric and the canonical metric. Intuitively, for the curve $U(t), t \in [0, 1]$ on $\mathbf{St}_{n,p}$, the embedded metric puts it in $\mathbb{R}^{n \times p}$ before measuring the length while the canonical metric puts it in \mathbf{SO}_n with some $Q(t) = [U(t) \ U_\perp(t)]$ and the length is measured by the length of $Q(t)$.

This chapter investigates the quotient structure in the \mathbf{SO}_n that defines the canonical metric on the Stiefel manifold and proposes a novel algorithm for solving the endpoint geodesic problem with the tools and primitives developed in previous chapters. In particular, the endpoint geodesic problem on the Stiefel manifold is equivalent to solving the following matrix equation

$$\exp\left(\begin{bmatrix} A & -B^T \\ B & \mathbf{0} \end{bmatrix}\right) = [V \ V_\perp] \quad (6.1)$$

where $V \in \mathbf{St}_{n,p}$ is taken as a parameter and the $A \in \mathbf{Skew}_p$, $B \in \mathbb{R}^{n-p \times p}$ and $V_\perp \in \mathbf{St}_{n,n-p}$ as a special orthogonal completion to V are variables that need to be solved from (6.1). Note that the partitioning of a skew symmetric $S = \begin{bmatrix} A & -B^T \\ B & C \end{bmatrix} \in \mathbf{Skew}_n$ that splits S into the blocks $A \in \mathbf{Skew}_p$, $B \in \mathbb{R}^{(n-p) \times p}$ and the remaining block in $C \in \mathbf{Skew}_{(n-p)}$ is often used in this chapter. For simplicity, the notation is introduced as follows

$$M = M_{[A,B,C]} := \begin{bmatrix} A & -B^T \\ B & C \end{bmatrix} \in \mathbf{Skew}_n \quad (6.2)$$

where $A \in \mathbf{Skew}_p$ and $C \in \mathbf{Skew}_{n-p}$ are assumed to be skew symmetric. Under this notation, (6.1) becomes $\exp(X_{[A,B,0]}) = [V \quad V_\perp]$ with the constrained unknown X and V_\perp .

6.2 Preliminaries

6.2.1 Riemannian Submersion

The canonical metric on the Stiefel manifold is constructed by a Riemannian submersion from \mathbf{SO}_n . A mapping $\varphi : \mathcal{M} \rightarrow \mathcal{N}$ from a manifold \mathcal{M} to another manifold \mathcal{N} is a submersion if it is onto, differentiable and its differential $D\varphi_x : T_x\mathcal{M} \rightarrow T_y\mathcal{N}$ is also onto for any $x \in \mathcal{M}$ and $y = \varphi(x) \in \mathcal{N}$. In other words, a submersion map φ specifies a linear operator $T\mathcal{M} \rightarrow T\mathcal{N}$, such that any motion on \mathcal{M} can be translated to a motion on $T\mathcal{N}$. When \mathcal{M} and \mathcal{N} have the same dimension, i.e., when $T_x\mathcal{M}$ and $T_y\mathcal{N}$ have the same dimension for any $x \in \mathcal{M}$ and any $y \in \mathcal{N}$, the submersion becomes a diffeomorphism. In this case, the two manifolds \mathcal{M} and \mathcal{N} share the identical differentiable structure and objects can be translated back and forth. In general, the dimension $d_{\mathcal{M}}$ of \mathcal{M} is larger than the dimension $d_{\mathcal{N}}$ of \mathcal{N} . In this case, the differentiable structure in \mathcal{M} determines a differentiable structure but not vice versa, as the motion at $y \in \mathcal{N}$, characterized as a tangent vector $\eta_y \in T_y\mathcal{N}$, cannot describe all motion at some $x \in \mathcal{M}$ where $\varphi(x) = y$. Those lost motions at x describe a submanifold structure as follows.

Proposition 6.2.1 (Inverse Theorem on Manifolds). *[4] Consider a submersion $\varphi : \mathcal{M} \rightarrow \mathcal{N}$ with dimensions $d_{\mathcal{M}} > d_{\mathcal{N}}$. For any $y \in \mathcal{N}$, the preimage*

$$\mathcal{F}(y) := \{x \in \mathcal{M} : \varphi(x) = y\} \quad (6.3)$$

forms a submanifold in \mathcal{M} with dimension $d_{\mathcal{F}(y)} = d_{\mathcal{M}} - d_{\mathcal{N}}$, which is denoted as a fiber over y .

It is clear that a fiber $\mathcal{F}(y)$ is specified by $y \in \mathcal{N}$ or any $x \in \mathcal{F}(y)$ in itself as $y = \varphi(x)$ if the submersion φ is given in the context. For simplicity, the fiber is denoted as \mathcal{F} with the y term

dropped if both the submersion and the y or the $x \in \mathcal{F}(y)$ are given in the context. The fact that a fiber is formed by preimage under φ immediate leads to the following observation.

Corollary 6.2.2. Consider a submersion $\varphi : \mathcal{M} \rightarrow \mathcal{N}$ and a fiber \mathcal{F} over $y \in \mathcal{N}$ with $x \in \mathcal{M}$. Then the tangent space $T_x\mathcal{F}$ is a subspace of $T_x\mathcal{M}$ and it is also the null space of $D\varphi_x : T_x\mathcal{M} \rightarrow T_y\mathcal{N}$, i.e., $D\varphi_x[\xi_x] = \mathbf{0}, \forall \xi_x \in T_x\mathcal{F}$.

A horizontal structure on \mathcal{M} further equips a submersion structure $\varphi : \mathcal{M} \rightarrow \mathcal{N}$ with a set complementary subspaces to $T_x\mathcal{F}$. These complementary subspaces uniquely defines a pseudo inverse $D\varphi_x^\dagger : T_y\mathcal{N} \rightarrow T_x\mathcal{M}$ such that the motion at y in \mathcal{N} can be uniquely identified by a unique motion at $x \in \mathcal{F}$ in \mathcal{M} .

Definition 6.2.3. Consider a submersion $\varphi : \mathcal{M} \rightarrow \mathcal{N}$ with the subspaces $T\mathcal{F} \subset T\mathcal{M}$ that assigns $T_x\mathcal{F}$ at any $x \in \mathcal{M}$. A horizontal structure \mathbb{H} comprises a set of complementary subspaces \mathbb{H}_x to $T_x\mathcal{F}$ at any $x \in \mathcal{M}$ (that is smooth with respect to the foot x), i.e.,

$$\forall \xi_x \in \mathcal{M}, \exists! v_\xi \in T_x\mathcal{F}, h_\xi \in \mathbb{H}_x, \text{ such that } \xi_x = v_\xi + h_\xi.$$

As a complementary notation to the notation of horizontal, the tangent space to the fiber $T_x\mathcal{F}$ is denoted as the vertical space $\mathbb{V}_x := T_x\mathcal{F}$. □

Definition 6.2.4. Given a submersion $\varphi : \mathcal{M} \rightarrow \mathcal{N}$ and a horizontal structure \mathbb{H} , a horizontal lift (to \mathbb{H}) is the unique pseudo inverse $D\varphi_x^\mathbb{H} : T_y\mathcal{N} \rightarrow \mathbb{H}_x$, i.e., the unique linear operator $T_y\mathcal{N} \rightarrow \mathbb{H}_x$ satisfying the following conditions.

$$\begin{cases} \left(D\varphi_x^\mathbb{H} \circ D\varphi_x \circ D\varphi_x^\mathbb{H} \right) [\eta_y] = \eta_y, \forall \eta_y \in T_y\mathcal{N} \\ \left(D\varphi_x \circ D\varphi_x^\mathbb{H} \circ D\varphi_x \right) [h_x] = h_x, \forall h_x \in \mathbb{H}_x \end{cases} \quad (6.4)$$

□

When the manifold \mathcal{M} is equipped with a Riemannian metric g , the orthogonal complementary subspace to $T_x\mathcal{F}$ becomes a natural choice of a horizontal structure. When there are further consistencies in the Riemannian metric evaluated at the lifted vectors among a fiber, the submersion becomes a *Riemannian submersion* as defined below.

Definition 6.2.5. Consider a submersion $\varphi : \mathcal{M} \rightarrow \mathcal{N}$ from a Riemannian manifold (\mathcal{M}, g) where g is a Riemannian metric. Then, the submersion is a *Riemannian submersion* if the horizontal lift to

$$\mathbb{H} = \mathbb{V}_\perp := \bigcup_{x \in \mathcal{M}} \{h_x \in T_x\mathcal{M} : g_x(v_x, h_x) = 0, \forall v_x \in \mathbb{V}_x\} \quad (6.5)$$

satisfies

$$g_{x_1} \left(D \varphi_{x_1}^{\vee\perp}[\eta_y], D \varphi_{x_1}^{\vee\perp}[\zeta_y] \right) = g_{x_2} \left(D \varphi_{x_2}^{\vee\perp}[\eta_y], D \varphi_{x_2}^{\vee\perp}[\zeta_y] \right) \quad (6.6)$$

$$\forall y \in \mathcal{N}, \forall \eta_y, \zeta_y \in T_y \mathcal{N}, \forall x_1, x_2 \in \mathcal{F}(y).$$

The value in (6.6) that depends on η_y and ζ_y but not $x \in \mathcal{F}(y)$ is denoted as

$$g_y^\varphi(\eta_y, \zeta_y) := g_x \left(D \varphi_x^{\vee\perp}[\eta_y], D \varphi_x^{\vee\perp}[\zeta_y] \right), \forall x \in \mathcal{F}(y). \quad (6.7)$$

□

There are many nice properties induced from a Riemannian submersion and those that are relevant to this dissertation are summarized as follows.

Proposition 6.2.6. *Given a Riemannian submersion $\varphi : \mathcal{M} \rightarrow \mathcal{N}$ from (\mathcal{M}, g) and consider $y \in \mathcal{M}$ and $x \in \mathcal{F}$, then the following statements hold.*

1. *The function $g_y^\varphi(\eta_y, \zeta_y) : T_y \mathcal{N} \times T_y \mathcal{N} \rightarrow \mathbb{R}$ is an inner product operator and the collection of them for all $y \in \mathcal{N}$ forms a Riemannian metric g^φ on \mathcal{N} .*
2. *For any $y \in \mathcal{N}$, any $\eta_y \in T_y \mathcal{N}$ and any $x \in \mathcal{F}(y)$, there is*

$$\varphi \left(\text{Exp}_x(t \cdot D \varphi_x^{\vee\perp}[\eta_y]) \right) = \text{Exp}_y^\varphi(t \cdot \eta_y), \forall t \in [0, 1], \quad (6.8)$$

where Exp_x and Exp_y^φ are the respective Riemannian exponential in (\mathcal{M}, g) and (\mathcal{N}, g^φ) . In other words, any geodesic in \mathcal{M} with horizontal velocity $D \varphi_x^{\vee\perp}[\eta_y] \in \mathbb{H}_x$ maps to a geodesic in \mathcal{N} under g^φ and vice versa.

3. *The geodesic $\text{Exp}_x(t \cdot D \varphi_x^{\vee\perp}[\eta_y]), t \in [0, 1]$ is a minimal geodesic in (\mathcal{M}, g) if and only if the geodesic $\text{Exp}_y^\varphi(t \cdot \eta_y), t \in [0, 1]$ is a minimal geodesic in (\mathcal{N}, g^φ) .*

6.2.2 Stiefel Manifold with the Canonical Metric

The canonical metric on the Stiefel manifold is a textbook-example on constructing a Riemannian structure from a submersion. This part summarizes some important objects and statements relevant to the dissertation. More details can be found in [11].

Consider the special orthogonal group \mathbf{SO}_n equipped with the classic metric

$$g_Q(QS, QX) = \frac{1}{2} \text{tr}(S^T X), \forall QX, QS \in T_Q \mathbf{SO}_n = \{QX : X \in \mathbf{Skew}_n\}$$

then the map that takes the first p columns out of a special orthogonal matrix as

$$\varphi : \mathbf{SO}_n \rightarrow \mathbf{St}_{n,p}$$

$$Q \mapsto QI_{n,p} = Q \begin{bmatrix} I_p \\ \mathbf{0} \end{bmatrix}$$

is a Riemannian submersion. The fiber over U is given by $\mathcal{F}(U) = \{Q \in \mathbf{SO}_n : QI_{n,p} = U\}$. More importantly, it can be characterized by the Q in itself as

$$\mathcal{F} = \left\{ Q \exp \left(\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Z \end{bmatrix} \right) : Z \in \mathbf{Skew}_{n-p} \right\},$$

which induces the vertical and horizontal spaces as follows.

$$\begin{cases} \mathbb{V}_Q = \left\{ Q \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & C \end{bmatrix} : C \in \mathbf{Skew}_{n-p} \right\} \\ \mathbb{H}_Q = \left\{ Q \begin{bmatrix} A & -B^T \\ B & \mathbf{0} \end{bmatrix} : A \in \mathbf{Skew}_p, B \in \mathbb{R}^{(n-p) \times p} \right\}. \end{cases}$$

The horizontal spaces further yields a characterization on the tangent space $T_U \mathbf{St}_{n,p}$ when there is a special orthogonal completion U_\perp to U is given in $Q = [U \ U_\perp] \in \mathbf{SO}_n$, such that

$$\begin{aligned} T_U \mathbf{St}_{n,p} &= \{D\varphi_Q[h_Q] : h_Q \in \mathbb{H}_Q\} \\ &= \left\{ UA + U_\perp B : A \in \mathbf{Skew}_p, B \in \mathbb{R}^{(n-p) \times p} \right\} \end{aligned}$$

where $D\varphi_Q[h_Q] = h_Q I_{n,p} = [U \ U_\perp] \begin{bmatrix} A & -B^T \\ B & \mathbf{0} \end{bmatrix} I_{n,p} = UA + U_\perp B$. Then, the horizontal lift as an pseudo inverse order is $D\varphi_Q^\perp[UA + U_\perp B] = Q \begin{bmatrix} A & -B^T \\ B & \mathbf{0} \end{bmatrix}$.

The Riemannian geodesic in $(\mathbf{St}_{n,p}, g^\varphi)$ is given by

$$\text{Exp}_U^\varphi(t \cdot (UA + U_\perp B)) = \text{Exp}_Q \left(t \cdot Q \begin{bmatrix} A & -B^T \\ B & \mathbf{0} \end{bmatrix} \right) = Q \exp \left(t \cdot \begin{bmatrix} A & -B^T \\ B & \mathbf{0} \end{bmatrix} \right)$$

where the Riemannian geodesic in (\mathbf{SO}_n, g) is given by $\text{Exp}_Q(t \cdot QX) = Q \exp(t \cdot X)$ with the matrix exponential $\exp : \mathbf{Skew}_n \rightarrow \mathbf{SO}_n$.

6.2.3 Related Work

There has been previous work done on solving the endpoint geodesic problem under both metrics. For the embedded metric, [6] provides a nice solution. For the canonical metric, [42][**Algorithm 4**] proposes the state-of-the-art algorithm that has great performance for sufficiently close endpoints. There is also a recent work done in [33] with more details given in [32] that takes the shooting approach with Newton direction obtained in a system solve. The work in [29] takes a similar approach but proposes a Newton algorithm on the shooting approach with the Hessian operator approximated by difference of Fréchet derivatives.

The state-of-the-art algorithm proposed in [42] is referred to as the BCH algorithm, which is named due to the essential Baker-Campbell-Hausdorff expansion utilized in the algorithm. Among

existing algorithms on solving the endpoint geodesic problem on the Stiefel manifold under the canonical metric, it is the main competitor with the algorithm developed in this chapter. The BCH solver exploits the fiber structure of $\varphi^{-1}(V)$ and tries to find a horizontal initial velocity starting from chosen $Q \in \varphi^{-1}(U)$ so that it ends at the fiber $\varphi^{-1}(V)$.

The BCH algorithm tries to get a global exact step update to bring $\bar{\xi}_i$ directly to a solution on $\mathbb{H}_{\bar{\tau}}$ by solving

$$\exp\left(\begin{bmatrix} A_i & -B_i^T \\ B_i & C_i \end{bmatrix} + \begin{bmatrix} X & -Y^T \\ Y & -C_i \end{bmatrix}\right) = \exp\left(\begin{bmatrix} A_i & -B_i^T \\ B_i & C_i \end{bmatrix}\right) \exp\left(\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Z \end{bmatrix}\right) \quad (6.9)$$

where $X \in \mathbf{Skew}_p$, $Y \in \mathbb{R}^{(n-p) \times p}$ and $Z \in \mathbf{Skew}_{n-p}$ are unknowns.

Unfortunately, there is no general algorithm for solving (6.9) and [42] proposes using the Baker-Campbell-Hausdorff infinite series (BCH series) on the right hand side to obtain

$$\begin{aligned} \exp\left(\begin{bmatrix} A_i & -B_i^T \\ B_i & C_i \end{bmatrix} + \begin{bmatrix} X & -Y^T \\ Y & -C_i \end{bmatrix}\right) &= \exp\left(\begin{bmatrix} A_i & -B_i^T \\ B_i & C_i \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Z \end{bmatrix} + \dots\right) \\ \stackrel{?}{\iff} \begin{bmatrix} A_i & -B_i^T \\ B_i & C_i \end{bmatrix} + \begin{bmatrix} X & -Y^T \\ Y & -C_i \end{bmatrix} &= \begin{bmatrix} A_i & -B_i^T \\ B_i & C_i \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Z \end{bmatrix} + \dots \end{aligned}$$

By selecting different terms from the BCH series consisting $\begin{bmatrix} A_i & -B_i^T \\ B_i & C_i \end{bmatrix}$ and $\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Z \end{bmatrix}$, approximated solution Z_i to (6.9) can be obtained and it is used to get the next guess $Q_{i+1} = Q_i \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Z_i \end{bmatrix}$.

However, there are 3 concerns of the BCH algorithm. First or all, the BCH series may not be converging for $\|S_{A_i, B_i, C_i}\|_2 > 1$ or $\|Z_i\|_2 > 1$. Secondly, when it converges, there is no guarantee that the converged sequence $\|S_{A_i, B_i, C_i} + S_{\mathbf{0}, \mathbf{0}, Z_i}\|_2 < \pi$, i.e., the converged sequence may not be on the principal branch of the matrix logarithm, so the matrix exponential on the equation cannot be drop arbitrarily. Finally, [42] does not collect the entire first order truncation in terms of Z , i.e., there are other (infinitely many) terms with Z in the BCH series that are not included in the truncation. Due to the simple truncation, when BCH series converges and the matrix exponential can be dropped, the quality of approximated Z_i still depends significantly on the size A , B and Z in the solution. Note that the size of Z is determined by the quality of the initial guess.

All of these concerns can be dropped when the endpoints U and V are sufficiently close together, as all terms appeared in (6.9) are tiny in this case, which is validated by its excellent numerical performance. As U and V increase in separation, the theoretical support of the BCH algorithm becomes empirical and the numerical performance is significantly compromised.

6.3 Problem Formulation

The endpoint geodesic problem on a Riemannian manifold seeks for any geodesic connecting the given points. This is a weaker version than the Riemannian logarithm problem and the smoothly evolving geodesic problem, as the former asks for a shortest Riemannian geodesic between the given endpoint and the latter asks for a Riemannian geodesic in a smoothly evolving manner as one of the endpoint moves. Therefore, the geometric insights obtained in developing the algorithm of the endpoint geodesic problem as well as the solution itself helps in solving these more complicated geodesic problems. This section gives the formulation of the endpoint geodesic problem on the Stiefel manifold with the canonical metric along with the necessary preprocessing strategy that simplifies the problem.

6.3.1 Matrix Equation

On the Stiefel manifold with the canonical metric, the endpoint geodesic problem with the given $U, V \in \mathbf{St}_{n,p}$ takes the following form

$$\text{Exp}_U^\varphi(UA + U_\perp B) = \begin{bmatrix} U & U_\perp \end{bmatrix} \exp(X_{[A,B,0]}) I_{n,p} = V \quad (6.10)$$

where the skew symmetric $A \in \mathbf{Skew}_p$, $B \in \mathbb{R}^{(n-p) \times p}$ and the respective special orthogonal complementary bases U_\perp are unknown variables. According to the Riemannian submersion that defines the canonical metric on the Stiefel manifold, any geodesic $\text{Exp}_U^\varphi(t \cdot UA + U_\perp B)$ that satisfies (6.10) can be lifted to the special orthogonal group emanating at $\forall Q = \begin{bmatrix} U & U_\perp \end{bmatrix} \in \mathcal{F}$. This invertible lifting process translates (6.10) of finding a geodesic in $\mathbf{St}_{n,p}$ into the equivalent problem of finding a geodesic with horizontal velocity in \mathbf{SO}_n as

$$Q \exp(X_{[A,B,0]}) = \begin{bmatrix} V & V_\perp \end{bmatrix} \quad (6.11)$$

where the unknown variable U_\perp degenerates to a known constant and the additional unknown variable V_\perp as a special orthogonal completion to V is introduced. Although the number of variables remains the same as expected, the matrix equation (6.11) characterized in \mathbf{SO}_n is more convenient. For example, by the transitive property on \mathbf{SO}_n discussed before, any geodesic emanating from any $Q \in \mathbf{SO}_n$ can be translated to I_n back and forth without losing anything. In particular, (6.11) is further equivalent to the following matrix equation

$$\exp(X_{[A,B,0]}) = Q^T \begin{bmatrix} V & V_\perp \end{bmatrix} = \begin{bmatrix} Q^T V & Q^T V_\perp \end{bmatrix}.$$

Notice that $Q^T V$ is a known parameter with the Q and V given in (6.11) and $Q^T V_\perp$ remains a unknown special orthogonal completion to the new $Q^T V$. Then, substitute $Q^T V$ as a given parameter $V \in \mathbf{St}_{n,p}$ and substitute $Q^T V_\perp$ as a unknown special orthogonal completion V_\perp to the new V , the formulation (6.1) proposed at the beginning is obtained.

6.3.2 Preprocessing for Rank Reduction

Lifting the endpoint geodesic problem on the Stiefel manifold to \mathbf{SO}_n and exploiting the transitive property in \mathbf{SO}_n resulted in the convenient matrix equation in (6.1) in size of $n \times n$. In practice, however, the $\mathbf{St}_{n,p}$ may have $p \ll n$ with a huge n , which makes it still very difficult to operate on \mathbf{SO}_n . Fortunately, there exists a smaller problem in $\mathbf{St}_{d,p}$ converted from $\mathbf{St}_{n,p}$ with $d \leq 2p$ discussed in this section. Although not all solutions to the original problem are kept in this smaller problem, any solution to the smaller problem recovers a solution to the original problem, which makes the conversion feasible for the endpoint geodesic problem. This conversion can be done with a rank-revealing factorization to $n \times 2p$ matrix as follows.

Proposition 6.3.1. [29] Consider the endpoint geodesic problem on $\mathbf{St}_{n,p}$ searching for Q, A, B in

$$P \exp \left(\begin{bmatrix} A & -B^T \\ B & \mathbf{0} \end{bmatrix} \right) = Q,$$

the rank-revealing QR decomposition

$$[U \ V]_{n \times 2p} \xrightarrow{\text{rank-revealing QR}} [U \ W]_{n \times d} [I_{d,p} \ \hat{V}]_{d \times 2p} \quad (6.12)$$

constructs endpoint geodesic problem on smaller $\mathbf{St}_{d,p}$ where the d bounded by $p \leq d \leq \min(n, 2p)$ is the rank of $[U \ V]$. The converted problem takes the following specific forms depended on d .

1. When $d > p$, i.e., U and V span different column spaces, then $\hat{V} \in \mathbf{St}_{d,p}$ and the solution $\hat{A} \in \mathbf{Skew}_p, \hat{B} \in \mathbb{R}^{(d-p) \times p}$ of

$$\exp \left(\begin{bmatrix} \hat{A} & -\hat{B}^T \\ \hat{B} & \mathbf{0} \end{bmatrix} \right) = [\hat{V} \ \hat{V}^\perp]$$

forms a solution $A = \hat{A}$ and $B = \begin{bmatrix} \hat{B} \\ \mathbf{0} \end{bmatrix}$ to the original problem with $P = [U \ W \ W^\perp]$.

2. When $d = p$, i.e., U and V span the same column space, W vanishes and $\hat{V} \in \mathbf{O}_p$.

- (a) When $\det(\hat{V}) = 1$, i.e., $\hat{V} \in \mathbf{SO}_p$, the solution $\hat{A} \in \mathbf{Skew}_p$ of

$$\exp(\hat{A}) = \hat{V}$$

forms a solution $A = \hat{A}$ and $B = \mathbf{0}$ solves the original problem with $P = [U \ U^\perp]$.

(b) When $\det(\hat{V}) = -1$, the solution $\hat{A} \in \mathbf{Skew}_p, \hat{B} \in \mathbb{R}^{1 \times p}$ of

$$\exp\left(\begin{bmatrix} \hat{A} & -\hat{B}^T \\ \hat{B} & 0 \end{bmatrix}\right) = \begin{bmatrix} \hat{V} & \mathbf{0} \\ \mathbf{0} & -1 \end{bmatrix}$$

forms a solution $A = \hat{A}$ and $B = \begin{bmatrix} \hat{B} \\ \mathbf{0} \end{bmatrix}$ to the original problem with $P = [U \ U^\perp]$.

Proof. It is easy to see that such rank revealing QR is always possible and

$$V = [U \ W] \hat{V}$$

where $\hat{V} \in \mathbf{St}_{d,p}$. Then it remains to show that a solution to the problem with $I_{d,p}$ and \hat{V} always determine a solution to the problem with U and V , as shown below.

Let \hat{A}, \hat{B} be a solution to the problem with $\hat{P} = I_d$ and \hat{V} , then

$$\hat{Q} = [\hat{V} \ \hat{Q}^\perp] = \exp\left(\begin{bmatrix} \hat{A} & \hat{B}^T \\ \hat{B} & 0 \end{bmatrix}\right) \in \mathbb{R}^{d \times d}$$

and one can write

$$\begin{bmatrix} \hat{Q} & \mathbf{0} \\ \mathbf{0} & I_{n-d} \end{bmatrix}_{n \times n} = \exp\left(\begin{bmatrix} \hat{A} & -\hat{B}^T & \mathbf{0} \\ \hat{B} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}\right).$$

Let $P = [U \ W \ W^\perp]_{n \times n}$, then there is

$$\begin{aligned} P \begin{bmatrix} \hat{Q}_{d \times d} & 0 \\ 0 & I_{n-d} \end{bmatrix}_{n \times n} &= P \exp\left(\begin{bmatrix} \hat{A} & -\hat{B}^T & \mathbf{0} \\ \hat{B} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}\right) \\ &= [U_{n \times k} \ W_{n \times (d-k)} \ W_{n \times (n-d)}^\perp] \begin{bmatrix} \hat{V}_{d \times k} & \hat{Q}_{d \times (d-k)}^\perp & \mathbf{0} \\ \mathbf{0} & & I_{n-d} \end{bmatrix} \\ &= \left[[U \ W]_{n \times d} \hat{V}_{d \times k} \ ; \ [U \ W]_{n \times d} \hat{Q}_{d \times (d-k)}^\perp \ ; \ W^\perp \right] \\ &= \left[V \ ; \ [U \ W]_{n \times d} \hat{Q}_{d \times (d-k)}^\perp \ ; \ W^\perp \right] \\ &:= Q \end{aligned}$$

By construction, there is $PI_{n,k} = U$ and $QI_{n,k} = V$, i.e., $P, Q, \hat{A}, \begin{bmatrix} \hat{B} \\ \mathbf{0} \end{bmatrix}$ solves the matrix equation (6.1).

For the $d = p$ case, $\hat{V} \in \mathbf{SO}_p$ is straightforward. When $\hat{V} \notin \mathbf{SO}_p$, simply notice that $\begin{bmatrix} \hat{V} & \mathbf{0} \\ \mathbf{0} & -1 \end{bmatrix} \in \mathbf{SO}_{p+1}$ and the zero constraint from the stiefel endpoint problem in $\mathbf{St}_{p+1,p}$ on the lower right 1×1 partition is guaranteed by the skew symmetric structure. \square

Note that this conversion is originally proposed in [41] without exploiting the shared dimensions between U and V . It is later developed in [29] but in a very compact expression without a clear algorithmic instruction in how to execute the conversion. **Proposition 6.3.1** summarizes the technique and point out that a rank-revealing QR factorization suffices the conversion task. In addition, it includes the extreme cases when $d = n$, which are not discussed in [29].

6.3.3 Manifold Root-Finding Formulation

The feasible set of solving (6.1) is the fiber over V with $QI_{n,p} = V$ as

$$\mathcal{F}(V) = \{Q \in \mathbf{SO}_n : QI_{n,p} = V\} = \{Q \exp(M_{[0,0,Z]}) : Z \in \mathbf{Skew}_n\}$$

where the second characterization fits in the co-manifold characterization $\mathfrak{C}_{S,\mathfrak{B}}$ in (4.8) where $\exp(S) = Q$ and $\mathfrak{B} = \{M_{[0,0,Z]} : Z \in \mathbf{Skew}_{n-p}\}$. Further notice that the co-manifold characterization collects a set of skew symmetric matrices as smooth submanifold in \mathbf{Skew}_n that emanates geodesics arriving at the fiber \mathcal{F} .

Recall that a geodesic in $\mathbf{St}_{n,p}$ under the canonical metric g^φ is equivalent to a geodesic in \mathbf{SO}_n with a horizontal velocity. Given a Riemannian submersion $\varphi : \mathcal{M} \rightarrow \mathcal{N}$, a geodesic in \mathcal{N} is equivalent to a geodesic in \mathcal{M} that emanates along a horizontal velocity and vice versa. Then the search on velocity $\xi \in T_x\mathcal{N}$ from $\{\text{Exp}_x^\varphi(t \cdot \xi), t \in [0, 1]\} \subset \mathcal{N}$ such that is equivalent to searching the $\bar{\xi} \in T_{\bar{x}}\mathcal{M}$ from $\{\text{Exp}_{\bar{x}}(t \cdot \bar{\xi})\} \subset \mathcal{M}$ constrained with

$$\text{Exp}_{\bar{x}}(\bar{\xi}) = \bar{y} \in \mathcal{F}(y) \tag{6.13}$$

$$\bar{\xi} \in \mathbb{H}_{\bar{x}} \tag{6.14}$$

Notice that the lifted geodesic (6.8) emanating from \bar{x} depends smoothly on the velocity $\bar{\xi} \in T_{\bar{x}}\mathcal{M}$. Consider the linear system

$$F : T_{\bar{x}}\mathcal{M} \rightarrow \mathbb{V}_{\bar{x}}$$

$$\bar{\xi} \mapsto \text{Proj}_{\mathbb{V}}(\bar{\xi})$$

where $\text{proj}_{\mathbb{V}}$ is the orthogonal projector onto $\mathbb{V}_{\bar{x}}$. Then, the horizontal condition (6.14) can be expressed as $\bar{\xi} \in \mathbb{H}_{\bar{x}} \iff F(\bar{\xi}) = \mathbf{0}$. Together, the searching on the $\bar{\xi}$ from (6.8) with conditions (6.13) and (6.14) is equivalent to the system solving with the constrained feasible set

$$F(\bar{\xi}) = \mathbf{0}, \text{Exp}_{\bar{y}}(\bar{\xi}) \in \mathcal{F}(y)$$

In the setup of Stiefel manifold, there is $\bar{x} = I_n \in \mathbf{SO}_n$, $\bar{\xi} = X_{[A,B,C]} \in \mathbf{Skew}_n$, $\text{Exp}_{\bar{y}}(\bar{\xi}) = \exp(X)$ and the orthogonal projector is given by $F(X_{[A,B,C]}) = \text{Proj}_{\mathbb{V}}(X_{[A,B,C]}) = C$. Further

notice that the constraint $\exp(X) \in \mathcal{F}(V)$ is equivalent to co-manifold characterization $X \in \mathfrak{C}_{S, \mathfrak{B}}$ where $\mathfrak{B} = \{M_{[0, \mathbf{0}, Z]} : Z \in \mathbf{Skew}_{n-p}\}$.

Then, the constrained root-finding problem becomes a manifold root-finding problem

$$F(X) = \mathbf{0}, \text{ for } X \in \mathfrak{C}_{S, \mathfrak{B}} \quad (6.15)$$

with some $S \in \mathbf{Skew}_n$ satisfying $\exp(S) = Q \in \mathcal{F}(V)$. In other words, the endpoint geodesic problem on $(\mathbf{St}_{n,p}, g^\alpha)$ is converted to a system solver on a manifold with $F : \mathfrak{C}_{S, \mathfrak{B}} \rightarrow \mathbf{Skew}_{n-p}$ defined on $\mathfrak{C}_{S, \mathfrak{B}}$, where $\mathfrak{C}_{S, \mathfrak{B}}$ is an embedded submanifold in \mathbf{Skew}_n around any given S . Since the horizontal and vertical spaces, $\mathbb{H}_{I_n} = \{X_{[A, B, \mathbf{0}]}\}$ and $\mathbb{V}_{I_n} = \{X_{[\mathbf{0}, \mathbf{0}, C]}\}$, are orthogonal complementary subspaces in $T_{I_n} \mathbf{SO}_n = \mathbf{Skew}_n$ by construction, the system solver problem is further equivalent the root-finding problem that seeks the intersection between $\mathfrak{C}_{S, \mathfrak{B}}$ and \mathbb{H}_{I_n} .

6.4 R-Newton Method of Solving a System on Manifold

Given a system defined on a manifold, the Riemannian Newton method for solving the system with an output is a simple generalization of the Newton-Raphson method in the Euclidean setting that solves $F(x) = \mathbf{0}, F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with the update formula

$$\begin{aligned} x_{i+1} &= x_i + \alpha \Delta_i \\ \text{D} F_{x_i}[\Delta_i] &= -F(x_i) \end{aligned}$$

where α is a step size and $\Delta_i \in \mathbb{R}^n$ is the *Newton update*. The Newton update in the Euclidean setting is an infinitesimal change on x_i that produce the negation of the current output $F(x_i)$ as the infinitesimal change to the system. Then, the generalization of Newton-Raphson methods in a manifold setting is obtained by restricting the infinitesimal change Δ_i to the tangent space at x_i as

$$\begin{aligned} x_{i+1} &\leftarrow \text{Exp}_{x_i}(\alpha \cdot \Delta_i) \\ \text{D} F_{x_i}[\Delta_i] &= -F(x_i), \Delta_i \in T_{x_i} \mathcal{M} \end{aligned} \quad (6.16)$$

where $\text{Exp}_{x_i} : T_{x_i} \mathcal{M}$ is the Riemannian exponential on \mathcal{M} and it can be replaced by any computationally tractable retraction on \mathcal{M} . Note that the Newton direction generalized from the Newton-Raphson method in (6.16) is consistent with the Newton update in the classic Riemannian Newton method on a real-valued function

$$\Delta_i = -\text{Hess}_{x_i}^{-1}[f] (\text{Grad}_{x_i}[f])$$

where $f : \mathcal{M} \rightarrow \mathbb{R}, x \mapsto \|F(x)\|_2^2$ is a real-valued function, $\text{Grad}_{x_i}[f]$ is the gradient vector of f at x_i and $\text{Hess}_{x_i}[f] : T_{x_i}\mathcal{M} \rightarrow T_{x_i}\mathcal{M}$ is the Hessian operator of f at x_i . For more details about the variation of the Newton method generalized to the Riemannian setting, please refer to the textbook [1]. Note that the $\|\cdot\|_2$ in f is the vector 2 norm is different from the matrix 2-norm encountered in other context. It accumulates all squared entries in the system output before taking the square root of the sum.

6.4.1 Newton Direction

The following **Proposition 6.4.1** applies the Riemannian Newton method of solving a system to the system in the root-finding formulation of the endpoint geodesic problem to obtain a Newton direction.

Proposition 6.4.1. *Consider the endpoint geodesic problem on the Stiefel manifold converted the root-finding formulation with the nonlinear system*

$$F : \mathfrak{C}_{\mathcal{F}(V)} \rightarrow \mathbf{Skew}_{n-p}, S_{[A,B,C]} \mapsto C,$$

where $\mathcal{F}V = \{Q \in \mathbf{SO}_n : QI_{n,p} = V\}$. The Newton direction $\Delta \in T_S\mathfrak{C}_{\mathcal{F}(V)}$ to the system is characterized by the matrix equation

$$D \exp_S [\Delta_{[X,Y,-C]}] = QS_{[0,0,Z]}$$

where $S = S_{[A,B,C]} \in \mathfrak{C}_{\mathcal{F}(V)}$, $Q = \exp(S)$ and X, Y, Z are unknown variables. This matrix equation can be further simplified as

$$\mathcal{L}_S(\Delta_{[X,Y,-C]}) = S_{0,0,Z}. \quad (6.17)$$

Proof. In order to make the Riemannian Newton method applicable, one needs to verify that the nonlinear system is smoothly defined on a Riemannian manifold. This follows from 2 observation. First of all the system F can be smoothly extended to a smooth system on the embedding Euclidean space as $F : \mathbf{Skew}_n \rightarrow \mathbf{Skew}_{n-p}, S_{[A,B,C]} \mapsto C$. Secondly, $\mathfrak{C}_{\mathcal{F}(V)}$ around S is a Riemannian manifold $\mathfrak{C}_{S, \{M_{[0,0, \mathbf{Skew}_{n-p}]}\}}$ that is diffeomorphism with $\mathcal{F}(V)$ locally. In conclusion, $F : \mathfrak{C}_{\mathcal{F}(V)} \rightarrow \mathbf{Skew}_{n-p}$ around $S \in \mathfrak{C}_{\mathcal{F}(V)}$ is a smooth system restricted to an embedded submanifold, which results in a smooth system on the embedded submanifold. Therefore, the Riemannian Newton method is applicable.

Then, the characterization of the Newton direction follows from the simple differential

$$D F_S[\Delta_{[\Delta_A, \Delta_B, \Delta_C]}] = \Delta_C, \forall S \in \mathfrak{C}_{\mathcal{F}(V)}.$$

Therefore, $D F_S[\Delta] = -C$ only if $\Delta_C = -C$. \square

Note that it is necessary to construct the co-manifold structure in \mathbf{Skew}_n such that the smooth Riemannian structure in \mathbf{SO}_n can be translated to $\mathfrak{C}_{\mathcal{F}(V)}$ in \mathbf{Skew}_n , on which the smooth nonlinear system $F : \mathfrak{C}_{\mathcal{F}(V)} \rightarrow \mathbf{Skew}_{n-p}$ is constructed. Otherwise, it is impossible to construct a smooth function on \mathbf{SO}_n directly that measures the “non-horizontalness” of the arriving endpoint Q , as the notion of horizontal is inherently defined in \mathbf{Skew}_n that cannot be translated to \mathbf{SO}_n via a global diffeomorphism. With the Newton direction characterized in (6.17), it can be solved as follows.

Corollary 6.4.2. The Newton direction characterized in (6.17) can be solved from two different systems that are derived from (6.17) as the *forward system*

$$\mathcal{L}_S^{\text{Forward}}(X, Y, Z) := \mathcal{L}_S(S_{X, Y, \mathbf{0}}) - S_{\mathbf{0}, \mathbf{0}, Z} = \mathcal{L}_S(S_{\mathbf{0}, \mathbf{0}, C}). \quad (6.18)$$

or the *backward system*

$$\mathcal{L}_S^{\text{Backward}}(Z) := [\mathbf{0} \quad I_{n-p}] \mathcal{L}_S^{-1}(S_{\mathbf{0}, \mathbf{0}, Z}) \begin{bmatrix} \mathbf{0} \\ I_{n-p} \end{bmatrix} = -C. \quad (6.19)$$

Since both linear actions $\mathcal{L}_S^{\text{Forward}}$ and $\mathcal{L}_S^{\text{Backward}}$ are computationally tractable, the matrix-free solver like the GMRES is applicable to both systems. Concerning the complexity brought by the triple dimensions, when $n \leq 2p$ in the forward system (6.18) while the linear action in both systems are similar, it is recommended to solve the backward system (6.19) for the direction Z .

6.4.2 Algorithm

With the Newton direction characterized in (6.17) and solved in (6.18) or (6.19), it remains to handle some technical details before applying the Riemannian Newton algorithm.

First of all, similar to the Newton-Raphson method in the Euclidean setting that fails when $D F_x[\Delta] = -F(x)$ has no solution that happens when the differential $D F$ is rank deficient. In the Riemannian setting (6.17). There is also a similar mechanism to handle such a failure. When the Newton direction does not exist at some x_* and there is a sequence x_i approaching to x_* , the system $D F_{x_i}[\Delta_i] = -F(x_i)$ is more and more ill-conditioned with a diverging Δ_i that has its norm diverge to infinity. Therefore, a huge Newton direction is a flag to that detect the non-existing Newton direction.

On the other hand, the huge Newton direction is also a flag that indicates the failure of a local model, as $\mathfrak{C}_{S,\mathfrak{B}}$ is not expected to be extended infinitely. Therefore, it is necessary to introduce a restart mechanism activated when a huge Newton direction is returned from solving (6.17). With a reasonable scale of $\Delta \in T_S \mathfrak{C}_{S,\mathfrak{B}}$, the Riemannian geodesic on $\mathfrak{C}_{S,\mathfrak{B}}$ is the smoothly evolving geodesic $S(t)$ solved from the $Q(t) = \exp(S) \exp(t \cdot \mathcal{L}_S[\Delta])$.

Algorithm 7: The Newton Algorithm of the Stiefel Endpoint Geodesic Problem

Data: $V \in \mathbf{St}_{n,p}$ where $n \leq 2p$
Input: Initial special orthogonal completion Q_0 where $Q_0 I_{n,p} = V$
Output: Solution $S_* = S_{[A_*, B_*, \mathbf{0}]} \in \mathbf{Skew}_n$ satisfying

```

1 Return  $\log(Q_0)$  ; // Principal Logarithm
2  $i \leftarrow 0$ ;
3  $S_i = S_{[A_i, B_i, C_i]} \leftarrow \log(Q_i)$ ; // Principal Logarithm
4 while  $\|C_i\|_F > \varepsilon$  do
5   Solve  $Z_i$  from  $\mathcal{L}_{S_i}(M_{[A_i, B_i, -C_i]}) = N_{[\mathbf{0}, \mathbf{0}, Z_i]}$  ; // (6.19) or (6.18)
6   if  $\|\mathcal{L}_{S_i}^{-1}(S_{\mathbf{0}, \mathbf{0}, Z_i})\| > (2\pi - \theta_1 - \theta_2)/2$  then
7      $S_0 \leftarrow \log(Q_i \exp(S_{\mathbf{0}, \mathbf{0}, Z_i}))$ ;
8      $i \leftarrow 0$ ; // Restart.
9     Goes to line 4;
10  Line search on step size  $\alpha_i$  along geodesic  $X(t)$ ;
11   $Q_{i+1} \leftarrow Q_i \exp(\alpha_i S_{\mathbf{0}, \mathbf{0}, Z_i})$ ;
12   $S_{i+1} \leftarrow X(\alpha_i)$ ;
13   $i \leftarrow i + 1$ ;
14 Return  $S_* = S_i$ ;

```

Note lines 10 – 11 come from the [42][**Algorithm 4**] that corresponds to the 5 terms truncation in the BCH series, and it is used here as the numerical empirical evidence indicates that it is better than the 1 term truncation of the BCH solver.

6.5 Numerical Experiments

As investigated in the previous chapter, the geometry of the special orthogonal group is more complicated around $Q = \exp(S)$ where S is near the conjugate locus. In the Stiefel manifold $\mathbf{St}_{n,p}$ where the geodesic is characterized by $\gamma(t) = \exp(t \cdot S_{[A, B, \mathbf{0}]}) I_{n,p}$ that emanates from $I_{n,p}$, the matrix 2-norm of the generating $S_{[A, B, \mathbf{0}]}$ is closely related to the difficulty in solving the Stiefel endpoint geodesic problem as demonstrated in this chapter.

Consider a Stiefel manifold on $\mathbf{St}_{n,p}$ with a randomly sampled skew symmetric matrix $S_{[A, B, \mathbf{0}]}$ that has $\|S_{[A, B, \mathbf{0}]}\|_2 = 1$. Then, this skew symmetric matrix emanates the following Riemannian

geodesic arriving at $U(\sigma)$ as follows

$$U(\sigma) = \exp(\sigma \cdot S_{[A,B,0]})I_{n,p} \in \mathbf{St}_{n,p}.$$

By construction, the $\sigma \cdot S_{[A,B,0]}$ is a solution to the endpoint geodesic problem between $I_{n,p}$ and $U(\sigma)$. On the other hand, the matrix 2-norm of this solution is designed to be $\|\sigma \cdot S_{[A,B,0]}\|_2 = \sigma$. Then, the BCH method in [42] and the R-Newton method proposed in **Algorithm 7** are used to find the set of endpoint geodesic problem between $I_{n,p}$ and $U(\sigma)$ generated by various $\sigma \in [0.1, 4.0]$. The performances of these method for solving the σ -labelled problem are collected and reported in below.

Note there are algorithmic alternatives in the BCH method with different implementations. Extensive empirical testing has been performed to identify the implementation with the best performance among the existing BCH variants. As a result of the BCH competition, the BCH method with the 5 terms update is selected as the competitor of the **Algorithm 7**.

Figure 6.1 reports the experimental results in $\mathbf{St}_{20,10}$ by plotting the performances in the σ -labelled problem with $U(\sigma) = \exp(\sigma \cdot S_{[A,B,0]})I_{n,p}$ against $\sigma = \|\sigma \cdot S_{[A,B,0]}\|_2$. It collects 3 important characteristics, the executed number of iteration until terminations, the elapsed time consumed in computing the update directions for both methods and the total elapsed time until termination. Both methods are set to terminate when it finds $\exp(S_{[X,Y,Z]})I_{n,p} = U(\sigma)$ with $\|Z\|_F < 10^{-6}$. For the R-Newton method, each update direction solved from (6.19) is computed by a matrix-free GMRES. This GMRES is set to only accepts solution with absolute error below 10^{-7} . It is clear that the required number of iterations for the Newton method is significantly smaller than those needed for the BCH method. Although the cost of computing the Newton direction (6.17) is more expensive compared to the BCH method, the fewer iterations compensate the elapsed time for computing the direction as well as the total time. The R-Newton method reduces the elapsed time by a factor of 1.925 on average. For the complicated problem with $\sigma \in [1.8, 3.2]$, the R-Newton method is 3.025-times faster the BCH method.

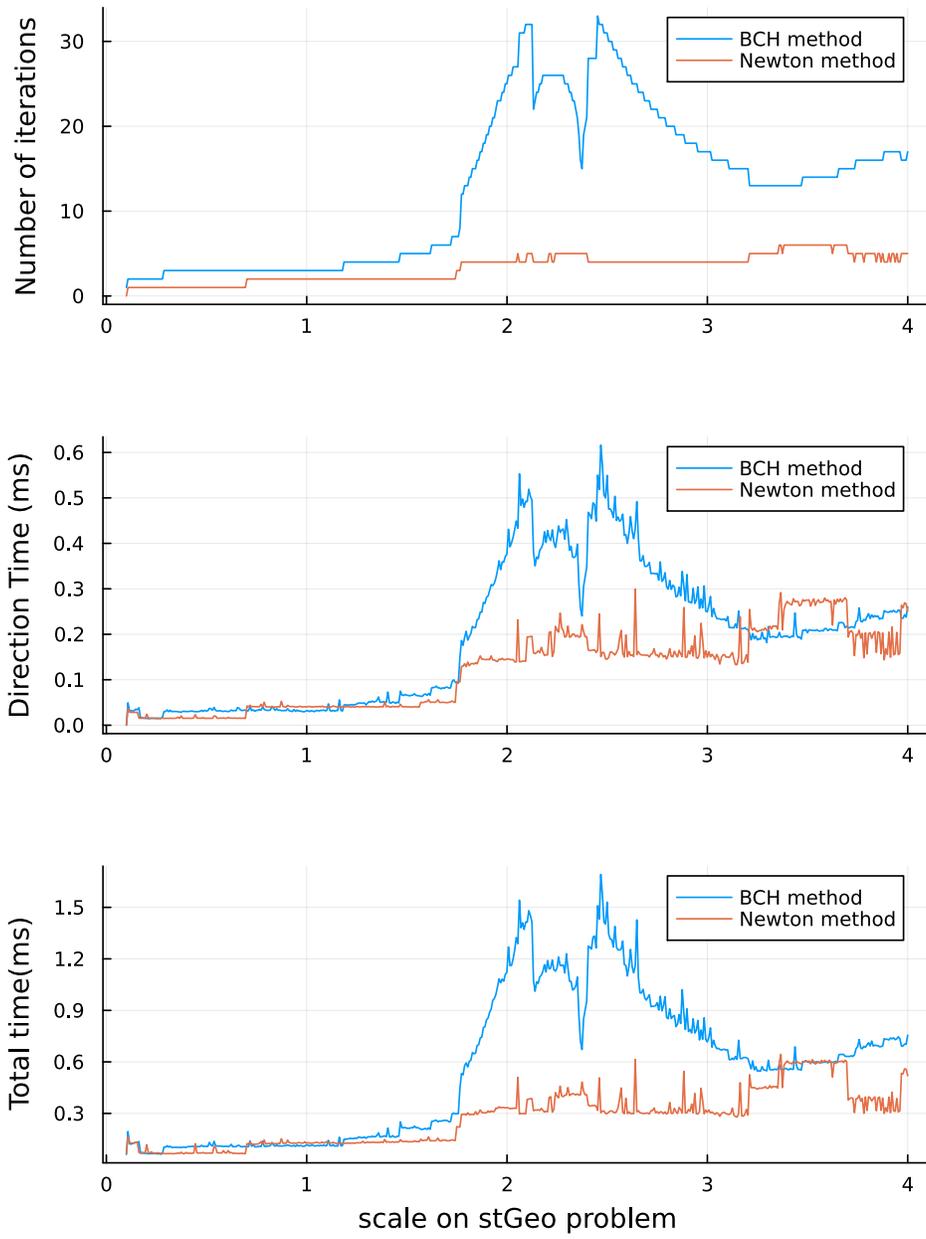


Figure 6.1: Performances on Solving the Stiefel Endpoint Geodesic Problem on $\mathbf{St}_{20,10}$

CHAPTER 7

QUOTIENT STRUCTURE ON THE FIXED RANK POSITIVE SEMI-DEFINITE MANIFOLD

7.1 Introduction

The fixed rank positive semi-definitive (FRPSD) matrix manifold

$$\mathcal{S}_{n,p}^+ := \{X \in \mathbb{R}^{n \times n} : X^T = X, \text{rank}(X) = p, X \succeq 0\}$$

arises in many applications with physics background or computer vision background, e.g., [34] and [12]. In [3], Bonnabel and Sepulchre construct a submersion from the product manifold $\mathcal{M} = \mathbf{St}_{n,p} \times \mathcal{S}_{p,p}^+$ to the FRPSD manifold $\mathcal{N} = \mathcal{S}_{n,p}^+$ as

$$\begin{aligned} \varphi : \mathcal{M} &\rightarrow \mathcal{N} \\ (U, C) &\mapsto UCU^T \end{aligned} \tag{7.1}$$

with an α -parameterized Riemannian metric family $g^{\alpha\text{-BS}}, \alpha > 0$ proposed on \mathcal{M} . Here, $\mathcal{S}_{p,p}^+$ is the well-studied manifold of positive definite $p \times p$ matrices.

This metric family $\{g^\alpha\}_{\alpha>0}$ is designed for the special horizontal structure

$$\mathbb{H}_{(U,C)} = \left\{ (U_\perp B, H) : B \in \mathbb{R}^{(n-p) \times p}, H \in \mathbf{Symm}_p \right\}. \tag{7.2}$$

where $\mathbf{Symm}_p := \{X \in \mathbb{R}^{p \times p} : X^T = X\}$ is the set of all $p \times p$ symmetric matrices. Although every metric $g^{\alpha\text{-BS}}$ does not form a Riemannian submersion in (7.1) nor make $\mathbb{H}_{(U,C)}$ in (7.2) orthogonal to the vertical space $\mathbb{V}_{(U,C)} = T_{(U,C)}\mathcal{F}(X)$ where $X = UCU^T$, it yields invariant metric evaluation on vectors lifted by (7.2), and the orthogonal complement $\mathbb{V}_\perp^{\alpha\text{-BS}}$ converges to the \mathbb{H} specified in (7.2) as $\alpha \rightarrow 0$.

Unfortunately, the metric family $\{g^{\alpha\text{-BS}}\}_{\alpha>0}$ proposed in [3] does not form a Riemannian submersion in (7.1). The lack of a Riemannian submersion makes the application of the metric family in [3] only interpretable in the limiting behaviors as $\alpha \rightarrow 0$, while any individual metric $g^{\alpha\text{-BS}}$ is less meaningful. This chapter further adapts the horizontal structure (7.2) in [3] to construct a different metric family such that a Riemannian submersion is obtained.

7.2 Preliminaries

7.2.1 Geometric Interpretations in the FRPSD manifold

The horizontal structure in (7.2) is chosen to decompose the curve in $\mathcal{S}_{n,p}^+$ into a curve of subspaces and a curve of ellipsoids. In particular, it consists of the classic horizontal space of the Grassmann manifold on the Stiefel manifold $\mathbb{H}_U^{\text{St}_{n,p}} = \{U \perp B : B \in \mathbb{R}^{(n-p) \times p}\}$ and the tangent space of the SPD manifold $T_R \mathcal{S}_{p,p}^+ = \mathbf{Symm}_p := \{X \in \mathbb{R}^{n \times n} : X = X^T\}$ as $\mathbb{H}_{(U,R)} = \mathbb{H}_U^{\text{St}_{n,p}} \times T_R \mathcal{S}_{p,p}^+$, where the Grassmann manifold is the manifold of p dimensional subspaces in \mathbb{R}^n . For a given $\{(U(t), C(t)) : t \in [0, 1]\} \in \mathcal{M}$ that forms $X(t) = U(t)C(t)U(t)^T \in \mathcal{S}_{n,p}^+$ at $U = U(0)$ and $C = C(0)$, the U specifies an orthonormal basis in the p -dimensional subspace in \mathbb{R}^n and the C specifies a p -ellipsoid in the subspace that aligns with the basis given in U .

Note that the motion in the ellipsoid can be further decomposed into the rotations in axes and the deformations in axis-length. Consider the spectral decomposition of $C(t) = Q(t)\Lambda(t)Q(t)^T$ that is smooth with respect to t . Then, $Q(t)$ characterizes the rotations and the $\Lambda(t)$ characterizes the deformations, [13]. For example, consider two ellipses $X = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ and $Y = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. The curve $Q(t) = I_2$ and $\Lambda(t) = \begin{bmatrix} 2-t & 0 \\ 0 & 1+t \end{bmatrix}$ connects X and Y with no rotation, while the curve $Q(t) = \begin{bmatrix} \cos(t\pi/2) & -\sin(t\pi/2) \\ \sin(t\pi/2) & \cos(t\pi/2) \end{bmatrix}$ and $\Lambda(t) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ connects them with no deformation. This is illustrated in **Figure 7.1**.

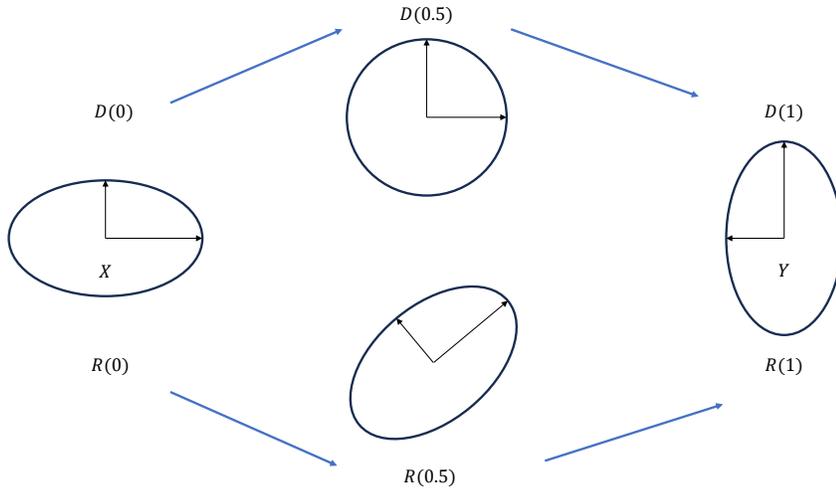


Figure 7.1: Illustration of deformation $Q(t)$ and rotation $\Lambda(t)$ in ellipse.

Note that such a rotation in axes, $Q(t)$, is entangled with the motion in $U(t)$ as one can insert arbitrary $O(t) \in \mathbf{SO}_p$ with $O(0) = I_n$ as

$$\begin{aligned} X(t) &= U(t)C(t)U(t)^T = (U(t)O(t)) (O(t)^T C(t) O(t)) (U(t)O(t))^T \\ &= \tilde{U}(t) (O(t)^T Q(t) \Lambda(t) Q(t)^T O(t)) \tilde{U}(t)^T = \tilde{U}(t) \left(\tilde{Q}(t) \Lambda(t) \tilde{Q}(t)^T \right) \tilde{U}(t)^T \end{aligned}$$

where $\tilde{U}(t) = U(t)O(t)$ and $\tilde{Q}(t) = O(t)^T Q(t)$. Through the arbitrary $O(t)$, the same motion $X(t)$ can be realized by different curves in \mathcal{M} .

7.2.2 Objects in the Submersion

This section collects relevant results about the objects in the submersion (7.1), which are available in the literature, c.f. [36], [23] and [3].

For the point $X \in \mathcal{N} = \mathcal{S}_{n,p}^+$, consider its representation $(U, C) \in \mathcal{M} = \mathbf{St}_{n,p} \times \mathcal{S}_{p,p}^+$ and the horizontal space specified as in (7.2). Then, the following objects are available.

1. The tangent space to \mathcal{M} at (U, C) is given by

$$T_{(U,C^2)}\mathcal{M} := \{(US + U_\perp K, M) : S \in \mathbf{Skew}_p, K \in \mathbb{R}^{(n-p) \times p}, M \in \mathbf{Symm}_p\}. \quad (7.3)$$

2. The fiber over $X = UCU^T$ is given by

$$\mathcal{F}(X) = \{(UQ, Q^T C Q) : Q \in \mathbf{SO}_p\}, \quad (7.4)$$

in particular, the spectral decomposition $X = U\Lambda U^T$ with eigenvectors in U and eigenvalues in Λ is in the fiber over Y , i.e., $\mathcal{F}(Y) = \{(UQ, Q^T \Lambda Q) : Q \in \mathbf{SO}_p\}$.

3. The vertical space at (U, C) is given by

$$\mathbb{V}_{(U,C)} = T_{(U,C)}\mathcal{F}(X) = \{(US, CS - SC) : S \in \mathbf{Skew}_p\}. \quad (7.5)$$

It yields the projection of any $\xi := (US + U_\perp K, M) \in T_{(U,C)}\mathcal{M}$ to $\mathbb{H}_{(U,C)}$ against $\mathbb{V}_{(U,C)}$ as

$$\begin{aligned} \text{Proj}_{\mathbb{H}}^{\mathbb{H} \oplus \mathbb{V}} : T_{(U,C)}\mathcal{M} &\rightarrow \mathbb{H}_{(U,C)} \\ (US + U_\perp K, M) &\mapsto (U_\perp K, M - CS + SC) \end{aligned} \quad (7.6)$$

such that the $\xi_{\mathbb{H}} := \text{Proj}_{\mathbb{H}}^{\mathbb{H} \oplus \mathbb{V}}(\xi) \in \mathbb{H}_{(U,C)}$ and the $\xi_{\mathbb{V}} := \xi - \xi_{\mathbb{H}} \in \mathbb{V}_{(U,C)}$ is the unique decomposition of $\xi = \xi_{\mathbb{H}} + \xi_{\mathbb{V}}$ into the two complement subspaces.

4. The differential to the submersion map is given by

$$\begin{aligned} D\varphi_{(U,C)}[(US + U_\perp K, M)] &= D\varphi_{(U,C)}[(U_\perp K, M - CS + SC)] \\ &= U_\perp KCU^T + UCK^T U_\perp^T + U(M - CS + SC)U^T \end{aligned} \quad (7.7)$$

where the first equation follows from projecting $(US + U_\perp K, M) \in T_{(U,C)}\mathcal{M}$ to $\mathbb{H}_{(U,C)}$.

7.3 Riemannian Metric by Riemannian Submersion

In order to obtain make the submersion φ Riemannian, a metric on \mathcal{M} , denoted as $g^{\alpha\text{-Hor}}$, with the following features is needed.

1. The vertical space and the horizontal space are orthogonal under $g^{\alpha\text{-Hor}}$, i.e.,

$$\bar{g}_{(U,C)}^{\alpha\text{-Hor}}((US, CS - SC), (U_{\perp}K, M)) = 0$$

for any $S \in \mathbf{Skew}_p$, $K \in \mathbb{R}^{(n-p) \times p}$ and $M \in \mathbf{Symm}_p$.

2. The inner products under $g^{\alpha\text{-Hor}}$ between the horizontally lifted vectors $D\varphi_{(U,C)}^{\mathbb{H}}[\xi]$ and $D\varphi_{(U,C)}^{\mathbb{H}}[\eta]$ for any $\xi, \eta \in T_X\mathcal{N}$ and any $(U, C) \in \mathcal{F}(X)$ equal to each other, i.e.,

$$g_{(U_1, C_1)}^{\alpha\text{-Hor}}(D\varphi_{(U_1, C_1)}^{\mathbb{H}}[\xi], D\varphi_{(U_1, C_1)}^{\mathbb{H}}[\eta]) = g_{(U_2, C_2)}^{\alpha\text{-Hor}}(D\varphi_{(U_2, C_2)}^{\mathbb{H}}[\xi], D\varphi_{(U_2, C_2)}^{\mathbb{H}}[\eta])$$

for any $X \in \mathcal{N}$, $\xi, \eta \in T_X\mathcal{N}$ and $(U_1, C_1), (U_2, C_2) \in \mathcal{F}(X)$.

7.3.1 Horizontal Lifting

The first feature can be satisfied by designating bases in the respective subspaces $\mathbb{H}_{(U,C)}$ and $\mathbb{V}_{(U,C)}$ as an orthonormal basis of the total space $T_{(U,C)}\mathcal{M} = \mathbb{H}_{(U,C)} \oplus \mathbb{V}_{(U,C)}$ which totally characterizes an inner product on it. The second feature depends on the horizontal lift operator $D\varphi_{(U,C)}^{\mathbb{H}} : T_X\mathcal{N} \rightarrow \mathbb{H}_{(U,C)}$. Such a horizontal lift specified to (7.2) is not available in the literature and the proposition below derives an expression from (7.6).

Proposition 7.3.1. *For any $X \in \mathcal{N} = \mathcal{S}_{n,p}^+$ and a tangent vector attached to it, denoted as a $\Delta \in T_X\mathcal{N} \subset \mathbb{R}^{n \times n}$, the horizontal lift of Δ to $(U, C) \in \mathcal{F}(X)$, i.e., $UCU^T = X$, is given by $D\varphi_{(U,C)}^{\mathbb{H}}[\Delta] = (U_{\perp}K, M)$ where*

$$\begin{cases} M = U^T \Delta U \\ K = U_{\perp}^T \Delta U C^{-1} \end{cases} \quad (7.8)$$

It is easily verified that (U_{\perp}, M) is horizontal and $D\varphi_{(U,C)}^{\mathbb{H}}[(U_{\perp}K, M)] = \Delta$. More importantly, notice that any SPD matrix can be written as the square of a unique SPD matrix, which is known as its unique square root. That means the representation $(U, C) \in \mathcal{F}(X)$ can also be re-parameterized as $(U, C^2) \in \mathcal{F}(X)$ such that $UC^2U^T = X$ and $C \in \mathcal{S}_{p,p}^+$. Then, the horizontal lift (7.8) is converted to the expression in below.

Lemma 7.3.2. *For any $X = UC^2U^T \in \mathcal{N} = \mathcal{S}_{n,p}^+$ where $C \in \mathcal{S}_{p,p}^+$, the horizontal lift of $\Delta \in T_X\mathcal{N}$ to $(U, C^2) \in \mathcal{F}(X)$ is given by $D\varphi_{(U,C)}^{\mathbb{H}}[\Delta] = (U_{\perp}K, CHC)$ where*

$$\begin{cases} H = C^{-1}U^T \Delta U C^{-1} \\ K = U_{\perp}^T \Delta U C^{-2} \end{cases} \quad (7.9)$$

Unless otherwise specified, the horizontal tangent vector at (U, C^2) in this chapter is denoted as $\bar{\xi}_{K,H} := (U_\perp K, CHC)$ with the U_\perp available from the context. Then, the horizontal lift can be written as $D\varphi_{(U,C^2)}^{\mathbb{H}}[\Delta] = \bar{\xi}_{U,C^2}^{(K,H)}$ where K and H are computed from (7.9). This special notation is introduced for the following convenient observation.

Proposition 7.3.3. *For any $X = UC^2U^T \in \mathcal{N} = \mathcal{S}_{n,p}^+$ and the horizontal lifted vector*

$$D\varphi_{U,C^2}^{\mathbb{H}}[\Delta] = \bar{\xi}_{U,C^2}^{(K,H)} = (U_\perp K, RHR),$$

consider another representation $(UQ, Q^T C^2 Q) \in \mathcal{F}(X)$ on the fiber with a $p \times p$ orthogonal Q . Then, the horizontal lift of the same Δ to the different representation $(UQ, Q^T C^2 Q)$ is given by

$$D\varphi_{UQ, Q^T C^2 Q}^{\mathbb{H}}[\Delta] = \bar{\xi}_{UQ, Q^T C^2 Q}^{(KQ, Q^T H Q)} = (U_\perp KQ, Q^T CHCQ), \quad (7.10)$$

i.e., the parameterization K and H varies by Q in a consistent fashion.

Proof. This follows from 2 observations. First of all the U_\perp remains orthogonal to all UQ , which simplifies the first component by setting $\tilde{U}_\perp = U_\perp$. Then, the unique SPD square root of $Q^T C^2 Q$ is given by $\tilde{C} := Q^T C Q$. Let $\tilde{U} = UQ$ and applies (7.9) to get $D\varphi_{(\tilde{U}, \tilde{C}^2)}^{\mathbb{H}}[\Delta] = \bar{\xi}_{(\tilde{U}, \tilde{C}^2)}^{(\tilde{K}, \tilde{H})}$ as

$$\begin{cases} \tilde{H} = \tilde{C}^{-1} \tilde{U}^T \Delta \tilde{U} \tilde{C}^{-1} = Q^T H Q \\ \tilde{K} = \tilde{U}_\perp^T \Delta \tilde{U} \tilde{C}^{-2} = K Q \end{cases}$$

□

Recall that a spectral decomposition of $X \in \mathcal{S}_{n,p}^+$ naturally yields a representation of X in the fiber over it. For the spectral decomposition $X = U\Lambda^2 U^T$ where Λ is diagonal with the square root of the eigenvalues in X , (7.9) and (7.10) characterizes all lifted vectors as $D\varphi_{(U,\Lambda^2)}^{\mathbb{H}}[\Delta] = \bar{\xi}_{U,\Lambda^2}^{K,H}$ and $D\varphi_{(UQ, Q^T \Lambda^2 Q)}^{\mathbb{H}}[\Delta] = \bar{\xi}_{UQ, Q^T \Lambda^2 Q}^{KQ, Q^T H Q}$.

7.3.2 Riemannian Metric Family Constructed by Designated Basis

Recall that a Riemannian metric at a point $(U, C^2) \in \mathcal{M}$ is an inner product operator on the tangent space $T_{(U,C)}\mathcal{M}$. As a linear space equipped with an inner product $\langle \cdot, \cdot \rangle$, it yields the notion of an orthonormal basis $B = \{B_i\}_{i=1}^d \subset T_{(U,C)}\mathcal{M}$ such that

$$\langle B_i, B_j \rangle = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

where d is the dimension of the linear space. On the other hand, any designated linear independent basis $B = \{B_i\}_{i=1}^d$ defines an inner product operator

$$\langle \xi, \eta \rangle = \left\langle \sum_{i=1}^d x_i B_i, \sum_{i=1}^d y_i B_i \right\rangle := \sum_{i=1}^d x_i y_i$$

where $\xi = \sum_{i=1}^d x_i B_i$ and $\eta = \sum_{i=1}^d y_i B_i$ are the unique linear decomposition in the basis B . Note that any subspaces spanned by the distinct sub-bases from B are orthogonal to each other under the induced inner product. For example, let $\mathbb{H} := \text{col}(\{B_{j_i}\}_{i=1}^m)$ and $\mathbb{V} := \text{col}(\{B_{k_i}\}_{i=1}^n)$ such that $\{B_{j_i}\}_{i=1}^m \cap \{B_{k_i}\}_{i=1}^n = \emptyset$. Then, any $\xi \in \mathbb{H}$ has the unique linear decomposition

$$\xi = \sum_{i=1}^m x_{j_i} B_{j_i} + \sum_{l \neq j_i} 0 \cdot B_l,$$

i.e., x_i not in the $\{j_i\}_{i=1}^m$ positions are all zeros. Similarly, the linear decomposition of η has y_i that are zero if they are not in the $\{k_i\}_{i=1}^n$ positions. Then ξ and η are orthogonal as

$$\langle \xi, \eta \rangle = \sum_{i=1}^m x_{j_i} \cdot 0 + \sum_{i=1}^n 0 \cdot y_{k_i} + \sum_{l \neq j_i, l \neq k_i} 0 \cdot 0 = 0.$$

According to this observation, any inner product in $T_{(U, C^2)} \mathcal{M}$ that is constructed from the bases in $\mathbb{V}_{(U, C^2)}$ and $\mathbb{H}_{(U, C^2)}$ together makes them orthogonal to each other naturally. For the

$$\begin{cases} \mathbb{V}_{(U, C^2)} = \{(US, C^2 S - SC^2) : S \in \mathbf{Skew}_p\} \\ \mathbb{H}_{(U, C^2)} = \{(U_{\perp} K, CHC) : K \in \mathbb{R}^{(n-p) \times p}, H \in \mathbf{Symm}_p\} \end{cases},$$

the parameterization S , K and H have designated bases for each subspaces. A more intuitive way to see this basis is to recall the canonical metric on the Stiefel manifold

$$g_U^{\mathbf{St}}(US_1 + U_{\perp} K_1, US_2 + U_{\perp} K_2) = \text{tr}(S_1^T S_2) + \text{tr}(K_1^T K_2), \forall US_1 + U_{\perp} K_1, US_2 + U_{\perp} K_2 \in T_U \mathbf{St}_{n,p}.$$

It is clear that the subspaces $\{US : S \in \mathbf{Skew}_p\}$ and $\{U_{\perp} K : K \in \mathbb{R}^{(n-p) \times p}\}$ are orthogonal under $g^{\mathbf{St}}$. Then, let

$$\begin{cases} B_U^{\mathbf{St}, S} = \{US_i\}_{i=1}^{d_S} \subset \{US : S \in \mathbf{Skew}_p\} \\ B_U^{\mathbf{St}, K} = \{U_{\perp} K_i\}_{i=1}^{d_K} \subset \{U_{\perp} K : K \in \mathbb{R}^{(n-p) \times p}\} \end{cases}$$

be the respective d_S and d_K dimensional bases. They form an orthonormal basis in $T_U \mathbf{St}_{n,p}$ together and, more importantly, they can be found in $\mathbb{V}_{(U, C^2)}$ and $\mathbb{H}_{(U, C^2)}$ respectively

$$\begin{cases} B_{U, C^2}^S := \{(US_i, C^2 S_i - S_i C^2)\}_{i=1}^{d_S} \subset \mathbb{V}_{(U, C^2)} \\ B_{U, C^2}^K := \{(U_{\perp} K_i, \mathbf{0})\}_{i=1}^{d_K} \subset \mathbb{H}_{(U, C^2)} \end{cases}. \quad (7.11)$$

where $\{US_i\}_{i=1}^{d_S} \cup \{U_\perp K_i\}_{i=1}^{d_K}$ is an orthonormal basis in $T_U \mathbf{St}_{n,p}$ under the canonical metric. Notice that B_{U,C^2}^S is already a basis that spans $\mathbb{V}_{(U,C^2)}$ but B_{U,C^2}^K is not enough to cover the $\mathbb{H}_{(U,C^2)}$. The remaining part is provided by the affine-invariant metric on the SPD manifold proposed in [14]

$$g_{C^2}^{S^+}(CH_1C, CH_2C) = \text{tr}(H_1^T H_2), \forall CH_1C, CH_2C \in T_{C^2} \mathcal{S}_{p,p}^+.$$

Similarly, one can find a basis in $\mathbb{H}_{(U,C^2)}$

$$B_{U,C^2}^H := \{(\mathbf{0}, CH_iC)\}_{i=1}^{d_H} \subset \mathbb{H}_{(U,C^2)}. \quad (7.12)$$

where $\{CH_iC\}_{i=1}^{d_H}$ are the orthonormal basis in $T_{C^2} \mathcal{S}_{p,p}^+$ under the affine-invariant metric.

Together, $B_{U,C^2}^S \cup B_{U,C^2}^K \cup B_{U,C^2}^H$ forms a linear independent basis in $T_{(U,C^2)} \mathcal{M}$ such that $\mathbb{V}_{(U,C^2)} = \text{col}(B_{(U,C^2)}^S)$ and $\mathbb{H}_{(U,C^2)} = \text{col}(B_{(U,C^2)}^K \cup B_{(U,C^2)}^H)$. By rescaling the $B_{(U,C^2)}^H$ basis with a parameter $\alpha > 0$, a parameterized metric is obtained as follows.

Proposition 7.3.4. *Consider any $(U, C^2) \in \mathcal{M} = \mathbf{St}_{n,p} \times \mathcal{S}_{p,p}^+$, then a parameterized linear independent basis of $T_{(U,C^2)} \mathcal{M}$ is given by*

$$B_{U,C^2}^{\alpha\text{-Hor}} := B_{(U,C^2)}^S \cup B_{(U,C^2)}^K \cup \frac{1}{\sqrt{\alpha}} \cdot B_{(U,C^2)}^H, \alpha > 0. \quad (7.13)$$

It defines a Riemannian metric on \mathcal{M} with

$$\begin{aligned} \bar{g}_{(U,C^2)}^{\alpha\text{-Hor}}(\bar{\xi}_1, \bar{\xi}_2) &= g_U^{\mathbf{St}}(US_1 + U_\perp K_1, US_2 + U_\perp K_2) + \alpha \cdot g_{C^2}^{S^+}(CH_1C, CH_2C) \\ &= \frac{1}{2} \text{tr}(S_1^T S_2) + \text{tr}(K_1^T K) + \alpha \cdot \text{tr}(H_1^T H_2). \end{aligned} \quad (7.14)$$

where $\bar{\xi}_i = (US_i + U_\perp K_i, CH_iC + C^2 S_i - S_i C^2)$, for $i = 1, 2$.

Proof. The fact that the proposed metric satisfies the properties of an inner product operator is easily verified. This proof only points out that the orthonormal basis $B_{(U,C^2)}^{\alpha\text{-Hor}}$ depends smoothly on the foot (U, C^2) varying on \mathcal{M} , which makes the inner product it defines in (7.14) also smoothly depending on the foot. A smoothly dependent inner product defined on all tangent space forms a Riemannian metric. \square

The horizontal space (7.2) and the vertical space (7.5) are orthogonal under the proposed Riemannian metric (7.14) by design. In order to show that the submersion $\varphi : (U, C^2) \mapsto UC^2U^T$ is Riemannian, it remains to show that the horizontal spaces on a fiber $\mathcal{F}(X)$ are isometric. Recall that any point on $\mathcal{F}(X) \subset \mathcal{M}$ can be characterized by the spectral decomposition $X = U\Lambda^2U^T$ and any

special orthogonal Q as $\mathcal{F}(X) = \{(UQ, Q^T \Lambda^2 Q) : Q \in \mathbf{SO}_p\}$. Consider any $\xi_1, \xi_2 \in T_X \mathcal{S}_{n,p}^+ \subset \mathbb{R}^{n \times n}$ and the lifted vectors $\bar{\xi}_{(U, \Lambda^2)}^{K_1, H_1}$ and $\bar{\xi}_{(U, \Lambda^2)}^{K_2, H_2}$ determined by (7.9). Then, the inner product between the lifted vectors are given by

$$\bar{g}_{(U, \Lambda^2)}^{\alpha\text{-Hor}} \left(\bar{\xi}_{(U, \Lambda^2)}^{K_1, H_1}, \bar{\xi}_{(U, \Lambda^2)}^{K_2, H_2} \right) = \text{tr}(K_1^T K_2) + \alpha \text{tr}(H_1^T H_2).$$

For the vectors lifted to a different representation $(UQ, Q^T \Lambda^2 Q)$, the inner product remains constant according to (7.10) as

$$\begin{aligned} \bar{g}_{(U, \Lambda^2)}^{\alpha\text{-Hor}} \left(\bar{\xi}_{(UQ, Q^T \Lambda^2 Q)}^{K_1, H_1}, \bar{\xi}_{(UQ, Q^T \Lambda^2 Q)}^{K_2, H_2} \right) &= \text{tr}(Q^T K_1^T K_2 Q) + \alpha \text{tr}(Q^T H_1^T Q Q^T H_2 Q) \\ &= \text{tr}(K_1^T K_2) + \alpha \text{tr}(H_1^T H_2). \end{aligned}$$

Proposition 7.3.5. *The submersion $\varphi : \mathcal{M} \rightarrow \mathcal{N}, (U, C^2) \mapsto UC^2U^T$ is Riemannian given the metric $\bar{g}^{\alpha\text{-Hor}}$ defined in (7.14).*

It is important to note that the metric on the total space

$$\bar{g}_{(U, C^2)}^{\alpha\text{-BS}}((US_1 + U_\perp K_1, CH_1 C), (US_2 + U_\perp K_2, CH_2 C)) = \text{tr}(S_1^T S_2) + \text{tr}(K_1^T K_2) + \text{tr}(H_1^T H_2)$$

proposed in [3] is not a member of the metric family in (7.14), not only due to the one-half scale in the $\text{tr}(S_1, S_2)$, but also due to the offset $C^2 S - SC^2$ that results from US not being reduced from RHR . Therefore, there are limited results on the Riemannian geodesic is derived in [3].

7.3.3 Riemannian Geodesic with Motions in Subspaces and Ellipsoids

Note that the choice of the Stiefel canonical metric and the SPD affine-invariant metric selected for constructing the metric family (7.14) is not mandatory but it yields the geometric insight of the induced Riemannian geodesic discussed in this section.

Consider two given Stiefel points $U, V \in \mathbf{St}_{n,p}$ that span p -dimensional subspaces in \mathbb{R}^n , let $\{\gamma_U(t) : t \in [0, 1]\} \subset \mathbf{St}_{n,p}$ be a smooth curve that connects U and V as $\gamma_U(0) = U$ and $\gamma_U(1) = V$. Then, this curve generates a curve of subspaces in forms of $\text{col}(\gamma_U(t))$ that connects the subspaces spanned by U and V . Suppose the differential to the curve

$$\frac{d}{dt} \gamma_U(t) = \begin{bmatrix} U(t) & U_\perp(t) \end{bmatrix} \begin{bmatrix} S(t) & -K(t)^T \\ K(t) & \Omega(t) \end{bmatrix},$$

then, its length and energy in the Grassmann manifold is given by

$$\begin{cases} l_{\text{col}(\gamma_U)} = \int_0^1 \sqrt{\text{tr}(K(t)^T K(t))} dt \\ E_{\text{col}(\gamma_U)} = \int_0^1 \text{tr}(K(t)^T K(t)) dt \end{cases}$$

In this setup, the curve that yields the minimal length in the Grassmann manifold is solved in [19] as follows.

Lemma 7.3.6 ([19]). The curve $\{\gamma_U(t) : t \in [0, 1]\} \subset \mathbf{St}_{n,p}$ connecting two given $U, V \in \mathbf{St}_{n,p}$ that yields minimal length and energy in the Grassmann manifold takes the form of

$$\gamma_U(t) = [U \quad U_\perp] \exp\left(t \cdot \begin{bmatrix} S & -K^T \\ K & \Omega \end{bmatrix}\right) \exp\left(t \cdot \begin{bmatrix} -S & \mathbf{0} \\ \mathbf{0} & -\Omega \end{bmatrix}\right) I_{n,p} \quad (7.15)$$

where $S \in \mathbf{Skew}_p$, $\Omega \in \mathbf{Skew}_{n-p}$ and $K \in \mathbb{R}^{(n-p) \times p}$.

On the other hand, the Riemannian geodesic in the SPD manifold under the affine-invariant metric is solved in [14] as follows.

Lemma 7.3.7 ([14]). The Riemannian geodesic $\{\gamma_{C^2}(t) : t \in [0, 1]\} \subset \mathcal{S}_{p,p}^+$ with initial velocity $\dot{\gamma}_{C^2}(0) = CHC \in T_{C^2}\mathcal{S}_{p,p}^+$ under the affine-invariant metric is given by

$$\gamma_{C^2}(t) = C \exp(tHC) \quad (7.16)$$

where $H \in \mathbf{Symm}_p$.

The curve (7.15) describes the minimal change in terms of subspaces in the Grassmann manifold and the curve (7.16) describes the minimal change in terms of the ellipsoids in the SPD manifold. Together, they form the horizontal Riemannian geodesic in $\mathcal{M} = \mathbf{St}_{n,p} \times \mathcal{S}_{p,p}^+$ as follows.

Theorem 7.3.8. *The Riemannian geodesic $\bar{\gamma}^{\alpha\text{-Hor}}(t)$ in $(\mathcal{M}, \bar{g}^{\alpha\text{-Hor}})$, $\alpha > 0$ emanating from $(U, C^2) \in \mathcal{M}$ along horizontal velocity $\left.\frac{d}{dt}\bar{\gamma}(t)\right|_{t=0} = (U_\perp K, CHC) \in \mathbb{H}_{(U, C^2)}$ is given by*

$$\bar{\gamma}^{\alpha\text{-Hor}}(t) = \left([U \quad U_\perp] \exp\left(t \cdot \begin{bmatrix} S & -K^T \\ K & \Omega \end{bmatrix}\right) \exp\left(t \cdot \begin{bmatrix} -S & \mathbf{0} \\ \mathbf{0} & -\Omega \end{bmatrix}\right), C \exp(tHC) \right) \quad (7.17)$$

where $S \in \mathbf{Skew}_p$ and $\Omega \in \mathbf{Skew}_{n-p}$ depends on the values of $\alpha > 0$. In other words, the horizontal Riemannian geodesic is the simple composition of the (7.15) and the (7.16).

Proof. Consider the energy functional of the any smooth curve $\bar{\tau}(t) = (\tau_U(t), \tau_{C^2}(t)) = (U(t), C(t)^2)$ with velocity.

$$\left.\frac{d}{dt}\bar{\tau}(t)\right|_{t=0} = (U(t)S(t) + U_\perp(t)K(t), C(t)H(t)C(t) + C(t)^2S(t) - S(t)C^2(t))$$

Then, its energy is given by

$$\begin{aligned} E^\alpha(\bar{\tau}) &= \int_0^1 \bar{g}_{\bar{\tau}(s)}^{\alpha\text{-Hor}} \left(\left.\frac{d}{dt}\bar{\tau}(t)\right|_{t=s}, \left.\frac{d}{dt}\bar{\tau}(t)\right|_{t=s} \right) ds \\ &= \int_0^1 \frac{1}{2} \text{tr}(S(t)^T S(t)) dt + \int_0^1 \text{tr}(K(t)^T K(t)) dt + \alpha \int_0^1 \text{tr}(H(t)^T H(t)) dt \end{aligned}$$

Notice that the first integral vanishes if the curve $\bar{\tau}(t)$ stays horizontal, i.e., $S(t) = 0, \forall t \in [0, 1]$. Then, the second integral is the energy functional in the Grassmann manifold of the curve $\tau_U(t)$, which obtains its minimal energy when $\tau_U(t)$ takes the form in (7.15). The third integral is the energy functional in the SPD manifold under the affine-invariant metric, which obtains its minimal energy when $\tau_{C^2}(t)$ takes the form in (7.16).

Finally, for a Riemannian submersion $\varphi : \mathcal{M} \rightarrow \mathcal{N}$, a geodesic emanating with horizontal velocity stays horizontal and it obtains the (local) minimal energy among all curves $\bar{\tau}(t) \in \mathcal{M}$ connecting the same endpoints. As the two integral are independent with each other, the (local) minimal energy is obtained with the composition of the (7.15) and the (7.16). \square

Note that how the parameter $\alpha > 0$ determines the two skew symmetric matrices $S \in \mathbf{Skew}_p$ and $\Omega \in \mathbf{Skew}_{n-p}$ remains unknown and it is left as future work. Another useful property follows from the Riemannian submersion is the fact that the horizontal Riemannian geodesic $\bar{\gamma}^{\alpha\text{-Hor}}$ must end up at some point on the fiber over its endpoint on \mathcal{N} . It derives the following formulas and bounds on the length of the lifted Riemannian geodesic.

Proposition 7.3.9. *Let $\gamma^\alpha(t)$ be a minimal Riemannian geodesic that connects $U\Lambda^2U^\top$ and $V\Sigma^2V^\top$ where Λ^2 and Σ^2 are diagonal matrices. Lift the geodesic horizontally so that $\hat{\gamma}^\alpha(0) = (U, \Lambda^2)$ and $\hat{\gamma}^\alpha(1) = (VQ, R^2 = Q^\top\Sigma^2Q)$. The length of γ^α under metric g^α is given by the length $l_{U \rightarrow VQ}$ of the Grassmann horizontal curve under the Stiefel metric (two well-known two metrics give the same length) and the length $l_{\Lambda^2 \rightarrow R^2}$ under the affine invariant metric as*

$$l_{\gamma^\alpha} = \sqrt{l_{U \rightarrow VQ}^2 + \alpha l_{\Lambda^2 \rightarrow R^2}^2}. \quad (7.18)$$

The two components in (7.18) are curve lengths in the Stiefel manifold and the symmetric positive definite (SPD) manifold respectively, which yields following the convenient bounds.

1. The length $l_{U \rightarrow VQ}$ of the horizontal curve connecting U and VQ is bounded from below as

$$d_{\mathbf{Gr}}(\text{col}(U), \text{col}(V)) \leq d_{\mathbf{St}}(U, VQ) \leq l_{U \rightarrow VQ}$$

where $d_{\mathbf{Gr}}(\text{col}(U), \text{col}(V))$ is the distance between $\text{col}(U)$ and $\text{col}(V)$ in the Grassmann manifold and $d_{\mathbf{St}}(U, VQ)$ is the distance between U and VQ in the Stiefel manifold.

2. The length $l_{\Lambda^2 \rightarrow C^2}$ of the Riemannian geodesic connecting Λ^2 and C^2 in the SPD manifold is determined by the generalized eigenvalues $\omega_i(\Lambda^2, C^2)$ of Λ^2 and C^2 as

$$l_{\Lambda^2 \rightarrow C^2} = \sqrt{\sum_{i=1}^k \log(\omega_i(\Lambda^2, C^2))^2}.$$

Let $C^2 = Q\Sigma^2Q^T$ be a spectral decomposition and let Λ^2 and Σ^2 be ordered in the decreasing order of magnitude, i.e., $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_k^2$ and so does σ_i^2 's, the sum of the squared generalized eigenvalues are bounded as follows, assuming.

(a) The generalized eigenvalues are bounded as $\lambda_i < \omega_i < \sigma_i$, which yields

$$l_{\Lambda^2 \rightarrow C^2} \geq \sqrt{\sum_{i=1}^k (\log(\lambda_i^2) - \log(\sigma_i^2))^2}$$

where the equal sign holds with $C^2 = \Sigma^2$.

(b) There exists a permutation P so that $P^T \Sigma P$ reorders σ_i^2 's into $\{\sigma_{i_j}^2\}_{j=1}^k$ and yields a bound from above as

$$l_{\Lambda^2 \rightarrow C^2} \leq \sqrt{\sum_{i=1}^k (\log(\lambda_i^2) - \log(\sigma_{i_j}^2))^2} \leq \sqrt{k(\log(\lambda_1^2) - \log(\sigma_k^2))^2}$$

where the first equal sign holds when $C^2 = P^T \Sigma^2 P$ and the second equal sign holds when $\lambda_1^2 = \lambda_2^2 = \dots = \lambda_k^2$ and $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$.

Combining the two different bounds yields that the lower bound on FRPSD

$$\begin{aligned} d^\alpha(X, Y) &\geq \sqrt{d_{\mathbf{St}}^2(U, VQ) + \alpha l_{\Lambda^2 \rightarrow C^2}^2} \\ &\geq \sqrt{d_{\mathbf{Gr}}^2(\text{col}(U), \text{col}(V)) + \alpha \sum_{i=1}^k (\log(\lambda_i^2) - \log(\sigma_i^2))^2} \end{aligned}$$

and the upper bounds

$$\begin{aligned} d^\alpha(X, Y) &\leq l_{\tau_{Q^*}} = \sqrt{d_{\mathbf{Gr}}^2(\text{col}(U), \text{col}(V)) + \alpha l_{\Lambda^2 \rightarrow Q^T \Sigma^2 Q^*}^2} \\ &\leq \sqrt{d_{\mathbf{Gr}}^2(\text{col}(U), \text{col}(V)) + \alpha \sum_{i=1}^k (\log(\lambda_i^2) - \log(\sigma_{i_j}^2))^2} \\ &\leq \sqrt{d_{\mathbf{Gr}}^2(\text{col}(U), \text{col}(V)) + k\alpha (\log(\lambda_1^2) - \log(\sigma_k^2))^2} \end{aligned}$$

for $X, Y \in \mathcal{S}_{n,p}^+$ with eigenvalues Λ^2 and Σ^2 . The equal sign in the lower bound is obtained when U and V are a canonical pair. The first equal sign in the upper bound is obtained when τ_{Q^*} is the geodesic. Based on these more accurate upper and lower bounds, the behavior of the Riemannian geodesic γ^α as α goes to 0 or to infinite can be better described as follows.

Proposition 7.3.10. *Let $X = U\Lambda^2U^T, Y = V\Sigma^2V^T \in \mathcal{S}_{n,p}^+$ be two points on fixed rank manifold where $U^T V$ is non-singular. Then the Riemannian geodesics $\hat{\gamma}^\alpha(t)$ with initial point $\forall(U, \Lambda^2) \in \hat{\varphi}^{-1}(X)$ lifted from the minimal Riemannian geodesic γ^α from X to Y converge to the special curve*

$$\hat{\gamma}^0(t) = \left([U \quad U_\perp] \exp \left(t \begin{bmatrix} \mathbf{0} & -K^T \\ K & \mathbf{0} \end{bmatrix} \right) I_{n,p}, \gamma_{C^2}(t) \right) := (\gamma_U^0(t), \gamma_{C^2}^0(t)) \quad (7.19)$$

as $\alpha \rightarrow 0$.

Proof. For curve $\hat{\gamma}^{\alpha\text{-Hor}}$ with $\alpha = 0$ or $\alpha > 0$, consider the energy functional under the metric \hat{g}^α

$$E_{\alpha\text{-Hor}}(\hat{\gamma}^\alpha) := E_U(\hat{\gamma}_U^\alpha) + \alpha E_{C^2}(\hat{\gamma}_{C^2}^\alpha), \alpha \geq 0$$

where E_U and the E_{C^2} are the energy functionals under Stiefel canonical metric and the affine-invariant metric of the respective curves.

Notice that $\gamma^{\alpha\text{-Hor}}$ is the minimal Riemannian geodesic under $g^{\alpha\text{-Hor}}$, it has energy no larger than the energy of γ^0 under g^α , i.e., $E_{\alpha\text{-Hor}}(\gamma^\alpha) \leq E_{\alpha\text{-Hor}}(\gamma^0)$. It yields

$$\limsup_{\alpha \rightarrow 0} E_{\alpha\text{-Hor}}(\gamma^{\alpha\text{-Hor}}) \leq \lim_{\alpha \rightarrow 0} E_{\alpha\text{-Hor}}(\gamma^0)$$

Then, $\lim_{\alpha \rightarrow 0} E_{\alpha\text{-Hor}}(\gamma^0) = E_U(\gamma_U^0)$ and $E_U(\gamma_U^{\alpha\text{-Hor}}) < E_{\alpha\text{-Hor}}(\gamma^{\alpha\text{-Hor}})$. It further yields

$$\limsup_{\alpha \rightarrow 0} E_U(\gamma_U^{\alpha\text{-Hor}}) \leq E_U(\gamma_U^0).$$

On the other hand, since the energy functional E_U is also the path energy functional of curve connecting U with fiber $[V] = \{VQ : Q \in \mathbf{SO}_n\}$. By definition $\gamma_U(t)$ is the unique minimal Riemannian geodesic connecting U with $[V]$ under the canonical metric on Stiefel, i.e., $E_U(\gamma_U^0) \leq E_U(\gamma_U^{\alpha\text{-Hor}})$ which further implies

$$E_U(\gamma_U^0) \leq \liminf_{\alpha \rightarrow 0} E_U(\gamma_U^{\alpha\text{-Hor}}).$$

The bounds conclude that $\lim_{\alpha \rightarrow 0} E_U(\gamma_U^{\alpha\text{-Hor}}) = E_U(\gamma_U^0)$. Since $E_U(\bar{\gamma}_U^0)$ is obtained only for the unique minimal geodesic $\bar{\gamma}_U^0$, $\gamma_U^{\alpha\text{-Hor}}(t)$ converges to $\gamma_U^0(t)$ as $\alpha \rightarrow 0$ for all $t \geq 0$. $\gamma_{C^2}^{\alpha\text{-Hor}}(t) \rightarrow \gamma_{C^2}^0(t)$ follows immediately as the $\gamma_U^{\alpha\text{-Hor}}(t)$ have fully characterized the endpoints of $\gamma_{C^2}^{\alpha\text{-Hor}}(t)$. \square

Proposition 7.3.11. *Let $X = U\Lambda^2U^\top, Y = V\Sigma^2V^\top \in \mathcal{S}_{n,p}^+$ be two points on fixed rank manifold where $U^\top V$ is nonsingular. Then the Riemannian geodesics $\hat{\gamma}^\alpha(t)$ with initial point $\forall (U, \Lambda) \in \hat{\varphi}^{-1}(X)$ lifted from the minimal Riemannian geodesic γ^α from X to Y converge to the special curve*

$$\gamma^\infty(t) = (\gamma_U(t), \Lambda^{1-t}\Sigma^t) := (\gamma_U^\infty(t), \gamma_{C^2}^\infty(t)) \quad (7.20)$$

where Λ and Σ has eigenvalues sorted by magnitude on the diagonals as $\alpha \rightarrow \infty$.

Proof. Similar prove can be generated with the energy functional

$$E_{\alpha\text{-Hor}}(\gamma^{\alpha\text{-Hor}}) := \frac{1}{\alpha} E_U(\hat{\gamma}_U^\alpha) + E_R(\hat{\gamma}_R^\alpha), \alpha \in (0, \infty]$$

and the lower bound of the $E_{C^2}(\gamma_{C^2}^{\alpha\text{-Hor}}) \geq \sum_i^p (\log(\lambda_i^2) - \log(\sigma_i^2))$. \square

CHAPTER 8

CONCLUSION AND FUTURE RESEARCH

Motivated by the smoothly evolving geodesic problem on the special orthogonal group, this dissertation closely investigates the differentiable geometry within the matrix exponential restricted to the set of skew symmetric matrices. Multiple important notions are presented and developed based upon the new geometric understanding on the set of skew symmetric matrices and the special orthogonal group, including the nearby matrix logarithm, the co-manifold characterization and the velocity-based Karcher mean on a Riemannian manifold. With the carefully designed and implemented routines, numerical experiments have demonstrated the value of these studies of these Riemannian objects. The major contributions of this dissertation are:

1. **The characterization of the conjugate locus in the special orthogonal group;**

This dissertation gives the first explicit characterization of the conjugate locus in the special orthogonal group that is expressed in the set of skew symmetric matrices. While there is a general characterization for the conjugate locus on the Lie group [21], the characterization (2.21) reveals more structures specific to the special orthogonal group and helps develop other novel results in this dissertation.

2. **The nearby matrix logarithm on the special orthogonal group constructed on the local diffeomorphism in the set of skew symmetric matrices;**

The local diffeomorphisms with the inscribed ball construction given in (3.11) are the immediate applications of the conjugate locus in the special orthogonal group. These local diffeomorphisms not only clarify the definition of the nearby matrix logarithm that was proposed before in a less rigorous description, but also extend its usage to skew symmetric matrices beyond the principal branch of the matrix logarithm. The nearby matrix logarithm redefined in **Definition 3.4.1** and computed in **Algorithm 3** or **Algorithm 4**, as the first reliable toolset, is essential to those applications that work with the skew symmetric matrices beyond the principal branch.

3. **The smoothly evolving geodesic problem on \mathbf{SO}_n ;**

The smoothly evolving geodesic problem aims to recover the smooth structure in the Riemannian exponential, which is lost in the Riemannian logarithm due to the shortest condition on geodesics. Having this problem solved on \mathbf{SO}_n facilitates the introduction of additional smooth structures on \mathbf{SO}_n for various applications, including the Karcher mean problem.

4. **The co-manifold characterization on \mathbf{Skew}_n ;**

The co-manifold characterization (4.8) and (4.10) on \mathbf{Skew}_n is a smooth structure developed on the smoothly evolving geodesic problem. It is in a locally diffeomorphism with the special orthogonal group and constructed by further restricting endpoints in the smoothly evolving geodesic problem to vary along geodesics in \mathbf{SO}_n . This smooth structure is especially useful for iterative algorithm as it guarantees that the moving along geodesics in \mathbf{SO}_n executed within each step of an algorithm is equivalent with moving on a smooth structure in \mathbf{Skew}_n . The algorithmic formulation on \mathbf{SO}_n can then be translated to an algorithmic formulation on \mathbf{Skew}_n with smoothness maintained.

5. **The velocity-based Karcher mean on \mathbf{SO}_n ;**

The velocity-based Karcher mean on a Riemannian manifold proposed in **Definition 5.3.8** is the first generalization of the Karcher mean formulation that addresses the non-smooth objective issues in the classic Karcher mean formulation on a Riemannian manifold. Having this problem solved and reliably computed in **Algorithm 6** significantly extends the application of Karcher mean on \mathbf{SO}_n with widely separated data sets and with backgrounds with smoothness constraints.

6. **The endpoint geodesic problem on $\mathbf{St}_{n,p}$;**

This dissertation is the first study that address the issues of widely separated endpoints issues in the current algorithms of computing endpoint geodesic problem on the Stiefel manifold. It proposes a root-finding formulation on \mathbf{Skew}_n to solve this endpoint geodesic problem and the resulting Newton algorithm implemented in **Algorithm 7** have obtained superior performance against the state-of-the-art algorithm proposed in [42].

7. **The new Riemannian metric in the FRPSD manifold;**

The FRPSD manifold has been studied in the literature. The Riemannian manifold proposed in [36] have disireable and useful properties including completeness in the metric, but it is at times too complicated in practice. The Riemannian manifold proposed in [23] is convenient to use but not complete in the metric space. The work in [3] attempts to construct a Riemannian metric that induces interpretable geodesics, e.g., in terms of a physics background, but it fails at building a Riemannian submersion structure. This dissertation continues the work in [3] and constructs a Riemannian submersion structure that induces meaningful Riemannian geodesics in FRPSD.

8. **Efficient implementations of low-level primitive associated to skew symmetric matrices, special orthogonal matrices;**

There are rich structures and primitives in the set of skew symmetric matrices and the special orthogonal matrices that can be exploited to accelerate computations. The characterization of these low level structures and primitives, e.g., (2.5) for Schur decompositions and the parameters in (2.14) and (2.16) for the differential, significantly accelerates the basic computation like the matrix exponential and its differential in the scope of $2 \sim 4$.

There are directions of the future research in both theoretical analysis and applications. For theoretical analysis on the smoothly evolving geodesic problem, there are still unknown structures of the co-manifold characterization it induces, especially at the conjugate locus. Having this smooth structure studied and understood will lead to further algorithmic analysis and design on the co-manifold characterization. In particular, the answer to the uniqueness of the velocity-based Karcher mean problem, the better answer to the existence of the solution to the endpoint geodesic problem and more depend on expanding the knowledge of this co-manifold characterization. Also, there are many Riemannian objects on the new Riemannian metric on the FRPSD manifold that are not identified. For applications, the new velocity-based Karcher mean can be applied to problems with smooth constraints or with separated data set that was poorly handled before. The faster endpoint geodesic computation in $\mathbf{St}_{n,p}$ with reliable performance for further separated endpoints is also useful in many applications.

BIBLIOGRAPHY

- [1] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.
- [2] Silvere Bonnabel, Anne Collard, and Rodolphe Sepulchre. “Rank-preserving geometric means of positive semi-definite matrices.” In: *Linear Algebra and Its Applications* 438.8 (2013), pp. 3202–3216.
- [3] Silvere Bonnabel and Rodolphe Sepulchre. “Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank.” In: *SIAM Journal on Matrix Analysis and Applications* 31.3 (2009), pp. 1055–1070.
- [4] William M Boothby. *An introduction to differentiable manifolds and Riemannian geometry*. Vol. 120. Academic press, 1986.
- [5] Nicolas Boumal. “Interpolation and regression of rotation matrices.” In: *International Conference on Geometric Science of Information*. Springer. 2013, pp. 345–352.
- [6] Darshan Bryner. “Endpoint Geodesics on the Stiefel Manifold Embedded in Euclidean Space.” In: *SIAM Journal on Matrix Analysis and Applications* 38.4 (2017), pp. 1139–1159.
- [7] KS Chan and B Munoz-Hernandez. “A generalized linear model for repeated ordered categorical response data.” In: *Statistica Sinica* (2003), pp. 207–226.
- [8] Baoline Chen and Peter A Zadrozny. “Analytic derivatives of the matrix exponential for estimation of linear continuous-time models.” In: *Journal of Economic Dynamics and Control* 25.12 (2001), pp. 1867–1879.
- [9] Luca Dieci et al. “On real logarithms of nearby matrices and structured matrix interpolation.” In: *Applied numerical mathematics* 29.1 (1999), pp. 145–165.
- [10] Shaoyi Du et al. “Affine iterative closest point algorithm for point set registration.” In: *Pattern Recognition Letters* 31.9 (2010), pp. 791–799.
- [11] Alan Edelman, Tomás A Arias, and Steven T Smith. “The geometry of algorithms with orthogonality constraints.” In: *SIAM journal on Matrix Analysis and Applications* 20.2 (1998), pp. 303–353.
- [12] Masoud Faraki, Mehrtash T Harandi, and Fatih Porikli. “Image set classification by symmetric positive semi-definite matrices.” In: *2016 IEEE Winter conference on applications of computer vision (WACV)*. IEEE. 2016, pp. 1–8.

- [13] Aasa Feragen and Andrea Fuster. “Geometries and interpolations for symmetric positive definite matrices.” In: *Modeling, Analysis and Visualization of Anisotropy*. Springer, 2017, pp. 85–113.
- [14] Wolfgang Förstner and Boudewijn Moonen. “A metric for covariance matrices.” In: *Geodesy-The Challenge of the 3rd Millennium*. Springer, 2003, pp. 299–309.
- [15] Yue Gao et al. “3-D object retrieval and recognition with hypergraph analysis.” In: *IEEE transactions on image processing* 21.9 (2012), pp. 4290–4303.
- [16] Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013.
- [17] Karsten Grove and Hermann Karcher. “How to conjugate C 1-close group actions.” In: *Mathematische Zeitschrift* 132.1 (1973), pp. 11–20.
- [18] Nicholas J. Higham. “The Scaling and Squaring Method for the Matrix Exponential Revisited.” In: *SIAM Journal on Matrix Analysis and Applications* 26.4 (2005), pp. 1179–1193.
- [19] Velimir Jurdjevic, Irina Markina, and F Silva Leite. “Extremal curves on Stiefel and Grassmann manifolds.” In: *The Journal of Geometric Analysis* 30.4 (2020), pp. 3948–3978.
- [20] Krzysztof Krakowski, Knut Hüper, and J Manton. “On the computation of the Karcher mean on spheres and special orthogonal groups.” In: *Conference Paper, Robomat*. Citeseer, Coimbra, Portugal. 2007.
- [21] John M Lee and John M Lee. *Smooth manifolds*. Springer, 2012.
- [22] Ruonan Li et al. “Differential geometric representations and algorithms for some pattern recognition and computer vision problems.” In: *Pattern Recognition Letters* 43 (2014), pp. 3–16.
- [23] Estelle Massart and P. A. Absil. “Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices.” In: (2018).
- [24] Simon Matabaigne et al. “The eigenvalue decomposition of normal matrices by the decomposition of the skew-symmetric part with applications to orthogonal matrices.” In: *arXiv preprint arXiv:2410.12421* (2024).
- [25] Awad H Al-Mohy and Nicholas J Higham. “A new scaling and squaring algorithm for the matrix exponential.” In: *SIAM Journal on Matrix Analysis and Applications* 31.3 (2010), pp. 970–989.
- [26] Awad H Al-Mohy and Nicholas J Higham. “Computing the Fréchet derivative of the matrix exponential, with an application to condition number estimation.” In: *SIAM Journal on Matrix Analysis and Applications* 30.4 (2009), pp. 1639–1657.

- [27] Awad H Al-Mohy, Nicholas J Higham, and Samuel D Relton. “Computing the Fréchet derivative of the matrix logarithm and estimating the condition number.” In: *SIAM Journal on Scientific Computing* 35.4 (2013), pp. C394–C410.
- [28] Igor Najfeld and Timothy F Havel. “Derivatives of the matrix exponential and their computation.” In: *Advances in applied mathematics* 16.3 (1995), p. 321.
- [29] Du Nguyen. “Closed-form geodesics and optimization for Riemannian logarithms of Stiefel and flag manifolds.” In: *Journal of Optimization Theory and Applications* 194.1 (2022), pp. 142–166.
- [30] Linyu Peng et al. “The geometric structures and instability of entropic dynamical models.” In: *Advances in Mathematics* 227.1 (2011), pp. 459–471.
- [31] Wulf Rossmann. *Lie groups: an introduction through linear groups*. Vol. 5. Oxford University Press on Demand, 2006.
- [32] Marco Sutti. “Riemannian algorithms on the Stiefel and the fixed-rank manifold.” PhD thesis. Ph. D. thesis, Université de Geneve, 2020, <https://archive-ouverte.unige...>, 2020.
- [33] Marco Sutti. “Shooting methods for computing geodesics on the Stiefel manifold.” In: *arXiv preprint arXiv:2309.03585* (2023).
- [34] Joel A Tropp et al. “Fixed-rank approximation of a positive-semidefinite matrix from streaming data.” In: *Advances in Neural Information Processing Systems* 30 (2017).
- [35] Pavan Turaga et al. “Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition.” In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.11 (2011), pp. 2273–2286.
- [36] Bart Vandereycken, P. A. Absil, and Stefan Vandewalle. “A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank.” In: *IMA Journal of Numerical Analysis* 33.2 (2013), pp. 481–514.
- [37] Ralph M Wilcox. “Exponential operators and parameter differentiation in quantum physics.” In: *Journal of Mathematical Physics* 8.4 (1967), pp. 962–982.
- [38] Shihui Ying et al. “Compute Karcher means on $SO(n)$ by the geometric conjugate gradient method.” In: *Neurocomputing* 215 (2016), pp. 169–174.
- [39] Xinru Yuan et al. “A Riemannian quasi-Newton method for computing the Karcher mean of symmetric positive definite matrices.” In: *Florida State University (FSU17-02)* (2017).
- [40] Yi Zhou et al. “On the continuity of rotation representations in neural networks.” In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 5745–5753.

- [41] Ralf Zimmermann. “A matrix-algebraic algorithm for the Riemannian logarithm on the Stiefel manifold under the canonical metric.” In: *SIAM Journal on Matrix Analysis and Applications* 38.2 (2017), pp. 322–342.
- [42] Ralf Zimmermann and Knut Hüper. “Computing the Riemannian logarithm on the Stiefel manifold: metrics, methods and performance.” In: *arXiv preprint arXiv:2103.12046* (2021).

BIOGRAPHICAL SKETCH

Zhifeng Deng, currently pursuing a Ph.D. in Applied and Computational Mathematics at Florida State University, has demonstrated a profound commitment to advancing the field of manifold optimizations and biomathematics computing. His research, deeply rooted in the Riemannian structure and submersion structure, focuses particularly on the Stiefel manifold and the fixed rank semi-symmetric-positive-definite (SPSD) manifold. Deng's work extends beyond theoretical exploration, bridging the gap between differential geometry and practical computational applications. This includes his innovative approach to computing a range of curves in endpoint format on manifolds, notably the Riemannian geodesic on the Stiefel manifold and the Riemannian geodesics on the fixed rank SPSPD manifold.

Deng's contributions to high-performance computing are equally noteworthy. He has been instrumental in developing a Julia library for manifold optimization, addressing the lack of complete wrappers and interfaces for essential low-level computations in this emerging programming language. His efforts in this area promise to lay a foundational framework for future developments in manifold optimizations. Additionally, his role as a Ph.D. Research Assistant in a collaborative research project, focusing on the structure of phylogenetic tree space, has led to significant advancements in TreeScaper software. Here, Deng's expertise in algorithm and software engineering has resulted in substantial improvements in computational efficiency and the development of new methodologies for phylogenetic analysis.

Deng's academic journey, marked by a dual B.S. in Mathematics and Applied Mathematics, and Electronic Science and Technology from Nankai University (Tianjin, China), reflects his multidisciplinary expertise. His teaching experience at Florida State University, coupled with his distinguished award in teaching, further underscores his ability to disseminate complex concepts effectively. Deng's blend of theoretical acumen, practical skill in high-performance computing, and dedication to teaching positions him as a rising scholar and innovator in his field.

ProQuest Number: 31632815

INFORMATION TO ALL USERS

The quality and completeness of this reproduction is dependent on the quality and completeness of the copy made available to ProQuest.



Distributed by
ProQuest LLC a part of Clarivate (2025).
Copyright of the Dissertation is held by the Author unless otherwise noted.

This work is protected against unauthorized copying under Title 17,
United States Code and other applicable copyright laws.

This work may be used in accordance with the terms of the Creative Commons license
or other rights statement, as indicated in the copyright statement or in the metadata
associated with this work. Unless otherwise specified in the copyright statement
or the metadata, all rights are reserved by the copyright holder.

ProQuest LLC
789 East Eisenhower Parkway
Ann Arbor, MI 48108 USA