ABSTRACT

Title of dissertation: MEANS AND AVERAGING ON RIEMANNIAN MANIFOLDS

Bijan Afsari, Doctor of Philosophy, 2009

Dissertation directed by: Professor P. S. Krishnaprasad
Department of Electrical and Computer Engineering, and Program of Applied Mathematics and Scientific Computations
Professor K. Grove
Department of Mathematics

Processing of manifold-valued data has received considerable attention in recent years. Standard data processing methods are not adequate for such data. Among many related data processing tasks finding means or averages of manifold-valued data is a basic and important one. Although means on Riemannian manifolds have a long history, there are still many unanswered theoretical questions about them, some of which we try to answer. We focus on two classes of means: the Riemannian L^p mean and the recursive-iterative means. The Riemannian L^p mean is defined as the solution(s) of a minimization problem, while the recursive-iterative means are defined based on the notion of Mean-Invariance (MI) in a recursive and iterative process. We give a new existence and uniqueness result for the Riemannian L^p mean. The significant consequence is that it shows the local and global definitions of the Riemannian L^p mean coincide under an uncompromised condi-

tion which guarantees the uniqueness of the local mean. We also study smoothness, isometry compatibility, convexity and noise sensitivity properties of the L^p mean. In particular, we argue that positive sectional curvature of a manifold can cause high sensitivity to noise for the L^2 mean which might lead to a non-averaging behavior of that mean. We show that the L^2 mean on a manifold of positive curvature can have an averaging property in a weak sense. We introduce the notion of MI, and study a large class of recursive-iterative means. MI means are related to an interesting class of dynamical systems that can find Riemannian convex combinations. A special class of the MI means called pairwise mean, which through an iterative scheme called Perimeter Shrinkage is related to cyclic pursuit on manifolds, is also studied. Finally, we derive results specific to the special orthogonal group and the Grassmannian manifold, as these manifolds appear naturally in many applications. We distinguish the 2-norm Finsler balls of appropriate radius in these manifolds as domains for existence and uniqueness of the studied means. We also introduce some efficient numerical methods to perform the related calculations in the specified manifolds.

MEANS AND AVERAGING ON RIEMANNIAN MANIFOLDS

by

Bijan Afsari

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:
Professor P. S. Krishnaprasad, Co-Chair and Advisor
Professor Karsten Grove, Co-Chair and Co-Advisor
Professor Pamela Abshire (Dean's Representative)
Professor Ramalingam Chellappa
Professor Ricardo Nochetto

Dedication

This dissertation is dedicated to my parents.

Acknowledgments

I feel privileged to have the opportunity to learn from two fine scientists: Professors P. S. Krishnaprasad and Karsten Grove. Prof. Krishnaprasad introduced me to geometry and geometrical thinking in studying many applied problems. His methodology of thinking will have a lasting impact on me. I am also thankful for his supports of all kind during the past several years, as well as for his patience in dealing with a not-so-patient student like me. I have learnt a lot about Riemannian geometry from Prof. Grove, and I have become an admirer of his thinking style. I certainly enjoyed my trips to the University of Notre Dame and discussions with him there. I have to thank my advisors for giving me the freedom and independence to think about problems that I found interesting.

As far as this dissertation is concerned, realistically, after my advisors the next entity to thank is the Google search engine, for obvious reasons!

I should also thank my Ph.D. committee members for their insightful comments.

I cannot forget the helps and cares from the people in the math department especially Alverda McCoy and Celeste Regaldo. Another group of people whom I have to thank is the nice people of International Education Services (IES) office at UMD.

Throughout my undergraduate studies in Sharif University, and my graduate studies in Tehran Polytechnic and the University of Maryland, I have benefited from attending many classes taught by excellent teachers. Among them I would

like to thank Professors: H. Amindavar, M. Dehghan, M. Freidlin, H. Hashemi, M. Jahanbegloo, M. Machedon, D. Madan, R. Nochetto, H. Sheikhzadeh, and A. Tits.

My list of friends who made my graduate life almost always enjoyable is too long, all of whom I thank. Fortunately, the list of those who directly influenced this dissertation is short: I thank Arash for his helps with LaTeXand Pedram for discussions about consensus algorithms. I also thank Ali Hirsa for several reasons.

I thank my family for their constant support. I cannot imagine finishing this dissertation without the helps from brother Bahman.

This research was supported in part by the Army Research Office under the ODDR&E MURI01 Program Grant No. DAAD19-01-1-0465 to the Center for Communicating Networked Control Systems (through Boston University), by NSF-NIH Collaborative Research in Computational Neuroscience Program (CRCNS2004) NIH-NIBIB grant 1 R01 EB004750-1, and by NSF grants DMS-0204671 and DMS-0706791.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

It is an irony of the history of mathematics that Gauss's efforts in the analysis of astronomical measurements led to the theory of statistics in Euclidean space rather than to the theory of statistics in manifolds [65, p. 12][1]. Gauss was interested in predicting spherical quantities but since the measurements were concentrated in a small region he approximate the sphere by its tangent space around that point of concentration, and hence reached at the celebrated Linear Mean Square estimation technique. In fact, the approach of linearization of a manifold in order to process a data set which takes values on the manifold is the most natural approach, and is used in everyday life. A very common example is the case of directional data, e.g, the heading of a vehicle moving in a plane or the direction of wind with respect to the north. In both cases the data values live on the unit circle S^1, but are represented by a single number θ in an interval such as $[0, 2\pi)$. One immediate problem is

[1]Of course, we should not ignore the fact that it was almost 30 years later than Gauss's initial calculations that mathematicians obtained a better understanding of the geometry of non-Euclidean spaces. The efforts of Gauss and many others in this direction culminated in the 1854 introduction of Riemannian Geometry by his student B. Riemann. The interesting point is that statistics in Euclidean space developed and flourished independently, while statistics in Riemannian manifolds is still in its development stage. One reason is that Riemannian geometry itself was developed mostly in the 20th century.

that mathematically there is no preferred reference point on S^1 and it is not clear that if we choose another point of reference how the result of our subsequent data processing will be affected. In other words the choice of the point around which we linearize the manifold as well as the actual linearization could affect our calculations. Of course, with a smart linearization and especially if the data points are very close to the point of linearization, we can get away with potential problems, as Gauss did. However, a more systematic way is to consider the geometry of the underlying manifold and try to define data processing tasks in an intrinsic fashion which depend only on the geometry of the manifold and not specific linearizations.

In recent years there has been an increasing interest in geometric methods for processing of manifold-valued due to both diverse applications and the understanding that such methods can be beneficial. Manifold-valued data appear naturally in many diverse areas such crystallography [44], diffusion tensor medical imaging [12, 75], computer graphics [17, 83], robotics [74, 92], multi-input multi-out wireless communications systems [64, 48], statistical analysis of shapes [62], array processing [84, 85] and computer vision [87, 7]. The general approach to define geometric signal processing tasks for such data has been to redefine or generalize the standard tasks via appropriate geometrical concepts. A very basic signal processing task is to find the mean or average of a set of data points on a manifold. The definition of the weighted average via the Riemannian center of mass is an example of the mentioned redefinition. Several other data processing tasks also have been redefined e.g., the Principal Component Analysis as in [29]. Actual numerical implementation of these geometrical methods requires specific algorithms for computations on Riemannian

manifolds which also have received considerable attention from statistics, signal processing and numerical analysis communities e.g., [84, 26, 2, 30, 45, 61, 3]. We should also mention several interesting Ph.D. dissertations such as [10],[56] and [6] which study statistics and data processing on Riemannian manifolds.

1.1 Scope and Main Contributions of this Dissertation

Although, there has been a great number of interesting works related to processing of manifold-valued data, still the literature on this subject lacks many fundamental pieces and there are many un-answered theoretical and applied questions in this regard. The effort of this dissertation is to answer some of these questions which relate to the notion of averaging of data points on a Riemannian manifold [1]. For example, the notion of L^2 center of mass or Riemannian L^2 mean which dates back to E. Cartan has been used extensively in applied research. The definition Cartan used for such a mean was a global definition which in a manifold of non-positive curvature gives a well defined unique center of mass for a probability measure with compact support. On the other hand, so far (with the exception of [61] and [17]) available existence and uniqueness results for the Riemannian L^2 mean on manifolds of positive curvature have been about the local L^2 mean or center of mass [34, 50, 52]. In Chapter 3 of this dissertation, we give an existence and uniqueness result for the global Riemannian mean, which in particular shows the equivalence of the local and global definitions under an uncompromised condition (see Section 3.1.1 for more

[1]The focus of this dissertation is on the so-called intrinsic means and intrinsic averaging on Riemannian manifolds. The other approach which is called the extrinsic or embedding approach has been popular in the statistics community and is much more studied [65]. We shall touch upon the notion of extrinsic mean in few places such as Sections 5.5.1 and 6.7.

details)[1]. We treat the larger class of Riemannian $L^p (1 \leq p \leq \infty)$ means together and in particular we pay special attention to the L^∞ center of mass (see Section 4.2). Another theoretical question is the relation between the Riemannian center of mass of a set of points and the convex hull of the set. In Section 4.5 we address this issue in detail, and answer a related question by [32]. Another fundamental question is about the nature of the operation of averaging in a Riemannian manifold. By this we mean, to what extent averaging reduces the effect of noise? Above all the main reason to average a set of data points is to reduce the effect of noise! In Section 4.6, we define different forms of averaging properties for the Riemannian L^2 mean (see Definition 4.6.1), and show that in a manifold of positive curvature averaging might in some occasions, unfortunately, amplify the noise! This is the direct consequence of positive curvature and it seems to be unavoidable. However, we show that in a small enough domain in a manifold of nonnegative curvature the L^2 mean can have an averaging property in a weak sense (see Theorem 4.6.1). Although the Riemannian L^2 mean is the most popular mean used, interestingly, there are many other possibilities for defining the notion of mean in a Riemannian manifold. For example in [6] and [5] two different forms of mean have been introduced which have certain desirable properties. In particular, in [5], for points on the cone of positive definite matrices, a constructive recursive-iterative approach has been given to define the geometric mean of N points based on the mean of $N-1$ points, and recursively down to $N = 2$ points. In Chapter 5, we analyze a wide class of means which are defined in a

[1] Although the intended goal of current study has been data analysis on manifolds, some of the derived results and specifically Theorems 3.2.1 and 4.5.1 might be of broad interest to researchers in Riemannian geometry, as well.

similar or more general recursive-iterative manner. In particular, we give existence and uniqueness results for such means. Our results are more general than those obtained in [5], since we introduce and use different tools. In particular, we introduce the notion of a primitive mean function, which generalizes the notion of a primitive stochastic matrix (see Section 5.2 and Definition 5.2.3). The study of such recursive-iterative means leads to the analysis of an interesting class of dynamical systems that find Riemannian convex combinations. We use the diameter of the smallest closed ball containing the points comprising the state as a Lyapunov function. We call that ball the minimal ball (a.k.a. circumscribed ball, see Definition 4.2.1). Such a system finds the mean of the set of initial data points by iteratively replacing each data point with a "convex combination" of some or all data points and the state of the system converges to a single point which will be the mean of the initial state. In each step the mean of the points which comprise the state of the system remains unchanged. We call this property Mean Invariance (MI). A particular class of the mentioned means is called the pairwise mean (see Section 5.3), which can be implemented efficiently in certain symmetric (matrix related) manifolds such as the special orthogonal group $SO(n)$ and the Grassmannian $G_k(n)$. In Chapter 6 we give more specific (i.e., stronger) results for these two manifolds. In particular, we place emphasis on identifying large domains (compared with Riemannian balls, see Section 6.4 and [36] for more details) for the existence and uniqueness of the means introduced in the previous chapters. Such larger domains are not balls in the Riemannian metric but rather balls in a Finsler metric which is derived from the matrix 2-norm (instead of the trace or Frobenius norm) in the Lie algebra

of $SO(n)$. We also introduce some efficient techniques to construct (and extend) the geodesic between two points in $SO(n)$ or $G_k(n)$ without the use of matrix exponential and logarithm -which are computationally very demanding-. We apply the derived methods to two applications.

1.2 Outline of this Dissertation

Each of the subsequent chapters (except for Chapter 2) in this dissertation has its own introduction, contribution and outline sections (see Sections 3.1.2, 4.1.1, 5.1.1 and 6.1.1 for more details). Here, we briefly give an outline of the chapters for ease of use and navigation. Throughout the text, it is assumed that the reader is familiar with the basics of Riemannian geometry at the level of a first year graduate course on the subject. Still, in Chapter 2 we have collected some less commonly used results, mostly from global Riemannian geometry, which are critical for our later developments. In particular, results on the structure of convex sets, bounds on the Hessian of the distance function and triangle comparison theorems are very essential for our developments, and are collected in this chapter. In Chapter 3 after some historical background, Theorem 3.2.1 is proved which gives a new existence and uniqueness result for the Riemannian L^p mean. In Chapter 4 further properties of the Riemannian L^p mean such as smooth dependence on data points, isometry compatibility, convexity, and averaging properties are studied. Also a simple example is given regarding numerical calculation of the mean. In Chapter 5 a large class of recursive-iterative means based on the notion of Mean-Invariance (MI) is studied. Interesting relations to the Birkhoff Curve Shortening Scheme, Perron-

Frobenius theorem, consensus algorithms [71] and cyclic pursuit on manifolds are also unveiled. In Chapter 6, after reviewing the geometries of $SO(n)$ and $G_k(n)$ we study existence and uniqueness of different means in Riemannian and 2-norm Finsler balls; and we also derive some efficient geodesic calculations which can be used to efficiently calculate the MI means. At the end of this chapter we also give two applications in which our methods fit very well. We conclude the dissertation in Chapter 7 with a summary of the dissertation, and discussions on open problems and possible future research directions.

Chapter 2

Preliminaries from Riemannian Geometry

2.1 Introduction and Notations

In this chapter we briefly review some specific ingredients needed from Riemannian geometry to deal with our problems of interest. We assume the reader is familiar with basic concepts of Riemannian geometry. We state theorems with references for their proofs. Also since there is little consensus about the exact definition of some terms, especially those related to convexity, we take the opportunity to set forth the definitions.

2.1.1 Some Notational Conventions

Throughout this manuscript M is an n-dimensional (or, if specified, m-dimensional) complete Riemannian manifold equipped with a Riemannian metric g. We might write M^n to show the dimension of M. We only deal with the unique Levi-Civita connection derived from g. We denote the tangent space of M at $q \in M$ by T_qM. We denote the Riemannian metric at q by g_q. Recall that $g_q(.,.) : T_qM \times T_qM \to \mathbb{R}$ is a positive definite inner product which changes smoothly with q. Usually, instead

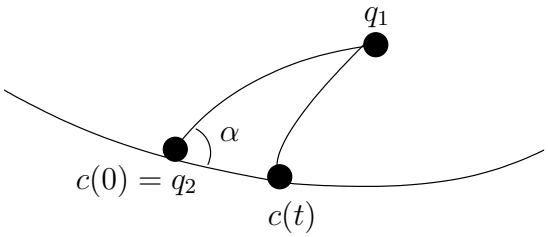
of $g_q(X,Y)$ which indicates the inner product of two tangent vector $X, Y \in T_qM$ we write $\langle X, Y\rangle_q$ and also when the base point q is understood unambiguously we omit the subscript and write $\langle X, Y\rangle$. Similarly and for convenience instead of $g_q(X,X)$ we write $\|X\|_q$ and even $\|X\|$. We call a unit-speed geodesic a normal geodesic. Such a geodesic is parameterized by its arc-length. By a minimizing, a minimal geodesic or a segment between two points we mean a geodesic whose length is equal to the distance between the two points. The letter B is reserved for metric balls and a ball $B(o,\rho)$ is an open ball with center $o \in M$ and radius ρ. We also denote a ball of radius ρ in T_qM with center $0 \in T_qM$ by $B(0,\rho)$. Unless otherwise stated, by a ball we mean an open ball. $\bar{A}$, int$A$ and ∂A denote the closure, interior and boundary of the set $A \subset M$, respectively. The exponential map of M at $q \in M$ is denoted by $\exp_q : T_qM \to M$ and whenever it is invertible its inverse is denoted by $\exp_q^{-1}$. We assume the injectivity radius of M to be nonzero; and we denote it by inj$M$. Note that $\exp_q$ is invertible on any ball $B(q,\rho)$ with $\rho < \text{inj}M$. We denote the curvature tensor of g by R. At point $q \in M$, $K_q(X,Y) = g_q(X, R(X,Y)Y)$ is the sectional curvature of the section spanned by orthonormal tangent vectors $X, Y \in T_qM$. If a is a matrix or a vector a^T denotes its transpose. We usually use subscripts and less often superscripts to denote indices of a sequence of points in a set. We also denote a sequence whose k^{th} term is x_k by $\langle x_k\rangle_{k=1}^{\infty}$ and when there is no confusion by $\langle x_k\rangle_k$ or $\langle x_k\rangle$.

2.2 First Variation of Arc Length

The calculation of the first variation of arch length functional is the standard method to characterize geodesics as stationary points (curves) of the arch length functional (e.g., [18, p.74]). Without going into further details we mention, for future use, the following useful immediate consequence of this calculation.

Proposition 2.2.1. Consider two points $q_1, q_2 \in M$ with $d(q_1, q_2) < \text{inj}M$ and let $\gamma : [0, 1] \to M$ be the minimizing geodesic connecting the two points with $\gamma(0) = q_1$ and $\gamma(1) = q_2$. Let $c : (-\epsilon, +\epsilon) \to M$ with $c(0) = q_2$ be a smooth curve passing through q_2 such that it makes an acute angle with γ at q_2, i.e., the angle between velocity vectors $-\dot{\gamma}(1)$ and $\dot{c}(0)$ is less than $\frac{\pi}{2}$. Then, there exists $\delta > 0$ such that for all $0 < t < \delta$, $d(q_1, c(t)) < d(q_1, q_2)$.

The angle condition ensures that the arc length functional for curves connecting q_1 to $c(t)$ is strictly decreasing for small $t > 0$. Figure 2.1 explains the proposition. We will use this result in few occasions.



$$\alpha < \tfrac{\pi}{2} \Rightarrow d(q_1, c(t)) < d(q_1, q_2),\ 0 < t < \delta$$

Figure 2.1: The first variation of arc length formula implies that, provided $\alpha < \frac{\pi}{2}$, by moving along $c(t)$ the distance to q_1 decreases at least for small times.

2.3 Structure of Convex Sets in Riemannian Manifolds

In order to study means in a Riemannian manifold M in the subsequent chapters, we need to have a good understanding of the structure of convex sets and balls in a Riemannian manifold. Theorems 2.3.1 and 2.3.2 below contain the main facts we need. We also prove some simple but useful corollaries of these theorems. For sake of completeness we recall some definitions first.

Definition 2.3.1 ([78, p.168])**.** Let C be a non-empty subset of a Riemannian manifold M.

1. C is strongly convex if for any two points $q_1, q_2 \in C$, there exists a unique minimizing geodesic γ in M connecting q_1 and q_2, and γ is contained in C.

2. If for any $q \in \bar{C}$, the closure of C, there exists an $\epsilon(q) > 0$ such that $C \cap B(q, \epsilon(q))$ is strongly convex, then C is said to be locally convex.

If C is strongly convex, then it is connected and locally convex. However, local convexity does not imply strong convexity. For example, on the unit sphere in $\mathbb{R}^3$, a closed hemisphere is locally convex but it is not strongly convex; since any two antipodal points on the boundary can be connected by two minimizing geodesics which lie entirely in the hemisphere. Small enough balls are strongly convex and the injectivity radius and upper bound on the sectional curvature of M determine a lower bound on the radius of largest strongly convex balls ([76, p. 177] and [18, p.404]):

Theorem 2.3.1. Let M be a complete Riemannian manifold with injectivity radius of $\mathrm{inj}M$ and sectional curvature upper bound of Δ. Let

$$r_{\mathrm{cx}} := \frac{1}{2}\min\{\mathrm{inj}M, \frac{\pi}{\sqrt{\Delta}}\}, \tag{2.1}$$

we interpret $\frac{1}{\Delta}$ when $\Delta \leq 0$ as ∞[1]. Then for any $o \in M$ the open ball $B(o,\rho)$ with $\rho \leq r_{\mathrm{cx}}$ and closed ball $B(o,\rho)$ with $\rho < r_{\mathrm{cx}}$ are strongly convex, and the distance function $x \mapsto d(o,x)$ is strictly convex along non-radial directions[2].

This theorem and the definition of r_{cx} are very important to us, as Theorem 3.2.1 which describes existence and uniqueness conditions concerns with strongly convex balls. We shall use the following simple corollary:

Corollary 2.3.1. A closed ball of radius $\rho < r_{\mathrm{cx}}$ is strongly convex; and the interior of the geodesic segment connecting any two points on the boundary lies in the interior of the ball.

Proof. Since $\overline{B(o,\rho)} \subset B(o,\rho')$ for $\rho < \rho' < r_{\mathrm{cx}}$; and since $x \mapsto d(o,x)$ is strictly convex in $B(o,\rho')$ along non-radial directions, it follows that if $q_1 \neq q_2$ are two points on the boundary of $B(o,\rho)$ and $\gamma : [0,1] \to M$ is a geodesic connecting them (with $\gamma(0) = q_1, \gamma(1) = q_2$), then $d(\gamma(t), o) < \rho$ for $t \in (0,1)$. □

One can show that the intersection of any family of strongly convex (locally convex) sets in a complete Riemannian manifold is strongly convex (locally convex).

[1] We shall follow this convention in all similar bounds we derive.

[2] Recall that the exponential map $\exp_o : T_oM \to M$ is a radial isometry, therefore $d(o,\gamma(\frac{1}{2})) = \frac{d(o,\gamma(0))+d(o,\gamma(1))}{2}$ for a geodesic $\gamma : [0,1] \to M$ connecting o to $x = \gamma(1)$ any x inside the injectivity ball of o.

We are mainly interested in strongly convex sets, so we define:

Definition 2.3.2. Let M be a complete Riemannian manifold. Let $A \subset M$ lie in a strongly convex subset of M. The convex hull of A denoted by ConvHull(A) is the intersection of all strongly convex sets containing A.

Definition 2.3.3 ([76] p.145)**.** A submanifold S of the complete Riemannian manifold M is called totally geodesic if geodesics in S are also geodesics in M and any geodesic in M which is tangent to S at some point lies in S for some time. S is globally totally geodesic if such a geodesic stays in S for all time.

According to this definition, a strongly convex open ball in M is a totally geodesic submanifold. This holds true for more general convex sets as the following theorem suggests. We state some results from [19] which are relevant.

Theorem 2.3.2 (Cheeger and Gromoll)**.** Let C be a connected locally convex subset of a Riemannian n-dimensional manifold M. Then the closure of C, $\overline{C}$, carries the structure of an embedded k dimensional ($0 \leq k \leq n$) topological submanifold N of M with smooth totally geodesic interior int$C$ and (possibly non-smooth or empty) boundary ∂C. For each $x \in C$ define the tangent cone

$$C_x = \{X \in T_xM | \exp_x \big(t\frac{X}{\|X\|}\big) \in N \text{ for some positive } t\} \cup \{0 \in T_xM\}. \qquad (2.2)$$

Then $C_x \backslash \{0\}$ is a k-dimensional open convex cone in T_xM. In case $x \in \text{int}C$, $C_x = T_x\text{int}C$. In particular, if there is a point $q \in \text{int}C$ and a normal minimal geodesic $\gamma : [0, d(x, d(x, q)] \to C$ from $x \in \partial C$ to q such that the distance between

q and ∂C is equal to $d(x,q)$ then $C_x\backslash\{0\}$ is a half-space. Moreover, if $q \in \text{int}C$ and $\gamma : [0, d(x,q)] \to M$ is the normal minimizing geodesic connecting $x \in \partial C$ to q, then $\gamma((0, d(x,q)]) \subset \text{int}C$.

The term "interior" here is not meant in its topological meaning; as for example a line segment in $\mathbb{R}^3$ has no topological interior yet it is closed and convex in $\mathbb{R}^3$. It is used in the same sprit as the "relative interior" or "interior" of a convex set in Euclidean space (see Appendix 4.8 for definitions). Therefore, we have the following useful corollary:

Corollary 2.3.2. Let x be point on the boundary of $B(o,\rho) \subset M$ with $\rho < r_{\text{cx}}$. Then $B_x\backslash\{0\} \subset T_xM$ is an open half-space and there exists a vector $\mathbf{n}_x \in C_x$ pointing inward B such that for any point $q \in B$ the (velocity vector of the) minimal geodesics connecting connecting x to q makes an angle less than $\frac{\pi}{2}$ with $\mathbf{n}_x$.

Proof. Note that since $B_x\backslash\{0\}$ is an open convex set it must have a supporting hyperplane at $0 \in T_xM$ and there should be a vector $\mathbf{n}_x \in T_xM$ pointing inward the cone which makes angles less than $\frac{\pi}{2}$ with all vectors pointing inside the cone $B_x\backslash\{0\}$ (see Appendix 4.8). □

The convex hull of a set is not necessarily closed. Moreover, in contrast to the Euclidean case, it is unknown whether the convex hull of a set of finite points in a Riemannian manifold of non-constant curvature is closed [9, p. 231]. Therefore, when dealing with points on the boundary of the convex hull of a set we need to make sure that the closure of the convex hull is strongly convex.

Proposition 2.3.1. Let $C \subset M$ be a strongly convex set such that $\overline{C}$ belongs to another bounded strongly convex set U. Then $\overline{C}$ is strongly convex. Therefore, if $A \subset M$ is such that the closure of its convex hull, ConvHull(A), belongs to a bounded strongly convex set, then $\overline{\text{ConvHull}(A)}$ is strongly convex.

Proof. It suffices to show that the unique minimal geodesic segment connecting $q_1, q_2 \in \partial C$ belongs to $\overline{C}$. Note that $\overline{C}$ is compact. Let $\langle q_1^k \rangle_k$ and $\langle q_2^k \rangle_k$ be two sequences in int$C$ with limits q_1 and q_2, respectively. Consider constant speed minimizing geodesics $\gamma_k : [0,1] \to \text{int}C$ such that $\gamma_k(0) = q_1^k$ and $\gamma_k(1) = q_2^k$. Note that the speed of γ_k is equal to $d(q_1^k, q_2^k)$. Therefore, speeds of these geodesics are uniformly bounded; and hence the sequence γ_k becomes an equicontinuous sequence on $\overline{C}$. By the Arzela-Ascoli theorem we can extract a subsequence which converges uniformly to a continuous curve $\gamma : [0,1] \to \bar{C}$. We know from the theory of length spaces (see e.g., [15, Proposition 2.5.17 on p. 48]) that γ is a shortest continuous curve connecting q_1 and q_2 in U. Therefore, due to strong convexity of U, γ must be the unique minimizing geodesic segment connecting q_1 and q_2. □

Note that on the unit sphere $S^2 \subset \mathbb{R}^3$ an open hemisphere is the largest strongly convex set and its closure is not strongly convex. Therefore, the condition on $\overline{C}$ to belong to another strongly convex set is necessary to guarantee the conclusion.

2.4 Bounds on the Hessian of the Distance Function and Its Powers

It is obvious that the convexity properties of the Riemannian distance function are closely related to existence and uniqueness properties of the L^p Riemannian means. Therefore, we review the bounds on the Hessian of the distance function and its powers. The gist of these comparison results is that in an appropriate comparison setting, as curvature increases the Hessian of the distance function (or any power of it) decreases. Recall that the Hessian of $f : M \to \mathbb{R}$ at $x \in M$ is the symmetric bilinear form $\text{Hess} f\big|_x : T_xM \times T_xM \to \mathbb{R}$, such that $\text{Hess} f\big|_x(\dot{\alpha}(t), \dot{\alpha}(t)) = \frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\alpha(t))$, where $\alpha : \mathbb{R} \to M$ is any smooth curve or geodesic passing through x at t. The treatment presented here is a standard one and is borrowed collectively from [78, 50]. Figure 2.2 shows the framework. Let $q \in M$ be a given point and $x \mapsto d(q, x)$ the distance function from that point. For convenience, we also denote this function by $d_q(x)$. Consider $\gamma(\cdot)$ a curve (or possibly a geodesic) such that at specific time t, $\gamma(t)$ is in the injectivity ball (or domain) around q, i.e., there is a unique minimizing geodesic connecting q to $\gamma(t)$. The goal is calculate the second derivative of $f(\gamma(t)) = d_q(\gamma(t))$. Consider the family of geodesics $c(s, t) = c_t(s)$ such that for each t, $c_t(s) : [0, d(q, \gamma(t))] \to M$ is a normal (i.e., unit speed) minimizing geodesic connecting q to $\gamma(t)$. Denote the covariant derivative along $c_t(s)$ by $\frac{D}{\partial s}$. Also we denote differentiation with respect to t by $\dot{}$ and with respect to s by $'$. Then $s \mapsto c_t(s)$ satisfies the geodesic equation for every t, i.e., $\frac{D}{\partial s} c_t'(s) = 0$. The geodesic equation is a second order nonlinear differential equation. In order to linearize it around $s \mapsto c_t(s)$ one finds a linear differential equation for the derivative

$J_t(s) = \frac{\partial c(s,t)}{\partial t} \in T_{c_t(s)}$. $s \mapsto J_t(s)$ is called the Jacobi field along geodesic $s \mapsto c_t(s)$. The Jacobi field satisfies the Jacobi equation:

$$\frac{D^2}{\partial s^2} J_t + R(J_t, c')c' = 0, \tag{2.3}$$

where R is the curvature tensor of M. There are two initial conditions $J_t(0)$ and $J_t'(0)$ associated with this equation. Note that the Jacobi equation is invariant

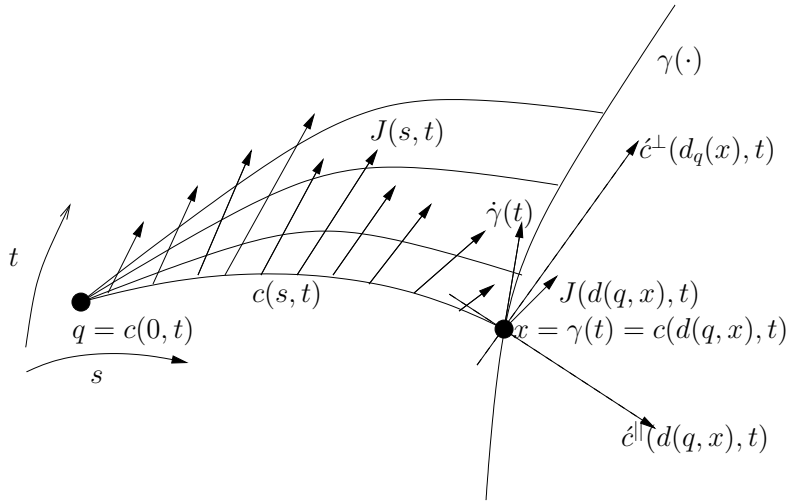


Figure 2.2: For each t the curve $s \mapsto c(s,t)$ is a normal minimal geodesic connecting $q = c(0,t)$ to $\gamma(t)$, where $\gamma(\cdot)$ is a curve passing through $x \in M$. $J(s,t) = \frac{\partial}{\partial t} c(s,t)$ is the associated Jacobi vector at $c(s,t)$. At each point on the geodesic $s \mapsto c_t(s)$, $J(s,t)$ has two components: one parallel to $c_t(s)$, denoted by $J^{\parallel}(s,t)$ and the other perpendicular to that, denoted by $J^{\perp}(s,t)$. These two components are not shown in the picture, but the two mentioned directions are shown.

under linear reparametrization of the geodesic $s \mapsto c_t(s)$. In many occasions, it is more convenient to assume that the original geodesic corresponds to $t = 0$; and as t deviates from 0, $c_t(s)$ deviates from $c_0(s)$. We also omit the subscripts from J_t and from $c_0(s)$. In order to be able to compare $c(s)$ with a nearby geodesic leaving q but with a different small velocity vector we need to assume $J(0) = 0$.

At any point $c(s)$ we can decompose $J(s) \in T_{c(s)}M$ to a component along the geodesic $s \mapsto c(s)$ denoted by $J^{\|}(s)$ and another one perpendicular to the geodesic denoted by $J^{\perp}(s)$. The parallel component is not so interesting since it can be computed explicitly as $J^{\|}(s) = (a + ts)c'(s)$ for constants $a, b \in \mathbb{R}$. However, the behavior of $J^{\perp}(s)$ depends on the curvature and is more interesting. It is possible to show that $\langle J'(s), c'(s)\rangle = \langle J'(0), c'(0)\rangle s + \langle J(0), c'(0)\rangle$ (see e.g., [24, p. 118]). Therefore, once we have $J(0) = 0$ and choose $J'(0)$ perpendicular to $c'(0)$, $J(s)$ remains perpendicular to $c'(s)$. A conjugate point of q along geodesic $c(s)$ is a point $q_c = c(s_c)$ and not equal to q, such that $J(s_c) = 0$ for some initial conditions of $J'(0)$. Let q_{s_1} be the first such a point. A crucial fact is that for $s \in (0, s_1)$ and any $X \in T_{c(s)}M$ one can find a unique Jacobi field with $J(0) = 0$ such that $J(s) = X$, where X is a given tangent vector in $T_{c(s_1)}M$ ([78, Lemma 2.4 p. 37]). Now let $x \in M\backslash\big(C_q \cup \{q\}\big)$ where C_q is the cut locus of q, i.e., the set of cut points of q[1]. For such an x let $c_s : [0, d(q, x)] \to M$ be the unique minimizing normal geodesic connecting q to x, i.e., $c(0) = q$ and $c(d(q, x)) = x$. For every $X \in T_xM$ we can find a Jacobi field $J(s)$ along $c(s)$ with $J(d(q, x)) = X$ with $J(0) = 0$. Then one can show that (see e.g., [78, Lemma 4.10 p. 109])

$$\mathrm{Hess} d_q|_x(X, X) = \langle J^{\perp'}(d_q(x)), J^{\perp}(d_q(x))\rangle. \tag{2.4}$$

Note that $J^{\perp'} = (J')^{\perp}$ and if $J(d_q(x)) = X$ is parallel to $c(d_q(x))$ then $J^{\perp}(d_q(x)) = 0$ and $\mathrm{Hess} d_q|_x(X, X) = 0$. This means that the Hessian of $x \mapsto d_q(x)$ is indefinite

[1]Recall that a point $z \in M$ is a cut point of q along a geodesic γ emanating from q, if z is the first point along γ after which γ ceases to be length minimizing.

along the geodesic connecting q to x although it might be positive definite along other directions. Equation (2.4) shows why we needed Jacobi fields to describe the Hessian of the distance function.

Combining (2.3) and (2.4) is not straightforward. But one can find lower and upper bounds on the Hessian of $x \mapsto d_q(x)$ based on the upper and lower bounds on the sectional curvatures of M, respectively. For that we need to solve (2.3) for constant curvature. If the sectional curvature is a constant κ, then (2.3) reads as

$$\frac{D^2}{\mathrm{d}s^2} J^\kappa(s) + \kappa J^\kappa(s) = 0. \tag{2.5}$$

As a matter uniform notations it is convenient to define:

$$\mathrm{sn}_\kappa(s) = \begin{cases} \frac{1}{\sqrt{\kappa}} \sin(\sqrt{\kappa} s) & \kappa > 0 \\ s & \kappa = 0 \\ \frac{1}{\sqrt{|\kappa|}} \sinh(\sqrt{|\kappa|} s) & \kappa < 0 \end{cases}, \quad \mathrm{cs}_\kappa(s) = \begin{cases} \cos(\sqrt{\kappa} s) & \kappa > 0 \\ 1 & \kappa = 0 \\ \cosh(\sqrt{|\kappa|} s) & \kappa < 0 \end{cases}. \tag{2.6}$$

It is easy to see that the solution to (2.5), can be written as $J^\kappa(s) = J^\kappa(0)\mathrm{cs}_\kappa(s) + (J^\kappa)'(0)\mathrm{sn}_\kappa(s)$.

Let $\delta \leq K \leq \Delta$ be the lower and upper bounds on the sectional curvatures of M. One can bound $\langle J^{\perp'}(s), J^\perp(s) \rangle$ in terms of $\langle (J^\Delta)^{\perp'}(s), (J^\Delta)^\perp(s) \rangle$ and $\langle (J^\delta)^{\perp'}(s), (J^\delta)^\perp(s) \rangle$ and this results in the following (see e.g., [78, Lemma 2.9 p. 153]):

Theorem 2.4.1. If $x \in B(q, \rho)$, where $\rho < \min\{\mathrm{inj}_q M, \frac{\pi}{\sqrt{\Delta}}\}$ if $\Delta > 0$ and $\rho < \mathrm{inj}_p M$

if $\Delta \leq 0$, then

$$\frac{\mathrm{cs}_\Delta(d_q(x))}{\mathrm{sn}_\Delta(d_q(x))}\langle X^\perp, X^\perp\rangle \leq \text{Hess } d_q\big|_x(X,X) \leq \frac{\mathrm{cs}_\delta(d_q(x))}{\mathrm{sn}_\delta(d_q(x))}\langle X^\perp, X^\perp\rangle, \tag{2.7}$$

where $X^\perp$ is the component of $X \in T_xM$ orthogonal to the geodesic connecting q to x, which is the same as the direction of the gradient of $x \mapsto d_q(x)$ at x.

From this theorem we observe that the upper curvature bound is the crucial factor in determining the lower bound on the Hessian (and its positiveness) of the distance function. In particular, if M is of non-positive curvature, then $x \mapsto d_q(x)$ is convex within injectivity balls around q. In addition, if M is simply connected and hence its injectivity radius is infinity, then the distance function in M is globally convex. Along directions other than the direction of the geodesic connecting q to $\gamma(t)$, the Hessian of $t \mapsto d_q(\gamma(t))$ is positive definite and the distance function is strictly convex. Note that the excluded or degenerate direction is exactly the direction of the gradient of $x \mapsto d_q(x)$ at $x = \gamma(t)$. We recall the same property for the distance function in $\mathbb{R}^n$. If M is of constant curvature κ, then the lower and upper bounds are equal, hence the Hessian of $x \mapsto d_q(x)$ is $\text{Hess}d_q|_x(X,X) = \frac{\mathrm{cs}_\kappa(d_q(x))}{\mathrm{sn}_\kappa(d_q(x))}\langle X^\perp, X^\perp\rangle$ for $X \in T_xM$.

Let S^n_κ be a simply connected complete Riemannian manifold of constant curvature κ and dimension $n = \dim M$. It is well-known that these conditions determine S^n_κ together with a Riemannian structure uniquely, up to isometry [78, p.137]. Therefore, if $\kappa > 0$ we can assume that S^n_κ is the standard sphere in $\mathbb{R}^{n+1}$ with radius $\frac{1}{\sqrt{\kappa}}$ whose injectivity radius is $\frac{\pi}{\sqrt{\kappa}}$. When $\kappa \leq 0$, as mentioned before,

the injectivity radius is infinity. Now again consider our manifold M with upper and lower curvature bounds of $\Delta > 0$ and $\delta \leq 0$, respectively. The corresponding spaces of constant curvature are $(S^n_\Delta, \tilde{g})$ and $(S^n_\delta, \tilde{\tilde{g}})$ with distances functions $\tilde{d}$ and $\tilde{\tilde{d}}$. Consider points $q, x \in M$ with $d(q,x) < \rho$ as in Theorem 2.4.1. Also consider pairs of points $\tilde{q}, \tilde{x} \in S^n_\Delta$ and $\tilde{\tilde{q}}, \tilde{\tilde{x}} \in S^n_\delta$, counterpart to q and x, such that $d(q,x) = \tilde{d}(\tilde{q}, \tilde{x}) = \tilde{\tilde{d}}(\tilde{\tilde{q}}, \tilde{\tilde{x}})$. Let γ,$\tilde{\gamma}$ and $\tilde{\tilde{\gamma}}$ be unit speed geodesics passing through x, $\tilde{x}$ and $\tilde{\tilde{x}}$, respectively; such that the angles between each of them and the geodesics connecting q ($\tilde{q}$ and $\tilde{\tilde{q}}$) to x ($\tilde{x}$ and $\tilde{\tilde{x}}$) are equal. Then we write:

$$\mathrm{Hess}\tilde{d}_{\tilde{q}} \leq \mathrm{Hess}d_q \leq \mathrm{Hess}\tilde{\tilde{d}}_{\tilde{\tilde{q}}} \tag{2.8}$$

meaning that:

$$\mathrm{Hess}\tilde{d}_{\tilde{q}}\big|_{\tilde{\gamma}(t)}(\dot{\tilde{\gamma}}(t), \dot{\tilde{\gamma}}(t)) \leq \mathrm{Hess}d_q(\dot{\gamma}(t), \dot{\gamma}(t)) \leq \mathrm{Hess}\tilde{\tilde{d}}_{\tilde{\tilde{q}}}\big|_{\tilde{\tilde{\gamma}}(t)}(\dot{\tilde{\tilde{\gamma}}}(t), \dot{\tilde{\tilde{\gamma}}}(t)). \tag{2.9}$$

We have this corollary:

Corollary 2.4.1. Let M be a complete Riemannian manifold of upper bounds on sectional curvature Δ. Define $f(x) = \frac{1}{p}d^p(q,x)$ which we also write as $f(x) = \frac{1}{p}d^p_q(x)$. Let $\gamma : \mathbb{R} \to M$ be a unit speed geodesic. Assume the angle between the geodesic connecting q to $x = \gamma(t)$ and $\gamma(t)$ be α then

$$\frac{\mathrm{d}f}{\mathrm{d}t}(\gamma(t)) = d^{p-1}(q, \gamma(t)) \cdot \cos\alpha \tag{2.10}$$

and

$$\frac{\mathrm{d}^2 f}{\mathrm{d}t^2}(\gamma(t)) \geq (p-1)\cdot d^{p-2}(q,\gamma(t))\cdot\cos^2\alpha + d^{p-1}(q,\gamma(t))\cdot\frac{\mathrm{cs}_\Delta d(q,\gamma(t))}{\mathrm{sn}_\Delta d(q,\gamma(t))}\cdot\sin^2\alpha \quad (2.11)$$

The latter is equivalent to

$$\mathrm{Hess} f|_q \geq \mathrm{Hess}\tilde{f}|_{\tilde{q}}. \quad (2.12)$$

If $M \equiv S^n_\Delta$ we have equality in the above.

Proof. We just need to calculate the first and second order derivatives of $f(\gamma(t)) = \frac{1}{p}d^p_q(\gamma(t))$ by the chain rule:

$$\frac{\mathrm{d}f}{\mathrm{d}t}(\gamma(t)) = d^{p-1}_q(\gamma(t))\frac{\mathrm{d}}{\mathrm{d}t}d_q(\gamma(t)) = d^{p-1}_q\langle\nabla d_q(\gamma(t)),\dot{\gamma}(t)\rangle \quad (2.13)$$

from which we have:

$$\frac{\mathrm{d}^2 f}{\mathrm{d}t^2}(\gamma(t)) = (p-1)d^{p-2}_q(\gamma(t))\langle\nabla d_q(\gamma(t)),\dot{\gamma}(t)\rangle^2 + d^{p-1}_q(\gamma(t))\frac{\mathrm{d}^2}{\mathrm{d}t^2}d_q(\gamma(t)). \quad (2.14)$$

Now the claims follow from Theorem 2.4.1. □

2.4.1 A Useful Length Comparison Estimate

Let $q, x, y \in B(o,\rho) \subset M$ where $\rho < r_{\mathrm{cx}}$ and assume $X, Y \in T_qM$ are such that $x = \exp_q X$ and $y = \exp_q Y$. Assume that M has lower and upper bounds of $\delta \leq 0$ and $\Delta \geq 0$ on its sectional curvature. We want to find upper and lower bounds on $d(x,y)$ in terms of $\|X - Y\|$ and the curvature bounds. Let $\gamma : [0,1] \to M$

be any curve in $B(o,\rho)$ connecting x to y such that $\gamma(0) = x$ and $\gamma(1) = y$. Let $c_t(s) : [0,1] \to M$, for fixed t, be a geodesic connecting q to $\gamma(t)$ such that $c_t(0) = q$ and $c_t(1) = \gamma(t)$. Denote by $J(s,t)$ the Jacobi field along $c_t(s)$ with $J(0,t) = 0$ and $J'(0,t)$ arbitrary. We are using the same conventions established before in this section. From Rauch's comparison theorem (e.g., [18, p. 390]) we have

$$\mathrm{sn}_{\Delta l_t^2}(s)\|J'(0,t)\| \leq \|J(s,t)\| \leq \mathrm{sn}_{\delta l_t^2}(s)\|J'(0,t)\|, \tag{2.15}$$

where $l_t = d(q,\gamma(t))$ (see (2.6)). The above holds for arbitrary $J'(0,t)$ since we assumed $\Delta \geq 0$ and $\delta \leq 0$, otherwise it would only hold for $J'(0,t)$ orthogonal to $c_t'(0)$ [18, p. 390]. Note that since $\gamma(t)$ with velocity $\dot\gamma(t) = J(1,t)$ is a curve connecting x to y we have $d(x,y) \leq \int_0^1 \|\dot\gamma\| = \int_0^1 \|J(1,t)\|\mathrm{d}t$. Let $J(t,s)$ be a particular Jacobi field such that $J'(0,t) = Y - X$, then we have $d(x,y) \leq \max_t \mathrm{sn}_{\delta l_t^2}(1) \cdot \|Y - X\|$. The best upper bound estimate for l_t is twice the radius of the ball i.e., $l_t < 2\rho$. So we have

$$\max_t \mathrm{sn}_{\delta l_t^2}(1) \leq \frac{1}{2\rho\sqrt{|\delta|}} \sinh 2\rho\sqrt{|\delta|}. \tag{2.16}$$

Now let $\gamma : [0,1] \to M$ be the geodesic connecting x to y. We have $\|\dot\gamma(t)\| = d(x,y)$. Back in T_qM there is a curve $\alpha : [0,1] \to T_qM$ with velocity vector $\dot\alpha(t) = J'(0,t)$ which induces the Jacobi field $J(t,s)$ with $\|J(1,t)\| = d(x,y)$. Now α is a curve connecting X to Y in T_qM, therefore

$$\|X - Y\| \leq \int_0^1 \|\dot\alpha\| = \int_0^1 \|J'(0,t)\|\mathrm{d}t \leq \frac{1}{\min_t \mathrm{sn}_{\Delta l_t^2}(1)} d(x,y). \tag{2.17}$$

Note that since the the distances $d(q,x)$ and $d(q,y)$ can be larger than r_{cx} we have no restriction on l_t based on convexity of the distance function; that is we cannot conclude that $l_t \leq \max\{\|X\|, \|Y\|\}$. However, as before we have $l_t < 2\rho$. As a result we have

$$\min_t \text{sn}_{\Delta l_t^2}(1) \geq \frac{1}{2\rho\sqrt{\Delta}} \sin 2\rho\sqrt{\Delta}. \tag{2.18}$$

Therefore, in a manifold of nonnegative curvature with curvature bounded from above by $\Delta > 0$ we have:

$$\frac{\sin 2\rho\sqrt{\Delta}}{2\rho\sqrt{\Delta}} \cdot \|X - Y\| \leq d(x,y) \leq \|X - Y\| \tag{2.19}$$

and in a manifold of non-positive curvature with curvature bounded from below by $\delta < 0$ we have:

$$\|X - Y\| \leq d(x,y) \leq \frac{\sinh 2\rho\sqrt{|\delta|}}{2\rho\sqrt{|\delta|}} \cdot \|X - Y\|. \tag{2.20}$$

We will use these estimates in Section 4.6.2 when we consider the sensitivity properties of the L^2 mean.

2.5 Cartan-Alexandrov-Toponogov (C.A.T.) Comparison Theorems

The material presented here can be found collectively in [15, pp.238-240], [51, pp.197-203] and [18, pp.399-403 and p.420].

Definition 2.5.1. A (geodesic) triangle in (M,d) consists of three vertices $q_1, q_2, q_3 \in M$ and three minimal geodesics segments connecting the three points. The angle

between the geodesic side connecting q_1 to q_2 and the geodesic side connecting q_1 to q_3 is denoted by $\angle q_2 q_1 q_3$. We denote the side between q_1 and q_2 by $\overline{q_1 q_2}$. The two sides $\overline{q_1 q_2}$ and $\overline{q_1 q_3}$ and angle $\angle q_2 q_1 q_3 = \alpha$ constitute a hinge at q_1 with angle α and sides $\overline{q_1 q_2}$ and $\overline{q_1 q_3}$. If the vertices are such that the sides of the triangle are determined uniquely, we denote the triangle by $\triangle q_1 q_2 q_3$. In referring to the vertices of the triangle, sometimes it is convenient to use modulo 3 cyclic indexing, e.g., $q_4 \equiv q_1$ and so forth. The perimeter of a triangle with vertices q_1, q_2, q_3 is $\sum_{i=1}^{3} d(q_i, q_{i+1})$.

As noted in the definition, a triangle is not necessarily determined uniquely by its vertices. However, if for a triangles with vertices $q_1, q_2, q_3 \in M$, we have $d(q_i, q_{i+1}) < \text{inj}M$ for $1 = 1, 2, 3$, then the geodesic sides are uniquely determined. In particular, if the perimeter of that triangle is smaller than $2\text{inj}M$, then the sides are defined uniquely. Since, otherwise, say $d(q_1, q_2) \geq \text{inj}M$, then the triangle inequality requires $d(q_2, q_3) + d(q_1, q_3) \geq \text{inj}M$, which contradicts the condition on the perimeter.

Definition 2.5.2. Let q_1, q_2 and q_3 be vertices of a triangle in M. A comparison triangle in $(S_\kappa^n, \tilde{d})$, if it exists, is a triangle with vertices $\tilde{q}_1, \tilde{q}_2, \tilde{q}_3 \in S_\kappa^n$ such that $\tilde{d}(\tilde{q}_i, \tilde{q}_{i+1}) = d(q_i, q_{i+1}), i = 1, 2, 3$. A comparison hinge in $(S_\kappa^n, \tilde{d})$ is a hinge with vertex $\tilde{q}_1 \in S_\kappa^n$ and sides $\overline{\tilde{q}_1 \tilde{q}_2}$ and $\overline{\tilde{q}_1 \tilde{q}_3}$ such that $\angle \tilde{q}_2 \tilde{q}_1 \tilde{q}_3 = \angle q_2 q_1 q_3$, $\tilde{d}(\tilde{q}_1, \tilde{q}_2) = d(q_1, q_2)$ and $\tilde{d}(\tilde{q}_1, \tilde{q}_3) = d(q_1, q_3)$.

When the lengths of two sides and the angle between them from a triangle in a manifold of constant curvature are given we can calculate the length of the third side and the two other angles. The calculation is carried out using the well-

known spherical, hyperbolic or Euclidean laws of cosines. For a triangle with vertices $q_1q_2q_3$ in the model space (S^n_κ, d), with $\kappa \neq 0$, the Law of Cosines for finding the side opposite to q_1 reads as [18, p. 103]:

$$\mathrm{cs}_\kappa d(q_2, q_3) = \mathrm{cs}_\kappa d(q_1, q_2) \cdot \mathrm{cs}_\kappa d(q_1, q_3) + \kappa \cdot \mathrm{sn}_\kappa d(q_1, q_2) \cdot \mathrm{sn}_\kappa d(q_1, q_3) \cdot \cos \angle q_2 q_1 q_3, \tag{2.21}$$

where functions sn_κ and cs_κ are defined in (2.6). However, in a general Riemannian manifold M^n, where the sectional curvatures are not necessarily constant, we cannot perform this calculation so neatly and explicitly. Nevertheless, one can find upper and lower bounds on the length of the third side based on lower and upper bounds on the sectional curvature of the manifold, respectively. Let $\delta \leq \Delta \in \mathbb{R}$ denote upper and lower sectional curvature bounds of M^n, respectively. The main idea is to compare the unknown length of the third side with the lengths of the third sides of comparison triangles (more accurately hinges, see below) in the corresponding model spaces, i.e., S^n_δ and S^n_Δ. A class of very useful triangle comparison theorems known as Toponogov or Cartan-Alexandrov-Toponogov (C.A.T) comparison theorems serve this purpose. These theorems can also be used to perform other forms comparisons: angle or secant comparisons. The statement given in the following is a very comprehensive one which we will use in different occasions [51, pp. 197-198]:

Theorem 2.5.1 (C.A.T Comparison Theorems)**.** Let (M^n, d) be a complete Riemannian manifold. For K, the sectional curvature of M assume $\delta \leq K \leq \Delta$. Denote the injectivity radius of M by $\mathrm{inj}M > 0$. Let r_{cx} be as in (2.1).

1. (Upper Bound on Curvature) If the perimeter of a triangle with vertices $q_1, q_2, q_3 \in$ is smaller than $4r_{\text{cx}}$, i.e., $\sum_{i=1}^{3} d(q_i, q_{i+1}) < 4r_{\text{cx}}$, then a comparison triangle $\triangle\tilde{q}_1\tilde{q}_2\tilde{q}_3$ in S^n_Δ exists and is unique up to congruence.

   (a) (Angle version) The angles of $\triangle\tilde{q}_1\tilde{q}_2\tilde{q}_3$ are not smaller than the corresponding angles of $\triangle q_1q_2q_3$, i.e., $\angle\tilde{q}_{i-1}\tilde{q}_i\tilde{q}_{i+1} \geq \angle q_{i-1}q_iq_{i+1}$.

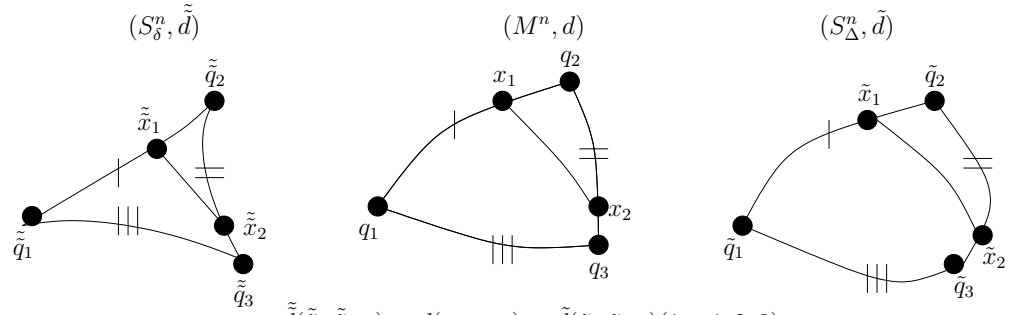
   (b) (Secant version) Let x and y be two points on two different sides of $\triangle q_1q_2q_3$. Via the obvious 1-1 correspondence between the points on the perimeters of the two triangles find two points $\tilde{x}$ and $\tilde{y}$ (corresponding to x and y, respectively) on the perimeter of $\triangle\tilde{q}_1\tilde{q}_2\tilde{q}_3$. Then we have that the secant connecting $\tilde{x}$ to $\tilde{y}$ is not shorter than the secant connecting x to y, i.e., $\tilde{d}(\tilde{x}, \tilde{y}) \geq d(x, y)$

   (c) (Hinge version) Corresponding to the hinge $(\angle q_2q_1q_3, \overline{q_1q_2}, \overline{q_1q_3})$ a comparison hinge $(\angle\tilde{q}_2\tilde{q}_1\tilde{q}_3, \overline{\tilde{q}_1\tilde{q}_2}, \overline{\tilde{q}_1\tilde{q}_3})$ in S^n_Δ exists and $d(q_2, q_3) \geq \tilde{d}(\tilde{q}_2, \tilde{q}_3)$.

2. (Lower Bound on Curvature: Toponogov Theorems) With curvature bounded from below there is no need for limiting the size of a triangle in M and for any triangle with vertices $q_1q_2q_3$ a comparison triangle in S^n_δ exists. Moreover, with $\delta > 0$ we have $\sum_i d(q_i, q_{i+1}) \leq 2\frac{\pi}{\sqrt{\delta}}$ and equality holds only if $M = S^n_\delta$. The Angle, Secant and Hinge versions in this part are exactly as in the previous part except that the directions of the inequalities are reversed.

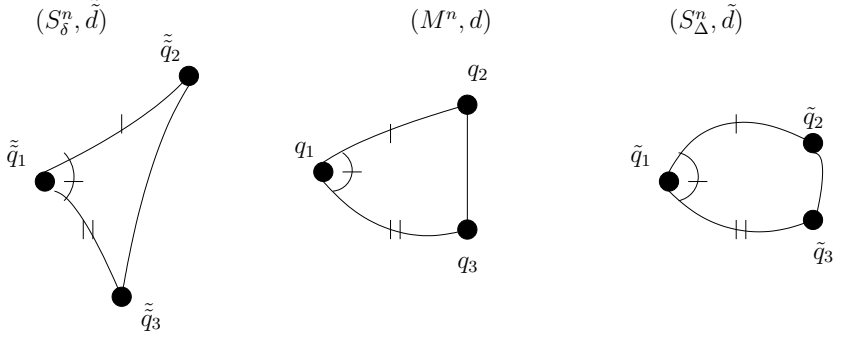
If the bounds on the curvature are strict bounds (e.g., $\delta < K$), then the corresponding inequalities can be replaced by strict inequalities.

One can think of the comparison triangle in S^n_Δ as fatter than our original triangle in M and the comparison triangle in S^n_δ as thinner than our original one in M. As a side, we mention that the remarkable property that the lower positive bound on curvature limits the perimeter of a triangle is the main theme in a class problems in global Riemannian geometry known as "pinching problems". Figure 2.3 pictorially explains the theorem.



$$\tilde{\tilde{d}}(\tilde{\tilde{q}}_i, \tilde{\tilde{q}}_{i+1}) = d(q_i, q_{i+1}) = \tilde{d}(\tilde{q}_i, \tilde{q}_{i+1})(i = 1, 2, 3)$$
$$\tilde{\tilde{d}}(\tilde{\tilde{x}}_i, \tilde{\tilde{q}}_i) = d(x_i, q_i) = \tilde{d}(\tilde{x}_i, \tilde{q}_i)$$
$$\Downarrow$$
$$\angle\tilde{\tilde{q}}_{i-1}\tilde{\tilde{q}}_i\tilde{\tilde{q}}_{i+1} \leq \angle q_{i-1}q_iq_{i+1} \leq \angle\tilde{q}_{i-1}\tilde{q}_i\tilde{q}_{i-1}(i = 1, 2, 3),$$
$$\tilde{\tilde{d}}(\tilde{\tilde{x}}_1, \tilde{\tilde{x}}_2) \leq d(x_1, x_2) \leq \tilde{d}(\tilde{x}_1, \tilde{x}_2)$$

[a]



$$\tilde{\tilde{d}}(\tilde{\tilde{q}}_1, \tilde{\tilde{q}}_i) = d(q_1, q_i) = \tilde{d}(\tilde{q}_1, \tilde{q}_i) \quad (i = 1, 2)$$
$$\angle\tilde{\tilde{q}}_2\tilde{\tilde{q}}_1\tilde{\tilde{q}}_3 = \angle q_2q_1q_3 = \angle\tilde{q}_2\tilde{q}_1\tilde{q}_3$$
$$\Downarrow$$
$$\tilde{\tilde{d}}(\tilde{\tilde{q}}_2, \tilde{\tilde{q}}_3) \geq d(q_2, q_3) \geq \tilde{d}(\tilde{q}_2, \tilde{q}_3)$$

[b]

Figure 2.3: C.A.T comparison theorems (Theorem 2.5.1). Part [a] describes the angle and secant versions and part [b] describes the hinge version. δ and Δ are lower and upper bounds on the sectional curvature of M^n, respectively. we construct comparison triangles in $(S^m_\delta, \tilde{\tilde{d}})$ and $(S^m_\Delta, \tilde{d})$ with vertices $\tilde{\tilde{q}}_1, \tilde{\tilde{q}}_2, \tilde{\tilde{q}}_3$ and $\tilde{q}_1, \tilde{q}_1, \tilde{q}_1$, respectively. Corresponding sides of the triangles have equal lengths. Corresponding to a secant $\overline{x_1x_2}$ we find secants $\overline{\tilde{\tilde{x}}_1\tilde{\tilde{x}}_2}$ and $\overline{\tilde{\tilde{x}}_1\tilde{\tilde{x}}_2}$ in other two triangles. The triangle in S^m_Δ is fatter and the triangle in S^m_δ is thinner than the original one in M, i.e., the angles and secants in M are not smaller than those in S^m_δ and not larger than those in S^m_Δ.

Chapter 3

Existence and Uniqueness of the Riemannian L^p Mean

3.1 Introduction

We start with the following definition:

Definition 3.1.1. Let M be a complete Riemannian manifold of dimension n with Riemannian structure g and corresponding distance function d. Depending on the situation we might write (M, g) or (M, d). Assume $\mathfrak{w}$ is a probability measure on M with support $\text{supp}(\mathfrak{w})$. Define

$$f_p(x) = \begin{cases} \frac{1}{p}\int d^p(x,s)\mathrm{d}\mathfrak{w}(s) & 1 \leq p < \infty \\ \max_{s\in\overline{\text{supp}(\mathfrak{w})}} d(x,s) & p = \infty. \end{cases} \tag{3.1}$$

The Riemannian L^p center of mass or mean with respect to measure $\mathfrak{w}$ is the set of (global) minimizers of f_p on M. Most of the times we assume that $\mathfrak{w}$ is a finite support probability mass function with support $\{x_i\}_{i=1}^N$, where $\mathfrak{w}(x_i) = w_i > 0$; and because of our intended applications we might refer to x_i's as data points. We call vector $\mathbf{w} = [w_1, \ldots, w_N]^T$ a weight vector. Unless otherwise stated, we always assume a weight vector is a vector of positive entries which add up to one. We

rewrite (3.1) as

$$f_p(x) = \begin{cases} \frac{1}{p}\sum_{i=1}^{N} w_i d^p(x, x_i) & 1 \leq p < \infty \\ \max_i d(x, x_i) & p = \infty. \end{cases} \tag{3.2}$$

The L^p Riemannian center of mass or mean for the set $\{x_i\}_{i=1}^N \subset M$ with corresponding weight vector $\mathbf{w}$ is the set of (global) minimizers of f_p in M. We might use alternative terms such as intrinsic L^p mean or Riemannian L^p average, as well. When we refer to the L^p mean of $\{x_i\}_{i=1}^N$, we implicitly assume that the weight vector is known. We denote the L^p mean of $\{x_i\}_{i=1}^N$ by $\bar{x}_p$; and in case there is no ambiguity we might drop the subscript p.

Remark 3.1.1. Although we write $\{x_i\}_{i=1}^N$ as a set we are not ruling out the possibility that some of x_i's might coincide; but obviously we are not interested in and rule out the trivial case where all of them coincide.

This definition in (M, d) is inspired by the standard corresponding definition in the Euclidean space $(\mathbb{R}^n, \|.\|)$ with the standard distance function $\|.\|$. Perfunctorily, one excepts the two to have similar properties, including existence and uniqueness properties, provided one confines the points $\{x_i\}_{i=1}^N$ in "small enough domains" in M. Note that the L^p mean in M is not necessarily a single point. The most immediate and valid questions are the accurate existence and uniqueness conditions for the L^p Riemannian means and their other properties. The main difficulty in answering these questions arises when M is a manifold of positive sectional curvature, where the Riemannian distance function to a point looses its convexity in distances far away. As it will become more evident the notion of convexity plays a central with

regards to the L^p means and their properties. Our goal in this chapter is to prove or more accurately to improve the existence and uniqueness theorems for the L^p means; and in the next chapter we shall investigate some important properties of the means. Before further exposition of details about the results we shall give a historical introduction and perspective about the subject which will help us relate our results to the previous ones. Even before that we introduce some notations and conventions.

3.1.1 On the History of the Riemannian Center of Mass

Here, we give a brief chronological history of the notion of Riemannian center of mass and related existence and uniqueness results. Our account is based on the most widely used and available texts; and it might not be complete. According to many sources (e.g., [9, p. 235]) Cartan was the first to define and use the center of mass in the context of Riemannian geometry. In 1920's, he defined the L^2 center of mass in complete simply connected manifolds of nonpositive curvature (a.k.a. Hadamard manifolds) and proved the existence and uniqueness of the center. He used the L^2 center of mass to prove that any compact subgroup of the isometry group of an Hadamard manifold has a fixed point. A proof of Cartan's theorem using the L^∞ center of mass instead, can be found in [78, p. 225] or [76, p. 164][1]. Aside from an unpublished work by Calabi pointed out in [9, p. 233], the next usage is by Grove and Karcher in [34], where motivated by solving pinching problems, they

[1]The benefit of resorting to the L^∞ mean instead of the L^2 mean is that one does not need to resort to the existence of a bi-invariance probability measure on the isometry group. I thank Prof. P. Petersen for this remark.

define the L^2 center of mass in general Riemannian manifolds but for probability measures with support in small enough balls. As a result, Grove and Karcher give an extension of the result of Cartan to arbitrary manifolds. As one expects and it will become clear later here, positive curvature and finite injectivity radius of M are what might bring about non-uniqueness of the center. In a series of papers [34, 35, 36] Grove, Karcher and Ruh improve i.e., enlarge the domain of existence and uniqueness and consider new applications of the L^2 center of mass. In [36], for compact Lie groups, the domain of uniqueness is enlarged and is shown to be a strongly convex non-ball set. Next, Karcher in [50] gives new proofs, applications and improvements of the previous results.

Two points are worth of elaboration at this time: First, the definition used by Grove and Karcher for the L^2 center of mass is what we call the "local definition", i.e., given a probability measure $\mathfrak{w}$ with support in a small ball $B(o,\rho)$ the local L^2 center of mass is defined as a point in $\overline{B(o,\rho)}$ at which f_2 is minimized. In other words, the function under minimization is $f_2|_{\bar{B}}$, the restriction f_2 to $\overline{B(o,\rho)}$. This definition serves the purposes of finding a unique zero of the gradient vector field of f_2 in B and deriving useful related estimates. However, as pointed out by Groisser in [32] , the center defined in this fashion is susceptible to depend on the candidate ball chosen; and the local definition does not immediately rule out the annoying possibility that a fixed probability measure might have different centers depending on which candidate ball we use. We refer the reader to [32] for more on this and alternative definitions and ways to ameliorate this problem. On the other hand, evidently the global definition does not suffer from this problem and

as soon as its uniqueness is established, the center is independent of the candidate ball[1]. Another possible benefit of resorting to global minimization of f_p is that one might expect the global minimizer to be a "better mean" than a local minimizer. For these reasons the global definition seems to be a more natural one. As it has become standard, the global mean is also referred to as the Frechét mean (e.g., [61]). Second, the detailed techniques used and the Jacobi field estimates derived to show the uniqueness of the local L^2 center of mass in [34, 35, 36] and in [50] are rather different, but the grand strategies used are very similar. Assume $\mathrm{supp}(\mathfrak{w}) \subset B(o, \rho)$. The strategy in [34, 35, 36] is to make sure that ρ is small enough such that the Hessian of $x \mapsto d^2(s, x)$ is positive definite at any critical point x of f_2 in $B(o, \rho)$ and for all $s \in B(o, \rho)$. The strategy in [50] is to choose ρ small enough such that the Hessian of $x \mapsto d^2(x, s)$ is positive definite for all $x, s \in B(o, \rho)$. In other words, both strategies use the fact that in order for the Hessian of f_2 to be positive definite at $x \in B(o, \rho)$, it is sufficient (but not necessary) for the Hessian of $x \mapsto d^2(x, s)$ to be positive definite for all $s \in B(o, \rho)$. One expects that the upper bounds on ρ derived based on the mentioned sufficient condition might not be very sharp and can be improved. Before delving into this issue, we briefly recall the result in [50]. Let $\mathrm{inj}M$ denote the injectivity radius of M and Δ an upper bound on the sectional curvatures of M. Let

$$\rho < r_{\mathrm{cx}} = \frac{1}{2} \min\{\mathrm{inj}M, \frac{\pi}{\sqrt{\Delta}}\}. \tag{3.3}$$

Denote the gradient vector field of f_2 by ∇f_2. Since $B(o, \rho)$ is strongly convex

[1]As we shall see, under a uniqueness condition, the two means coincide; therefore, the mentioned unappealing scenario will not happen.

(see Theorem 2.3.1), $-\nabla f_2$ will be pointing inward the ball on its boundary (see Theorem 3.2 for more details). Therefore, the minimum of $f_2|_{\bar{B}}$ cannot happen on the boundary of the ball and the minimizer must lie inside the ball. In order to guarantee the uniqueness of the minimizer, as explained before, Karcher requires $x \mapsto d^2(x,s)$ to be strictly convex for all $x, s \in B(o,\rho)$. Based on the Jacobi field estimates (see Section 2.4.1), this happens when $d(x,s) < \frac{\pi}{2\sqrt{\Delta}}$, which in turn is guaranteed if $\rho < \frac{\pi}{4\Delta}$. Therefore, the local L^2 center of mass in $B(o,\rho)$ is unique if

$$\rho < \frac{1}{2}\min\{\mathrm{inj}M, \frac{\pi}{2\sqrt{\Delta}}\}. \tag{3.4}$$

Unfortunately, as pointed out by Karcher [50], for the unit sphere in $\mathbb{R}^3$ the above gives $\frac{\pi}{4}$ as the upper bound on the radius of the largest ball containing $\mathrm{supp}(\mathfrak{w})$ to ensure the uniqueness of the center of mass. Nevertheless, it is obvious that for at least two point masses with equal weights lying in a hemisphere (i.e., if $\rho < \frac{\pi}{2}$), the midpoint of the shortest geodesic between them is the unique L^2 center of mass. Roughly speaking, in most occasions, bound (3.4) underestimates the radius of uniqueness by a factor of two. It is shown by W. S. Kendall [52], that $\rho < \frac{\pi}{4\sqrt{\Delta}}$ is not necessary to ensure the uniqueness of the local L^2 center of mass in $B(o,\rho)$. In [52], he initially calls any local minimizer of f_2 on M a "Karcher mean". However, later in the paper Kendall extends his definition to the situation where in the definition of f_2 the distance function d is replaced by its restriction to $B(o,\rho)$, i.e., $d|_B$. Groisser calls this latter mean the "solipsistic Karcher mean" [32]. Kendall shows that given $\mathrm{supp}(\mathfrak{w}) \subset B(o,\rho)$, if $\rho < \min\{\mathrm{inj}M, \frac{\pi}{2\sqrt{\Delta}}\}$, then

there exists a unique solipsistic Karcher mean in $B(o,\rho)$ (see Theorem (7.3) and the remark above it in [52]). A ball whose radius ρ satisfies Kendall's bound is weakly convex but not necessarily strongly convex [18, p.405], i.e., any two points in it can be connected by a unique geodesic which is the unique shortest curve between the two points among all curves in the ball and not entire M; and that is why ρ can be larger than $\frac{1}{2}\text{inj}M$ in Kendall's bound. It is interesting to note that the mentioned extended definition had appeared in [16], although probably it was not noticed by Kendall. The definition of the center of mass used in [18, p. 407] is as such. For our applications this definition seems to add no clear benefit. Strong convexity of the ball is needed to ensure the existence of the Riemannian center of mass. Therefore, as also alluded to in [32], when one uses the global distance in defining f_2 (as we do in (3.1)), Kendall's result yields the bound in (3.3) to ensure both the existence and uniqueness of the local L^2 center of mass in $B(o,\rho)$ for a probability measure whose support is in $B(o,\rho)$. Observe that on the unit sphere in $\mathbb{R}^3$, by (3.3) the local L^2 center of mass is unique if $\rho < \frac{\pi}{2}$. As mentioned long time ago in [34], interest in Riemannian center of mass also exists among probabilists who investigate Brownian motions in manifolds and their relations to harmonic maps. Interestingly, Kendall's work belongs to this area of research. His proof of his uniqueness result requires some preparations and is rooted in the mentioned literature.

The next step in this journey is the result of Buss and Fillmore [17]. Their work is motivated by the use of the Riemannian L^2 center of mass in spherical interpolation for many applications including computer graphics and animation. Buss and Fillmore show that for a probability mass function with finite support on

the unit sphere in $\mathbb{R}^n$, if the support is within a hemisphere, the global L^2 center of mass is unique and belongs to the hemisphere. They use an interesting reflection argument to show that the center must lie inside the ball. They also use an elegant technique of pairing the points in order to show that any stationary point of f_2 inside the hemisphere is a local minimizer; hence, they prove the uniqueness of the L^2 center. Their uniqueness proof can be considered as an alternative proof for Kendall's result on a constant curvature manifold and for finite support probability mass functions. Although, they are not specific about it, Buss and Fillmore show that for such a measure on the unit sphere the "local" and "global" L^2 centers of mass coincide. We mention that it is usually possible to obtain a compromised global existence and uniqueness result from a local existence and uniqueness result. For example, assume we know that $\rho < r_*$ for some $r_* > 0$ guarantees the existence and uniqueness of the local L^p center of mass for a probability measure with support in $B(o, \rho)$ and that the center is the only stationary point of f_p in $B(o, \rho)$. Then just by the triangle inequality we obtain the existence and uniqueness of the global L^p center of mass for a measure with support in $B(o, \rho)$ where $\rho < \frac{1}{2} r_*$. Le in [61] uses a rather similar argument to show that in a general (i.e., nonconstant curvature) manifold $\rho < \frac{1}{2} r_{\text{cx}}$ guarantees existence and uniqueness of the global L^2 mean for a probability measure with support in $B(o, \rho)$. Le's work is motivated by applications in the statistical analysis of shapes.

We should mention that statisticians have dealt with data points on circles and spheres since early 20th century in a context known as "statistical analysis of directional and spherical data" [28, p. xi]. A survey of references such as [65,

28] shows that the framework of Riemannian geometry and especially the related existence-uniqueness issues, however, have not appeared explicitly in the mentioned context. As this body of literature is more concerned with some specific parametric distributions (e.g., von Mises distribution) on the circle and the sphere. Another reason is that usually what is considered as the mean, is the so-called extrinsic mean which arises from the standard embedding of the sphere or any other manifold in a related Euclidean space of higher dimension. In contrast, what we call the Riemannian mean can also be called "intrinsic mean." The extrinsic mean in most cases is the projection of the standard Euclidean mean of the data points computed in the ambient space onto the (sub)manifold; and in the case of the sphere or few other manifolds it is much easier to numerically compute than the Riemannian or intrinsic mean. Until recently with works such as [11] the Riemannian mean had not have gained much attention in the statistics literature.

3.1.2 Contributions and Outline of the Chapter

In this chapter we provide a unified and extensive analysis of the existence and uniqueness of global Riemannian L^p means for $1 \leq p \leq \infty$. The main contributions of this chapter are as follows:

1. In Theorem 3.2.1 (or Theorem 3.2.2[1]) we derive a new bound on the radius

[1]We derive our results in this and other chapters for the case of finite number of points with positive weights which we have introduced and dealt with so far. The reason is that in an applied and computation-oriented setting this is the more realistic framework. However, most of the results in this and the next chapter carry over to the case of arbitrary probability measures in M with no or little extra work. Since in the related literature the existence-uniqueness results are stated for such measures, in Section 3.2.2 we state Theorem 3.2.2 which is the counterpart of Theorem 3.2.1 for arbitrary probability measures.

of the ball containing the support of a probability measure on M to ensure uniqueness of the (global) L^p mean with respect to that measure. In particular, for $p = 2$ our existence and uniqueness result, improves the result by Le in [61] regarding the global L^2 mean. More significantly, our result leads to the discovery that under the best available existence and uniqueness condition for the local L^p mean, the local and global L^p means coincide; therefore, in the sense described before, ours is an uncompromised global existence and uniqueness result. On the route to prove Theorem 3.2.1, we extend (with some modifications) the method of Buss-Fillmore in [17], via comparison arguments, to arbitrary manifolds.

2. As a part the proof of Theorem 3.2.1 (or Theorem 3.2.2) we give an alternative and more geometrical proof for the uniqueness bound by W. S. Kendall for the L^2 mean.

The organization of the chapter is as follows: Entire Section 3.2 is devoted to the proof of the existence and uniqueness theorems for the L^p means. In Section 3.3 we discuss the uniqueness issue when $1 \leq p < 2$ and open problems.

3.2 Existence and Uniqueness of the Riemannian L^p Means

As mentioned before, the existence and uniqueness of properties of the L^p means in manifolds of non-positive curvature especially if they are simply connected, look like their counterparts in Euclidean space in many aspects. However, the case of a manifold of positive curvature is more complicated as can be seen in the example

of the sphere. In fact, the sphere gives us a guideline as what to expect. This fact will be our guide in using comparison theorems to prove the following existence and uniqueness theorem:

Theorem 3.2.1. Let (M, d) be a complete Riemannian manifold with sectional curvature upper bound of Δ and injectivity radius of $\mathrm{inj}M$. Define

$$\rho_{\Delta,p} = \begin{cases} \frac{1}{2}\min\{\mathrm{inj}M, \frac{\pi}{2\sqrt{\Delta}}\} & 1 \leq p < 2 \\ \frac{1}{2}\min\{\mathrm{inj}M, \frac{\pi}{\sqrt{\Delta}}\} (= r_{\mathrm{cx}}) & 2 \leq p \leq \infty. \end{cases} \tag{3.5}$$

Let $\{x_i\}_{i=1}^N \subset \overline{B(o,\rho)} \subset M$ with $\rho < \rho_{\Delta,p}$, then this set of points has a unique L^p mean for $1 < p \leq \infty$, which lies in $B(o,\rho)$. For $p = 1$, unless all the data points lie on a geodesic, again the mean exists, is unique and lies in $\overline{B(o,\rho)}$; and in the degenerate case the mean might not be unique but it lies in the interior of the geodesic. Moreover, for $1 < p < \infty$ the L^p mean is the only stationary point of f_p (see (3.2)) in $\overline{B(o,\rho)}$. In addition, for $1 < p < 2$, f_p is strictly convex in $\overline{B(o,\rho)}$; and for $p = 1$ it is again strictly convex unless all x_i' lie on a geodesic in which case f_1 will be only convex.

The following subsections include the proof of this theorem with some extra descriptions.

3.2.1 Proof of Theorem 3.2.1

The following subsections constitute the proof of Theorem 3.2.1. Before starting the proof notice that we can write the expression for the gradient vector field of

$f_p(1 < p < \infty)$ at $x \in \overline{B(o,\rho)}$ as

$$\nabla f_p(x) = -\sum_{i=1}^{N} w_i d^{p-2}(x, x_i) \exp_x^{-1} x_i. \tag{3.6}$$

Note that for $p = 1$ the above remains valid at $x \neq x_i (1 \leq i \leq N)$.

3.2.1.1 Insideness

We show that a minimum of f_p, if it exists, should happen in $B(o,\rho)$ if $1 < p \leq \infty$ and in $\overline{B(o,\rho)}$ if $p = 1$. First, let us resolve the situation for points on the boundary of the ball. By Corollary 2.3.1 the interior of a minimizing geodesic connecting any two points in $\overline{B(o,\rho)}$ belongs to the interior of the ball; therefore, $\exp_x^{-1} x_i \in T_x M$ belongs to the interior of B_x for any $x \neq x_i$ on the boundary of $B(o,\rho)$, where B_x is the tangent cone of B at x (see Theorem 2.3.2). Hence, $\nabla f_p(x)(1 < p < \infty)$ which is a linear combination of $\exp_x^{-1} x_i$'s also belongs to the interior of B_x and will be pointing inward the ball at any point x on the boundary of the ball; therefore, x cannot be a minimizer of f_p. This also is valid for $p = 1$, unless when $x = x_i$ for some i in which case the gradient of f_1 at x is not defined. Note that at x, since the boundary of the ball is smooth and $\exp_x^{-1} x_i$ lies in the interior of B_x it makes an acute angle with the radial geodesic connecting x to o. Hence, by a first order variation of the arc-length (see Proposition 2.2.1), if we move along this radial geodesic towards the center, we can still reduce $d(x, x_i)$ for $1 \leq i \leq N$ and consequently we can reduce $f_\infty(x)$. Therefore, for $p = \infty$ also a point on the boundary of $B(o,\rho)$ cannot minimize f_∞. Next, we show that for any point outside

$\overline{B(o,\rho)}$ like q there is q' inside the ball such that $f_p(q) > f_p(q')$. In order to show that, of course, we show that $d(q,x_i) > d(q',x_i)$ for $1 \leq i \leq N$. If $d(o,q) \geq 2\rho$, then from the triangle inequality $d(q,x_i) \geq d(o,q) - d(o,x_i) > \rho$, hence $f_p(q) > f_p(o)$ and q cannot be a minimizer. So we only consider a q which is in the annular region $A = \{x \in M | \rho < d(o,x) < 2\rho\}$. Since $d(q,o) < 2\rho < \text{inj}M$, there is a unique minimal geodesic that connects q to o. This geodesic meets the boundary of $B(o,\rho)$ perpendicularly at a unique point, q_c, which lies between q and q_c.Denote the corresponding normal vector pointing inward $B(o,\rho)$ by $\mathbf{n}_{q_c}$ (see Corollary 2.3.2). We choose a point q', the reflection of q inside the ball, at distance $d(q,q_c)$ between q_c and o on the geodesic connecting q to o. A minimizing geodesic emanating from q_c to any of x_i's makes an angle less than $\frac{\pi}{2}$ with $\mathbf{n}_{q_c}$ at q_c (by Theorem 2.3.2). Note that since $d(q_c,x_i) < 2\rho < \text{inj}M$, the mentioned geodesic is unique. Now, we perform two sets of comparisons: one for the case $\Delta \leq 0$ and one for the case $\Delta > 0$.

The Case $\Delta \leq 0$

Our model or comparison space is $(S_0^n \equiv \mathbb{R}^n, \tilde{d})$, where $\tilde{d}$ is the standard Euclidean distance function. Consider the same configuration as the one in M, in the model space: a ball of radius ρ centered at $\tilde{o}$ and a point $\tilde{q}$ outside the ball at distance $d(o,q)$ from the center, i.e., $\tilde{d}(\tilde{o},\tilde{q}) = d(o,q)$. Figure 3.1 shows the configuration in both M and S_Δ^n. In that figure, corresponding equal angles or sides in M and S_Δ^n are marked. From $\tilde{q}'$ choose a geodesic which makes an angle equal to $\angle oq'x_i$ with the geodesic connecting o to q' and choose the point $\tilde{x}_i$ such that $\tilde{d}(\tilde{q}',\tilde{x}_i) = d(q',x_i)$. Then by applying the hinge version of Theorem 2.5.1 with upper bound on curvature

to the triangles $\triangle q'ox_i$ and $\triangle \tilde{q}'\tilde{o}\tilde{x}_i$ we see that $d(\tilde{o},\tilde{x}_i) < d(o,x_i)$, hence $x_i \in B(\tilde{o},\rho)$. Therefore, from Corollary 2.3.2, $\tilde{\alpha}_i = \angle \tilde{q}'\tilde{q}_c\tilde{x}_i < \frac{\pi}{2}$. Now applying the usual Law of Cosines formula to two triangles $\triangle \tilde{q}_c\tilde{q}'\tilde{x}_i$ and $\triangle \tilde{q}\tilde{q}_c\tilde{x}_i$ yields

$$\tilde{d}(\tilde{q},\tilde{x}_i) > \tilde{d}(\tilde{q}',\tilde{x}_i). \tag{3.7}$$

Next, we apply Theorem 2.5.1 to triangles $\triangle qq'x_i$ and $\triangle \tilde{q}\tilde{q}'\tilde{x}_i$. To verify that we can apply the theorem, we note that:

$$d(q,q') = 2d(q_c,q),\ d(x_i,q') \le d(x_i,o) + d(o,q') < 2\rho - d(q_c,q). \tag{3.8}$$

If $d(x_i,q) \ge 2\rho - d(q_c,q)$, then we already have the claim that $d(x_i,q') < d(x_i,q)$, so we assume $d(x_i,q) < 2\rho - d(q_c,q)$. Therefore, the perimeter of $\triangle qq'x_i$ is smaller than $4\rho < 4r_{\text{cx}}$. Note that by construction $\angle x_iq'q = \pi - \angle oq'x_i$ and $\angle \tilde{x}_i\tilde{q}'\tilde{q} = \pi - \angle \tilde{o}\tilde{q}'\tilde{x}_i$; therefore $\angle x_iq'q = \angle \tilde{x}_i\tilde{q}'\tilde{q}$. Applying Theorem 2.5.1 to the two triangles together with (3.7) yield $d(x_i,q) \ge \tilde{d}(\tilde{x}_i,\tilde{q}) > d(x_i,q')$.

The Case $\Delta > 0$

In this case we proceed exactly as the case $\Delta \le 0$, except that we need to use the Spherical Law of Cosines to get (3.7). We elaborate on that. Let $\angle \tilde{q}'\tilde{q}_c\tilde{x}_i = \tilde{\alpha}_i$. Then $\angle \tilde{q}\tilde{q}_c\tilde{x}_i = \pi - \tilde{\alpha}_i$. The Law of Cosines (2.21) in the triangles $\triangle \tilde{q}'\tilde{q}_c\tilde{x}_i$ and $\triangle \tilde{q}\tilde{q}_c\tilde{x}_i$

reads as

$$\begin{aligned}\cos\sqrt{\Delta}\tilde{d}(\tilde{q}',\tilde{x}_i) &= \cos\sqrt{\Delta}\tilde{d}(\tilde{q}',\tilde{q}_c).\cos\sqrt{\Delta}\tilde{d}(\tilde{x}_i,\tilde{q}_c)\\ &+ \sin\sqrt{\Delta}\tilde{d}(\tilde{q}',\tilde{q}_c).\sin\sqrt{\Delta}\tilde{d}(\tilde{x}_i,\tilde{q}_c).\cos\tilde{\alpha}_i \qquad (3.9)\end{aligned}$$

and

$$\begin{aligned}\cos\sqrt{\Delta}\tilde{d}(\tilde{q},\tilde{x}_i) &= \cos\sqrt{\Delta}\tilde{d}(\tilde{q},\tilde{q}_c).\cos\sqrt{\Delta}\tilde{d}(\tilde{x}_i,\tilde{q}_c)\\ &- \sin\sqrt{\Delta}\tilde{d}(\tilde{q},\tilde{q}_c).\sin\sqrt{\Delta}\tilde{d}(\tilde{x}_i,\tilde{q}_c).\cos\tilde{\alpha}_i, \qquad (3.10)\end{aligned}$$

respectively. Noting that $\tilde{d}(\tilde{q},\tilde{q}_c) = \tilde{d}(\tilde{q}',\tilde{q}_c) < \frac{\pi}{\sqrt{\Delta}}$, $\tilde{d}(\tilde{x}_i,\tilde{q}_c) < \frac{\pi}{\sqrt{\Delta}}$, $\tilde{\alpha}_i < \frac{\pi}{2}$ and comparing these two equations we see that (3.7) is valid in this case, too. The rest of the proof is as the previous case.

3.2.1.2 Existence

By the previous part we have $\min_{\overline{B(o,\rho)}} f_p \leq \inf_M f_p$. Considering continuity of f_p, this shows existence of a minimizer of f_p on M. Note that the minimum of f_p on M happens in $B(o,\rho)$ for $1 < p \leq \infty$, while it can happen inside the ball as well as on the boundary when $p = 1$, in which case it must coincide with one of the data points. We observe that for $1 < p < \infty$ the gradient of f_p is continuous, and the L^p mean is a stationary point of ∇f_p in $B(o,\rho)$; while, for $p = 1$ the center is a critical point of f_1 in $\overline{B(o,\rho)}$.

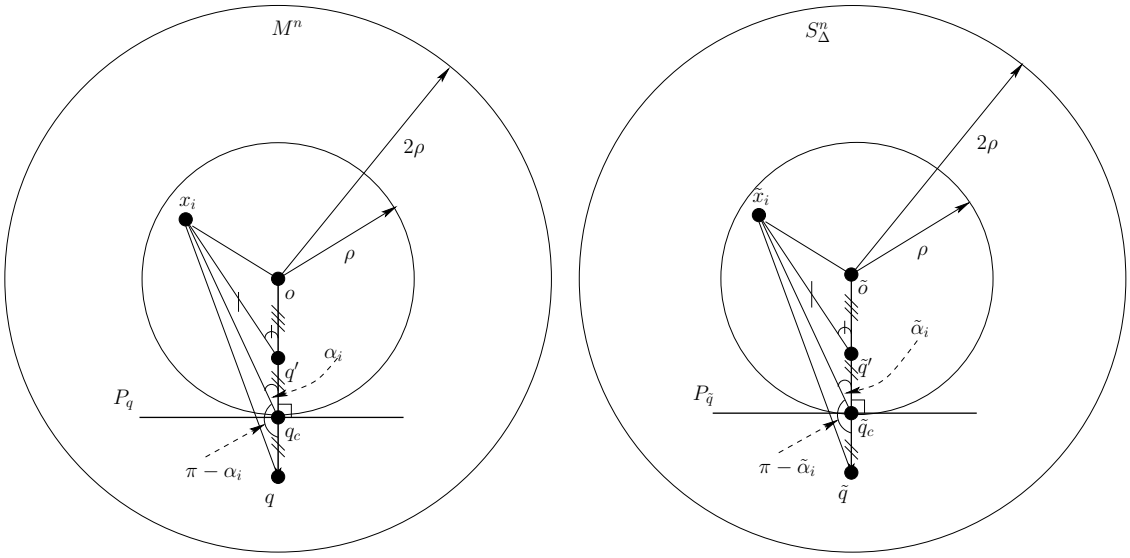


Figure 3.1: We apply comparison theorems to show that for any point q in the annular region $A = \{x \in M | \rho < d(o,x) < 2\rho\}$ there is the point q' inside $B(o,\rho)$ which is the reflection of q across the boundary of $B(o,\rho)$, such that $d(q,x_i) > d(q',x_i)$. This is essentially the result of the fact that $\alpha_i < \frac{\pi}{2}$. The left and right pictures show the configurations in M and the model space S^n_Δ, respectively. Corresponding equal angles or sides in M and S^n_Δ are marked.

3.2.1.3 Uniqueness

We assume $1 \leq p < \infty$ and postpone the cases of $p = \infty$ to the end of the section.

The Case $\Delta \leq 0$

Recall that since $\rho < \rho_{\Delta,p} = \frac{\text{inj}M}{2}$ each $x \in B(o,\rho)$ can be connected by a unique geodesic to all x_i's. For $\Delta \leq 0$ and $1 < p < \infty$, from (2.6) and Corollary 2.4.1, we see that the Hessian of $x \mapsto \frac{1}{p}d^p(x,x_i)$ is positive definite at $x \neq x_i$. From this and since $x \mapsto \frac{1}{p}d^p(x,x_i)$ is continuous at $x = x_i$ and $d(x_i,x_i) = 0 < d(x,x_i)$, we conclude that $x \mapsto \frac{1}{p}d^p(x,x_i)$ (and hence f_p) is a strictly convex function inside $B(o,\rho)$ for $1 < p < \infty$. If $p = 1$ again unless all x_i lie on a geodesic segment again $x \mapsto f_1(x)$ will

be strictly convex, otherwise, it will be only convex. It remains to consider $p = \infty$. Let $q, q' \in B(o, \rho)$ be two distinct minimizers of f_∞. Let $r = f_\infty(q) = f_\infty(q')$. Note that we must have $r < r_{\text{cx}}$. The closed balls $\overline{B(q, r)}$ and $\overline{B(q', r)}$ must contain $\{x_i\}_{i=1}^N$. By the strict convexity of the distance function along non-radial directions, the distance between any point in $\{x_i\}_{i=1}^N$ and the midpoint of the unique minimizing geodesic connecting q and q' is smaller than r. This contradicts the minimality of both q and q'. Hence, the L^∞ mean must be unique. This is all we want.

The Case $\Delta > 0$

Recall Corollary 2.4.1 and Definition 2.6 with positive Δ. Unfortunately, for $\Delta > 0$, the previous rather un-tuned method, which is based on showing the strict convexity of each individual terms in f_p, gives convexity only when $\rho < \min\{\frac{1}{2}\text{inj}M, \frac{\pi}{4\sqrt{\Delta}}\}$. This is because $\cot \sqrt{\Delta} d(x, x_i)$ is positive only when $d(x_i, x) < \frac{\pi}{2\sqrt{\Delta}}$. Therefore, in order to guarantee strict convexity of $x \mapsto \frac{1}{p} d^p(x, x_i)$ (and hence f_p), one is forced to have $\rho < \frac{\pi}{4\sqrt{\Delta}}$. Note that this is enough to prove our claim for $1 \leq p < 2$. For $p = \infty$ also the argument used for $\Delta \leq 0$ works for $\Delta > 0$. Henceforth, we assume $2 \leq p < \infty$ and $\Delta > 0$. The goal is to show that any stationary point of f_p in $B(o, \rho)$ is a local minimizer of f_p. Let $q \in B(o, \rho)$ be a stationary point of f_p i.e., $\nabla f_p(q) = 0$. Note that if q coincides with the center o, then we have $d(q, x_i) < \frac{\pi}{2\sqrt{\Delta}}$ for $1 \leq i \leq N$. In this special case the Hessian of each individual term $x \mapsto \frac{1}{p} d^p(x, x_i)$ at $x = q = o$ will be positive definite; and hence q will be a local minimizer of f_p. Henceforth, we assume that $q \neq o$ or equivalently $d(q, o) > 0$. Let $t \mapsto \gamma(t)$ be a unit speed geodesic leaving $q \in B(o, \rho)$ at $t = 0$. We denote the angle between the geodesic connecting

$x_i \in \overline{B(o,\rho)}$ to q and γ at q by α_i. For convenience let us write $d(x_i, q)$ by $d_i(q)$ or even d_i. By Corollary 2.4.1 we have

$$\frac{\mathrm{d}^2 f_p}{\mathrm{d}t^2}(\gamma(t))|_{t=0} \geq \sum_{i=1}^{N} w_i \bigg((p-1)\cdot d_i^{p-2}(q)\cdot \cos^2\alpha_i + d_i^{p-1}(q)\cdot\sqrt{\Delta}\cdot\cot\sqrt{\Delta}d_i(q)\cdot\sin^2\alpha_i \bigg). \tag{3.11}$$

The summand can be written as

$$(p-2)\cdot d_i^{p-2}\cdot\cos^2\alpha_i + d_i^{p-2}\cdot\big(1-\sin^2\alpha_i(1-\sqrt{\Delta}.d_i\cdot\cot\sqrt{\Delta}d_i)\big). \tag{3.12}$$

Since $\sqrt{\Delta}d_i(q)\cot\sqrt{\Delta}d_i(q) \leq 1$, we have[1]

$$\frac{\mathrm{d}^2 f_p}{\mathrm{d}t^2}(\gamma(t))|_{t=0} \geq \sum_{i=1}^{N} w_i d_i^{p-1}(q)\sqrt{\Delta}\cot\sqrt{\Delta}d_i(q). \tag{3.13}$$

We need to show that the above is positive if $\nabla f_p(q) = 0$. Let γ_{qo} denote the unit speed minimal geodesic connecting q to o. Let β_i denote the angle between the geodesic connecting q to $x_i \in \overline{B(o,\rho)}$ and γ_{qo}. The projection of $\nabla f_p(q)$ along the direction of γ_{qo} at q is zero. Therefore, we have

$$\sum w_i d^{p-1}(q, x_i)\cos\beta_i = 0. \tag{3.14}$$

[1] For $1 \leq p < 2$, we cannot get a relation similar to (3.13), i.e., in (3.12) we cannot get rid of the effect of the angle α_i and reach at a lower bound independent of α_i which only involves d_i and β_i. We will elaborate on the implications of this fact in Section 3.3. It is possible that a finer analysis for $1 \leq p < 2$ will result in a better bound on ρ, possibly depending on p. This can be a topic for further research.

Our goal is to show that this is enough to guarantee that the Hessian of f_p is positive definite at q. We do this by comparison with the spherical case. Consider a ball $B(\tilde{o}, \rho) \subset S^n_\Delta$. Choose a point $\tilde{q}$ inside the ball such that $\tilde{d}(\tilde{o}, \tilde{q}) = d(o, q)$. Then choose the point $\tilde{x}_i$ such that $\tilde{d}(\tilde{q}, \tilde{x}_i) = d(q, x_i)$ and $\angle \tilde{x}_i \tilde{q} \tilde{o} = \angle x_i q o = \beta_i$. Note that by this construction $\tilde{x}_i$ might not belong to $B(\tilde{o}, \rho)$. However, applying Theorem 2.5.1 to the triangles $\triangle \tilde{x}_i \tilde{q} \tilde{o}$ and $\triangle x_i q o$ we observe that $\tilde{d}(\tilde{o}, \tilde{x}_i) \leq d(x_i, o) < \rho$; and therefore, $\tilde{x}_i \in B(\tilde{o}, \rho)$. Also note that since $d(o, x_i), d(o, q) < \rho$ and $d(q, x_i) < 2\rho$, the perimeter of $\triangle x_i q o$ is less than $4r_{\text{cx}}$ and we are allowed to use Theorem 2.5.1. By construction we have

$$\sum_{i=1}^{N} w_i \tilde{d}^{p-1}(\tilde{q}, \tilde{x}_i) \cos \beta_i = 0, \tag{3.15}$$

where as mentioned before $\angle \tilde{o} \tilde{q} \tilde{x}_i = \beta_i$ for $1 \leq i \leq N$.

Note that $B(\tilde{o}, \rho)$ is inside the larger hemisphere $H = B(\tilde{o}, \frac{\pi}{\sqrt{\Delta}})$. The boundary of a hemisphere is a "global" totally geodesic submanifold of S^n_Δ, i.e., a geodesic connecting two points on the boundary belongs to the boundary for all times [9, p.208]. We use this fact as follows. Continue the radial geodesic connecting $\tilde{q}$ to $\tilde{o}$ till it meets the boundary of the hemisphere at $\tilde{z}$. Also continue the geodesic connecting $\tilde{q}$ to $\tilde{x}_i$ to meet the boundary of H at $\tilde{y}_i$ $(1 \leq i \leq N)$. This geodesic meets the boundary of $B(\tilde{o}, \rho)$ at $\tilde{y}'_i$. Figure (3.2) depicts the situation for a single data point $\tilde{x}_i$. For the reason explained before we have $\angle \tilde{q} \tilde{z} \tilde{y}_i = \frac{\pi}{2}$ $(1 \leq i \leq N)$. Then the application of the Spherical Law of Cosines twice in the triangle $\triangle \tilde{q} \tilde{y}_i \tilde{z}$, once for the angle $\angle \tilde{q} \tilde{z} \tilde{y}_i = \frac{\pi}{2}$ and once for the angle $\angle \tilde{y}_i \tilde{q} \tilde{z} = \beta_i$ yields the crucial

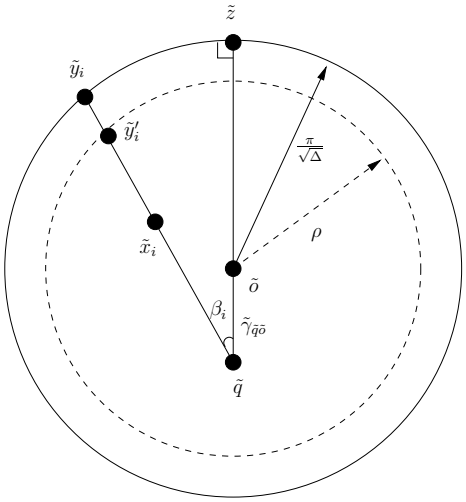


Figure 3.2: $\tilde{o}$ is the center of a hemisphere H of radius $\frac{\pi}{\sqrt{\Delta}}$ in S^n_Δ. Every data point $\tilde{x}_i$ belongs to $B(\tilde{o}, \rho) \subset H$, and $\tilde{q} \in B$ is a critical point of $\nabla \tilde{f}_p$. The minimal geodesic connecting $\tilde{q}$ to $\tilde{x}_i$ hits the boundary of the hemisphere perpendicularly at $\tilde{z}$. The minimal geodesic connecting $\tilde{q}$ to $\tilde{x}_i$ hits the boundaries of B and H at $\tilde{y}'_i$ and $\tilde{y}_i$, respectively.

relation

$$\cot \sqrt{\Delta} \tilde{d}(\tilde{y}_i, \tilde{q}) = \cos \beta_i \cot \sqrt{\Delta} \tilde{d}(\tilde{z}, \tilde{q}), \quad 1 \leq i \leq N. \tag{3.16}$$

This relation enables us to relate β_i to $\tilde{d}(\tilde{x}_i, \tilde{q})$ via a constant (i.e., $\tilde{d}(\tilde{z}, \tilde{q})$ which is independent of i. Since the derivative of $t \mapsto \cot t$ is smaller than -1 on $(0, \pi)$ it is easy to see that

$$\cot t_1 - \cot t_2 \leq -(t_1 - t_2), \quad t_1, t_2 \in (0, \pi). \tag{3.17}$$

Note that

$$\tilde{d}(\tilde{y}_i, \tilde{q}) - \tilde{d}(\tilde{x}_i, \tilde{q}) = \tilde{d}(\tilde{x}_i, \tilde{y}_i). \tag{3.18}$$

Also note that $\tilde{d}(\tilde{x}_i, \tilde{y}_i) \geq \tilde{d}(\tilde{y}'_i, \tilde{y}_i)$ and $\tilde{d}(\tilde{y}'_i, \tilde{y}_i) \geq \frac{\pi}{2\sqrt{\Delta}} - \rho$ (see Figure (3.2)). There-

fore we have

$$\tilde{d}(\tilde{y}_i, \tilde{q}) - \tilde{d}(\tilde{x}_i, \tilde{q}) \geq \frac{\pi}{2\sqrt{\Delta}} - \rho. \tag{3.19}$$

So from (3.17) we have

$$\cot \sqrt{\Delta}\tilde{d}(\tilde{x}_i, \tilde{q}) \geq \cot \sqrt{\Delta}\tilde{d}(\tilde{y}_i, \tilde{q}) + (\frac{\pi}{2} - \rho\sqrt{\Delta}), \quad 1 \leq i \leq N. \tag{3.20}$$

Combining this together with (3.16) we have $\tilde{d}(\tilde{x}_i, \tilde{q}) < \tilde{d}(\tilde{y}_i, \tilde{q}) < \frac{\pi}{\sqrt{\Delta}}$ we have

$$\cot \sqrt{\Delta}\tilde{d}(\tilde{x}_i, \tilde{q}) \geq \cot \sqrt{\Delta}\tilde{d}(\tilde{z}, \tilde{q}) \cos \beta_i + (\frac{\pi}{2} - \rho\sqrt{\Delta}) > \cot \sqrt{\Delta}\tilde{d}(\tilde{z}, \tilde{q}) \cos \beta_i, \ 1 \leq i \leq N. \tag{3.21}$$

Recall that $\rho < \frac{\pi}{2\sqrt{\Delta}}$; hence we have the strict inequality. By our construction, the above implies

$$\cot \sqrt{\Delta} d(q, x_i) > \cot \sqrt{\Delta}\tilde{d}(\tilde{q}, \tilde{z}) \cos \beta_i. \tag{3.22}$$

Plugging the above in (3.13) and using (3.14), we conclude

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} f_p(\tilde{\gamma}(t))|_{t=0} > \cot \sqrt{\Delta}\tilde{d}(\tilde{q}, \tilde{z}) \sum w_i d^{p-1}(q, x_i) \cos \beta_i = 0, \tag{3.23}$$

i.e., the Hessian of f_p is positive definite at q. Since on the boundary of $B(o, \rho)$, $-\nabla f_p$ is pointing inward, the Poincaré-Hopf Index Theorem implies that q is the only zero of ∇f_p in B. Hence, by our insideness argument in Subsection 3.2.1.1, q is the unique minimum of f_p in M. This completes the proof of the theorem.

3.2.2 The Case of an Arbitrary Probability Measure in M

Our existence and uniqueness result in Theorem 3.2.1 concerns with finite number of points in M with corresponding weights. For most data processing applications this is the natural framework. However, it is customary to consider existence and uniqueness of the mean for arbitrary probability measures in M. It is straightforward to extend Theorem 3.2.1 to this more general case. In fact, the proof we gave for Theorem 3.2.1 carries over to this case:

Theorem 3.2.2. Let $\mathfrak{w}$ be a probability measure with support inside the ball $\overline{B(o,\rho)}$ with $\rho < \rho_{\Delta,p}$ where $\rho_{\Delta,p}$ is defined in Theorem 3.2.1. Then for $1 < p \leq \infty$, the L^p center of mass with respect to $\mathfrak{w}$ is a unique point and lies inside B. For $p = 1$, unless the $\mathfrak{w}$-measure of a geodesic segment is 1, again the center exists, is unique and lies in B; and in the degenerate case the center might not be unique. Moreover, for $1 < p < \infty$ the L^p center is in fact the only stationary point of f_p (see (3.1)) in $\overline{B(o,\rho)}$. In addition, for $1 < p < 2$, f_p is strictly convex in $\overline{B(o,\rho)}$; and for $p = 1$ it is again strictly convex unless in the degenerate case mentioned before in which case f_1 will be only convex.

Proof. The proof is essentially the same as that of Theorem 3.2.1. □

3.3 Some Comments and Non-uniqueness of the L^1 Mean in Manifolds with Positive Curvature

As mentioned before, we could not get rid of the effect of the angles α_i in (3.12) for $1 \leq p < 2$; hence, $\rho_{\Delta,p}$ for $1 \leq p < 2$ is smaller than the one for $p \geq 2$. It might

be possible to improve the bound for $1 \leq p < 2$. Nonetheless, a simple example shows that the bound for $p \geq 2$ is not valid for at least values of p equal or close to 1. Note that since $x \mapsto d(q, x)$ in not C^1 at q, every x_i is a critical point of f_1, and for this reason the mean might be a critical point of f_1 which is not a stationary point, i.e., it can be one of the x_i's. In this situation $\bar{x}_1$ will not satisfy $\nabla f_1(\bar{x}_1) = 0$, necessarily. Now, consider the unit upper hemisphere in $\mathbb{R}^3$ and let o denote the north pole. Consider three points x_1, x_2 and x_3 which are located at (arc) distance r from o such that the geodesics connecting o to x_i's make equal angles of $\frac{\pi}{3}$ with each other. Figure 3.3 shows the top view of the upper hemisphere and the triangle $\triangle x_1x_2x_3$. By symmetry, the triangle $\triangle x_1x_2x_3$ is an equilateral triangle whose L^2 (and in fact $L^p, p \geq 2$) center of mass is o for $r < \frac{\pi}{2}$ by Theorem 3.2.1. At the same time, as long as $r < \frac{\pi}{4}$, again by symmetry and the same theorem o is the only L^p mean for $1 \leq p < 2$. Denote the side length of the triangle by a. The Spherical Law of Cosines determines the dependence of a on r as:

$$\cos a = \cos^2 r - \frac{1}{2} \sin^2 r. \tag{3.24}$$

Note that $f_1(o) = 3r$ and $f_1(x_i) = 2a$ for $i = 1, 2, 3$. It is easy to check that there is a number $r_1 (\approx 0.4248\pi)$, such that if $r = r_1$, then $f_1(x_i) = f_1(o)$ for $i = 1, 2, 3$; hence, o can no longer be the unique L^1 mean. A direct numerical evaluation of f_1 shows that the four points o, x_1, x_2 and x_3 minimize f_1. Note that at $x_i (i = 1, 2, 3)$ the gradient of f_1 is not defined; whereas, at o we have $\nabla f_1(o) = 0$. For r little bit larger than r_1, we have $f_1(x_i) < f_1(o)$ $(i = 1, 2, 3)$. The same holds for values of

p close to 1; hence, o cannot be the unique L^p mean for those values of p. As an example, for $r = 0.4250\pi$, x_1, x_2 and x_3 (but not o) are the minimizers of f_1 on the sphere. For this value of r the sum of the internal angles of the triangle $\triangle x_1x_2x_3$ is approximately 408°! Despite this example, it is left open whether the derived bound for $1 \leq p < 2$ can be improved. Nevertheless, from the proof of Theorem 3.2.1, we maintain that although there might be more than one $L^p(1 \leq p < 2)$ center of mass when $\rho < r_{\text{cx}}$, they still must belong to the interior of $B(o, \rho)$.

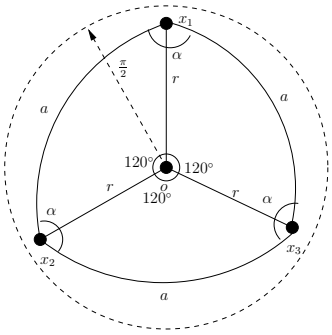


Figure 3.3: Top view of the upper unit hemisphere in $\mathbb{R}^3$. o is the north pole and $\triangle x_1x_2x_3$ is an equilateral triangle whose vertices are symmetrically located at distance r from o. At o the gradient of f_1 is zero (we assume equal weights). However, for values of r close to $\frac{\pi}{2}$ we have $f_1(x_i) < f_1(o)$ and the three vertices themselves constitute their L^1 center of mass.

Chapter 4

Properties of the Riemannian L^p Mean

4.1 Introduction

In this chapter we study some important properties of the Riemannian L^p means. One expects the Riemannian L^p means restricted to small enough strongly convex balls have more or less similar properties as their Euclidean counterparts. In most occasions that is the case; however, there are exceptions, mainly in manifolds of positive curvature.

4.1.1 Contributions and Outline of the Chapter

Two main contributions of this chapter are:

1. In Theorem 4.5.1 we give an affirmative answer to the question asked by Groisser in [32] that whether finding the global L^2 (or more generally $L^p(1 < p < \infty)$) mean of $\{x_i\}_{i=1}^N$ automatically finds a zero of the gradient vector field of f_p which is in the convex hull of $\{x_i\}_{i=1}^N$? In Theorem 4.5.2 we give a refined result for manifolds of constant curvature.

2. In Definition 4.6.1 of Section 4.6 we define some notions which help us bet-

ter understand the sensitivity or averaging properties of the L^2 mean, and in Proposition 4.6.1 and Theorem 4.6.1 discover some averaging and non-averaging properties of the L^2 mean in positively curved manifolds. The results of this section are important, since they invite us to be cautious in averaging data on positively curved manifolds.

The organization of this chapter is as follows: In Section 4.2 we define the minimal ball of $\{x_i\}_{i=1}^N$ (also often ambiguously referred to as the "circumscribed ball"), describe its relation to the L^∞ mean of $\{x_i\}_{i=1}^N$ and study its other properties. The minimal ball is crucial to our developments in Chapter 5. In Section 4.3 we derive smoothness properties of the L^p mean. Continuous dependence of the L^p mean on the data, weights and p are very important facts which we shall use to derive other results. In Section 4.4 we study the important property of isometry compatibility of the L^p means. In Section 4.5 the convexity properties of the L^p mean, by which we mean the relative position of the mean with respect to the convex hull of the data points are studied. As mentioned before, in Section 4.6 sensitivity and averaging properties of the L^2 mean are investigated. In Section 4.7 we consider gradient flow and steepest descent algorithms for finding the L^p mean, and give a numerical example for finding the L^p mean on a sphere. In Appendix 4.8 we have gathered some facts about convex sets and cones in Euclidean space which we need especially in Section 4.5.

4.2 The L^∞ Center of Mass and the Minimal Ball

In Theorem 3.2.1 we gave the existence and uniqueness condition for the L^∞ center of mass of $\{x_i\}_{i=1}^N \subset M$. Since the L^∞ mean or center of mass has significant geometrical meaning which we shall use in the next chapter for our further developments, it is useful to study it further. An important observation is that the L^∞ mean of $\{x_i\}_{i=1}^N$ is the center of the smallest closed ball containing the set.

Definition 4.2.1. Let $\{x_i\}_{i=1}^N \subset M$. We call a closed ball of minimum radius that contains $\{x_i\}_{i=1}^N$ a minimal ball of $\{x_i\}_{i=1}^N$.

Remark 4.2.1. A minimal ball is neither unique nor convex. However, if $\{x_i\}_{i=1}^N$ lies in a strongly convex ball of radius $\rho < r_{\text{cx}}$, then by Theorem 3.2.1 the minimal ball is both strongly convex and unique.

Remark 4.2.2. The minimal ball of a bounded set in an Hadamard manifold or the related Bruhat-Tits space has been studied e.g., in [58]. Also, the minimal ball in manifolds of positive curvature seems to have been studied or used very little. The only reference that we are aware of which studies the minimal ball of a closed set in a manifold of positive curvature is [1] where under the name of "circumscribed ball" it is defined for a particular simply connected manifold with positive upper curvature. Very often, in Euclidean space also the minimal ball is referred to as the "circumscribed ball." However, note that e.g., for $N = 3$ points in the plane the minimal ball of $\{x_i\}_{i=1}^3$ might not coincide with the usual "circumscribed circle" of $\{x_i\}_{i=1}^3$. To avoid any confusions we choose the name "minimal ball".

Theorem 4.2.1. Let $\rho < r_{\mathrm{cx}}$. If $\{x_i\}_{i=1}^N$ lies in the closed ball $\overline{B(o,\rho)}$, then $\overline{B(\bar{x}_\infty, f_\infty(\bar{x}_\infty))}$ where $\bar{x}_\infty \in B(o,\rho)$ is the unique minimal ball of $\{x_i\}_{i=1}^N$ which is strongly convex. Moreover, at least two of x_i's belong to the boundary of the ball $B(\bar{x}_\infty, f_\infty(\bar{x}_\infty))$. If exactly two of them lie on the boundary of the ball then they are at antipodal positions. $\overline{B(\bar{x}_\infty, f_\infty(\bar{x}_\infty))}$ is the smallest closed strongly convex ball containing $\{x_i\}_{i=1}^N$.

Proof. The fact that $\bar{x}_\infty$, the center of the minimal ball, is unique follows from Theorem 3.2.1. By definition the radius of the minimal ball cannot be large than ρ; hence, the ball is strongly convex.

Now by a gradient-like method we establish the properties for the points on the boundary of the minimal ball. Obviously one of x_i's, say x_1, lies on the boundary of $\overline{B(\bar{x}_\infty, f_\infty(\bar{x}_\infty))}$. In addition to x_1 another point, say x_2, should lie on the boundary of the minimal ball. Otherwise, we could find a smaller ball by moving the center toward x_1 and then shrinking the radius, thereby reducing $f_\infty(x) = d(x, x_1)$ in a small neighborhood of $\bar{x}_\infty$. Now assume that there are exactly two points x_1 and x_2 on the boundary of $B(\bar{x}_\infty, f_\infty(\bar{x}_\infty))$. In this case $f_\infty(\bar{x}_\infty) = d(x_1, \bar{x}_\infty) = d(x_2, \bar{x}_\infty)$ and in a small neighborhood of $\bar{x}_\infty$, $f_\infty(x) = \max_{i=1,2} d(x, x_i)$. If $\angle x_1 x_\infty x_2 = \alpha < \pi$, then we could reduce $f_\infty(\bar{x}_\infty)$ further. To see this, let $\gamma : [0,1] \to M$ with $\gamma(0) = \bar{x}_\infty$ be a geodesic leaving $\bar{x}_\infty$ whose velocity vector is such that it makes an acute angle $\frac{1}{2}\alpha$ with the two sides of the hinge $(\angle x_1 \bar{x}_\infty x_2, \overline{\bar{x}_\infty x_1}, \overline{\bar{x}_\infty x_2})$. Since the mentioned angle is acute, from Proposition 2.2.1, we see that for small enough nonzero t, $d(\gamma(t), x_i) < d(\bar{x}_\infty, x_i) = f_\infty(\bar{x}_\infty)$, where $i = 1, 2$. Therefore, by moving $\bar{x}_\infty$

infinitesimally along γ we could reduce the radius of the minimal ball further, which contradicts the minimality assumption. Hence, x_1 and x_2 ought to be at antipodal positions. Notice, that we can do this radius reduction until x_1 and x_2 become antipodal on the ball or a third point lies on the ball. This completes the proof.

□

Figure (4.1) shows the minimal ball of $N = 5$ points. Three of the points are on the boundary and two of them are in antipodal positions. This need not be the case in general, i.e., we can have three points on the boundary and no two of them in antipodal positions. As it can be seen the center or the L^∞ mean is blind to the points in the interior. However, if x_1 moves either inside or outside the ball or if x_4 moves outside the ball, then the center of the minimal ball will change.

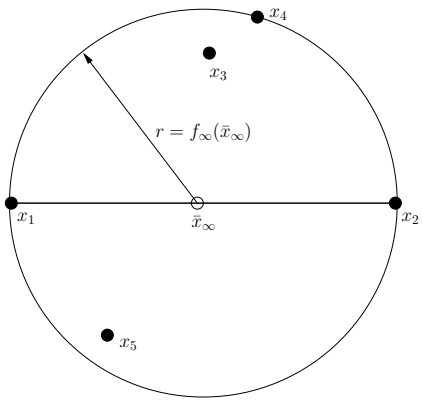


Figure 4.1: The minimal ball for $\{x_i\}_{i=1}^5$. At least two of the points (here three of them) are on the boundary of the ball. The center of the ball, denoted by $\bar{x}_\infty$ is the L^∞ mean of the points. The radius of the ball is $f_\infty(\bar{x}_\infty)$. The center is blind to the data points inside the ball, i.e., small movements of these points have no effect on the center or mean.

4.2.1 Lipschitz Continuity of $f_\infty(x)$

Except for $p = 1$ and $p = \infty$, f_p is at least continuously differentiable everywhere in $\overline{B(o, \rho)}$ with $\rho < r_{\text{cx}}$ (see (2.1)). For $p = 1$ or $p = \infty$, however, f_p is only Lipschitz with Lipschitz constant of $L = 1$. Recall that $f : A \subset M \to \mathbb{R}$ is Lipschitz with constant L if

$$|f(x) - f(y)| \leq L\ d(x, y), \tag{4.1}$$

for all $x, y \in A$. The fact that given the weights w_i, f_1 is Lipschitz with constant $L = 1$ is immediate by the triangle inequality. To see the Lipschitz continuity for $p = \infty$ we need little extra work. Let us define:

$$\check{f}_p = \big(\sum_{i=1}^{N} w_i d^p(x, x_i)\big)^{\frac{1}{p}}, \tag{4.2}$$

where the weights are kept fixed independent of p. Note that $\check{f}_p(x) \to f_\infty(x)$ as $p \to \infty$ for all x; and that given the conditions in Theorem 3.2.1, both f_p and $\check{f}_p$ have $\bar{x}_p$ as their unique minimizers. For our purposes and in this part we can assume that all the functions of interest have the compact ball $A = \overline{B(o, \rho)}$, $(\rho < \rho_{\Delta,p})$, as their fixed domains. Next, we show that $\langle \check{f}_p(x) \rangle_p$ is a uniformly bounded and equicontinuous sequence. Note that $\check{f}_p(x) \leq 2\rho$. The gradient of $\check{f}_p$ at $x \in A$ is

$$\nabla \check{f}_p(x) = -(\sum_i w_i d^p(x, x_i))^{\frac{1}{p}-1} \sum_{i=1}^{N} w_i d^{p-2}(x, x_i) \exp_x^{-1} x_i. \tag{4.3}$$

One can easily check that $\|\nabla \check{f}_p\|_x \leq 1$ for all $x \in A$. Therefore, $\check{f}_p$ is Lipschitz continuous with $L = 1$ for $1 \leq p < \infty$. We use this fact to prove the following proposition for our later use:

Proposition 4.2.1. Let $\{x_i\}_{i=1}^N \subset B(o, \rho)$ with $\rho < \frac{1}{2}\text{inj}M$. Then the sequence $\langle \check{f}_p \rangle_p$ has a subsequence $\langle \check{f}_{p_j} \rangle_{p_j}$, which converges to f_∞ uniformly on $\overline{B(o, \rho)}$ as $p \to \infty$. This implies that f_∞ is Lipschitz on $\overline{B(o, \rho)}$ with Lipschitz constant $L = 1$.

Proof. Since $\check{f}_p$ is Lipschitz with constant $L = 1$ for all $1 < p < \infty$, $\langle \check{f}_p \rangle_p$ is an equicontinuous and uniformly bounded sequence. The Arzela-Ascoli Theorem guarantees the existence of a subsequence $\langle \check{f}_{p_j} \rangle$ which converges to f_∞ uniformly on $\overline{B(o, \rho)}$. Since the convergence is uniform, the limit function f_∞ is Lipschitz with constant $L = 1$, as well. To see this observe that for every $\epsilon > 0$ and large p_j

$$|f_\infty(x) - f_\infty(y)| \leq |f_\infty(x) - \check{f}_{p_j}(x)| + |\check{f}_{p_j}(x) - \check{f}_{p_j}(y)| + |\check{f}_{p_j}(y) - f_\infty(y)| \leq \epsilon + d(x, y). \tag{4.4}$$

□

4.3 Smoothness Properties of the L^p Means

Note that $\bar{x}_p$ depends on the data points $\{x_i\}_{i=1}^N$, on the weights w_i and on the order p. We show that $\bar{x}_p$ depends continuously on these factors; however, higher orders of smoothness do not hold always as we shall elaborate on. Unless otherwise stated, by saying "f is a smooth function", we mean "f has continuous derivatives of all orders." Before proving the next theorem, we recall some smoothness properties of

the Riemannian distance function $x \mapsto d(q, x)$:

Proposition 4.3.1. Let (M, d) be a complete Riemannian manifold of injectivity radius inj$M$. Denote by C_q the cut-locus of q. Then we have:

1. The distance function $x \mapsto d(q, x)$ is C^∞ on $M \backslash \big(\{q\} \cup C_q\big)$. In particular it is C^∞ on $B(q, \text{inj}M) \backslash \{q\}$.

2. At $x = q$, $x \mapsto d^p(x, q)$ has the same smoothness properties as $\tilde{x} \mapsto \|\tilde{x} - \tilde{q}\|^p$ has at $\tilde{x} = \tilde{q}$ where $\tilde{x}, \tilde{q} \in \mathbb{R}^n$ and $\|.\|$ is the standard Euclidean norm in $\mathbb{R}^n$.

Proof. The first part is a result of diffeomorphic properties of the exponential map [78, p. 108]. The second part is the immediate property of the exponential map and the Riemannian distance function.

□

Theorem 4.3.1. Let $\{x_i\}_{i=1}^N$ belong to a ball of radius ρ defined in Theorem 3.2.1. Then

1. For $1 \leq p < \infty$, if p is an even integer, then $\bar{x}_p$ is a smooth function of w_i's and x_i's; otherwise, $\bar{x}_p$ is at least a $C^{[p-1]}$ function of w_i's and x_i's (for $p = 1$ the non-degeneracy is assumed). Here, $[a]$ is the integer part of $a \in \mathbb{R}$. Moreover, if $\bar{x}_p$ does not lie on any x_i's, then $\bar{x}_p$ changes smoothly as w_i's and x_i's change in a small neighborhood.

2. For $p = \infty$, $\bar{x}_p$ is a continuous function of w_i's and x_i's.

Proof. First, we assume $2 \leq p < \infty$. $\bar{x}_p$ is the solution to $F(x; x_1 \ldots, x_N, w_1, \ldots, w_N) = -\nabla f_p = 0$ (see (3.6)). We invoke the Implicit Function Theorem (IFT). The deriva-

tive of F_p with respect to x is nothing but the Hessian of f_p which is positive definite as shown before. However, note that F_p is not (local) C^∞ at $x = x_i$ and the degree of its smoothness at $x = x_i$ depends on p. For $x \notin \{x_i\}_{i=1}^N$, F_p is a local C^∞ function of x. Similar, to the Euclidean case, if p is an even integer, F_p is a local C^∞ function of x. For other values of p, again similar to the Euclidean case, F_p is only a $C^{[p-1]}$ function, because it is so at $x = x_i$'s. Now the result is immediate result of the Implicit Function Theorem. For $1 \leq p < 2$, note that F_p is not differentiable at x_i's. Therefore, we can not invoke the IFT[1]. For $1 \leq p < 2$, we prove continuous dependence of $\bar{x}_p$ on w_i's and x_i's directly. Let $\mathbf{v} = [x_1, \ldots, x_N, w_1, \ldots, w_N]^T$. We take a sequence $\mathbf{v}_k = [x_1^k, \ldots, x_N^k, w_1^k, \ldots, w_N^k]^T$ which converges to $\mathbf{v} = [x_1, \ldots, x_N, w_1, \ldots, w_N]^T$ in $M^N \times [0,1]^N$ and we show that the sequence $\bar{x}_p^k$, i.e., the minimizer of $\check{f}_p^{(k)} = \check{f}_p, (x, \mathbf{v}_k)$ converges to $\bar{x}_p$. Recall the definition of $\check{f}_p$ in (4.2) and that $\langle \check{f}_p^k \rangle_k$ is an equicontinuous and uniformly bounded sequence of functions on $\overline{B(o,\rho)}$. Therefore, by the Arzela-Ascoli theorem we can extract a subsequence $\langle \check{f}_p^{k_j} \rangle_{k_j}$ which converges to $\check{f}_p$ uniformly. Now by the uniform convergence we observe that given any $\epsilon > 0$ for large enough k_j

$$-\epsilon < \check{f}_p(x; \mathbf{v}_{k_j}) - \check{f}_p(x; \mathbf{v}) < \epsilon. \tag{4.5}$$

Therefore, $\min \check{f}_p(x; \mathbf{v}_{k_j}) \to \min \check{f}_p(x; \mathbf{v})$ as $\mathbf{v}_k \to \mathbf{v}$. Using this fact, the continuity

[1]The standard version of the Implicit Function Theorem, requires F_p to be at least C^1 in all variables and its continuity is not enough. In the current situation, i.e., $p < 2$, the maximum we can expect some form local Holder smoothness, which might require stronger versions of the IFT. However, here, we prove continuous dependence by a different approach

of $\check{f}_p(.;.)$ and noting that

$$|\check{f}_p(\bar{x}_p^{k_j};\mathbf{v}) - \check{f}_p(\bar{x}_p;\mathbf{v})| \leq |\check{f}_p(\bar{x}_p^{k_j};\mathbf{v}) - \check{f}_p(\bar{x}_p^{k_j};\mathbf{v}_{k_j})| + |\check{f}_p(\bar{x}_p^{k_j};\mathbf{v}_{k_j}) - \check{f}_p(\bar{x}_p;\mathbf{v})|, \quad (4.6)$$

we have $\lim \check{f}_p(\bar{x}_p^{k_j};\mathbf{v}) = f_p(\bar{x}_p;\mathbf{v})$. Due to compactness of $\overline{B(o,\rho)}$, $\langle \bar{x}_p^{k_j} \rangle$ has a converging subsequence. Denote this subsequence again by $\bar{x}_p^{k_j}$. We have $\check{f}_p(\lim \bar{x}_p^{k_j};\mathbf{v}) = \lim \check{f}_p(\bar{x}_p^{k_j};\mathbf{v}) = f_p(\bar{x}_p;\mathbf{v})$. Now, since $\bar{x}_p$ is the unique minimizer of $\check{f}_p$ we must have $\lim \bar{x}_p^{k_j} = \bar{x}_p$. We maintain that $\lim \bar{x}_p^{k} = \bar{x}_p$. Otherwise, we have an infinite subsequence of $\bar{x}_p^k$, whose all terms will stay away from $\bar{x}_p$, but by the previous argument this subsequence must have a subsequence converging to $\bar{x}_p$ which is a contradiction.

The above argument works for $p = 1$ as long as f_1 has a unique minimizer which is ensured if $\{x_i\}_{i=1}^N$ are not on a geodesic. It also works for $p = \infty$ because of Proposition 4.2.1, which implies the equicontinuity of the sequence $\langle f_\infty(x; x_1^k, \ldots, x_N^k)\rangle_k$. Note that for $p = \infty$ we have no weights. The proof is complete. □

Probably, lack of high order smoothness for $p = \infty$ is rather surprising, since as p increases smoothness of $\bar{x}_p$ in its dependence on x_i's increases. In Figure 4.1, note that if x_4 moves inside the minimal ball, the center will not change, but as it moves outside the ball the center has to change. This can explain the nonsmooth dependence of $\bar{x}_\infty$ on the points on the boundary of the minimal ball. The following simple corollary will be useful in Section 5.2:

Corollary 4.3.1. Let $\{x_i\}_{i=1}^N \subset \overline{B(o,\rho)}$ with $\rho < r_{\text{cx}}$. Then the radius of the minimal ball of the set $\{x_i\}_{i=1}^N$ is a continuous function of the data points x_i's.

The following theorem concerns the dependence of $\bar{x}_p$ on p.

Theorem 4.3.2. Let $1 \leq p_0 \leq \infty$ and $p \to p_0$ (for $p_0 = 1$ we only mean p approaching 1 from the right). Assume $\{x_i\}_{i=1}^N \subset B(o, \rho)$ where $\rho < \min\{\rho_{\Delta,p_0}, \rho_{\Delta,p}\}$ for p close to p_0. For $p_0 = 1$ assume the data points are not on a geodesic. Then $\bar{x}_p \to \bar{x}_{p_0}$.

Proof. The proof can be exactly as the one in the second part of the proof of the previous theorem, except here we take a sequence $\langle \check{f}_p \rangle_p$ as $p \to p_0$ which is again equicontinuous and uniformly bounded (as explained before). □

4.4 Isometry Compatibility

An important property of the L^2 center of mass used in [34] is its isometric compatibly (see also [50]). By this we mean that if $\phi : M \to M$ is an isometry of M, then the center of mass of $\{\phi(x_i)\}_{i=1}^N$ can be found by application of ϕ to the center of mass of $\{x_i\}_{i=1}^N$.

Theorem 4.4.1. Let $\{x_i\}_{i=1}^N \subset \overline{B(o, \rho)}$ with $\rho < \rho_{p,\Delta}$. Assume ϕ is an isometry of M. Then the L^p center of mass of $\{\phi(x_i)\}_{i=1}^N$ is $\phi(\bar{x}_p)$.

Proof. For $1 < p < \infty$, $\bar{x}_p$ is the only zero of $\nabla f_p(x; x_1, \ldots, x_N)$ inside $B(o, \rho)$. Since ϕ is an isometry, from the chain rule it follows that $\phi(\bar{x}_p)$ is a zero of the gradient of $f_p(x; , \phi(x_1), \ldots, \phi(x_N))$ in $B(\phi(o), \rho)$ (see (3.6) and note that $d(x_i, \bar{x}_p) = d(\phi(x_i), \phi(\bar{x}_p))$. Because $\{\phi(x_i)\}_{i=1}^N$ has a unique center in $B(\phi(o), \rho)$, we conclude that $\phi(\bar{x}_p)$ must be this unique center. For $p = 1, \infty$ the result follows from conti-

nuity with respect to p. Note that for $p = 1$ the claim holds even in the degenerate case. $\square$

4.5 Convexity Properties of the L^p Means

The L^p mean of $\{x_i\}_{i=1}^N$ exists uniquely as soon as $\{x_i\}_{i=1}^N$ lies in a strongly convex ball of radius ρ determined in Theorem 3.2.1. Note that the mean is independent of the mentioned ball and the existence of other candidate balls has no effect on the position of the mean. Therefore, the mean belongs to the intersection of all strongly convex balls containing $\{x_i\}_{i=1}^N$. One would like the mean to belong to the convex hull of $\{x_i\}_{i=1}^N$. The proof of this fact is not immediate, since the convex hull of a set is defined as the intersection of all strongly convex "sets" containing that set and not merely the "balls." To appreciate the difference, let M be an open disk in $\mathbb{R}^2$ and $\{x_i\}_{i=1}^3$ vertices of a triangle in M. The convex hull of $\{x_i\}_{i=1}^3$ is the triangle $\triangle x_1x_2x_3$ itself, but we cannot construct the triangle by intersection of any collection of open balls in M containing $\{x_i\}_{i=1}^3$, since these balls are of bounded radius. However, if $M = \mathbb{R}^2$, then we could construct the triangle as an intersection of infinitely many open circles of unbounded radius. Still, we can prove the following theorem using the uniqueness of the stationary points of f_p inside the candidate ball:

Theorem 4.5.1. If the set of points $\{x_i\}_{i=1}^N$ lies in a ball of radius $\rho < \rho_{\Delta,p}$ defined in Theorem 3.2.1, then the L^p mean for $1 \leq p \leq \infty$, in general, belongs to the closure of the convex hull of the points. More precisely,

1. For $1 < p < \infty$, if at least one of the x_i's belongs to the interior of the convex

hull then $\bar{x}_p$ also belongs to the interior of the convex hull; otherwise, it belongs the closure of the convex hull.

2. For $p = 1, \infty$, $\bar{x}_p$ belongs to the closure of the convex hull of $\{x_i\}_{i=1}^N$.

Moreover, for $1 < p \leq \infty$, $\bar{x}_p$ belongs to the interior of the minimal ball of $\{x_i\}_{i=1}^N$ and for $p = 1$ it might belong to the boundary of the ball in addition to its interior. The first part of the previous statement holds for the L^p mean of any K elements of $\{x_i\}_{i=1}^N$; more precisely, the L^p mean of any K elements of $\{x_i\}_{i=1}^N$ where at least two of them are distinct belongs to the interior of the minimal ball of $\{x_i\}_{i=1}^N$.

Proof. Let $B(o, \rho)$ be the candidate ball. We know that ConvHull$(\{x_i\}_{i=1}^N)$, the convex hull of $\{x_i\}_{i=1}^N$, belongs to $B(o, \rho)$. Since we deal with finite number of points we can assume that the closure of the convex hull also belongs to $B(o, \rho)$ (and in fact to its interior). First, let $1 < p < \infty$. We already know that f_p is strictly convex in $B(o, \rho)$ and hence in the closure of ConvHull$(\{x_i\}_{i=1}^N)$ which is a strongly convex subset of $B(o, \rho)$ by Proposition 2.3.1. About the vector field $-\nabla f_p$ on the boundary of ConvHull$(\{x_i\}_{i=1}^N)$ we have:

Lemma 4.5.1. Let $r < \rho_{\Delta,p}$ and $\{x_i\}_{i=1}^N$ belong to a ball of radius r. Then:

1. For $1 < p < \infty$, if at least one of x_i's, say x_1, belongs to the interior of the convex hull of $\{x_i\}_{i=1}^N$, then at every point on the boundary of the convex hull the vector field $-\nabla f_p$ is pointing inward.

2. For $1 < p < \infty$, at every point on the boundary of the minimal ball of $\{x_i\}_{i=1}^N$ the vector field $-\nabla f_p$ is pointing inward.

Proof. Denote the convex hull of $\{x_i\}_{i=1}^N$ by C. The condition on r guarantees that any pair of points of interest can be connected by a unique minimizing geodesic and that the convex hull and the minimal ball exist uniquely. Let x be a point on the boundary of C. From Theorem 2.3.2 we know that C_x, the tangent cone of int$C$ at x, is an open convex cone as a subset of T_xM. Let $\gamma_i : [0, d(x, x_i)] \to M$ be the normal geodesic that connects x to x_i. For γ_i two scenarios might happen:

1. If for some t, $\gamma_i(t) \in \text{int}C$ then, by definition, $\dot{\gamma}_i(0) \in C_x$. In particular, $\dot{\gamma}_1(0) \in C_x$.

2. If for all $t \in [0, d(x, x_i)]$, $\gamma_i(t) \notin \text{int}C$, then since by Proposition 2.3.1 $\bar{C}$ is strongly convex, $\gamma(t) \in \partial C$ for all $t \in [0, d(x, x_i)]$. Let $\langle y_k \rangle_{k=1}^{\infty}$ with $y_k \in \text{int}C$ be a sequence of points converging to x_i and $c_k : [0, d(x, y_k)] \to \bar{C}$ the sequence of geodesics that connect x to y_k's. Note that $\dot{c}_k(0) \in C_x$. The sequence of initial velocities $\langle \dot{c}_k(0) \rangle_k$ is a bounded sequence in T_xM; and it has a convergent subsequence which must converge to $\dot{\gamma}_i(0)$; therefore $\dot{\gamma}_i(0)$ belongs to the boundary of C_x.

Therefore, $\dot{\gamma}_i(0)$'s belong to $\bar{C}_x \subset T_xM$ and at least $\dot{\gamma}_1(0)$ belongs to the interior of the closed convex cone $\bar{C}_x$. Since $-\nabla f_p(x)$ is a linear combination of $\dot{\gamma}_i(0)$'s with positive weights, we conclude that $-\nabla f_p(x)$ belongs to the interior of C_x (see Appendix 4.8). Hence, along the boundary of the convex hull $-\nabla f_p(x)$ points inward the hull. This proves the first claim. The situation for the minimal ball is easier, since its boundary is smooth. In this case for any point x on the boundary of the ball, including those x_i's that lie on the boundary, each $\dot{\gamma}_i(0)$ (defined as before) is

pointing inward (see Corollary 2.3.1); so is $-\nabla f_p(x)$. $\square$

Recall that the interior of $C = \text{ConvHull}(\{x_i\}_{i=1}^N)$ is a smooth convex submanifold of M. Assuming x_1 is in the interior of the convex hull, the vector field $-\nabla f_p$ is pointing inward the convex hull, and the minimum of $f_p|_{\bar{C}}$, the restriction of f_p to the compact set $\bar{C}$ should happen in the interior of the hull. This point should be a stationary point of $f_p|_{\bar{C}}$, i.e., where the projection of $\nabla f_p(\bar{x}_p)$ onto $T_{\bar{x}_p}\text{int}C$ is zero. From Proposition 2.3.1, $\bar{C}$ is strongly convex; therefore, each unique minimizing geodesic connecting $\bar{x}_p$ to x_i's (which might lie on the boundary) belongs to $\bar{C}$. Using this and since the interior of C is totally geodesic by Theorem 2.3.2 each velocity vector $\exp_{\bar{x}_p}^{-1} x_i \in T_{\bar{x}_p}M$ $(1 \leq i \leq N)$ has no component out of $T_{\bar{x}_p}\text{int}C$. This means that at $\bar{x}_p$ actually $\nabla f_p(\bar{x}_p) = 0$ (see (3.6)). Now recall from Theorem 3.2.1 that ∇f_p has $\bar{x}_p$ as its only zero in $B(o, \rho)$. Therefore, $\bar{x}_p$ should belong to the interior of the convex hull of $\{x_i\}_{i=1}^N$. A similar argument shows that, for $1 < p < \infty$, $\bar{x}_p$ belongs to the interior of the minimal of $\{x_i\}_{i=1}^N$. This is independent of whether any of x_i's belong to the interior of the minimal ball, since $-\nabla f_p$ will be always pointing inward the minimal ball on the boundary of the minimal ball by Part 2 of Lemma 4.5.1.

The rest of the proof is comprised of limit arguments to address all the remaining cases:

1. First, for $1 < p < \infty$ and if none of x_i's belong to the interior of $\text{ConvHull}(\{x_i\}_{i=1}^N)$, we use the smoothness results from the previous section (Theorem 4.3.1). Let all x_i's belong to the boundary of the convex hull and let $\langle x_1^k \rangle_k$ be a sequence

of interior points that converges to x_1. Consider new sets of data points $\{x_1^k, x_1, \ldots, x_N\}$ comprised of $N+1$ points and corresponding weight vector $[\frac{1}{2}w_1, \frac{1}{2}w_1, w_2, \ldots, w_N]^T$ in defining the mean. All these sets have the same convex hull as our original set. Denote the L^p mean of each set by $\bar{x}_p^k$. By the preceded argument $\bar{x}_p^k$ belongs to the interior of ConvHull$(\{x_i\}_{i=1}^N)$ for every k. Since the L^p mean (of $N+1$ points) is a continuous function of the points we have $\bar{x}_p^k \to \bar{x}_p$ as $x_1^k \to x_1$. Therefore, $\bar{x}_p$ belongs to the closure of ConvHull$(\{x_i\}_{i=1}^N)$.

2. Second, for $p = \infty$ we have already shown that when $\{x_i\}_{i=1}^N$ (and the weights) are fixed, as $p \to \infty$, $\bar{x}_p \to \bar{x}_\infty$. This implies that $\bar{x}_\infty$ must belong to the closure of the convex hull of $\{x_i\}_{i=1}^N$. The same argument also applies to the case $p = 1$, provided we exclude the degenerate situation. However, for the degenerate situation even though the minimizer of f_1 might not be unique it belongs to the interior of the convex hull. This completes the proof of the theorem.

□

The proof of the theorem suggests that $\bar{x}_1$ or $\bar{x}_\infty$ might lie on the boundary of the convex hull although some of x_i's lie in the interior of the convex hull. One can see this easily in examples in the plane.

Remark 4.5.1. Note that the L^p mean of a subset of $\{x_i\}_{i=1}^N$ belongs to the interior of the minimal ball of $\{x_i\}_{i=1}^N$ according to this theorem, however, it will not belong

to the interior of the convex hull of $\{x_i\}_{i=1}^N$ even in the case of $M \equiv \mathbb{R}^n$. This shows a dramatic difference in inclusion properties of the convex hull and the minimal ball.

Remark 4.5.2. Our proof suggests that in the case we have countable number of points with positive weights also the theorem claims hold.

Remark 4.5.3. In [32], Groisser asks the question whether (global) minimization of f_2 (see (3.1)) automatically finds a zero of the vector field ∇f_2 which is inside the convex hull of the support of the underlying probability measure $\mathfrak{w}$. Theorem 4.5.1 gives an affirmative answer to this question with two technical caveats. The first one is that we assume $\mathfrak{w}$ is a probability mass function. Note that for a more general probability measure the question might become trivial. Second, we showed that the center belongs to the "closure of the convex hull" rather than to the convex hull itself. Resolving this technicality seems to be difficult, but in the next subsection we elaborate on that further.

4.5.1 A Refined Result for Constant Curvature Manifolds

The exact structure of the convex hull of a finite set of points in a manifold of non-constant is not known. Specifically it is not known whether the convex hull is closed or not [9, p. 231]. If it were a closed set, as in the case of Euclidean space, our statement in part 1 of Theorem 4.5.1 would include "convex hull" rather then "closure of the convex hull." Here, we show that in a manifold of constant curvature the L^p mean $(1 < p < \infty)$ of $\{x_i\}_{i=1}^N$ belongs to the interior of the convex hull of the set. The main property that helps us show this is the fact that these

manifolds posses generic totally geodesic submanifolds of lower dimension, which is rare in the case of manifolds of non-constant curvature. We recall the following axiom of plane from [78, p. 136]: Let M be a manifold of constant curvature of dimension n. Let W be any r dimensional subspace of T_qM, $q \in M$. Then the set $S_\rho(W) = \{\exp_q X | X \in W, \|X\| < \rho\}$ is an r-dimensional totally geodesic submanifold of M for every $\rho < \text{inj}M$ (see Definition 2.3.3).

Theorem 4.5.2. Let M be a complete Riemannian manifold of constant curvature Δ. Define $\rho_{\Delta,p}$ as in (3.5) with $1 < p < \infty$ and let $\rho < \rho_{\Delta,p}$. Then the L^p mean of $\{x_i\}_{i=1}^N \subset B(o,\rho)$ belongs to the interior of the convex hull of $\{x_i\}_{i=1}^N$.

Proof. As before denote the convex hull by C. Recall our definitions and arguments in the proof of Theorem 4.5.1 and Lemma 4.5.1. We only need to show that $-\nabla f_p(x)$ is pointing inward the convex hull. As mentioned before, $\exp_x^{-1} x_i$ belongs to $\overline{C}_x$ for $1 \leq i \leq N$. For convenience, let $\eta_i = d^{p-2}(x, x_i) \exp_x^{-1} x_i$. Note that $-\nabla f_p(x) = \sum_i w_i \eta_i$ (see (3.6)). We assume that all η_i's belong to the boundary of the cone C_x; otherwise, similar to Theorem 4.5.1, $-\nabla f_p(x)$ will be pointing inward C. First, assume that the dimension of C is n. Let all η_i's belong to a face F of dimension $k < n$. Consequently, η_i's and hence $\exp_x^{-1} x_i$'s belong to a k-dimensional subspace $W \subset T_pM$, which in turn implies that $\{x_i\}_{i=1}^N$ belongs to a totally geodesic submanifold of dimension smaller than n. This is a contradiction since we assumed the dimension of the convex hull to be n. Now, assume there is no face of dimension less than n which contains all η_i's. If $\sum_i w_i \eta_i$ does not belong to the interior of the hull, it belongs to a face F' of dimension $k' < n$. It follows from the properties of closed convex cones

and their faces that $\eta_i \in F'$ for $1 \leq i \leq N$ which is again a contradiction. Therefore, if the dimension of the convex hull is n, $-\nabla f_p(x)$ must be pointing inward the hull at every point x on the boundary of the hull. Next, if the dimension of C (or C_x) is less than n, $\{x_i\}_{i=1}^N$ belongs to a totally geodesic submanifold of lower dimension and we can apply the same argument to this submanifold. Note that our claim is trivial if the data points lie on a geodesic, i.e, a totally geodesic submanifold of dimension one. Therefore, we conclude that on the boundary of the convex hull of $\{x_i\}_{i=1}^N$, the vector field $-\nabla f_p$ is pointing inward the hull. □

We emphasize that the assumption of $w_i > 0$ in the definition of the mean is crucial for this result to hold. The main obstacle in extending the preceding argument to the non-constant curvature case is that a face of the cone C_x does not map to a k-dimensional $(1 < k < n)$ totally geodesic submanifold via the exponential map, necessarily. Whether constant curvature is necessary to ensure the mean to belong to the interior of the convex hull is not clear.

4.6 Sensitivity Properties of the L^2 Mean

The standard Euclidean mean enjoys an appealing averaging or smoothing property; and in fact that is the main reason for its use in everyday life. We distinguish three forms of averaging properties:

Definition 4.6.1 (Averaging properties)**.** Given $B(o, \rho) \subset M$, let $\mu_{\mathbf{w},B} : B^N = B \times \ldots B \to B$ denote the L^2 mean function on B, i.e., $\mu = \mu_{\mathbf{w}_N,B}(x_1, \ldots, x_N)$ is the L^2 mean of $\{x_i\}_{i=1}^N \subset B$ with corresponding weight vector $\mathbf{w}_N = [w_1, \ldots, w_N]^T$.

Then we say

1. $\mu_{\mathbf{w}_N,B}$ is strongly averaging if for any two sets of points $\{x_i\}_{i=1}^N$ and $\{x_i'\}_{i=1}^N$ in B we have

$$d(\mu_{\mathbf{w}_N,B}(x_1,\ldots,x_N),\mu_{\mathbf{w}_N,B}(x_1',\ldots,x_N')) \leq \sum_i w_i d(x_i,x_i'). \tag{4.7}$$

2. $\mu_{\mathbf{w}_N,B}$ is averaging if for any two sets of points $\{x_i\}_{i=1}^N$ and $\{x_i'\}_{i=1}^N$ in B we have

$$d(\mu_{\mathbf{w}_N,B}(x_1,\ldots,x_N),\mu_{\mathbf{w}_N,B}(x_1',\ldots,x_N')) \leq \max_i d(x_i,x_i'). \tag{4.8}$$

3. $\mu_{\mathbf{w}_N,B}$ is weakly averaging if for any set of points $\{x_i\}_{i=1}^N$ in B replacing a single point x_j by another point x_j' does not change the position of the mean by more than $d(x_j,x_j')$, i.e.,

$$d(\mu_{\mathbf{w}_N,B}(x_1,\ldots,x_N),\mu_{\mathbf{w}_N,B}(x_1,\ldots,x_{j-1},x_j',x_{j+1},\ldots,x_N)) \leq d(x_j,x_j'). \tag{4.9}$$

In occasions we might refer to the L^2 mean without being specific about $B(o,\rho)$, $\mathbf{w}_N$, or N; in which case we assume that $\rho < \rho_{\Delta,2}$, o, $\mathbf{w}$ or N are arbitrary.

Strong averaging implies averaging which in turn implies weak averaging. It is easy to see that the Euclidean L^2 mean in $\mathbb{R}^n$ is strongly averaging. More generally, in Theorem 4.6.1 we show that the Riemannian L^2 mean in a manifold of nonpositive curvature also has the same property. For manifolds of positive curvature the situation is completely different and the L^2 mean is not strongly averaging. The

interpretation is that, the mean can have high sensitivity with respect to the noise.

4.6.1 Lack of (Strong) Averaging Property for L^2 Mean in Manifolds of Nonnegative Curvature

Here, we show why the L^2 mean cannot be strongly averaging if M is of nonnegative sectional curvature [1]. Let $q \in M$ and $\gamma(t) = \exp_q tX$ be a unit speed geodesic emanating from q with velocity vector $X \in T_qM$. Take two points $x_1 = \exp_q t_1X$ and $x_2 = \exp_q t_1X$ on γ with $0 < t_1 < t_2 < \text{inj}M$. The L^2 mean of the set $\{x_1, x_2\}$ with weight vector $\mathbf{w}_2 = [w, 1-w]^T (0 < w < 1)$ is $\bar{x} = \exp_q(wt_1+(1-w)t_2)X$. Now, consider another geodesic $\eta(t) = \exp_q tX'$ emanating from q with velocity vector X' of unit norm. Take perturbed versions of x_1 and x_2 on η, as $x_1' = \exp_q t_1X'$ and $x_2' = \exp_q t_1X'$. Obviously, the L^2 mean of $\{x_1', x_2'\}$ with weight vector $\mathbf{w}_2 = [w, 1-w]^T$ is $\bar{x}' = \exp_q \big(wt_1 + (1-w)t_2\big)X'$. It is a standard matter to show that (e.g. [66])

$$d(\exp_q tX, \exp_q tX') = t\|X - X'\|_q\big(1 - \frac{K_q(X, X')}{12}(1 + \langle X, X'\rangle_q)t^2\big) + o(t), \quad (4.10)$$

where $K_q(X, X')$ is the sectional curvature of M at q along the plane spanned by unit norm tangent vectors $X, X' \in T_qM$ and $\frac{o(t)}{t^3} \to 0$ as $t \to 0$. If $K_q(X, X') > 0$, then $t \mapsto h(t) = d(\exp_q tX, \exp_q tX')$ is a strictly concave function on the interval $(0, \epsilon)$ for small enough ϵ. Therefore, for $0 < t_1 < t_2 < \epsilon$ and $0 < w < 1$ we have $wh(t_1) + (1-w)h(t_2) < h(wt_1 + (1-w)t_2)$. This means that $d(\bar{x}, \bar{x}') > wd(x_1, x_1') + (1-w)d(x_2, x_2')$, which implies that the L^2 mean for two points cannot

[1] Tacitly we assume that the sectional curvature is not identically zero.

be strongly averaging. By choosing repeated points we can conclude that the L^2 mean for more than two points also cannot be strongly averaging. Note that for small enough ϵ, $t \mapsto d(\exp_q tX, \exp_q tX')$ is strictly increasing on the interval $[0, \epsilon]$. Therefore, $d(\bar{x}, \bar{x}') < d(x_2, x_2')$ for small t_2. This suggests that although the L^2 mean is not strongly averaging it might be averaging in small domains. Next, we show that this is not possible if the sectional curvatures of M are bounded below by $\delta > 0$ and above by Δ. First, we consider the constant curvature case and specifically the unit sphere (in $\mathbb{R}^3$). Figure 4.2 explains the situation. Consider two geodesics (great circles) $t \mapsto \gamma(t) = \exp_{q_1} tX$ and $t \mapsto \eta(t) = \exp_{q_1} tX'$ with unit velocities X and X', respectively, that leave q_1 at angle α. The two geodesics meet again at q_2, the antipodal of q_1. As before we consider the function $t \mapsto d(\exp_{q_1} tX, \exp_{q_1} tX')$. The interesting point is that this function achieves a maximum at $\bar{t} = \frac{\pi}{2}$ which corresponds to midpoints $\bar{x} = \exp_{q_1} \bar{t}X$ and $\bar{x}' = \exp_{q_1} \bar{t}X'$. The two geodesics deviate from each other for $0 < t < \bar{t}$ and become closer for $\bar{t} < t \leq \pi$. Now, if we choose $x_1 = \exp_{q_1}(\bar{t} - wl)X$ and $x_2 = \exp_{q_1}(\bar{t} - (1-w)l)X$, then the L^2 mean of $\{x_1, x_2\}$ with weight vector $\mathbf{w}_2 = [w, 1-w]^T$ is $\bar{x}$ for $0 < w < 1$ and all $0 \leq l \leq \bar{t}$. On η we choose $x_1' = \exp_{q_1}(\bar{t} - wl)X'$ and $x_2' = \exp_{q_1}(\bar{t} - (1-w)l)X'$ whose L^2 mean is $\bar{x}'$. Obviously, $d(\bar{x}, \bar{x}') > \max_i d(x_i, x_i')$. This means that the L^2 mean of two points in a ball which contains our data points cannot be averaging. The same argument by choosing repeated points that coincide with x_1 and by varying w implies that the L^2 mean cannot be averaging for more than two points, as well. Note that by choosing α and l small enough we see that the L^2 mean cannot be averaging even in arbitrary small balls on S^2.

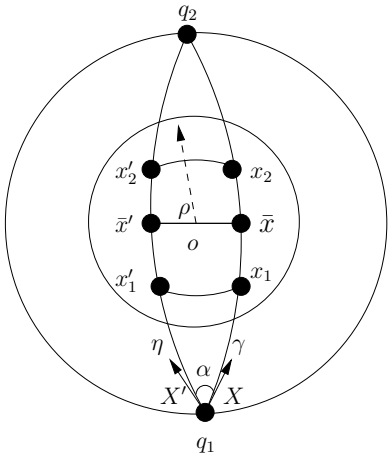


Figure 4.2: Top view for the upper unit hemisphere in $\mathbb{R}^3$. The pairs of points x_1, x_2 and x_1', x_2' lie on two geodesics separated by angle α and connecting the two poles q_1 and q_2, as shown. $\bar{x}$ and $\bar{x}'$ are the midpoints of the two geodesics. x_1 and x_2 (x_1' and x_2') are mirror images with respect to $\bar{x}(\bar{x}')$. For all values of $\alpha > 0$ and no matter how close x_1 and x_2 (and x_1' and x_2') to each other are, we have that $d(\bar{x}, \bar{x}') > d(x_i, x_i'),\ i = 1, 2$. Therefore, the L^2 mean on the sphere cannot be averaging, even in very small balls.

In M, we shall use a more general argument using the properties of Jacobi fields to show that one can always find data points $\{x_1, x_2\}$ and their perturbations $\{x_1', x_2'\}$ for which the L^2 mean is not averaging. Consider $q_1 \in M$ and the family of geodesics $t \mapsto f(t, s) = \exp_{q_1} t(X + sY)$ where $X \in T_{q_1}M$ and $Y \in T_X T_{q_1}M$ [1] . Denote the geodesic $t \mapsto f(t, 0)$ by $\gamma(t)$. In the following we denote all derivatives, either the Levi-Civita covariant derivative of a vector field or the standard derivative of function, with respect to t by $\dot{}$. The covariant derivatives of the related vector fields are taken along $t \mapsto \gamma(t)$. Assume that $\|\dot{\gamma}(0)\| = \|X\| = 1$. The distance between points $f(t, s)$ and $f(t, 0)$ can be written as:

$$d(\exp_{q_1} tX, \exp_{q_1} t(X + sY)) = \|J(t, 0)\|s + o(t, s), \tag{4.11}$$

[1] Here, to avoid confusion and for convenience we shall departure from our original notation on Jacobi fields introduced in Section 2.4 and interchange the role of the t and s variables.

where $t \mapsto J(t,0) = J_0(t) = J(t)$ is the Jacobi field along $t \mapsto \exp_{q_1} tX$ with initial conditions $J(0) = 0$ and $\dot{J}(0) = Y$; and $\lim_{s\to 0^+} \frac{o(t,s)}{s^2} = 0$. We know that there exists an orthonormal basis of parallel vector fields $\{\dot{\gamma}(t), E_1(t), \ldots, E_{n-1}(t)\}$ along γ. Recall, that a vector field is parallel along the geodesic $t \mapsto \gamma(t)$, if its covariant derivative along γ is identically zero. We take the two dimensional section along γ spanned by $\{\dot{\gamma}(t), E_1(t)\}$, and denote the sectional curvature of M along this section by $K(\dot{\gamma}(t), E_1(t)) = K(t)$. Note that $0 < \delta \leq K(t) \leq \Delta$. Consider a Jacobi field $J(t)$ with $J(0) = 0$ along this section, which is orthogonal to $\gamma'(t)$ along $\gamma(t)$:

$$J(t) = g(t)E_1(t), \tag{4.12}$$

where $g(0) = 0$, $\dot{g}(0) \neq 0$ and $g(t) : \mathbb{R} \to \mathbb{R}$ satisfies

$$\ddot{g}(t) + K(t)g(t) = 0. \tag{4.13}$$

This equation is derived from the Jacobi equation (2.3); and $\dot{J}(0) = \dot{g}(0)E_1(0) = Y$. The latter is the only restriction that we impose on Y introduced before. Due to linearity of (4.13) we can assume that $\dot{g}_1(0) > 0$. It is a well-known consequence of Sturm's comparison theorems for ordinary differential equations (e.g. [24, pp.210-211]) that

$$\dot{g}(0)\frac{\sin\sqrt{\Delta}t}{\sqrt{\Delta}} \leq g(t) \leq \dot{g}(0)\frac{\sin\sqrt{\delta}t}{\sqrt{\delta}} \tag{4.14}$$

for $t > 0$. In the above, each inequality is valid as long as its left side remains positive. It is easy to see that at some time $\frac{\pi}{\sqrt{\Delta}} \leq t_c \leq \frac{\pi}{\sqrt{\delta}}$, $g(t_c) = 0$ or $J(t_c) = 0$;

but for $t \in (0, t_c)$ we have $g(t) > 0$. Note that $q_2 = \gamma(t_c)$, is a conjugate point of q_1 along γ, but not necessarily the first conjugate point. Since $g(t) > 0$ on the interval $(0, t_c)$ we have $\|J(t)\| = g(t)$; and since $g(0) = g(t_c) = 0$ there should be a time $0 < \bar{t} < t_c$ at which $\|J(t)\|$ achieves its maximum on the interval $[0, t_c]$. At $\bar{t}$, we have $\dot{g}(\bar{t}) = 0$ and from (4.13) we have $\ddot{g}(t) = -K(t)g(t) < 0$ in the interval $(0, t_c)$ i.e., g is strictly concave in that interval. Therefore, $\bar{t}$ is the only maximizer of $\|J(t)\| = g(t)$ on $[0, t_c]$. The existence of such a $\bar{t}$ is the main ingredient we needed. We proceed very similarly to the sphere case now. Let $\bar{x} = \exp_{q_1} \bar{t}X$ and $\bar{x}' = \exp_{q_1} \bar{t}(X + sY)$ and choose $t_1, t_2 \in (0, t_c) \backslash \{\bar{t}\}$ such that $t_1 < \bar{t} < t_2$. We choose two points $x_1 = \exp_{q_1} t_1 X$ and $x_1 = \exp_{q_1} t_2 X$ on $t \mapsto \exp_{q_1} tX$. Similarly, we choose $x_1' = \exp_{q_1} t_1(X + sY)$ and $x_2' = \exp_{q_1} t_2(X + sY)$ on $t \mapsto \exp_{q_1} t(X + sY)$ and on both sides of $\bar{x}'$. Let $w = \frac{\bar{t} - t_1}{t_2 - t_1}$; we have $0 < w < 1$. Now, $\bar{x}$ is the L^2 mean of $\{x_1, x_2\}$ with the weight vector $\mathbf{w} = [w, 1 - w]^T$ and $\bar{x}'$ is the L^2 mean of $\{x_1', x_2'\}$ with the same weight vector $\mathbf{w}$. Also from (4.11) we conclude that for small enough s which depends on t_1 and t_2

$$d(\exp_{q_1} \bar{t}X, \exp_{q_1} \bar{t}(X + sY)) > d(\exp_{q_1} t_i X, \exp_{q_1} t_i(X + sY)) \tag{4.15}$$

for $i = 1, 2$. Therefore, we have:

$$d(\bar{x}, \bar{x}') > \max_{i=1,2} d(x_i, x_i'). \tag{4.16}$$

Note that this result holds even for t_1 and t_2 very close to each other and for small

s. Therefore, the L^2 mean of two points is not averaging even in small domains. This argument, by using repetitive points and changing the weight vector, can be extended to more than two points.

To summarize we state the following:

Proposition 4.6.1. The L^2 mean is not strongly averaging in a manifold of non-negative sectional curvature (excluding zero curvature). It also is not averaging in a manifold with positive lower sectional curvature bound.

Note that the sectional curvature of a 1-dimensional manifold is zero and the above excludes the case of unit circle in $\mathbb{R}^2$, for example. A few words are due at this stage. First, note that we have not ruled out the possibility of finding a set of points and their perturbations whose L^2 means satisfy the averaging or strong averaging properties: we just showed that (4.7) and (4.8) cannot hold for arbitrary sets of points in a manifold of positive curvature. Second, we emphasized on the lack of averaging property even in small domains, since one might suspect that this non-averaging behavior is a global phenomenon. On the other hand, as we show next, one can achieve weak averaging by restricting the data points to a small enough domain.

4.6.2 Weak Averaging Property of the L^2 Mean in Positively Curved Manifolds: An Example

In manifolds of positive curvature, one can attain weak averaging in small domains. First, we give an example which shows that the L^2 mean can not be weakly averaging

in large domains. Consider the unit sphere S^2 in $\mathbb{R}^3$. Let x_1 and x_2 belong to the upper hemisphere and assume that they are very close to the boundary of the hemisphere in almost antipodal positions as shown in Figure 4.3. In this figure x_1 is on the yz plane and x_2 is in a small neighborhood of the plane and close to the xy plane. The z axis is normal to the page and upward pointing. $\bar{x}$ is the midpoint of geodesic segment connecting x_1 and x_2. Now if x_2 gets reflected to the left of the yz plane to a new point x_2', which is still close to x_2, the midpoint of the new geodesic segment between x_1 and x_2 will be $\bar{x}'$ which will be very far from $\bar{x}$. This means that L^2 mean on the sphere is not weakly averaging in a "large" domain such as a hemisphere. This is a global phenomenon; and stems from the fact that x_1 and x_2 are close to the cut locus of each other. Recall that the lack of strong averaging property on the sphere was due to the local effect of curvature.

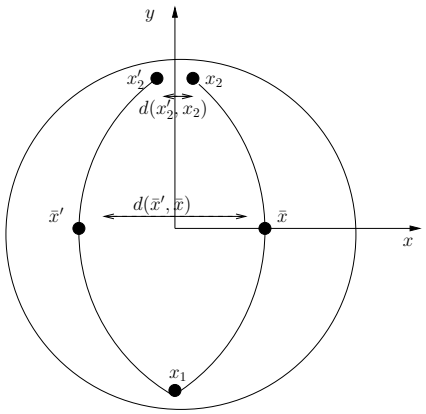


Figure 4.3: Top view of the upper unit hemisphere in $\mathbb{R}^3$. $\bar{x}$ and $\bar{x}'$ are the midpoints of the geodesics connecting x_1 to x_2 and x_1 to x_2', respectively. We have $d(\bar{x}, \bar{x}') \gg d(x_2', x_2)$ while x_2 and x_2' are very close to each other. Therefore, sadly, the L^2 mean is not even weakly averaging in the hemisphere!

We try to quantify this phenomenon. Let $x_1 \in M$ and $X_2, X_2' \in T_{x_1}M$ be two tangent vectors at x_1 such that $x_2 = \exp_{x_1}(X_2)$ and $x_2' = \exp_{x_1}(X_2')$. Assume that

the points belong to a ball $B(o,\rho)$ with $\rho < r_{\text{cx}}$. The midpoints of the geodesic segments joining x_1 and x_2 and joining x_1 and x_2' are $\bar{x} = \exp_{x_1}(\frac{X_2}{2})$ and $\bar{x}' = \exp_{x_1}(\frac{X_2'}{2})$, respectively. If M is of non-negative curvature with curvature upper bound Δ, then using (2.19) twice gives $d(\bar{x},\bar{x}') \leq \frac{1}{2} \cdot \frac{2\rho\sqrt{\Delta}}{\sin 2\rho\sqrt{\Delta}} \cdot d(x_2,x_2')$. Applying the hinge version of the C.A.T comparison theorem with $\delta = 0$, i.e., comparing with a triangle in $\mathbb{R}^m$ shows that $\frac{1}{2}d(x_2,x_2') \leq d(\bar{x},\bar{x}')$; and together with the previous relation we conclude

$$\frac{1}{2}d(x_2,x_2') \leq d(\bar{x},\bar{x}') \leq \frac{1}{2} \cdot \frac{2\rho\sqrt{\Delta}}{\sin 2\rho\sqrt{\Delta}} \cdot d(x_2,x_2'). \tag{4.17}$$

If M is of non-positive curvature with lower curvature bound of $\delta < 0$, applying (2.20) yields $\frac{1}{2} \cdot \frac{2\rho\sqrt{|\delta|}}{\sinh 2\rho\sqrt{|\delta|}} d(x_2,x_2') \leq d(\bar{x},\bar{x}')$. If we apply the hinge version of the C.A.T comparison theorem with curvature upper bound of $\Delta = 0$ and use the previous relation we have

$$\frac{1}{2} \cdot \frac{2\rho\sqrt{|\delta|}}{\sinh 2\rho\sqrt{|\delta|}} \cdot d(x_2,x_2') \leq d(\bar{x},\bar{x}') \leq \frac{1}{2}d(x_2,x_2'). \tag{4.18}$$

Therefore, in a manifold of negative curvature the equal weight L^2 mean (of two points) shows better sensitivity behavior than the L^2 mean in the Euclidean space. However, in a manifold of positive curvature we need to restrict the data to small domains in order to guarantee a weakly averaging behavior of the L^2 mean. For example, if we require

$$\rho < \min\{\frac{\text{inj}M}{2}, \frac{0.3\pi}{\sqrt{\Delta}}\}, \tag{4.19}$$

then $d(\bar{x},\bar{x}') < d(x_2,x_2')$. Note that this bound is more restrictive than $\rho_{\Delta,2} < r_{\text{cx}}$ but better than Karcher's bound (3.4).

4.6.3 A General Bound for Sensitivity

Here, we show that the L^2 mean in a manifold of nonpositive curvature is always strongly averaging, and in a manifold of positive curvature it can be only weakly averaging inside small enough balls as described below. The main result that we use is (4.24) which is essentially Corollary 1.6 in [50]. However, there the underlying constants are not given explicitly, while we would like to have them explicitly. To derive (4.24) we need the following result which is given as Theorem 1.5.1 in [50]: Let $\{x_i\}_{i=1}^N \subset B(o,\rho)$ with $\rho < \frac{1}{2}\min\{\text{inj}M, \frac{\pi}{2\sqrt{\Delta}}\}$. For $\bar{x}_2$ the L^2 mean of $\{x_i\}_{i=1}^N$ (with respect to a weight vector $\mathbf{w}$), and any $x \in B(o,\rho)$ we have

$$\|\nabla f_2(x)\| \geq d(x,\bar{x}_2)\cdot \min\{1, 2\rho\frac{\text{cs}_\Delta 2\rho}{\text{sn}_\Delta 2\rho}\} = d(x,\bar{x}_2)\cdot \begin{cases} 1 & \Delta \leq 0 \\ 2\rho\sqrt{\Delta}\cdot\cot(2\rho\sqrt{\Delta}) & \Delta > 0 \end{cases} \tag{4.20}$$

where sn_Δ and cs_Δ are defined in (2.6). Consider another set of points, i.e., the perturbations of x_i's, as $\{x_i'\}_{i=1}^N \subset B(o,\rho)$. We denote the L^2 mean of $\{x_i'\}_{i=1}^N$ by $\bar{x}_2'$. Since the weights in the definition of f_2 are fixed and only x_i's change, we use $f_2(x;x_1,\ldots,x_N)$ to show the dependence of f_2 on x_i's. Now, for $x = \bar{x}_2'$ in (4.20) we have

$$d(\bar{x},\bar{x}') \leq \|\nabla f_2(\bar{x}';x_1,\ldots,x_N)\|\cdot \begin{cases} 1 & \Delta \leq 0 \\ \frac{1}{2\rho\sqrt{\Delta}\cdot\cot(2\rho\sqrt{\Delta})} & \Delta > 0 \end{cases}. \tag{4.21}$$

Recall the definition of $\nabla f_2(x; x_1, \ldots, x_N)$ and the fact that $\nabla f_2(\bar{x}'; x'_1, \ldots, x'_N) = 0$; so we have

$$\begin{aligned} \nabla f_2(\bar{x}'; x_1, \ldots, x_N) &= \nabla f_2(\bar{x}'; x_1, \ldots, x_N) - \nabla f_2(\bar{x}'; x'_1, \ldots, x'_N) \\ &= \sum_{i=1}^{N} w_i \big(- \exp_{\bar{x}'}^{-1} x_i + \exp_{\bar{x}'}^{-1} x'_i \big). \end{aligned} \tag{4.22}$$

From (2.19) and (2.20) we have

$$\| \exp_{\bar{x}'}^{-1} x_i - \exp_{\bar{x}'}^{-1} x'_i \| \leq d(\bar{x}_i, \bar{x}'_i) \cdot \begin{cases} 1 & \Delta \leq 0 \\ \frac{2\rho\sqrt{\Delta}}{\sin(2\rho\sqrt{\Delta})} & \Delta > 0 \end{cases}. \tag{4.23}$$

Recall that the condition needed to use (2.19) and (2.20) is $\rho < r_{\text{cx}}$ which holds here. Now, (4.21), (4.22) and (4.23) yield:

$$d(\bar{x}, \bar{x}') \leq \left\{ \begin{array}{ll} 1 & \Delta \leq 0 \\ \frac{1}{\cos(2\rho\sqrt{\Delta})} & \Delta > 0 \end{array} \right\} \cdot \sum_{i=1}^{N} w_i d(x_i, x'_i). \tag{4.24}$$

Note that for this bound we need ρ to satisfy (3.4), which is the condition we assumed in (4.20). We state the following theorem:

Theorem 4.6.1. The L^2 mean of N points is strongly averaging in a manifold of nonpositive curvature. In a manifold with upper curvature bound of $\Delta > 0$, the L^2 mean is weakly averaging in a ball of radius ρ if ρ satisfies

$$\rho < \frac{1}{2} \min\{\text{inj}M, \frac{\cos^{-1}(\max_i w_i)}{\sqrt{\Delta}}\}. \tag{4.25}$$

Proof. From (4.24) we see that L^2 mean is strongly averaging in a manifold of nonpositive curvature. By restricting ρ we can achieve weak averaging in $B(o,\rho)$: if $\cos 2\rho\sqrt{\Delta} \geq \max_i w_i$, then $d(\bar{x},\bar{x}') \leq d(x_j,x_j')$, where in the data set $\{x_i\}_{i=1}^N$, x_j is replaced by x_j'. To satisfy the mentioned restriction ρ must obey (4.25). □

Remark 4.6.1. In the case of equal weights (4.25) reads as

$$\rho < \frac{1}{2}\min\{\mathrm{inj}M, \frac{\cos^{-1}(\frac{1}{N})}{\sqrt{\Delta}}\}. \tag{4.26}$$

Note as N increases the upper bound converges to $\frac{1}{2}\min\{\mathrm{inj}M, \frac{\pi}{2\sqrt{\Delta}}\}$, which is Karcher's bound (3.4). The bound (4.26) is more conservative than the bound we found in (4.19) for $N = 2$. The reason is that (4.20) requires $\rho < \frac{\pi}{4\sqrt{\Delta}}$. This is exactly the same restriction that we could surpass by using the method of Buss-Fillmore in Section 3.2. A look at the proof of (4.20) in [50] shows that in order to improve (4.20), i.e., to relate the gradient of $f_2(x)$ to $d(x,\bar{x})$ for larger values of ρ we need to find a better lower bound for $\mathrm{Hess} f_2$, which seems to be more complicated. The example in the previous subsection, shows that in the case $N = 2$ the worst displacement of the mean by displacing only one of the points happens when the two data points are close to the cut points of each other, i.e., where the exponential map is close to becoming singular. In the case of equal weights, as (4.26) suggests, one expects that as N increases the restriction on ρ to become looser. Therefore, we state the following conjecture which we already proved for $N = 2$:

Conjecture 4.6.1. The L^2 mean of N points with equal weights in $B(o,\rho) \subset M$ is weakly averaging if $\rho < \min\{\frac{\mathrm{inj}M}{2}, \frac{0.3\pi}{\sqrt{\Delta}}\}$.

4.6.3.1 A Lipschitz Property

The estimate in (4.24) can be used to establish the Lipschitz continuity of the L^2 mean as a function from $M^N = M \times \ldots \times M^N$ to M. Recall the product structure of M^N and denote the induced distance function by $d_{M^N}(.,.)$. We also use the convention of denoting $(x_1, \ldots, x_N) \in M^N$ by $\mathbf{x} \in M^N$. As before, the L^2 mean of $\{x_i\}_{i=1}^N$ with weight vector $\mathbf{w}$ is denoted by $\mu_{\mathbf{w},B}(\mathbf{x})$. The following proposition follows easily from (4.24):

Proposition 4.6.2. If $\rho < \frac{1}{2}\min\{\text{inj}M, \frac{\pi}{2\sqrt{\Delta}}\}$, then $\mu_{\mathbf{w},B} : M^N \to M$ is a Lipschitz function with Lipschitz constant of $L = \frac{1}{\cos 2\rho\sqrt{\Delta}}(\sum_i w_i^2)^{\frac{1}{2}}$, more precisely we have

$$d(\mu_{\mathbf{w},B}(\mathbf{x}), \mu_{\mathbf{w},B}(\mathbf{y})) \leq L d_{M^N}(\mathbf{x}, \mathbf{y}) \tag{4.27}$$

for any $\mathbf{x}, \mathbf{y} \in M^N$, such that $\{x_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^N$ lie in $B(o, \rho)$. Here, $\Delta \leq 0$ is treated as $\Delta = 0$.

Note that for equal weights and if M is of nonpositive curvature the Lipschitz constant is $L = \frac{1}{\sqrt{N}}$, and if M is of positive curvature the Lipschitz constant is $L > \frac{1}{\sqrt{N}}$; which again is a testimony of higher sensitivity in the case of positively curved manifolds. We shall see some consequences of this fact in our further developments in Section 5.4.

4.7 Computing the L^p Means and an Example for Points on a Sphere

In this section we consider the gradient descent method for finding the L^p means.

4.7.1 Gradient Descent Flow for Finding $\bar{x}_p$ $(1 < p < \infty)$

The following theorem shows that a (continuous-time) gradient flow with initial condition in a uniqueness ball will find $\bar{x}_p$ for $1 < p < \infty$:

Theorem 4.7.1. Let $1 < p < \infty$ and $\{x_i\}_{i=1}^N \subset B(o, \rho)$ where $\rho < \rho_{\Delta,p}$. Then the differential equation

$$\frac{\mathrm{d}z}{\mathrm{d}t}(t) = -\nabla f_p(z(t)),\ z(0) = z_0 \in B(o, \rho), \tag{4.28}$$

where ∇f_p is defined in (3.6), induces a gradient descent flow on M which will take z_0 to $\bar{x}_p$ as $t \to \infty$.

Proof. First, we need to establish existence and possibly uniqueness of a solution of (4.28) on the interval $[0, \infty)$. Since on the boundary of $B(o, \rho)$, $-\nabla f_p$ is pointing inward the ball any solution $t \mapsto z(t)$ cannot leave $B(o, \rho)$. Therefore, since $\rho < \mathrm{inj}M$, $B(o, \rho)$ is a local coordinate neighborhood in which the solution stays, and $\exp_o^{-1} : B(o, \rho) \subset M \to B(0, \rho) \subset T_pM \cong \mathbb{R}^n$ is the corresponding local coordinate function. Let $D \subset \mathbb{R}^n$ be a bounded open set. Recall that an ordinary differential equation of the form $\frac{\mathrm{d}z}{\mathrm{d}t} = f(z)$ with the initial condition $z(0) = z_0 \in D$ and $f : D \to \mathbb{R}^n$:

1. has a local solution, i.e., a solution that solves the equation only on a small interval $[0, \delta)$, if f is continuous [20, p. 178];

2. has a global solution, i.e., a solution which solves it on $[0, +\infty)$, but not

necessarily unique if f is continuous and has linear growth, that is there are constants $a, b > 0$ such that $\|f(x)\| \leq a\|x\| + b$ for all $x \in \mathbb{R}^n$ [20, p. 178].

3. has a global and unique solution if f is locally Lipschitz and any solution starting in the compact set $\bar{D} \subset \mathbb{R}^n$ stays in $\bar{D}$ for entire time [53, p. 94].

This result carries over to a local coordinate neighborhood in M such as $B(o, \rho)$. Now, note that for $2 \leq p < \infty$, ∇f_p is a locally Lipschitz vector field (see Proposition 4.3.1 also) and since $z(t)$ does not leave $\overline{B(o, \rho)}$ we can apply part 3 of the above. Therefore, when $2 \leq p < \infty$ the solution of (4.28) exists and is unique for $t \geq 0$. For $1 < p < 2$ the existence is guaranteed by the part 2 of the above. To see this in the Euclidean case note that, for $x \in B(o, \rho) \subset \mathbb{R}^n$ we have

$$\|x - x_i\|^{p-1} \leq (2\rho)^{p-2}(\|x\| + \|x_i\|); \tag{4.29}$$

and therefore, $\nabla f_p(x)$ has linear growth in $B(o, \rho)$ (see (3.6)). Since $z(t)$ will not leave $B(o, \rho)$, from part 2 we conclude that (4.28) in $\mathbb{R}^n$ has a solution defined on the interval $[0, +\infty)$ for $1 < p < 2$. As explained before this extends to $B(o, \rho) \subset M$, immediately. In summary, for $1 < p < \infty$ the gradient flow in (4.28) has a solution defined on the interval $[0, \infty)$. If $z(t)$ is such a solution, since $\frac{\mathrm{d}}{\mathrm{d}t} f_p(z(t)) < 0$ for $z(t) \neq \bar{x}_p$ and since $z(t) \in B(o, \rho)$, by LaSalle's invariance principle or Lyapunov's Theorem [42, 53] we have $z(t) \to \bar{x}_p$ as $t \to \infty$. This completes the proof. □

We just mention that for $p = 1$ and ∞ one might be able to define generalized gradient flows as discussed in [20]; however, since their numerical implementation

will impose extra difficulty we avoid this approach. In Section 4.3 we saw that $\bar{x}_1$ or $\bar{x}_\infty$ can be approximated by $\bar{x}_p$ where p is very large or p is close to 1, respectively.

4.7.2 Gradient Descent Method for Finding $\bar{x}_p$

Numerical computation of the L^p mean in a manifold can be challenging. One reason is that the explicit evaluation of the exponential map, except for some special manifolds, is complicated. As we mentioned in Theorem 3.2.1 and Section 4.7.1, if $\{x_i\}_{i=1}^N$ lies in $B(o,\rho)$ the continuous-time gradient flow (4.28) for minimization of f_p starting in $B(o,\rho)$ will converge to $\bar{x}_p$. Unfortunately, the continuous-time flow is not implementable on a digital computer; and in a more realistic situation we would either discretize the flow or use a gradient descent algorithm instead. In either case, a new iterate might not belong to $B(o,\rho)$ or even $B(o,\rho_{\Delta,p})$. One can impose an extra constraint to keep the iterates inside the ball as part of a gradient descent algorithm; however, keeping track of the iterates will be demanding. Note that in reality even though we might know apriori that the data points are close enough, we might not know the actual small ball which contains the points. Finding such a ball can be as difficult as finding the mean itself. Another complicated issue with a gradient descent algorithm is finding a good step-size such that the new iterate reduces the cost function by the most. Again, finding such a step-size can be very difficult or impossible in most cases. In Table 4.1 we present an idealistic gradient descent for finding the L^p mean for $1 < p < \infty$. Finding the step-size h_k as prescribed there and keeping the iterates inside a suitable ball make the algorithm idealistic. There are possible step-size selection methods, such as the Armjio step-

size selection, that make the algorithm realistic. However, in practice a method that might be the most efficient is to use a small constant step-size. In the following we give an application of such a gradient descent in finding the L^p means for points on the unit sphere in $\mathbb{R}^3$. Newton-type methods also have been proposed for finding the mean e.g., [17, 32].

Idealistic Gradient Descent Algorithm for finding the $L^p(1 < p < \infty)$ mean

1. `Let` $\{x_i\}_{i=1}^N \subset B(o, \rho)$ `and` $\rho < \rho_{\Delta,p}$. `Choose` $\epsilon > 0$.
2. `Set` $\bar{x}_p^0$ `to an arbitrary point in` $B(o, \rho)$.
3. `Find` $-\nabla f_p(\bar{x}_p^0)$ `from` (3.6)
4. `While` $\|\nabla f_p(\bar{x}^k)\| > \epsilon$
   (a) `Set` $\bar{x}_p^{k+1} = \exp_{\bar{x}_p^k}\big(-h\nabla f_p(\bar{x}_p^k)\big)$ `and find` $h_k = \text{argmin}_h f_p(\bar{x}_p^{k+1})$ `such that` $\bar{x}_p^{k+1} \in B(o, \rho_{\Delta,p})$.
   (b) `Find` $\bar{x}_p^{k+1} = \exp_{\bar{x}_p^k}\big(-h_k\nabla f_p(\bar{x}_p^k)\big)$
   (c) `Find` $-\nabla f_p(\bar{x}_p^{k+1})$ `from` (3.6)
   (d) $k \leftarrow k + 1$
5. `Set` $\bar{x}_p = \bar{x}_p^k$.

Table 4.1: Pseudo-code of a gradient descent algorithm for finding the L^p mean in M. This algorithm is idealistic but it is guaranteed to converge to the mean.

4.7.3 An Example for Points on the Sphere in $\mathbb{R}^3$

In this example, we find the equal weight L^p mean of $N = 4$ points $\{x_i\}_{i=1}^N$ which lie on the upper left part of the unit hemisphere in $\mathbb{R}^3$. The sphere S^2 is endowed with the Riemannian metric induced from the standard metric of $\mathbb{R}^3$. In this metric the distance between $x, y \in S^2 \subset \mathbb{R}^3$ is

$$d(x, y) = \cos^{-1}\langle x, y\rangle, \tag{4.30}$$

where $\langle .,. \rangle$ is the standard inner product in $\mathbb{R}^3$. We implement a gradient descent algorithm with fixed step-size $h_k = 1$. The exponential map corresponding to this metric is calculated as follows [17]. We describe $\exp_q : T_q S^2 \to S^2$ at the north pole $q = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. At other points it is found by a simple rotation with respect to the north pole. Identify $T_q S^2$ with $\mathbb{R}^2$, so $Z = \left[\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix} \right]$ is a tangent vector at q. Then

$$z = \exp_q Z = \begin{bmatrix} Z_1 \frac{\sin r}{r} \\ Z_2 \frac{\sin r}{r} \\ \cos r \end{bmatrix}, \tag{4.31}$$

where $r = \|Z\| = \sqrt{Z_1 + Z_2}$ and $\frac{\sin 0}{0} = 1$. We require $r < \pi$ which is the injectivity radius of S^2. The inverse of $\exp_q$ can be found easily. If $z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \in S^2 \subset \mathbb{R}^3$ is not antipodal to q then

$$Z = \left[\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix} \right] = \exp_q^{-1} z = \frac{r}{\sin r} \left[\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix} \right], \tag{4.32}$$

where $r = d(z, q) = \cos^{-1} z_3$. We denote by θ_i and ϕ_i the angles that $x_i \in \mathbb{R}^3$ makes with the z axis and the angle that the projection of x_i on the xy plane makes with the x axis, respectively. They are the standard polar angles. Then

$$x_i = \begin{bmatrix} \sin \phi_i \cos \theta_i \\ \sin \phi_i \sin \theta_i \\ \cos \phi_i \end{bmatrix}; \tag{4.33}$$

and we choose the following pairs of (ϕ_i, θ_i)

$$\begin{aligned}(\phi_1, \theta_1) &= (0, 0)\\ (\phi_2, \theta_2) &= (\tfrac{2\pi}{10}, 0)\\ (\phi_3, \theta_3) &= (\tfrac{\pi}{5}, \tfrac{\pi}{4})\\ (\phi_4, \theta_4) &= (\tfrac{2\pi}{5}, \tfrac{\pi}{2}).\end{aligned} \tag{4.34}$$

Figure 4.4 shows the four points and their L^p means for $p = 1.1, 2, 3, 10$. We used

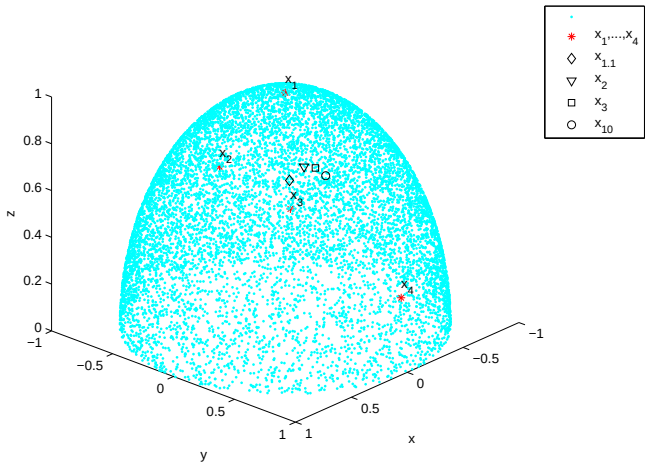


Figure 4.4: An example for equal weight L^p means of $N = 4$ points on the unit hemisphere for $p = 1.1, 2, 3, 10$. The means are found by a gradient descent algorithm. As p increases the mean moves towards the L^∞ center of $\{x_i\}_{i=1}^4$. For this data set three points x_1, x_2 and x_4 lie on the boundary of the minimal ball of $\{x_i\}_{i=1}^4$.

the gradient descent algorithm in Table 4.1 with fixed step-size $h_k = 1$. It is clear as p increases $\bar{x}_p$ moves towards x_4: $d(x_4, \bar{x}_{1.1}) = 0.316\pi$ while $d(x_4, \bar{x}_2) = 0.268\pi$ and $d(x_4, \bar{x}_{10}) = 0.220\pi$. This demonstrates the fact that as p increases, the outliers contribute more in determining the mean. Unfortunately, finding the L^p mean for larger values of p becomes numerically very sensitive, since it amounts to computing

large powers of $d(x, x_i)$ (see (3.6) for the formula of the gradient of f_p) which causes loss of numeric precision very fast. For example, in finding $\bar{x}_{20}$, with two different initial conditions we got two answers that are apart by 3.6°, while for $p = 10$ we got two answers that are apart by only 8.5377×10^{-7} degrees. Nevertheless, by approximating $\bar{x}_{20}$ we see that $d(x_1, \bar{x}_{20}) \approx d(x_2, \bar{x}_{20}) \approx d(x_4, \bar{x}_{20}) \approx 0.212\pi$. Recall that $\bar{x}_p \to \bar{x}_\infty$ as p increases. In fact, we have $d(\bar{x}_{1.1}, \bar{x}_{20}) \approx 17.5^\circ$ while $d(\bar{x}_{10}, \bar{x}_{20}) \approx 4.8^\circ$. For our configuration of points, from Theorem 4.2.1 we conclude that x_1, x_2 and x_4 must lie on the boundary of the minimal ball of $\{x_i\}_{i=1}^4$ and the center of the ball is located at approximate polar coordinates of $(\phi_\infty, \theta_\infty) = (0.220\pi, 0.337\pi)$. Hence, we see that the points belong to a ball of radius smaller than $\frac{\pi}{4}$ which Theorem 3.2.1 requires for uniqueness of $\bar{x}_{1.1}$. In running the algorithm for small p ($1 < p < 5$), we did not observe divergence or iterates leaving the hemisphere. For $p = 10$ and for some initial conditions the iterate leaves the upper hemisphere but it returns to it and finally converges to $\bar{x}_p$. Overall, this is rather remarkable since we have used a fixed step-size of $h_k = 1$. In [17] it is reported that no divergence has been observed for a gradient descent algorithm with fixed step-size of 1 for $p = 2$ on the unit sphere. In [61] it is shown that, in particular, in S^2 and for $p = 2$, if $\rho < \frac{3}{10} r_{\text{cx}}$, then the gradient descent algorithm with fixed step size of $h_k = 1$ converges to $\bar{x}_2$.

4.8 Appendix: Some Facts About Convex Sets in Euclidean Space

A reference for the following material is [82]. Let A be a subset of the vector space $\mathbb{R}^n$ which is endowed with the standard inner product $\langle ., . \rangle$. We say that A is convex if for all $v_1, v_2 \in A$ and $0 \leq \lambda \leq 1$ we have $\lambda v_1 + (1 - \lambda) v_2 \in A$. Note that if the

previous inclusion holds for one $0 < \lambda < 1$ it holds for all $\lambda \in [0, 1]$. We say that the convex set A is of dimension r if the smallest affine subspace of $\mathbb{R}^n$ which contains A is of dimension r. We denote this subspace by $\mathcal{S}_A$. A is closed (open) if it is closed (open) in $\mathcal{S}_A$. By the relative interior (relative boundary) or simply the interior (boundary) of A we mean the interior (boundary) of A as a subset of $\mathcal{S}_A$. If the dimension of A is r we can always embed A in an affine subspace of $\mathbb{R}^n$ of dimension $r < n$. A face of a closed convex set A is a convex subset $F \subset A$ such that for $\forall v_1, v_2 \in C$ and a $0 < \alpha < 1$, $\alpha v_1 + (1 - \alpha)v_2 \in F$ implies $v_1, v_2 \in F$. The only n-dimensional face of an n-dimensional closed set A is A itself. The rest of the faces are of lower dimensions. Any face $F \neq A$ lies on the boundary of A. Any two distinct faces of A have disjoint interiors and A can be decomposed to disjoint unions of the interiors of its faces from dimension $k = 0$ to $k = n$. A hyperplane P_x passing through $x \in A$ is called a supporting hyperplane of A if A lies completely in one side of P_x. Alternatively, this means that a normal vector of P_x, like $\mathbf{x}_x$, exists such that $\langle y - x, \mathbf{n}_x \rangle \geq 0$ for all $y \in A$, or equivalently $y - x$ makes an angle not larger than $\frac{\pi}{2}$ with $\mathbf{n}_x$. Any convex set has a supporting hyperplane at any of its boundary points.

A convex cone $C \subset \mathbb{R}^n$ is a convex set for which $x \in C$ implies $tx \in C$ for all $t \geq 0$. Note that the origin belongs to C. The closure of C is a convex cone again denoted by $\bar{C}$. The sum of a point on the boundary and a point in the interior of C belongs to the interior of C. $F \subset C$ is a face of C if and only if $\sum_{i=1}^{N} w_i v_i \in F$ with $w_i > 0$ and $\sum_{i=1}^{N} w_i = 1$ imply $v_i \in F$ for $1 \leq i \leq N$.

Chapter 5

Recursive-Iterative Approach in Defining Means

5.1 Introduction

The idea of minimizing f_p (see (3.2)) in order to define the center of mass or mean for a set of points on a Riemannian manifold is obviously based on our intuition about the properties of the standard Euclidean center of mass. As we explained in Subsection 4.6, in the case of a manifold of positive curvature, some rather counter-intuitive situations can happen which question our reliance on the properties of the Euclidean mean. Nevertheless, the standard Euclidean or arithmetic mean has also some other properties, which might be used to define other kinds of means. One of them is the property that the Euclidean arithmetic mean of e.g., three points is equal to the arithmetic mean of the pairwise means of the three points; and if we repeat this process, i.e., replace the points by their pairwise means, then the newly generated points will eventually converge to the arithmetic mean of the initial points. One interpretation is that all the intermediate sets of points generated in this process have the same mean, i.e., the mean remains invariant under the pairwise mean replacement. We call this the Mean-Invariance (MI) property of the arithmetic

mean. In this chapter we want to define new means based on similar ideas or their generalizations, i.e., to construct the mean of N points through recursion and iteration based on the mean of $K < N$ points. Here, we consider recursion based definitions as well as pairwise definitions. One can mix the given data points and produce convex combinations via pairwise or multi-point interactions in order to generate the mean of the data points. In most occasions we generate a sequence of points which lie in the closure of the convex hull and in the of minimal ball of the data points, and eventually converge to a single point, i.e., the mean of the data points. The analysis of these iterative schemes is very interesting and we develop powerful tools for that. As explained in Section 4.5, the common property of all the L^p Riemannian means for different values of p is that each of them assigns a point in the closure of convex of hull of a given set of data points. However, the actual form of this assignment depends on p. The recursive-iterative approach also results in similar assignments. The process to define an iterative mean for N points on a manifold M induces a very interesting dynamical system on M^N. In all these dynamics the diagonal of M^N is an invariant set, and we would like the orbit of a point in M^N to converge to the diagonal of M^N. If the data points belong to a small enough strongly convex ball or a suitable strongly convex set this type of convergence happens. However, if the data points are not close enough, a periodic behavior might occur. As an example which shows this periodic behavior, we investigate a Perimeter Shrinkage (PS) scheme whose invariant sets are equidistance points on closed or periodic geodesics in M.

5.1.1 Contributions and Outline of the Chapter

The main contribution of this chapter is introduction of the notions of convex, strictly convex and primitive vectorial mean functions (see Definitions 5.2.2 and 5.2.3), which generalize the notions of stochastic, positive stochastic and primitive stochastic matrices to Riemannian manifolds, respectively. Theorem 5.2.2 is a more general version of the Perron-Frobenius Theorem about the infinite power of a primitive matrix and enables us to prove our existence and uniqueness theorems for a large class of recursive-iterative means. We use the size or radius of the minimal ball of the iterates as a (so-called) Lyapunov function to prove Theorem 5.2.2. Section 5.2 is devoted to the development of the theory of primitive convex mean functions and their properties. After developing all required results we digress little, at the end of the section, and consider the dynamical systems aspect of the iterative applications of a primitive vectorial mean function and relations to infinite products of stochastic matrices in Subsections 5.2.5 and 5.2.6, respectively.

We study two subclasses of recursive-iterative means: the pairwise mean and Mean-Invariance (MI) based means[1]. We introduce the notion of pairwise mean in conjunction with the Primitive Shrinkage (PS) scheme in Section 5.3. As another contribution in Subsections 5.3.2 and 5.3.3 we study the problem of cyclic pursuit on a manifold and identify the equidistance points on closed geodesics as its invariant set (see Theorem 5.3.2 for details). The results in Section 5.3 are not entirely new by themselves. The PS scheme, in a rather different formulation, has been known

[1]The pairwise mean is also based on the notion of mean-invariance; however, we prefer to distinguish the two concepts since we use different tools to analyze them.

to geometers interested in studying the existence of closed geodesics as the Birkhoff Curve Shortening Scheme. In the course of this research some of the properties of the Birkhoff Curve Shortening were re-discovered. Nevertheless, to our knowledge the results of Theorem 5.3.2 are new in the context of cyclic pursuit schemes. In Section 5.4 we introduce the MI means and their generalizations, and investigate some of their properties. The main idea of building the class of means which we call the Mean Invariance (MI) based means (see Section 5.4) was given by Ando, Li and Mathias in [5]. They defined such means on the cone of positive (semi) definite matrices to define geometric for such matrices (see also [77]). Later Lawson and Lim in [59] extended this definition to manifolds of nonpositive curvature and specifically to Hadamard manifolds. However, the tools and techniques used in [59] do not allow for extension to the case of positive curvature manifolds [1]. Despite the generalities of our approach based on primitive mean functions, there is a special case studied in Subsection 5.4.2.2, where we could define the MI means in domains which are not Riemannian balls, necessarily. We do this by an extra assumption and departure from resorting to the radius of the minimal to prove the convergence. Finally, in Section 5.5 we give some numerical examples and provide related discussions.

[1] To be accurate, in order to define mean-invariance based means, Theorem 3.14 in [59] requires the base case (see Definitions 5.4.1, 5.4.2 and 5.4.3) to have a property which the authors call nonexpansiveness. This property is the same as what we called averaging property in Definition 4.6.1; and as we showed in Section 4.6.1 the L^2 mean lacks this property in manifolds of positive curvature (even locally). For this reason and few others, Lawson and Lim's theory cannot be employed to yield results as general as ours in Section 5.4.

5.2 Convex Mean Functions and Their Finite and Infinite Compositions

Our ultimate goal is to build the mean of $N > 2$ data points based on the mean of $N - 1$ or less number of points in a recursive and iterative process. To have a unified and coherent theory it is useful to introduce the notions of convex and strictly convex mean functions and their compositions.

Definition 5.2.1. A convex mean or averaging function of order $N \geq 2$ in M is a function $\mu : M^N \equiv M \times \ldots \times M \to M$ that assigns to an N-tuple $(x_1, \ldots, x_N) \in M^N$ a point $\mu(x_1, \ldots, x_N)$ called the μ-mean of $(x_1, \ldots, x_N)$ such that

1. $\mu(x_1, \ldots, x_N)$ belongs to the closure of the convex hull of $\{x_i\}_{i=1}^N$,

2. there exist a natural number $K \leq N$ and indices $1 \leq i_1 < \ldots < i_K \leq N$ such that $\mu(x_1, \ldots, x_N)$ belongs to the interior of any closed strongly convex ball containing $\{x_i\}_{i=1}^N$, unless when $x_{i_1} = \ldots = x_{i_K} = x$ in which case $\mu(x_1, \ldots, x_N) = x$,

3. $\mu(x_1, \ldots, x_N)$ is continuous in its arguments.

If in the above, item 2 is replaced by

2′. $\mu(x_1, \ldots, x_N)$ belongs to the interior of any closed strongly convex ball containing $\{x_i\}_{i=1}^N$, unless when $x_1 = \ldots = x_N = x$, in which case $\mu(x, \ldots, x) = x$,

then we call μ a strictly convex mean function of order N.

In many occasions we denote an element $(x_1, \ldots, x_N) \in M^N$ by the boldface letter $\mathbf{x}$. Then we write $\mathbf{x} = (x_1, \ldots, x_N)$ and $\mu(\mathbf{x})$ will be interpreted accordingly. Unless otherwise stated, we consider the domain of μ to be of the form

$$D_\mu = \{(x_1, \ldots, x_N) \in M^N | \exists o \in M \text{ and } \rho \leq \rho_\mu, \{x_i\}_{i=1}^N \subset \overline{B(o, \rho)}\}, \tag{5.1}$$

where $\rho_\mu < r_{\text{cx}}$ is a small enough real number such that μ is well defined for all $\mathbf{x} \in D_\mu$. A convex mean function is permutation invariant if for every $\mathbf{x} \in D_\mu$ we have $\mu(x_{\sigma(1)}, \ldots, x_{\sigma(N)}) = \mu(x_1, \ldots, x_N)$, where σ is any permutation of the numbers $1, \ldots, N$.

Figure 5.1 shows the difference between a convex and strictly convex mean function. In the left panel, μ is a convex mean function of order $N = 6$. The set $\{x_i\}_{i=1}^N$ lies in the closed strongly convex ball shown, and the data points x_1, x_2 and x_3 coincide and lie on the boundary of the ball containing, while the rest of the data points lie inside the ball. The mean lies on the boundary of the ball and $\mu(x_1, \ldots, x_6) = x_1 = x_2 = x_3$. In the left panel, μ is a strictly convex mean function, and for the same set of data points, the mean lies inside the ball, even though some of the data coincide and lie on the boundary of the ball.

Example 5.2.1 (L^p Means as Convex Mean Functions)**.** In Theorem 3.2.1, we showed that the L^p mean of $\{x_i\}_{i=1}^N$ (with positive weights) for $1 < p \leq \infty$ belongs to the interior of any closed ball containing $\{x_i\}_{i=1}^N$. Moreover, in Theorem 4.5.1 we showed that the mean also belongs to the closure of the convex hull of $\{x_i\}_{i=1}^N$. Therefore, the L^p mean is a strictly convex mean function of $(x_1, \ldots, x_N)$. In the

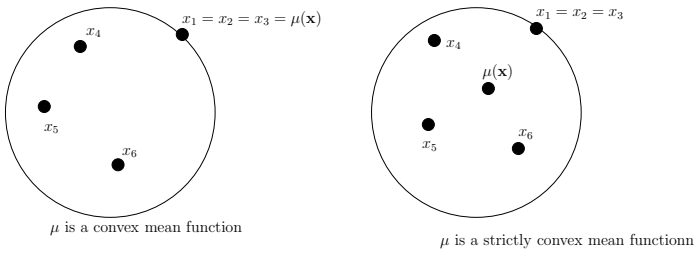


Figure 5.1: Left: A convex mean function. Right: A strictly convex mean function.

case of equal weights it is also permutation invariant. The $L^p (1 < p \leq \infty)$ mean of $\{x_i\}_{i=1}^N$, if some of the points appear with zero weight, is only a convex mean function of $(x_1, \ldots, x_N)$. On the other hand, the L^1 mean of $\{x_i\}_{i=1}^N$ is not a strictly convex mean function of $(x_1, \ldots, x_N)$, since the L^1 mean of $\{x_i\}_{i=1}^N$ does not necessarily belong to the interior of the minimal ball of $\{x_i\}_{i=1}^N$; but, the L^1 mean is a convex mean function of $(x_1, \ldots, x_N)$.

In the definition of a strictly convex mean function, a more natural requirement could have been to require $\mu(x_1, \ldots, x_N)$ to belong to the interior of the convex hull of $\{x_i\}_{i=1}^N$. However, this "more natural requirement" can be restrictive and bring about difficulties for two technical reasons. First, as explained in the previous example and we discussed before in Section 4.5, the $L^p (1 < p \leq \infty)$ mean does not satisfy this more natural requirement or at least we do not know whether it does or not. Second, our ultimate goal is to study infinite compositions of convex mean functions. Since the shape of the convex hull of a finite set of points in a non-constant curvature manifold M is not well understood, it would be difficult to define, in a sound way, the notion of reduction of the size of the convex of hull of the points in order prove convergence results. On the other hand, working with

balls, specifically the minimal ball, is quite straightforward. Also note that the requirement of $\mu(x_1, \ldots, x_N)$ belonging to the interior of the convex hull of $\{x_i\}_{i=1}^N$ is less general than the items 1 and 2 (or 2′) of Definition 5.2.1 combined together. Therefore, in summary, by having the two items 1 and 2 (or 2′) in Definition 5.2.1 we can develop a coherent theory, avoid unnecessary technicalities and consider a large class of means.

The following proposition will give us significant notational convenience:

Proposition 5.2.1. Let μ_1 be a strictly convex mean function of order $N_1 \geq 2$ and let $N_2 > N_1$ be another natural number. Then $\mu_2 : M^{N_2} \to M$, the extension of μ_1 to M^{N_2}, defined as

$$\mu_2(x_1, \ldots, x_{N_1}, \ldots, x_{N_2}) = \mu_1(x_1, \ldots, x_{N_1}) \tag{5.2}$$

is a convex (but not a strictly) mean function of order N_2.

Proof. The claim follows from the observation that the closure of the convex hull of $\{x_i\}_{i=1}^{N_1}$ is a subset of the closure of the convex hull of $\{x_i\}_{i=1}^{N_2}$ and that any closed ball that contains $\{x_i\}_{i=1}^{N_2}$ contains $\{x_i\}_{i=1}^{N_1}$, too. □

5.2.1 Finite Compositions of Strictly Convex Mean Functions

We are interested in compositions of mean functions to construct higher order mean functions from lower order ones. The simplest approach is to consider finite compositions of strictly convex mean function in the sense described in the following theorem (whose proof is obvious):

Theorem 5.2.1. Let $\mu(x_1, \ldots, x_N)$ be a strictly convex mean function. If one or more of the input arguments are replaced by strictly convex functions of $(x_1, \ldots, x_N)$, then the resultant function is still a strictly convex mean function of $(x_1, \ldots, x_N)$.

Next, we consider a more tangible construction:

Example 5.2.2 (Geodesic Averaging)**.** Let $\mu_\alpha(x_1, x_2)$ for $0 < \alpha < 1$ be a second order strictly convex mean function that assigns to x_1 and x_2, a point on the unique minimizing geodesic connecting x_1 to x_2 at distance $\alpha d(x_1, x_2)$ from x_1. Then we define

$$\mu_3(x_1, x_2, x_3) = \mu_{\frac{2}{3}}(\mu_{\frac{1}{2}}(x_1, x_2), x_3). \tag{5.3}$$

It is easy to see that $\mu_3(x_1, x_2, x_3)$ is a third order strongly convex mean. Note that in $\mathbb{R}^n$, $\mu_3(x_1, x_2, x_3) = \frac{1}{3}(x_1 + x_2 + x_3)$ which is the Euclidean L^2 average of x_1, x_2 and x_3 (with equal weights). However, in a general Riemannian manifold, $\mu_3(x_1, x_2, x_3)$ is not necessarily the same as the L^2 mean of x_i's with equal weights. More disturbing is that $\mu_3(.,.,.)$ is not necessarily a permutation-invariant strictly convex mean function either. We can factorize any convex combination $\sum_{i=1}^N w_i x_i$ to a series of nested pairwise convex combinations and translate that to a series of N nested applications of the geodesic averaging $\mu_{\alpha_i}(.,.)$'s ($0 < \alpha_i < 1$ and $1 \leq i \leq N$) defined above. A similar idea has been used in [88] to define a linear combination of $N > 2$ points on a manifold based on pairwise linear or more precisely geodesic combinations without convexity constraints. One can show that μ_N defined as

$$\mu_N(x_1, \ldots, x_N) = \mu_{\alpha_1}\big(x_1, \mu_{\alpha_2}(x_2, \mu_{\alpha_3}(\ldots, \mu_{\alpha_{N-1}}(x_{N-1}, x_N)))\big), \tag{5.4}$$

is a strictly convex mean function of order N.

Besides not being permutation invariant another valid criticism about the mean functions built in this example is that in their construction very little information about the global geometry of M has been used. As a concrete example, let $N = 3$ and $\{x_i\}_{i=1}^N$ belong to the unit sphere S^2 in $\mathbb{R}^3$. Assume that the points are in an open hemisphere. One can imagine another manifold M embedded in $\mathbb{R}^3$ such $\{x_i\}_{i=1}^3$ is in M, the geodesic segments connecting x_2 and x_3 in S^2 and in M coincide; and the geodesic segments connecting $\mu_{\alpha_2}(x_2, x_3)$ in S^2 and M also coincide. In this situation μ_3 will assign, in both manifolds, the same point as the mean. However, the corresponding triangle $\triangle x_1x_2x_3$ in S^2 and M can have very different shapes. In this sense the above mean functions do not involve much "search" or "sweeping" as compared to, e.g., the L^2 Riemannian mean which seeks the mean globally. On the other hand the nested pairwise mean function in (5.4) is certainly much easier to calculate. In the sequel, we extend the pairwise mean calculations in two directions by infinite iterative processes. In one direction, we construct a mean which in not permutation invariant for $K > 3$ points, however, it includes more search or sweeping of the manifold. In other direction, again via an infinite iterative process, we define a large class of permutation invariant means of higher orders based on means of lower orders. For that we need to introduce the idea of infinite combinations of strictly convex mean functions.

5.2.2 Infinite Compositions of Convex Vectorial Mean Functions

Now, we introduce the notions of convex and strictly convex vectorial mean functions from M^N to M^N, which can be considered as generalizations of the notions of stochastic and positive stochastic matrices, respectively.

Definition 5.2.2. A (strictly) convex vectorial mean function of order N in M is a function $\boldsymbol{\mu} : M^N \to M^N$ defined as

$$\boldsymbol{\mu}(\mathbf{x}) = (\mu_1(\mathbf{x}), \ldots, \mu_N(\mathbf{x})) \in M^N, \tag{5.5}$$

where each $\mu_i : M^N \to M$ is a (strictly) convex mean function of order N. We call μ_i the ith component of $\boldsymbol{\mu}$. The composition of two or more convex vectorial mean functions is defined the same way as the standard composition of functions on M^N. Let $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ be two convex vectorial mean functions of order N in M by $\boldsymbol{\mu}_3 = \boldsymbol{\mu}_1 \circ \boldsymbol{\mu}_2$ we mean

$$\boldsymbol{\mu}_3(\mathrm{x}) = \boldsymbol{\mu}_2(\boldsymbol{\mu}_1(\mathrm{x})) \tag{5.6}$$

for all $x \in M^N$. We denote the k^{th} power of $\boldsymbol{\mu}$ by $\boldsymbol{\mu}^k$.

The following important proposition is an immediate consequence of the above definition:

Proposition 5.2.2. Let $\boldsymbol{\mu}$ be a convex vectorial mean function of order N in M. If $\mathbf{y} = \boldsymbol{\mu}(\mathbf{x})$, then the radius of the minimal ball of $\{y_i\}_{i=1}^N$ is not larger than the radius of the minimal ball of $\{x_i\}_{i=1}^N$. Moreover, if $\boldsymbol{\mu}$ is strictly convex, then the

radius of the minimal ball of $\{y_i\}_{i=1}^N$ is less than the radius of the minimal ball of $\{x_i\}_{i=1}^N$, unless $x_1 = \ldots = x_N$, where both balls have radius of zero.

Therefore, applying a strictly convex vectorial mean function $\boldsymbol{\mu}$ to a set of N data points will generate new N points whose minimal ball radius is smaller than that of the input points, unless all the points coincide. However, note that the minimal ball of the output points is not necessarily inside the the minimal of the input points.

Example 5.2.3. Let μ_α be the geodesic averaging function with parameter α as in Example 5.2.2. Assume $\{x_i\}_{i=1}^3$ lies in a strongly convex ball. The mean function $\hat{\mu}_\alpha^{12}(x_1, x_2, x_3) = \mu_\alpha(x_1, x_2)$ is only a convex mean function of order 3. Now define $\boldsymbol{\mu}$ as

$$\boldsymbol{\mu}(\mathbf{x}) = (\mu_{\alpha_1}(x_1, x_2), \mu_{\alpha_2}(x_2, x_3), \mu_{\alpha_3}(x_3, x_1)), \tag{5.7}$$

where $0 < \alpha_i < 1$ for $i = 1, 2, 3$. Again $\boldsymbol{\mu}$ is a convex but not a strictly convex vectorial mean function of order 3. This is because if $x_1 = x_2$ and $\mathbf{y} = \boldsymbol{\mu}(\mathbf{x})$, then $y_1 = x_1$ and y_1 belongs to the boundary of the minimal ball of $\{x_i\}_{i=1}^3$. However, note that $\boldsymbol{\mu}^2$, the second power of $\boldsymbol{\mu}$, will be strictly convex.

This example leads us to the following definition (borrowed from the theory of positive matrices), which introduces a notion similar to the primitivity for stochastic matrices [43, p. 519] but for convex vectorial mean functions:

Definition 5.2.3. A convex vectorial mean function $\boldsymbol{\mu}$ of order N is called primitive, if there exists a natural number n_s such that $\boldsymbol{\mu}^{n_s}$ is strictly convex. We call the smallest value for n_s, the index of primitivity of $\boldsymbol{\mu}$.

Note that any finite composition of strictly convex vectorial mean functions of order N is a strictly convex vectorial mean function by Theorem 5.2.1. The following important theorem shows that the infinite power of a primitive vectorial mean function also is (or converges to) a strictly convex vectorial mean function, which has a special feature that all its components are the same function.

Theorem 5.2.2. Let $\boldsymbol{\mu}$ be a primitive vectorial mean function of order N with primitivity index $n_s \in \mathbb{N}$ (i.e., $\boldsymbol{\mu}^{n_s}$ is a strictly convex vectorial mean function). The sequence of convex vectorial mean functions $\langle \tilde{\boldsymbol{\mu}}_k \rangle_{k=1}^{\infty}$ where

$$\tilde{\boldsymbol{\mu}}_k(\mathbf{x}) = \boldsymbol{\mu}^k(\mathbf{x}) \tag{5.8}$$

converges pointwise to a strictly convex vectorial mean function $\bar{\boldsymbol{\mu}}$ whose all N components are equal, i.e., $\bar{\boldsymbol{\mu}} = (\bar{\mu}, \ldots, \bar{\mu})$, where $\bar{\mu}$ is a strictly convex mean function of order N.

Proof. Let $\mathbf{x} = (x_1, \ldots, x_N) \in M^N$ be a point in the domain of $\boldsymbol{\mu}$. We write

$$\mathbf{x}_{k+1} = \tilde{\boldsymbol{\mu}}_{k+1}(\mathbf{x}) = \boldsymbol{\mu}(\mathbf{x}_k), \tag{5.9}$$

with $\mathbf{x}_0 = \mathbf{x}$. We use the convention $\mathbf{x}_k = (x_1^k, \ldots, x_k^N)$. Denote the minimal ball of $\{x_i^k\}_{1=1}^N$ by $\overline{B(q_k, r_k)}$. Note that by the definition of strictly convex vectorial mean function, $\{x_i^k\}_{i=1}^N$ will stay in $\overline{\text{ConvHull}(\{x_i^k\}_{i=1}^N)} \subset B(q_0, r_0')$ for all $k \geq n_s$, where $r_0' < r_0$. Moreover, note that $\{x_i\}_{i=1}^N$ lies in a ball of radius smaller than r_{cx}; therefore, the minimal ball of $\{x_i\}_{i=1}^N$ exists and is unique. Also by Proposition

5.2.2 we have $r_{k+1} \leq r_k$; therefore, there exists $r_* \geq 0$ such $\lim_k r_k = r_*$. We show that $r_* = 0$. Since $\langle \mathbf{x}_k \rangle_k$ stays in a compact subset of M^N, it has a converging subsequence (in the product topology) $\langle \mathbf{x}_{k_j} \rangle_{k_j}$ with a limit $\bar{\mathbf{x}}$. Now by continuity of $\boldsymbol{\mu}$ we have

$$\lim_{k_j} \boldsymbol{\mu}^{n_s}(\mathbf{x}_{k_j}) = \boldsymbol{\mu}^{n_s}(\bar{\mathbf{x}}) \tag{5.10}$$

or

$$\lim_{k_j} \mathbf{x}_{k_j+n_s} = \boldsymbol{\mu}^{n_s}(\bar{\mathbf{x}}) = \bar{\bar{\mathbf{x}}}, \tag{5.11}$$

for a point $\bar{\bar{\mathbf{x}}} \in M^N$. Note that by continuity of the radius of the minimal ball with respect to the data points established in Corollary 4.3.1, the radius of the minimal ball of $\{\bar{x}_i\}_{i=1}^N$ is equal to r_*. The sequence of radii of minimal balls of $\{x_i^{k_j+n_s}\}_{i=1}^N$, also has r_* as its limit. On one hand, due to continuity, the minimal ball of $\{\bar{\bar{x}}_i\}_{i=1}^N$ must have radius r_*; on the other hand, from (5.11) and due to strict convexity of $\boldsymbol{\mu}^{n_s}$, the radius of the minimal ball of $\{\bar{\bar{x}}_i\}_{i=1}^N$ should be less than r_*. This is a contradiction and we must have $r_* = 0$, which means that $\bar{x}_1 = \ldots = \bar{x}_N = \bar{x}$. Note that for a large enough index K_j, $\{x_i^{K_j}\}_{i=1}^N$ will be in a small enough closed ball around $\bar{x}$. However, after that, due to convexity of $\boldsymbol{\mu}$, $\{x_i^k\}_{i=1}^N$ will stay in that ball for all $k \geq K_j$. Therefore, $\lim_k \mathbf{x}_k = \bar{\mathbf{x}}$.

Consider the assignment $\mathbf{x} \mapsto \bar{\boldsymbol{\mu}}(\mathbf{x}) = (\bar{\mu}(\mathbf{x}), \ldots, \bar{\mu}(\mathbf{x}))$. Note that due to strict convexity of $\boldsymbol{\mu}^{n_s}$, $\bar{x}$ must belong to the interior of any ball containing $\{x_i\}_{i=1}^N$, unless all the data points coincide. Also, by construction, $\bar{x}$ belongs to the closure of the convex hull of $\{x_i\}_{i=1}^N$. We only need to show the continuous dependence of $\bar{\boldsymbol{\mu}}(\mathbf{x})$ on $\mathbf{x}$ to prove that $\bar{\boldsymbol{\mu}}$ is a strictly convex mean function (see Definitions 5.2.1 and

5.2.2). To this end, first note that given $\epsilon > 0$ we can find large enough K such $\boldsymbol{\mu}^K(\mathbf{x}) \in B(\bar{\boldsymbol{\mu}}(\mathbf{x}), \frac{\epsilon}{2}) \subset M^N$. Next, note that when K, $\mathbf{x}$ and ϵ are fixed, we can find $\delta > 0$, such that $\mathbf{y} \in B(\mathbf{x}, \delta) \subset M^N$ implies $\boldsymbol{\mu}^k(\mathbf{y}) \in B(\boldsymbol{\mu}^k(\mathbf{x}), \frac{\epsilon}{2}) \subset M^N$. Therefore, we have that given $\epsilon > 0$ there are K and $\delta > 0$ such that if $\mathbf{y} \in B(\mathbf{x}, \delta)$, then $\boldsymbol{\mu}^K(\bar{\mathbf{y}}) \in B(\bar{\boldsymbol{\mu}}(\mathbf{x}), \frac{\epsilon}{2})$. However, because of the convexity of $\boldsymbol{\mu}$, $\boldsymbol{\mu}^k(\mathbf{y})$ will stay in $B(\bar{\boldsymbol{\mu}}(\mathbf{x}), \epsilon)$, for $k \geq K$. Hence, $\bar{\boldsymbol{\mu}}(\mathbf{y})$ must be in $\overline{B(\bar{\boldsymbol{\mu}}(\mathbf{x}), \epsilon)}$ and this shows continuity of $\boldsymbol{\mu}$.

□

5.2.3 Higher Orders of Smoothness

Assuming $\boldsymbol{\mu}$ is continuous, the strictly convex mean function $\bar{\mu}$ constructed in Theorem 5.2.2 is continuous. Whether $\bar{\mu}$ can achieve higher orders of smoothness, if $\boldsymbol{\mu}$ itself has higher orders of smoothness, is an interesting question. In general, it seems to be difficult to answer this question; the main reason is that we are dealing with infinite compositions of functions. However, if we could find uniform (in k) bounds on the norms of the derivatives of $\boldsymbol{\mu}^k$ we might be able to answer the question using standard tools. A special, yet important case is when $\boldsymbol{\mu}$ is Lipschitz with constant $L = 1$ i.e., there is $L = 1$ such that

$$d_{M^N}(\boldsymbol{\mu}(\mathbf{x}), \boldsymbol{\mu}(\mathbf{y})) \leq L d_{M^N}(\mathbf{x}, \mathbf{y}), \tag{5.12}$$

for $\mathbf{x}, \mathbf{y}$ such that $\{x_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^N$ lie in a ball $B(o, \rho_{\boldsymbol{\mu}})$.

Theorem 5.2.3. If in addition to conditions in Theorem 5.2.2, $\boldsymbol{\mu}$ is Lipschitz with

Lipschitz constant $L = 1$ (see 5.12) in its domain, then the new strictly convex mean functions $\bar{\boldsymbol{\mu}}$ and $\bar{\mu}$ also are Lipschitz with constants $L = 1$ and $L = \frac{1}{\sqrt{N}}$, respectively.

Proof. Observe that the compositions of Lipschitz functions result in a Lipschitz function whose constant is the product of the Lipschitz constants of the participating functions. As a result, all the members in the sequence $\langle \tilde{\boldsymbol{\mu}}_{\boldsymbol{k}} \rangle_k$ defined by $\tilde{\boldsymbol{\mu}}_k = \boldsymbol{\mu}^k$, are Lipschitz with Lipschitz constant $L = 1$. Therefore, $\langle \tilde{\boldsymbol{\mu}} \rangle_k$ is an equicontinuous sequence and by the Arzela-Ascolli theorem, there exists a subsequence $\langle \tilde{\boldsymbol{\mu}}_{k_j} \rangle_{k_j}$ which converges to $\bar{\boldsymbol{\mu}}$ uniformly. Hence, $\bar{\boldsymbol{\mu}}$ must be Lipschitz with constant $L = 1$. Note that because of the product metric on M^N, this means that $\bar{\mu} : M^N \to M$, will be Lipschitz with constant $L = \frac{1}{\sqrt{N}}$. □

The fact that Lipschitz constant of $\boldsymbol{\mu}^k$ is L^k, explains why we might not have higher order smoothness in this infinite process. For example, if $\boldsymbol{\mu}$ has its gradient bounded by 1 we could apply the above theorem, and conclude that $\bar{\mu}$ is differentiable almost everywhere, since it is Lipschitz. However, if the gradient of $\boldsymbol{\mu}$ was bounded by $L > 1$, we could not have the same conclusion, at least through this path.

5.2.4 Isometry Compatibility

We mentioned about the isometry compatibility of the L^p means in Section 4.4. For a convex mean function μ of order N, if $\mu(\phi(x_1), \ldots, \phi(x_N)) = \phi(\mu(x_1, \ldots, x_N))$ for any isometry ϕ of M, then μ is called isometry compatible. Similarly, we say that a convex vectorial mean function $\boldsymbol{\mu}$ of order N is isometry compatible if $\boldsymbol{\mu}(\phi(x_1), \ldots, \phi(x_N)) = \big(\phi(\mu_1(\mathbf{x})), \ldots, \phi(\mu_N(\mathbf{x}))\big)$. The following theorem shows

that if we start with isometry compatible means our higher order means also will be isometry compatible.

Theorem 5.2.4. If in Theorem 5.2.2, $\boldsymbol{\mu}$ is isometry compatible, then $\bar{\mu}$ also will be isometry compatible.

Proof. The proof follows from the fact that any isometry of M is a continuous function and that we have pointwise convergence in Theorem 5.2.2. □

Therefore, all the means that we construct are isometry compatible.

5.2.5 Dynamical System Point of View

The repetitive application of $\boldsymbol{\mu}$ induces an interesting dynamical system on M^N. Let us denote the diagonal of M^N by

$$\Delta(M^N) = \{(x_1, \ldots, x_N) \in M^N | x_1 = \ldots = x_N\}. \tag{5.13}$$

Define a cylinder with cubical or square cross section around the diagonal as follows:

$$C_{\boldsymbol{\mu}} = \{(x_1, \ldots, x_N) \in M^N | \exists \mathbf{q} = (q, \ldots, q) \in \Delta(M^N) \text{ such that } \max_i d(q, x_i) \leq \rho_{\boldsymbol{\mu}}\}, \tag{5.14}$$

where $\rho_{\boldsymbol{\mu}}$ is small enough such that all the components of $\boldsymbol{\mu}$ are well-defined. Notice the similarity between this definition and Definition 5.1. Now for any point $\mathbf{x} \in C_{\boldsymbol{\mu}}$ the iteration

$$\mathbf{x}_{k+1} = \boldsymbol{\mu}(\mathbf{x}_k) = \boldsymbol{\mu}^{k+1}(\mathbf{x}) \tag{5.15}$$

is well-defined. More interestingly if $\overline{B(q_k, r_k)}$ is the minimal ball of $\{x_i^k\}_{i=1}^N$, then the distance between $\mathbf{x}_k$ and $\Delta(M^N)$ measured in the L^∞ norm is minimized by $\mathbf{q}_k = (q_k, \ldots, q_k)$ and

$$r_k = \min_{\mathbf{q} \in \Delta(M^N)} \max_i d(q, x_i^k). \tag{5.16}$$

Note that $\mathbf{q}_k$ is the projection $\mathbf{x}_k$ onto $\Delta(M^N)$ under the distance induced by the L^∞ norm, and the center of the minimal ball of $\{x_i^k\}_{i=1}^N$ is the preimage of $\mathbf{q}_k$ under the diagonal map, i.e., the map which maps M to $\Delta(M^N)$. Now r_k decreases at least every n_s iterations and the L^∞ distance between $\mathbf{x}_k$ and $\Delta(M^N)$ decreases, accordingly. We showed in Theorem 5.2.2 that the orbit of $\mathbf{x}$, converges to the diagonal of M^N and that the diagonal of M^N is the only invariant set of this dynamical system. The orbit will eventually meet the diagonal at $\bar{\boldsymbol{\mu}}(\mathbf{x})$, which is the image of the mean of $\{x_i\}_{i=1}^N$ on $\Delta(M^N)$. Figure 5.2 helps us visualize the behavior of the dynamical system induced by a primitive mean function. One can visualize that the smallest cubical cylinder containing $\mathbf{x}_k$ shrinks by every n_s applications of $\boldsymbol{\mu}$. Note that the diagonal $\Delta(M^N)$ also coincides with the continuum of the equilibria of the system (5.15). Moreover, any equilibrium point of the system is stable in the sense of Lyapunov [53], but it is not asymptotically stable since a perturbation of $\mathbf{x} \in \Delta(M^N)$ will not eventually come back to $\mathbf{x}$, necessarily.

Remark 5.2.1. In what preceded we alluded to the L^∞-projection interpretation of the L^∞ mean. More generally, the L^p mean of $\{x_i\}_{i=1}^N$ (with weight vector $\mathbf{w}$) can be considered as the projection of $\mathbf{x} \in M^N$ onto $\Delta(M^N)$ under the distance

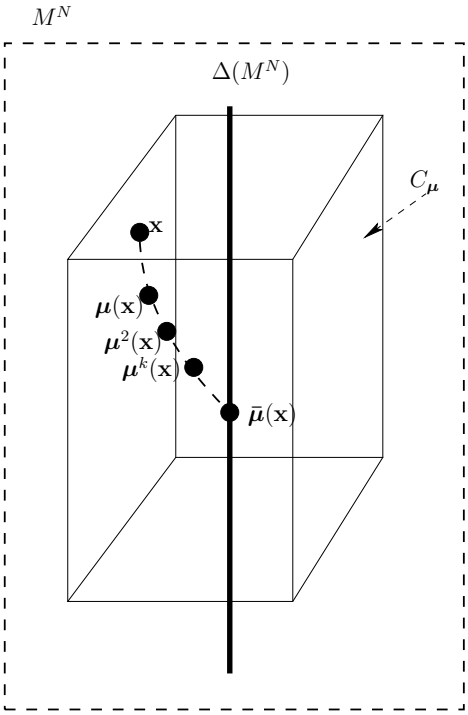


Figure 5.2: Repetitive application of the primitive mean function $\boldsymbol{\mu}$ brings a point $\mathbf{x}$ in the cylinder $C_{\boldsymbol{\mu}}$ closer (in the L^∞ sense) to $\Delta(M^N)$, the diagonal of M^N. In the limit the orbit of $\mathbf{x}$ meets the diagonal at $\bar{\boldsymbol{\mu}}(\mathbf{x})$.

function induced by the L^p norm in M^N. For $\mathbf{x}, \mathbf{y} \in M^N$ we define

$$d_{M^N,p}(\mathbf{x}, \mathbf{y}) = \big(\sum_{i=1} w_i d^p(x_i, y_i) \big)^{\frac{1}{p}}, \tag{5.17}$$

and for points which are close enough to $\Delta(M^N)$ define

$$\bar{\mathbf{x}}_p = \text{argmin}_{\mathbf{y} \in \Delta(M^N)} d_{M^N,p}(\mathbf{x}, \mathbf{y}). \tag{5.18}$$

Therefore, we have $\bar{\mathbf{x}}_p = (\bar{x}_p, \ldots, \bar{x}_p)$ where $\bar{x}_p$ is the L^p mean of $\{x_i\}_{i=1}^N$ with weights $\mathbf{w} = [w_1, \ldots, w_N]^T$. Theorem 3.2.1 states that if $\mathbf{x}$ belongs to a cylinder of the form in (5.14) whose radius is smaller than $\rho_{\Delta,p}$, then the L^p projection of $\mathbf{x}$ onto $\Delta(M^N)$

defined in (5.18) is well-defined and yields a unique point.

5.2.6 Relation to Infinite Products of Stochastic Matrices

Theorem 5.2.2 is very similar to the famous Perron-Frobenius theorem which gives a condition for the convergence of the powers A^k of an $N \times N$ stochastic matrix A. More general results about the behavior of the infinite product of different matrices are available, e.g. the result of Wolfowitz [90]. It seems that our method of using the radius of minimal ball to study the behavior of the infinite composition of a single convex vectorial mean function with itself can be useful in deriving similar results for the more general case of different convex vectorial mean functions or matrices. Since it is not of direct interest to us, we just give a simple example of such a result here, but we believe that deeper results are possible.

Theorem 5.2.5. Let $\langle \boldsymbol{\mu}_k \rangle_k$ be a sequence of convex vectorial mean functions of order N with common domains. Assume that there exists a subsequence of $\langle \boldsymbol{\mu}_k \rangle_k$ which converges uniformly to a strictly convex vectorial mean function. Then the infinite composition of $\langle \boldsymbol{\mu}_k \rangle_k$ defined by

$$\bar{\boldsymbol{\mu}}_k = \boldsymbol{\mu}_k \circ \cdots \circ \boldsymbol{\mu}_1 = \boldsymbol{\mu}_k \circ \bar{\boldsymbol{\mu}}_{k-1} \tag{5.19}$$

converges pointwise to a strictly convex vectorial mean function $\bar{\boldsymbol{\mu}}$, whose all components are equal.

Proof. Let $\langle \boldsymbol{\mu}_{k_j} \rangle_{k_j}$ be a subsequence which converges uniformly to $\tilde{\boldsymbol{\mu}}$. Due to compactness, the subsequence $\langle \mathbf{x}_{k_j} \rangle_{k_j}$ must in turn have a subsequence (denoted again

by $\langle \mathbf{x}_{k_j} \rangle_{k_j}$ for convenience) which converges to a point $\bar{\mathbf{x}} \in M^N$. We show that $\lim_{k_j} \boldsymbol{\mu}_{k_j}(\mathbf{x}_{k_j}) = \lim_{k_j} \tilde{\boldsymbol{\mu}}_{k_j}(\bar{\mathbf{x}}) = \tilde{\boldsymbol{\mu}}(\bar{\mathbf{x}})$. We denote the Riemannian distance function induced by the product structure of M^N by $d_{M^N}(.,.)$. We have

$$d_{M^N}(\boldsymbol{\mu}_{k_j}(\mathbf{x}_{k_j}), \tilde{\boldsymbol{\mu}}(\bar{\mathbf{x}})) \leq d_{M^N}(\boldsymbol{\mu}_{k_j}(\mathbf{x}_{k_j}), \tilde{\boldsymbol{\mu}}(\mathbf{x}_{k_j})) + d_{M^N}(\tilde{\boldsymbol{\mu}}(\mathbf{x}_{k_j}), \tilde{\boldsymbol{\mu}}(\bar{\mathbf{x}})). \tag{5.20}$$

Now given $\epsilon > 0$, due to the uniform convergence of $\boldsymbol{\mu}_{k_j}$ to $\tilde{\boldsymbol{\mu}}$, we can find a natural number $K_1(\epsilon)$ such that for all indices k_j larger than $K_1(\epsilon)$,

$$d_{M^N}(\boldsymbol{\mu}_{k_j}(\mathbf{x}), \tilde{\boldsymbol{\mu}}(\mathbf{x})) \leq \frac{\epsilon}{2}, \tag{5.21}$$

for all $\mathbf{x}$ in the common domain of $\boldsymbol{\mu}_k$'s. Also due to continuity of $\tilde{\boldsymbol{\mu}}$ and since $\mathbf{x}_{k_j} \to \bar{\mathbf{x}}$, there exists another natural number $K_2(\epsilon)$, such that for all indices k_j larger than $K_2(\epsilon)$ we have

$$d_{M^N}(\tilde{\boldsymbol{\mu}}(\mathbf{x}_{k_j}), \tilde{\boldsymbol{\mu}}(\bar{\mathbf{x}})) \leq \frac{\epsilon}{2}. \tag{5.22}$$

The above three relations mean that given $\epsilon > 0$, for all indices k_j larger than $\max\{K_1(\epsilon), K_2(\epsilon)\}$

$$d_{M^N}(\boldsymbol{\mu}_{k_j}(\mathbf{x}_{k_j}), \tilde{\boldsymbol{\mu}}(\bar{\mathbf{x}})) \leq \frac{\epsilon}{2}. \tag{5.23}$$

Therefore, $\lim_{k_j} \boldsymbol{\mu}_{k_j}(\mathbf{x}_{k_j}) = \tilde{\boldsymbol{\mu}}(\bar{\mathbf{x}})$. The rest of the proof is essentially the proof of Theorem 5.2.2 with $n_s = 1$, except that the above relation replaces (5.10) in that proof. □

Corollary 5.2.1. In Theorem 5.2.5 if all $\boldsymbol{\mu}_k$'s are Lipschitz continuous with Lipschitz constant of $L = 1$, then $\bar{\boldsymbol{\mu}}_k$ converges to $\bar{\boldsymbol{\mu}}$ where $\bar{\boldsymbol{\mu}}$ itself is a strictly convex mean function with equal components; and moreover, it is Lipschitz continuous with Lipschitz constant of $L = 1$.

Corollary 5.2.2. Let $\langle \boldsymbol{\mu}_k \rangle_k$ be a sequence of vectorial strictly convex mean functions of order N in M. If there exists a function which appears infinitely many often in that sequence, then the result of Theorem 5.2.5 holds.

Remark 5.2.2. In [69], Moreau introduces a general theory of infinite compositions of certain convex nonlinear operators and in particular stochastic matrices to analyze consensus algorithms with time-varying communication links or topologies. The goal of a consensus algorithm is to bring a set of autonomous agents to agreement on a state or quantity via local communications between the agents. The theory developed in [69] is based on the analysis of the graphs associated with the interaction of the agents and the properties of the convex linear or nonlinear operators employed. The convexity assumptions for the those functions mentioned in [69] are similar to the ones we made for strictly convex vectorial mean functions in Definition 5.2.2 except that the assumptions in [69] do not include (or are not equivalent to) the strict inclusion with respect to closed balls and only a strict inclusion with respect to the convex hull is assumed. Hence, one cannot use the convergence results introduced in [69] to define the class of means we define here. The main reason is that the proofs in [69] rely on the inclusion properties of the convex hull of finite points in Euclidean space which are not valid in a Riemannian manifold. Nevertheless, we

believe that the rest of the ideas used in [69] can be used to prove stronger versions of Theorem 5.2.5.

5.3 Perimeter Shrinkage Scheme, Cyclic Pursuit and a Pairwise Based Mean

In this section, we use the simplest mean function of order $N = 2$, i.e., the geodesic midpoint assignment or more generally the geodesic averaging, to design mean functions of higher orders. Although, it is possible to analyze this process using the methods from the previous section, it would more interesting to use other techniques.

5.3.1 Perimeter Shrinkage Scheme and a Pairwise-Iterative Mean

Let μ_α be the geodesic averaging function defined in Example 5.2.2. Consider the points $\{x_i\}_{i=1}^N \subset M$. We use a cyclic indexing: $x_{N+1} = x_1$ and $x_{N+2} = x_2$. Now assume that x_i and x_{i+1} for all $1 \leq i \leq N$ are close enough such that $\mu_\alpha(x_i, x_{i+1})$ is well defined i.e., is a single point. For example, if each pair of (x_i, x_{i+1}) lies in a ball of radius less than $\frac{\text{inj}M}{2}$, where inj$M$ is the injectivity radius of M, then $\mu_\alpha(x_i, x_{i+1})$ is well-defined. In particular, if all of the points lie in a strongly convex ball things become very simple, as we shall see. Consider the following iterative scheme:

$$x_i^{k+1} = \mu_\alpha(x_i^k, x_{i+1}^k), \quad 1 \leq i \leq N \tag{5.24}$$

where k is the iteration step and $x_i^0 = x_i$. Figure (5.3) schematically shows the first iteration of the process for $N = 4$ and $\alpha = \frac{1}{2}$. We call this iterative scheme a *perimeter shrinkage* scheme because of the next lemma. We define the perimeter shrinkage map $PS_\alpha : M^N \to M^N$ as

$$PS_\alpha(\mathbf{x}) = (\mu_\alpha(x_1, x_2), \ldots, \mu_\alpha(x_N, x_1)), \tag{5.25}$$

where $\mathbf{x} = (x_1 \ldots, x_N) \in M^N$. Here, we have implicitly assumed that the geodesic averagings are well-defined. We denote $(x_1, \ldots, x_N) \in M^N$ by $\mathbf{x} \in M^N$ and write (5.24) compactly as

$$\mathbf{x}_{k+1} = PS_\alpha(\mathbf{x}_k). \tag{5.26}$$

If we assume the domain of PS_α to be as in (5.1) with $\rho_{PS_\alpha} \leq r_{\text{cx}}$, then PS_α is a convex vectorial mean function of order N and it is easy to check that the $N - 1^{th}$ power of PS_α, is a strictly convex vectorial mean function of order N. Therefore, we could use Theorem 5.2.2 to show the convergence of the iteration, when initial data set $\{x_i\}_{i=1}^N$ lies in a strongly convex ball. However, since there is another interesting mechanism involved, i.e., the perimeter shrinkage mechanism, and we want to consider a domain different from the one mentioned above, we use a different approach.

Another function which we need is the perimeter function $P : M^N \to \mathbb{R}$ defined as:

$$P(x_1, \ldots, x_N) = \sum_{i=1}^{N} d(x_i, x_{i+1}), \tag{5.27}$$

which is also continuous. Similarly, the energy function $E : M^N \to \mathbb{R}$, which is the sum of the squares of the side lengths of the geodesic polygon $x_1 \dots x_N$, defined as

$$E(x_1, \dots, x_N) = \sum_{i=1}^{N} d^2(x_i, x_{i+1}) \tag{5.28}$$

is continuous, too. Also recall the definition of a closed geodesic:

Definition 5.3.1. On a Riemannian manifold a closed geodesic is a loop which is geodesic at all of its points.

Less formally, a closed geodesic is a geodesic that comes back to its initial point but at an angle which is equal to π. A geodesic that is only closed, i.e., it closes the loop at an angle different from π is called a *geodesic lasso* [24, pp. 255]. A closed geodesic is also called a *periodic geodesic* [9]. A simple closed geodesic is a closed geodesic that does not intersect itself. Closed geodesics, their existence and counting them have very old and deep roots in Riemannian geometry [9]. They correspond to the periodic solutions of the geodesic flow.

The following simple lemma is very useful:

Lemma 5.3.1. Let $\{x_i^k\}_{i=1}^N$ and $\{x_i^{k+1}\}_{i=1}^N$ be two consecutive sets of points generated by the PS process with parameter α. Let P_k and P_{k+1} be the perimeters of the closed geodesic polygons $x_1^k x_2^k \dots x_N^k$ and $x_1^{k+1} x_2^{k+1} \dots x_N^{k+1}$, respectively. Also let E_k and E_{k+1} denote the sums of the squares of the side lengths of the closed geodesic polygons $x_1^k x_2^k \dots x_N^k$ and $x_1^{k+1} x_2^{k+1} \dots x_N^{k+1}$, respectively. We have

1. $P_{k+1} \leq P_k$ with $P_{k+1} = P_k$ if and only if the geodesic polygon $x_1^k x_2^k \dots x_N^k$ is a

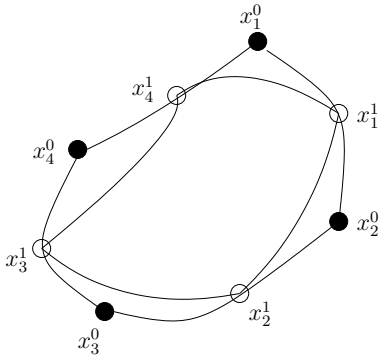


Figure 5.3: The initial geodesic polygon $x_1^0x_2^0x_3^0x_4^0$ is transformed to the new geodesic polygon $x_1^1x_2^1x_3^1x_4^1$ by inserting the new vertices at the midpoint of each side. The perimeter of $x_1^0x_2^0x_3^0x_4^0$ is strictly less than that of $x_1^1x_2^1x_3^1x_4^1$, unless the latter is a closed geodesic. We assume that the initial points are such that the sides of the initial geodesic can be determined uniquely, i.e, the shortest geodesics of interest are unique.

closed geodesic[1].

2. $E_{k+1} \leq E_k$ with $E_{k+1} = E_k$ if and only if the geodesic polygon $x_1^k \dots x_N^k$ is of equal side lengths and it is also a closed geodesic.

Here, a point is assumed to be a closed geodesic of length zero.

Proof. The proof of the first part is essentially applying the triangle identity. Recall that for $x, y, z \in M$:

$$d(x, y) \leq d(x, z) + d(z, y) \tag{5.29}$$

with equality if and only if z lies between x and y on the shortest geodesic connecting

[1]To be pedantic, we have to say that "the geodesical polygon $x_1^k x_2^k \dots x_N^k$ can be parameterized as a unit speed closed geodesic," since geodesics are curves, i.e., functions defined on the real line.

these two points. Therefore, we have:

$$d(x_i^{k+1}, x_{i+1}^{k+1}) \leq d(x_i^{k+1}, x_{i+1}^k) + d(x_{i+1}^k, x_{i+1}^{k+1}) = (1-\alpha) \cdot d(x_i^k, x_{i+1}^k) + \alpha \cdot d(x_{i+1}^k, x_{i+2}^k), \tag{5.30}$$

for $1 \leq i \leq N$; and hence $P_{k+1} \leq P_k$. The equality only happens when we have equalities in all the triangle inequalities employed. But that means that the consecutive geodesic segments $x_i^k x_{i+1}^k$ and $x_{i+1}^k x_{i+2}^k$ have to be geodesic at the connecting vertex x_{i+1}^k. As a result, the whole geodesic polygon $x_1^k x_2^k \dots x_N^k$ has to be a closed a geodesic. For the second part first apply the triangle inequality as before and expand as

$$d^2(x_i^{k+1}, x_{i+1}^{k+1}) \leq (d(x_i^{k+1}, x_{i+1}^k) + d(x_{i+1}^k, x_{i+1}^{k+1}))^2 \leq (1-\alpha)^2 \cdot d^2(x_i^k, x_{i+1}^k) + 2\alpha(1-\alpha) d(x_{i+1}^k, x_{i+1}^{k+1})) d(x_{i+1}^k, x_{i+2}^k) + \alpha^2 d^2(x_{i+1}^k, x_{i+2}^k), \tag{5.31}$$

for $1 \leq i \leq N$. Now adding both sides of the above for $1 \leq i \leq N$ yields:

$$E_{k+1} \leq (1-\alpha)^2 E_k + 2\alpha(1-\alpha) \sum_{i=1}^{N} d(x_{i+1}^k, x_{i+1}^{k+1})) d(x_{i+1}^k, x_{i+2}^k) + \alpha^2 E_k. \tag{5.32}$$

Now applying the Cauchy-Schwartz inequality gives

$$E_{k+1} \leq E_k((1-\alpha)^2 + 2\alpha(1-\alpha) + \alpha^2) = E_k, \tag{5.33}$$

where the equality holds if and only if the geodesic polygon $x_1^k \dots x_N^k$ is a closed

geodesic (as before) and the two vectors $[d(x_1, x_2), \ldots, d(x_N, x_{N+1})]^T$ and $[d(x_2, x_3), \ldots, d(x_{N+1}, x_{N+2})]^T$ are parallel in $\mathbb{R}^N$ (i.e., the equality condition for the Cauchy-Schwartz inequality). It is easy to see that the latter requires that $d(x_i, x_{i+1}) = d(x_{i+1}, x_{i+2})$ for $1 \leq i \leq N$. This completes the proof. □

The above lemma shows that the set

$$\Omega = \{(y_1, \ldots, y_N) \in M^N | \text{ geodesic polygon } y_1 \ldots y_N \text{ is a closed geodesic}\}, \quad (5.34)$$

is invariant under PS_α, i.e., $PS_\alpha(\Omega) \subset \Omega$. For the elements of Ω the perimeter of the corresponding geodesics remain invariant under the PS_α map, too. However, the energy function E in (5.28) is invariant under PS_α only on $\Omega_E \subset \Omega$:

$$\Omega_E = \{(y_1, \ldots, y_N) \in \Omega | d(y_1, y_2) = \ldots = d(y_N, y_{N+1})\}. \quad (5.35)$$

Neither of these two sets is a fixed point set of PS_α. One can see that $\mathbf{y} = (y_1, \ldots, y_N) \in M^N$ is a fixed point of PS_α if and only if $y_1 = \ldots = y_N$, i.e., the length of the corresponding closed geodesic is zero or equivalently $\mathbf{y}$ belongs to the diagonal of M^N. In order to define a mean we need to guarantee that the PS scheme converges to a fixed point of PS_α. If $\{x_i\}_{i=1}^N$ lies in a strongly convex set this will be guaranteed. A strongly convex set cannot contain a non-trivial closed geodesic, since if such a geodesic exists in that set, for two points on the geodesic whose distance from each other is half of the length of the closed geodesic, there are two minimizing geodesics connecting them in that set, which contradicts the defini-

tion of a strongly convex set. Also another consequence of the strong convexity is that if $\{x_i\}_{i=1}^N$ lies strongly convex set A, so does $\{x_i^k\}_{i=1}^N$ for all $k \geq 1$.

Theorem 5.3.1. Consider the perimeter shrinkage map PS_α defined in (5.25). If $\{x_i\}_{i=1}^N$ lies in a compact strongly convex set $A \subset M$, then the perimeter shrinkage iteration scheme induced by PS_α converges to a single point $\bar{x}$ in A, which also belongs to the closure of the convex hull of $\{x_i\}_{i=1}^N$. The assignment $(x_1, \ldots, x_N) \mapsto \bar{x}$ is a continuous function from A to A. Moreover, the function $\mu_{P,\alpha}(x_1, \ldots, x_N) = \bar{x}$ with domain defined in (5.1) and with parameter $\rho_\mu < r_{\text{cx}}$ is a strictly convex mean function of order N. For $\alpha = \frac{1}{2}$ we call $\bar{x}$ the pairwise mean of $(x_1, \ldots, x_N)$[1].

Proof. Denote the convex hull of $\{x_i\}_{i=1}^N$ by C and its closure by $\overline{C}$. Note that $\overline{C} \subset A$ is a compact strongly convex set by Proposition 2.3.1; hence, it includes no closed geodesic. Consider the sequence $\langle \mathbf{x}_k \rangle_{k=0}^\infty$ defined in (5.26). The sequence stays in the compact set $\mathcal{C} = \overline{C} \times \ldots \times \overline{C}$. By compactness, there is a subsequence $\langle \mathbf{x}_{k_j} \rangle$ which converges to a point $\bar{\mathbf{x}} = (\bar{x}_1, \ldots, \bar{x}_N) \in \mathcal{C}$. On the other hand, from Lemma (5.3.1) the sequence of positive numbers $\langle P_k \rangle_k$ and hance $\langle P_{k_j} \rangle_{k_j}$ should have a limit P_*; and also $P(PS_\alpha(\mathbf{x})) < P(\mathbf{x})$ in $\mathcal{C}$ unless $P(\mathbf{x}) = 0$. Note that $P(\bar{\mathbf{x}}) = P_*$ (by continuity of P). If $P_* \neq 0$ we have (by continuity of functions PS_α and P):

$$P_* = \lim_{k \uparrow \infty} P(PS_\alpha((\mathbf{x}_k)) = \lim_{k_j} P(PS_\alpha(\mathbf{x}_{k_j})) = P(PS - \alpha(\bar{\mathbf{x}})) < P(\bar{\mathbf{x}}), \tag{5.36}$$

which is a contradiction. Therefore, $P_* = 0$ and $\bar{\mathbf{x}} = (\bar{x}, \ldots, \bar{x})$, where $\bar{x} \in \overline{C}$.

[1]The pairwise mean in general is not permutation invariant, so to be accurate we use n-tuple $(x_1, \ldots, x_N)$ instead of $\{x_i\}_{i=1}^N$.

Now, after finite steps the sets $\{x_i^{k_j}\}_{i=1}^N$ will be in an arbitrary small strongly convex ball around $\bar{x}$; and hence does the next set of points $PS_\alpha(x_1^{k_j}, \ldots, x_N^{k_j})$ and so forth. Therefore, $\mathbf{x}_k \to \bar{\mathbf{x}} = (\bar{x}, \ldots, \bar{x})$ in the product topology. The proof of the continuity of $\mathbf{x} \mapsto \bar{x}$ is similar to the proof of a similar fact in Theorem 5.2.1.

Now assume $\{x_i\}_{i=1}^N$ lies in a closed ball $\overline{B(o,\rho)}$. If $(y_1, \ldots, y_N) = PS_\alpha^{N-1}(x_1, \ldots, x_N)$ then $\{y_i\}_{i=1}^N$ lies the smaller ball $\overline{B(o,\rho')}$ for some $\rho' < \rho$, unless x_i's coincide and they lie on the boundary of $\overline{B(o,\rho)}$. Therefore, excluding this case, by the previous part $\bar{x}$ belongs to the interior of $\overline{B(o,\rho)}$, which shows that $\mu_{P,\alpha}(\mathbf{x})$ is a strictly convex mean function of order N. □

Remark 5.3.1. It is appropriate here to mention both the strengths and limitations of the notion of a primitive vectorial mean function and Theorem 5.2.1. Note that if we define a more general version of the PS map PS_α in (5.25) as

$$\boldsymbol{\mu}_{\boldsymbol{\alpha}}(x_1, \ldots, x_N) = (\mu_{\alpha_1}(x_1, x_2), \ldots, \mu_{\alpha_N}(x_N, x_1)), \tag{5.37}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]^T$, then $\boldsymbol{\mu}_{\boldsymbol{\alpha}}(x_1, \ldots, x_N)$ is a primitive vectorial mean function with primitivity index of $N-1$, provided we consider its domain to be of the form (5.54) with parameter r_{cx}. Therefore, the iterative application of $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$, by Theorem 5.2.1, induces a strictly convex mean function of order N. A look at the proof of Lemma 5.3.1, shows that with α_i's being different from each other, contrary to PS_α, $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$ is not a perimeter shrinkage map, necessarily. Hence, the use of the radius of the minimal ball in Theorem 5.2.1, which is the consequence of the properties of the primitive vectorial mean functions, enabled us to analyze more general mean

functions than those we could analyze by just using the perimeter function. At the same time the use of perimeter function in the first part of Theorem 5.3.1, enabled us to prove the convergence of the PS scheme in a strongly convex domain whose shape is arbitrary, i.e., it is not a ball necessarily.

5.3.2 Cyclic Pursuit on Manifolds: Discrete-time Case

As soon as we make sure that a set of data points lies in a strongly convex ball Theorem 5.3.1 assures us that the perimeter shrinkage iteration converges to a single point. How about when the points are not in any strongly convex domain? Analysis of the problem in this case has relations with the problem of cyclic pursuit in which N ordered agents such as ants, frogs or moving vehicles follow each other in a cyclic fashion i.e., agent i follows agent $i+1$ (modulus N) with certain rules of pursuit [14]. The simplest rule of pursuit can be that agent i moves along the shortest path to agent $i+1$ either with speed proportional to its distance to agent $i+1$ or with unit speed. In the discrete-time setting we assume that each agent hops to the midpoint of the shortest geodesic connecting it to the agent it follows. Therefore, maybe the behavior of frogs will be more suitable to this situation. Now with this formulation the cyclic pursuit problem and the PS scheme are equivalent. It is well-known that the cyclic pursuit in the plane or in Euclidean space results in convergence to a single point [14]. Here, we consider the cyclic pursuit problem in a compact Riemannian manifold.

Theorem 5.3.2. Let M be a compact and connected (hence complete) Riemannian manifold. For the points $\{x_i\}_{i=1}^{N} \subset M$ assume the PS scheme (or the cyclic pursuit

scheme) iteration is well-defined at each step (in particular, if $d(x_i, x_{i+1}) < \text{inj}M$, then this would be guaranteed). Then $\mathbf{x}_k \to \Omega_E$ and $E(\mathbf{x}_k) \to E_* \geq 0$, where Ω_E is defined in (5.35). More precisely,

1. If $E_* = 0$, then $x_i^k \to x_* \in M$ for $1 \leq i \leq N$.

2. If $E_* > 0$ and $\mathbf{x}_* = (x_1^*, \ldots, x_N^*) \in \Omega_E$ is one of the cluster points of $\langle \mathbf{x}_k \rangle_k$, then the geodesic polygon $x_1^* \ldots x_N^*$ is a closed geodesic of length $\sqrt{NE_*}$ and $d(x_i^*, x_{i+1}^*) = \sqrt{\frac{E_*}{N}}$ for $1 \leq i \leq N$.

Moreover, the set

$$\Omega_{E_l} = \{(y_1, \ldots, y_N) \in \Omega_E | d(x_i, x_{i+1}) = \frac{l}{N},\ 1 \leq i \leq N\} \subset \Omega_E \tag{5.38}$$

is an asymptotically unstable invariant set of the PS scheme unless $l = 0$, by which we mean that for any point $\mathbf{x} \in \Omega_{E_l}$ there is a small perturbation of $\mathbf{x}$ such that the orbit of $\mathbf{x}$ never converges back to Ω_{E_l}. If the set of the lengths of the closed geodesics of length smaller than l is finite[1], then Ω_{E_l} is an unstable invariant set for the PS scheme, meaning that there exists a small perturbation of $\mathbf{x} \in \Omega_{E_l}$ whose orbit will not stay close to Ω_{E_l}.

Proof. From the proof of Lemma 5.3.1 we know that if $d(x_i^k, x_{i+1}^k) < \text{inj}M$ for $1 \leq i \leq N$, then $d(x_i^{k+1}, x_{i+1}^{k+1}) < \text{inj}M$ for $1 \leq i \leq N$. Hence, the mentioned condition is sufficient to guarantee that the PS process remains well defined at all steps. Now, the decreasing sequence $\langle E_k \rangle_k$ of real numbers must have a limit,

[1]This condition is satisfied automatically in the special orthogonal groups and the Grassmannian manifolds (see Remark 6.2.2).

denoted by E_*. If $E_* = 0$, then since after finite number of steps $PS_\alpha(\mathbf{x}_k)$ belongs to a small convex ball, the scheme converges to a single point by Theorem 5.3.1. Now let $E_* > 0$. The compactness of M^N requires the sequence $\langle \mathbf{x}_k \rangle_k$ to have a converging subsequence $\langle \mathbf{x}_{k_j} \rangle_{k_j}$, with the limit $\mathbf{x}_*$. By continuity of the functions involved we have:

$$E(\mathbf{x}_*) = E(\lim_{k_j} \mathbf{x}_{k_j}) = \lim_{k_j} E(\mathbf{x}_{k_j}) = E_* \tag{5.39}$$

and

$$E(PS_\alpha(\mathbf{x}_*)) = E(PS_\alpha(\lim \mathbf{x}_{k_j})) = \lim_{k_j} E(PS_\alpha(\mathbf{x}_{k_j})) = E_*. \tag{5.40}$$

Therefore, $\mathbf{x}_*$ belongs to Ω_E. This means that any cluster point of $\langle \mathbf{x}_k \rangle_k$ belongs to Ω_E. This together with the compactness of M^N implies that $\mathbf{x}_k \to \Omega_E$. Claims 1 and 2 follow from the definition of Ω_E.

Now, note that for any $\mathbf{x} \in \Omega_{E_l}$ with $l > 0$, there exists a small perturbation $\hat{\mathbf{x}}$ for which $E(\hat{\mathbf{x}})$, and hence, $E(\hat{\mathbf{x}}_k)$ for $k \geq 0$ is smaller than $E_* = \frac{l^2}{N}$. Therefore, obviously $\hat{\mathbf{x}}_k$ cannot come back to Ω_{E_l} because of the continuity of $\mathbf{x} \mapsto E(\mathbf{x})$. If there are only a finite number of possible values for the lengths of closed geodesics shorter than l, then since $E(\hat{\mathbf{x}}_k)$ eventually will differ from E_* by a constant $\epsilon > 0$, the distance of $\hat{\mathbf{x}}_k$ from Ω_{E_l} also will eventually be larger than a constant δ_ϵ. This means that $\Omega_{E_l}(l > 0)$ is an unstable invariant set with respect to PS_α. □

This theorem says that in a cyclic pursuit on M the agents will converge to a formation which consists of equidistance points on closed geodesics of fixed length l. However, only Ω_{E_0}, i.e., the diagonal of M^N is an asymptotically stable invariant

set of the PS scheme. Note that we did not claim that the formation corresponds to a single closed geodesic. Convergence to a single closed geodesic requires a deeper study of the properties of the PS scheme such as the one in [21].

5.3.3 Cyclic Pursuit Schemes in Continuous-time

The formulation of the PS scheme described before is in discrete-time. It is easy to formulate a continuous-time counterpart of the PS scheme: for all i at each instant of time x_i should pursue x_{i+1} along the minimizing geodesic connecting them with speed equal to the distance between them. We implicitly have assumed that always x_i and x_{i+1} are such that the minimizing geodesic connecting them is unique. This implies that x_i pursues x_{i+1} in the sense that it tries to reduce its distance from x_{i+1}. We can write this scheme as:

$$\dot{x}_i(t) = -\frac{1}{2}\nabla d^2(x_{i+1}, x)\big|_{x=x_i} = \exp^{-1}_{x_i(t)} x_{i+1}(t), \quad 1 \leq i \leq N. \tag{5.41}$$

The initial conditions for this set of equations are $x_i(0) = x_i$. A simple argument shows that similar to the discrete-time PS scheme, the continuous-time PS scheme also remains well-defined for all $t > 0$, if we start with initial conditions such that $d(x_i(0), x_{i+1}(0)) < \text{inj}M$ for $1 \leq i \leq N$. Now consider the energy E defined in (5.28) as a Lyapunov function for the dynamical system (5.41) defined on M^N.

$$\frac{\mathrm{d}}{\mathrm{d}t}E(x_1(t), \ldots, x_N(t)) = \sum_{i=1}^{N} \langle \dot{x}_i, \nabla d^2(x, x_{i+1})\big|_{x=x_i} \rangle_{x_i(t)} + \langle \dot{x}_{i+1}, \nabla d^2(x, x_i)\big|_{x=x_{i+1}} \rangle_{x_{i+1}(t)} \tag{5.42}$$

or equivalently

$$\frac{\mathrm{d}}{\mathrm{d}t}E(x_1(t),\ldots,x_N(t)) = -2\sum_{i=1}^{N}\Big(\langle \exp^{-1}_{x_i(t)} x_{i+1}(t), \exp^{-1}_{x_i(t)} x_{i+1}(t)\rangle_{x_i(t)} + \langle \exp^{-1}_{x_{i+1}(t)} x_{i+2}(t), \exp^{-1}_{x_{i+1}(t)} x_i(t)\rangle_{x_{i+1}(t)}\Big). \tag{5.43}$$

Now algebraic simplification and noting that $\| \exp^{-1}_{x_i} x_{i+1}\|_{x_i} = \| \exp^{-1}_{x_{i+1}} x_i\|_{x_{i+1}} = d(x_i, x_{i+1})$ yield

$$\frac{\mathrm{d}}{\mathrm{d}t}E(x_1(t),\ldots,x_N(t)) = -2\sum_{i=1}^{N} \| \exp^{-1}_{x_i(t)} x_{i-1}(t) + \exp^{-1}_{x_i(t)} x_{i+1}(t)\|^2_{x_i(t)}, \tag{5.44}$$

where $\|.\|_x$ is the norm induced by the Riemannian structure of M at $x \in M$. We conclude that along the trajectories of (5.41) we have $\frac{\mathrm{d}}{\mathrm{d}t}E(x_1(t),\ldots,x_N(t)) = 0$ if and only if $\mathbf{x} = (x_1,\ldots,x_N) \in \Omega_E$. In the language of LaSalle's invariance principle [53, 42], Ω_E is the largest invariant set of the system in (5.41) at which the total derivative of E is zero. Based on this observation we could state a theorem similar to Theorem 5.3.2 for the continuous-time case; however, for the sake of space we avoid that.

We should mention that since the closed geodesic formation is not a stable formation (see Theorem 5.3.2) one might think of stabilizing this formation by requiring the autonomous agents to apply some sort of control or feedback in a decentralized fashion. It might be possible that second order dynamics for the agents as introduced in [49] for the planar case is helpful in this case too. Also we add that a similar control problem for stabilization of the formation of agents on the sphere

S^2 embedded in $\mathbb{R}^3$ has been proposed in [73], with the caveat that the control described is not found intrinsically. By this we mean that the control depends on quantities which are measured and defined in the ambient space (i.e., $\mathbb{R}^3$). We also refer the reader to [79, 80, 81] for related works.

5.4 Means Based on Mean-Invariance

In this section we introduce a simple scheme to construct higher order permutation invariant strictly convex mean functions from simple geodesic midpoint assignment. This idea was first introduced in [5] for defining a geometrical mean for positive definite matrices. Here, we extend it to more general manifolds and also associate it with the concept of mean-invariance. First, we give the definition of the simpler equal weight Mean-Invariance (MI) based mean. We describe the procedure in the following definition, and in Section 5.4.2 we prove existence and uniqueness and some other properties of the derived means.

Definition 5.4.1. Let (M, d) be a complete Riemannian manifold and $\mu_{\frac{1}{2}}(.,.)$ denote the geodesic midpoint assignment map (see Example 5.2.2). Define the N^{th} order Mean-Invariance (MI) based mean function $\mu_{\mathrm{MI},N}$ recursively as:

1. If $N = 2$, then set $\mu_{\mathrm{MI},2}(.,.) = \mu_{\frac{1}{2}}(.,.)$,

2. Otherwise, in order to find $\mu_{\mathrm{MI},N}(\mathbf{x})$ for any $\mathbf{x} \in M^N$ set $x_1^0 = x_1, \ldots, x_N^0 = x_N$ and perform the following iteration:

$$x_i^{k+1} = \mu_{\mathrm{MI},N-1}(\mathbf{x}_k^{\downarrow N+1-i}), \quad 1 \leq i \leq N, \tag{5.45}$$

where for the N-dimensional vector $\mathbf{x} \in M^N$, $\mathbf{x}^{\downarrow j} \in M^{N-1}$ is an $N-1$-dimensional vector built from $\mathbf{x}$ by dropping the j^{th} component of $\mathbf{x}$ out. We write the above iteration more compactly as:

$$\mathbf{x}_{k+1} = \left(\mu_{\mathrm{MI},N-1}(\mathbf{x}_k^{\downarrow N}), \ldots, \mu_{\mathrm{MI},N-1}(\mathbf{x}_k^{\downarrow 1})\right) \tag{5.46}$$

for $k = 1, 2, \ldots$ or even more compactly we write

$$\mathbf{x}_{k+1} = \boldsymbol{\mu}_{\mathrm{MI},N}(\mathbf{x}_k), \tag{5.47}$$

where the components of $\boldsymbol{\mu}_{\mathrm{MI},N}(\cdot)$ are defined in (5.46). Assuming all the steps of this construction are well-defined and if the sequence $\mathbf{x}_k = (x_1^k, \ldots, x_N^k) \in M^N$ converges to a point $\bar{\mathbf{x}} = (\bar{x}, \ldots, \bar{x})$, then we call $\bar{x}$ the N^{th} order MI mean of $\{x_i\}_{i=1}^N$ and set $\mu_{\mathrm{MI},N}(\mathbf{x}) = \bar{x}$.

Assuming $\mu_{\mathrm{MI},N}(x_1, \ldots, x_N)$ is well-defined, we observe that by construction (i.e., the iteration in (5.45)) all the generated intermediate sets of points $\{x_i^k\}_{i=1}^N$ have $\bar{x}$ as their N^{th} order MI mean. Hence, the name "mean-invariance" is justified. In fact, the main idea in defining the N^{th} order MI mean, is nothing but replacing each point by the $N-1^{th}$ order mean of $N-1$ points including the point of interest itself and repeating this process. One can show, using the Perron-Frobenius theorem, that this process in Euclidean space produces the standard Euclidean mean with equal weights. Note that $\mu_{\mathrm{MI},2}(.,.) = \mu_{\frac{1}{2}}(.,.)$ is permutation invariant. Now, assume $\mu_{\mathrm{MI},N-1}$ is permutation invariant. Since all N combinations of size $N-1$ from the

set $\{x_i\}_{i=1}^N$ are included in defining $\boldsymbol{\mu}_{\mathrm{MI},N}$ (see (5.46)), by induction we conclude that $\mu_{\mathrm{MI},N}$ must be permutation invariant.

5.4.1 Weighted Mean-invariance Means

Here, we propose two methods to incorporate weights in the defining the MI mean. In the first method, given the set $\{x_i\}_{i=1}^N$ we can increase the number of points by introducing new points that are repetition of the original points. For example, for $\{x_1, x_2\}$, $\mu_{\mathrm{MI},2}(x_1, x_2)$ is the midpoint of the minimizing geodesic connecting them x_1 and x_2, while if we set $x_3 \equiv x_1$ then $\mu_{\mathrm{MI},3}(x_1, x_2, x_3)$ is a point on the geodesic at distance third of $\frac{1}{3}d(x_1, x_2)$ from x_1. More generally, we can introduce N_i repetitions of x_i for all i, and resemble rational weights of $w_i = \frac{N_i}{\sum_i N_i}$. We can approximate any weight by this process; however, obviously it has the drawback of dramatically increasing the number of total points to $N_{new} = \sum_{i=1}^N N_i$.

The second idea is to include the weights in the geodesic averaging step explicitly starting with $N = 2$. As an example, let $N = 3$ and weights w_1, w_2 and w_3 be given. We interpret these weights as global contributions of x_i's in the mean. However, for the pairwise mean of x_1 and x_2 we require the weight $\alpha_1 = \frac{w_1}{w_1+w_2}$ for x_1 and $1-\alpha_1$ for w_2. Similarly we set $\alpha_i = \frac{w_i}{w_i+w_{i+1}}$ with modulus 3 cyclic indexing.

Example 5.4.1. Let us consider an example with $N = 3$ points in $\mathbb{R}$. We have three weights $\{w_i\}_{i=1}^3$ from which we determine the α_i as $\alpha_i = \frac{w_i}{w_i+w_{i+1}} (i = 1, 2, 3)$. Denote $\mathbf{x}_k = [x_1^k, x_2^k, x_3^k]^T \in \mathbb{R}^3$. Then the iteration process is obtained from (5.45) by replacing the midpoint mean function $\mu_{\mathrm{MI},2} = \mu_{\frac{1}{2}}$ that gives x_1^{k+1}'s by $\mu_{\alpha_1}(.,.)$,

i.e., we set $x_1^{k+1} = \mu_{\alpha_1}(x_1^k, x_2^k)$ and so forth for $i = 2, 3$. Therefore, we have

$$\mathbf{x}_{k+1} = A\mathbf{x}_k, \tag{5.48}$$

where

$$A = \begin{bmatrix} \alpha_1 & 1-\alpha_1 & 0 \\ 0 & \alpha_2 & 1-\alpha_2 \\ 1-\alpha_3 & 0 & \alpha_3 \end{bmatrix}. \tag{5.49}$$

Now, from the Perron-Frobenius theorem we know that A^k converges to a rank one matrix A_∞ whose all rows are the same [43, p.524]. This common vector is the eigenvector corresponding to the largest eigenvalue of A^T which is 1. The eigenvector should be normalized so that the sum of its entries are 1. It is easy to calculate this eigenvector as $\tilde{\mathbf{w}} = [\tilde{w}_1, \tilde{w}_2, \tilde{w}_3]^T$ where

$$\tilde{w}_i = \frac{(1-\alpha_i)^{-1}}{\sum_{j=1}^{3}(1-\alpha_j)^{-1}}, \qquad i = 1, 2, 3. \tag{5.50}$$

More concretely, if $\mathbf{w} = [w_1, w_2, w_3]^T = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]^T$, then $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3]^T = [\frac{1}{2}, \frac{1}{2}, \frac{1}{2}]^T$ and $\tilde{\mathbf{w}} = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]^T$, hence $\mu_{MI}^3(x_1, x_2, x_3) = \frac{1}{3}(x_1 + x_2 + x_3)$. However, if $\mathbf{w} = [\frac{1}{4}, \frac{1}{8}, \frac{5}{8}]^T$, then $\boldsymbol{\alpha} = [\frac{2}{3}, \frac{1}{6}, \frac{1}{2}]^T$ and $\tilde{\mathbf{w}} = [\frac{30}{77}, \frac{12}{77}, \frac{35}{77}]^T$, hence the explicitly weighted MI mean of $\{x_i\}_{i=1}^3$ corresponding to $\mathbf{w}$ is $\mu_{\mathrm{WMI},3}(x_1, x_2, x_3) = \frac{30}{77}x_1 + \frac{12}{77}x_2 + \frac{35}{77}x_3$. As it can be seen in this simple example the relation between initial weights w_i's and the actual weights w_i's is a nonlinear relation. However, our intuition that putting a large w_1 for x_1 will result in large $\tilde{w}_1$ is correct. More accurately, we can see that

if $w_1 > w_2 > w_3$ then $\alpha_1 > \alpha_2 > \alpha_3$, and hence $\tilde{w}_1 > \tilde{w}_2 > \tilde{w}_3$.

Next, we give a formal definition for the explicitly Weighted MI (WMI) mean based on the second idea mentioned above, since it is more practical.

Definition 5.4.2. Let (M, d) be a complete Riemannian manifold and $\mu_\alpha(.,.)$ be the geodesic averaging function. Define the N^{th} order weighted Mean-Invariance based mean $\mu_{\text{WMI},\mathbf{w}_N}(.)$ corresponding to the positive weight vector $\mathbf{w}_N = [w_1, \ldots, w_N]^T$ recursively as:

1. If $N = 2$ then $\mu_{\text{WMI},\mathbf{w}_2}(.,.) = \mu_{w_1}(.,.)$ and $\mathbf{w}_2 = [w_1, w_2]^T$ is the weight vector.

2. Otherwise, to find $\mu_{\text{WMI},\mathbf{w}_N}(\mathbf{x})$ where $\mathbf{x} = (x_1, \ldots, x_N)$, set $\mathbf{x}_0 = \mathbf{x}$ and perform the following iteration:

$$\mathbf{x}_{k+1} = \big(\mu_{\text{WMI},\mathbf{w}^1_{N-1}}(\mathbf{x}_k^{\downarrow N}), \ldots, \mu_{\text{WMI},\mathbf{w}^N_{N-1}}(\mathbf{x}_k^{\downarrow 1})\big). \tag{5.51}$$

In the above the corresponding weight vector $\mathbf{w}^i_{N-1}$ is found as follows:

$$\mathbf{w}^i_{N-1} = \frac{\mathbf{w}_N^{\downarrow N+1-i}}{\mathbf{1}^T_{N-1}\mathbf{w}_N^{\downarrow N+1-i}} \quad i = 1, \ldots, N+1. \tag{5.52}$$

Here $\mathbf{1}_{N-1}$ is an $N-1$-dimensional vector all whose elements are 1. We write (5.51) more compactly as

$$\mathbf{x}_{k+1} = \boldsymbol{\mu}_{\text{WMI},\mathbf{W}_N}(\mathbf{x}_k), \tag{5.53}$$

where $\mathbf{W}_N$ is the $N \times N$ matrix of weights whose i^{th} row is equal to the

transpose of $(\mathbf{w}_N^i)^{\uparrow N+1-i}$. Here, for an $N-1$-dimensional vector $\mathbf{x}$, by $\mathbf{x}^{\uparrow i}$ we mean the N-dimensional vector whose i^{th} component is zero and the rest of it is filled in an increasing order of indices with the elements of $\mathbf{x}$ also in a increasing order of their indices (cf. $\mathbf{x}^{\downarrow i}$ in Definition 5.4.1). If the sequence of points $\langle \mathbf{x}_k \rangle_k$ in M^N converges to a point $\bar{\mathbf{x}} = (\bar{x}, \ldots, \bar{x})$, then we call $\bar{x}$ the N^{th} order WMI mean of $\{x_i\}_{i=1}^N$ with weight vector $\mathbf{w}_N = [w_1, \ldots, w_N]^T$ and set $\mu_{\mathrm{WMI},\mathbf{w}_N}(x_1, \ldots, x_N) = \bar{x}$.

In the case of equal weights the WMI mean reduces to the MI mean. Also note that the matrix $\mathbf{W}_N$ built in the definition, is a stochastic matrix which has exactly one zero at each row, and each column of it also has exactly one zero. This suggests that the convex vectorial mean function $\boldsymbol{\mu}_{\mathrm{WMI},\mathbf{w}_N}$ must be a primitive mean function of index $n_s = 2$, as we will see next.

5.4.2 Existence, Uniqueness and Other Properties

Next, we prove existence and uniqueness properties of the MI and WMI means. For this we use Theorem 5.2.2 about convex vectorial mean functions.

Theorem 5.4.1. Define

$$D = \{(x_1, \ldots, x_N) \in M^N | \exists o \in M \text{ and } \rho < r_{\mathrm{cx}}, \{x_i\}_{i=1}^N \subset \overline{B(o, \rho)}\}. \tag{5.54}$$

For any integer $N \geq 2$ and positive weight vector $\mathbf{w}_N$, the functions $\boldsymbol{\mu}_{\mathrm{MI},N}$ and $\boldsymbol{\mu}_{\mathrm{WMI},\mathbf{W}_N}$ defined in Definitions 5.4.1 and 5.4.2 are primitive mean functions with primitivity index of $n_s = 2$ on the domain D defined above. Therefore, if $\{x_i\}_{i=1}^N$ lies

in a ball $\overline{B(o,\rho)}$ with $\rho < r_{\text{cx}}$, then the MI mean and the WMI mean with weight vector $\mathbf{w}_N$ exist, are unique, belong to $B(o,\rho)$ and to the closure of the convex hull of $\{x_i\}_{i=1}^N$. Moreover, $\mu_{\text{MI},N}(\mathbf{x})$ and $\mu_{\text{WMI},\mathbf{w}_N}(\mathbf{x})$ are strictly convex mean functions of order N with domain D as above.

Proof. The proof is by induction over N. Obviously, all the claims hold for $N = 2$. Now, we assume they holds for $N \geq 2$, and we show they hold for $N + 1$. We only need to show that $\boldsymbol{\mu}_{\text{MI},N+1}$ and $\boldsymbol{\mu}_{\text{WMI},\mathbf{W}_{N+1}}$ are primitive mean functions. The rest of the claims follow from Theorem 5.2.2, immediately. First, note that $\boldsymbol{\mu}_{\text{MI},N+1}$ is a convex vectorial mean function of order $N + 1$. Its convexity follows from the fact that each of its components, i.e., $\mathbf{x} \mapsto \mu_{\text{MI},N}(\mathbf{x}^{\downarrow N+2-i})$ is convex by Proposition 5.2.1. Now suppose $\{x_i\}_{i=1}^{N+1} \subset \overline{B(o,\rho)}$ and $\mathbf{y} = \boldsymbol{\mu}_{\text{MI},N+1}(\mathbf{x})$. If of one y_i's, say y_1, lies on the boundary of $B(o,\rho)$, then by strict convexity of $\mu_{\text{MI},N}$, $x_1 = \ldots = x_N$ and they must lie on the boundary of the ball. Assume x_{N+1} does not lie on the boundary of the ball - otherwise we have nothing to prove-, then again by the strict convexity of $\mu_{\text{MI},N}$, y_i $(2 \leq i \leq N+1)$ lies inside the ball. Now since only one point among $\{y_i\}_{i=1}^{N+1}$ lies on the boundary of the ball, for the new set of points generated as $\mathbf{z} = \boldsymbol{\mu}_{\text{MI},N+1}(\mathbf{y})$ we see that all z_i's must belong to the interior of the ball. This means that $\boldsymbol{\mu}_{\text{MI},N+1}$ is primitive with primitivity index $n_s = 2$. The same argument also works for $\boldsymbol{\mu}_{\text{WMI},\mathbf{W}_{N+1}}$. □

This result, in particular gives a new proof for the existence, uniqueness and continuity results derived in [5, 59].

5.4.2.1 A Lipschitz Property in Manifolds of Nonpositive Curvature

In the construction of the MI mean from order $N-1$ to N, if each mean function of order $N-1$ has Lipschitz constant L, then the vectorial mean function constructed has Lipschitz constant $L' = \sqrt{N-1}L$. In order to guarantee the resultant mean function of order N to be Lipschitz, according to Theorem 5.2.3 we must have $L = 1$. Therefore, we need to have $L' = \frac{1}{\sqrt{N-1}}$. This only can happen if M is of nonpositive curvature.

Theorem 5.4.2. In a manifold of non-positive curvature the MI mean function $\mu_{\text{MI}} : M^N \to M$ of order N is Lipschitz with constant $L = \frac{1}{\sqrt{N}}$.

Proof. For the L^2 mean of $N = 2$ points with equal weights or the geodesic midpoint assignment we found the Lipschitz constant in Proposition 4.6.2 to be $L = \frac{1}{\sqrt{2}}$. Therefore, $\boldsymbol{\mu}_{\text{MI},3} : M^3 \to M^3$ defined according to (5.47) is Lipschitz with constant $L = 1$; hence, by Theorem 5.2.3 $\mu_{\text{MI},3}$ will be Lipschitz with constant $L = \frac{1}{\sqrt{3}}$. Now a simple induction proves the claim. □

It seems that, unfortunately, opposite to the Riemannian L^2 mean no form of Lipschitz property exists for the MI mean in the case of a manifold of positive curvature. From Proposition 4.6.2, recall that in a manifold of positive curvature (or more precisely in a manifold of nonnegative curvature where the curvature is not identically zero) the Riemannian L^2 mean, and in particular the geodesic midpoint assignment, are Lipschitz but with a Lipschitz constant larger than in the Euclidean case. This results in a Lipschitz constant larger than 1 for the vectorial mean

function which is used to construct the MI of order $N = 3$. Consequently, our iterative process, as explained in Section 5.2.3, annihilates Lipschitz property by multiplication of the constants larger than 1.

5.4.2.2 An Existence and Uniqueness Result for Non-ball Domains

In Theorem 5.4.1, in order to have a valid MI or WMI mean we required $\{x_i\}_{i=1}^N$ to belong a to strongly convex ball of radius smaller than r_{cx}. This in turn, was needed because of the requirements of Theorem 5.2.2 and our requirements about convex mean functions. As we shall see in Chapter 6 this assumption is restrictive for our applications in homogenous spaces such as the orthogonal group and the Grassmannian manifold, where we would like to have the same result but for points which belong to a special non-ball strongly convex set (see Sections 6.4 and 6.6 for more details). It is possible to prove existence and uniqueness of the MI and WMI means in a less restrictive situation as follows:

Theorem 5.4.3. Let (M, d) be a complete Riemannian manifold. Let $A \subset M$ be a strongly convex open set with $o \in A$ such that $x \mapsto d^2(x, o)$ is a strictly convex function in A. Also assume that there exists another strongly convex set A' such that $\bar{A} \subset A'$.[1] If $\{x_i\}_{i=1}^N \subset A$, then the MI and WMI means of $\{x_i\}_{i=1}^N$ exist, are unique, depend continuously on x_i's, and belong to the closure of the convex hull of $\{x_i\}_{i=1}^N$. Moreover, we have

$$d(o, \mu_{\text{MI}}(x_1, \ldots, x_N)) \leq \max_i d(o, x_i), \tag{5.55}$$

[1]This is a technical assumption which is automatically satisfied when A has a specific shape such as a Riemannian or a Finsler ball as we shall see in Section 6.6.

with equality if and only if x_i's coincide. The same holds for μ_{WMI}.

Proof. The proof is based on induction over the number of points N. The assumption about A' guarantees that the closure of the convex hull of $\{x_i\}_{i=1}^N$ is a strongly convex set (see Proposition 2.3.1); and the claim that the mean, if it exits and is unique, belongs to the closure of the convex hull of $\{x_i\}_{i=1}^N$ is the direct consequence of the construction and closed-ness of the closure of the convex hull of $\{x_i\}_{i=1}^N$. Also note that in case the MI or WMI mean of $\{x_i\}_{i=1}^N \subset A$ exists and is unique, after finite number of iterations the points x_i^k will belong to a small convex ball, hence the continuity of the mean with respect to the initial points follows from Theorem 5.4.1. Hence, we only need to prove the existence and uniqueness of the mean as well as (5.55). As before we only consider the MI scheme. For $N = 2$ due to strong convexity of A and strict convexity of $d^2(o, x)$, $\mu_\alpha(x_1, x_2)$ is unique, belongs to the geodesic connecting x_1 and x_2, $d^2(o, \mu_\alpha(x_1, x_2)) \leq \max_i d^2(o, x_i)$ with equality if and only if $x_1 = x_2$. Now assume the claim holds for $N - 1$ points. Define

$$h_k = \max_i d^2(o, x_i^k) \tag{5.56}$$

Denote by I_j^{N-1} the subset of indices $\{1, \ldots, N\}$ which is formed by indices that contribute in determining x_j^{k+1} from x_i^k's. Note that if $\max_{1 \leq i \leq N} d^2(o, x_i^{k+1})$ happens at $i = j$ then

$$\max_{1 \leq i \leq N} d^2(o, x_i^{k+1}) = d^2(o, x_j^{k+1}) \leq \max_{i \in I_j^{N-1}} d^2(o, x_i^k) \leq \max_{1 \leq i \leq N} d^2(o, x_i^k). \tag{5.57}$$

By the induction assumption, equality happens in the first inequality above if and only if all the $N-1$ x_i^ks contributing in determining x_j^{k+1} for $i \in I_j^{N-1}$ coincide. The equality in the second inequality above holds if and only if $\max_{1\leq i\leq N} d^2(o, x_i^k)$ happens at some $i = l \in I_j^{N-1}$. Therefore, $h_{k+1} \leq h_k$ with equality if and only if at least $N-1$ of x_i^k's coincide and these $N-1$ points are not closer to o than the other x_i^k. Assume this situation has happened, but not all the points coincide, say $x_1^k = \ldots = x_{N-1}^k$ but x_{N-1}^k is different from them with $d^2(o, x_N^k) \leq d^2(o, x_1^k)$. Then $d^2(o, x_1^{k+1}) = d^2(o, x_1^k)$ but $d^2(o, x_i^{k+1}) < d^2(o, x_1^k)$ for $1 < i \leq N$ since in determining each of these x_i^{k+1}'s not all of contributing x_i^k's coincide. Therefore, the farthest $x_i^{k+1} (1 \leq i \leq N)$ from o, which is x_1^{k+1} cannot coincide with any other points at this step. As a result at step $k+2$ we have $h_{k+2} < h_{k+1} = d^2(o, x_1^{k+1})$, unless all x_i^k's would have coincided. Hence, in any situation we have $h_{k+2} \leq h_k$ with equality if and only if all x_i^k's coincide. Therefore, noting the continuity of $(y_1, \ldots, y_N) \mapsto \max_i d^2(o, y_i)$ and the continuity of the means of order $N-1$, it easy to see that the subsequence $\langle \mathbf{x}_{2k-1} \rangle_k$ must have a converging subsequence which converges to $\bar{\mathbf{x}} = (\bar{x}, \ldots, \bar{x}) \in M^N$. This implies that after some finite steps k_0, $\{x_i^{2k_0-1}\}_{i=1}^N$ lies in a small enough convex ball. Therefore, from Theorem 5.4.1 this means that all sequences $\langle x_i^k \rangle_k$ converge to a common limit which coincides with $\bar{x}$. Also it is obvious that $\lim_k h_k = d^2(o, \bar{x}) \leq h_1 = \max_i d^2(o, x_i)$ with equality if and only if all x_i's coincide and this proves (5.55). $\square$

5.4.3 Generalized Weighted Mean-Invariance Means Resembling L^p Means

The MI or WMI means introduced before resemble the L^2 mean, as evidenced by the case of Euclidean space where the MI mean and the equal weight L^2 mean coincide. Note that by resemblance we do not necessarily mean closeness, rather we mean similar behavior with respect to outliers. Although, note that if the data points are close enough, then the equal weight L^2 mean and the MI mean in M will be close (at least up to the first order of approximation). It is possible to define recursive-iterative means which resemble other L^p means, provided we start with a base case different from $K = 2$. Note that the equal weight L^p mean of two close points coincides with the midpoint of the unique minimizing geodesic connecting them, for all $p > 1$. However, that is not the case for $K \geq 3$ points. The following definition generalizes Definitions 5.4.1 and 5.4.2 by choosing a base case at which the L^p mean of $K \geq 3$ points is found. We call such a mean the Generalized Weighted Mean Invariance (GWMI) based mean of type p. The interesting point is that the same method used in Theorem 5.4.1 can be used to prove the existence and uniqueness of the GWMI mean, as we see in Theorem 5.4.4.

Definition 5.4.3. Let (M, d) be a complete Riemannian manifold. Assume $\mu_{p,\mathbf{w}_N} : M^N \to M$ is the $L^p (1 < p \leq \infty)$ mean function with the positive weight vector $\mathbf{w}_N$, i.e., $\mu_{p,\mathbf{w}_N}(x_1, \ldots, x_N)$ is the L^p mean of the set $\{x_i\}_{i=1}^N$ with the corresponding weight vector $\mathbf{w}_N$, provided the mean exists uniquely. Let $K \geq 3$ and $1 < p \leq \infty$ be given. We define the N^{th} order Generalized Weighted Mean-Invariance Mean

(GWMI) of type p or resembling the L^p mean as follows:

1. If $N = K$ then $\mu_{\text{GWMI},p,\mathbf{w}_N}(.) = \mu_{p,\mathbf{w}_K}(.)$,

2. Otherwise, to find $\mu_{\text{GWMI},p,\mathbf{w}_N}(\mathbf{x})$ where $\mathbf{x} = (x_1, \ldots, x_N)$, set $\mathbf{x}_0 = \mathbf{x}$ and perform the following iteration:

$$\mathbf{x}_{k+1} = \big(\mu_{\text{GWMI},p,\mathbf{w}^1_{N-1}}(\mathbf{x}_k^{\downarrow N}), \ldots, \mu_{\text{GWMI},p,\mathbf{w}^N_{N-1}}(\mathbf{x}_k^{\downarrow 1})\big). \tag{5.58}$$

In the above the corresponding weight vector $\mathbf{w}^i_{N-1}$ is found from $\mathbf{w}_N$ according to (5.52) in Definition 5.4.2. We write (5.58) more compactly as

$$\mathbf{x}_{k+1} = \boldsymbol{\mu}_{\text{GWMI},p,\mathbf{W}_N}(\mathbf{x}_k), \tag{5.59}$$

where $\mathbf{W}_N$ is defined the same as in Definition 5.4.3. If the sequence of points $\mathbf{x}_k \in M^N$ converges to a point $\bar{\mathbf{x}} = (\bar{x}, \ldots, \bar{x}) \in M^N$, then we call $\bar{x}$ the N^{th} order GWMI mean of $\{x_i\}_{i=1}^N$ of type p with weight vector $\mathbf{w}_N = [w_1, \ldots, w_N]^T$ and set $\mu_{\text{GWMI},p,\mathbf{w}_N}(x_1, \ldots, x_N) = \bar{x}$. If the weights are equal we call the mean Generalized Mean Invariance (GMI) based mean.

We have the following existence and uniqueness theorem:

Theorem 5.4.4. For $1 < p \leq \infty$ define

$$D_p = \{(x_1, \ldots, x_N) \in M^N | \exists o \in M \text{ and } \rho < \rho_{\Delta,p}, \{x_i\}_{i=1}^N \subset \overline{B(o,\rho)}\}, \tag{5.60}$$

where $\rho_{\Delta,p}$ is defined in (3.5). Given any integer $N \geq K$ and positive weight vector

$\mathbf{w}_N$, the function $\boldsymbol{\mu}_{\mathrm{GWMI},p,\mathbf{w}_N}$ defined in Definition 5.4.3 is a primitive vectorial mean function with primitivity index of $n_s = 2$ on the domain D_p defined above. Therefore, if $\{x_i\}_{i=1}^N$ lies in a ball $\overline{B(o,\rho)}$ with $\rho < \rho_{\Delta,p}$, then the GWMI mean with weight vector $\mathbf{w}_N$ exists, is unique, belongs to $B(o,\rho)$ and to the closure of the convex hull of $\{x_i\}_{i=1}^N$. Moreover, $\mu_{\mathrm{GWMI},p,\mathbf{w}_N}(\mathbf{x})$ is a strictly convex mean function of order N with domain D_p as above.

Proof. The proof is the same as the proof of Theorem 5.4.1, except that the base case of the induction proof is $N = K$ instead of $N = 2$. □

Note that even in Euclidean space the GWMI mean of type $p \neq 2$ and the L^p mean are highly nonlinear functions of the data points and there might not be any direct relation type between them.

5.5 Discussion and Examples

Besides theoretical interest, our motivation in studying the recursive-iterative means and especially the pairwise and WMI mean is possible computational benefits over e.g., the gradient descent method for finding the L^2 mean. Recall that finding the L^p means in a general manifold requires knowing or using the exponential map of the manifold, which can be very difficult or computationally intensive. For certain homogenous manifolds such as the unit sphere S^n, the special orthogonal group $SO(n)$ and the Grassmannian manifold $G_k(n)$, if we use a suitable embedding of the manifold in a Euclidean space, then the midpoint of a geodesic between two points can be found without explicit use of the corresponding exponential maps. This fact,

which we shall explore further in Section 6.7, can be useful in computing pairwise, MI or WMI means in the mentioned manifolds. On the other hand, the MI and WMI means suffer from a curse of dimensionality in the sense that the computational load to calculate them increases exponentially with N. To see this, note that if C_N is the computation load for finding the MI mean of order N, then from 5.45 in Definition 5.4.1 we have $C_N = RNC_{N-1}$, where R is the number of iterations we perform till achieve convergence. We assume that R is fixed. Clearly, the relation between C_N and C_{N-1} will result in a prohibitive exponential dependence of C_N on N. Note that each of the N sets of iterations performed in order to find the MI and WMI means of order N can be implemented independently or in parallel which can result in faster implementation. Still this parallel implementation can become expensive for large N. Therefore, the practical use of MI and WMI or GWMI means are limited to the situation where the number of data points is not large. In contrast, the pairwise mean, which unfortunately is not permutation invariant (unless $N = 3$), is much cheaper to calculate.

5.5.1 Perimeter Shrinkage and Pairwise Mean: An Example

In our first example we apply the PS scheme to the same set of points as in the example studied in Section 4.7.3. Let $\{x_i\}_{i=1}^4$ be as in that example:

$$x_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, x_2 = \begin{bmatrix} 0.5878 \\ 0 \\ 0.8090 \end{bmatrix}, x_3 = \begin{bmatrix} 0.5 \\ 0 \\ 0.7071 \end{bmatrix}, x_4 = \begin{bmatrix} 0 \\ 0.9511 \\ 0.3090 \end{bmatrix}. \tag{5.61}$$

An important observation which we mentioned in the beginning of the section is that implementation of the PS scheme in this case is quite easy and does not require any complicated calculation. To see this note that the midpoint of the geodesic between two non-antipodal points $x_1, x_2 \in S^n$ where S^n is considered as an embedded submanifold of $\mathbb{R}^n$ is simply

$$\mu_{\frac{1}{2}}(x_1, x_2) = \frac{x_1 + x_2}{\|x_1 + x_2\|}, \tag{5.62}$$

where $\|.\|$ is the standard Euclidean norm of $\mathbb{R}^{n+1}$. This is due to the symmetry of S^n and the symmetry of the embedding.

Now, we apply the PS scheme to the set $\{x_i\}_{i=1}^N$. The points are fed to the PS scheme according to their original index order. Figure 5.4 shows the perimeter of the geodesic polygon $x_1^k x_2^k x_3^k x_4^k$ in terms of index iteration k. The bottom graph show $\log P_k$ in terms of k. This latter graph suggests that the convergence of the polygon or the the points to the pairwise mean is a linear convergence, in the sense that $P_{k+1} \leq hP_k$ where $h < 1$ is a constant. This should not be surprising since, after a while that the points become close enough, the behavior of the PS scheme in S^3 or in any manifold will become similar to its counterpart's in Euclidean space, which has this linear behavior.

Also we try to feed the PS scheme with the data points in different orders. Note that we have three distinct order possibilities to feed the PS scheme. They are 1234, 1324 and 1342. The rest of the possibilities are equivalent to one of the three mentioned. Associated to each of these three orders we have three different

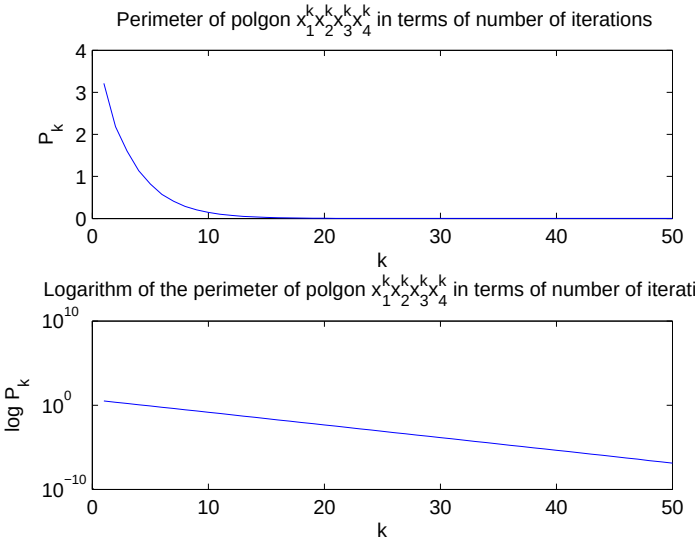


Figure 5.4: Perimeter of the polygon $x_1^k x_2^k x_3^k x_4^k$ in example in Section 5.5.1 under the PS iteration. The bottom graph shows that convergence of the polygon to a point is a linear convergence.

pairwise means μ_{1234}, μ_{1324} and μ_{1342}. It would be interesting to compare the distance between these three means[1]. We also add two other means to this comparison. First, is the equal weight L^2 mean of $\{x_i\}_{i=1}^4$ denoted by μ_{L^2}, which is found by the gradient descent algorithm. Second, is the so-called extrinsic mean μ_{ex} which is the projection of the Euclidean mean of $\{x_i\}_{i=1}^4$ as a subset of $\mathbb{R}^3$ onto S^2 and is quite easy to find:

$$\mu_{\text{ex}} = \frac{\sum_{i=1}^N x_i}{\| \sum_{i=1}^N x_i \|}, \tag{5.63}$$

whenever $\sum_{i=1}^N x_i$ is not zero. Table 5.1 shows the angular distances between all these means in Degrees. As one can see all the distances are very small. Nevertheless, on

[1]A valid question which we have not addressed is: what is the source of this permutation dependence? It seems that curvature is one factor. Also another factor is the number of data points. By knowing and understanding these factors we might be able to devise means which are simple to compute and yet their permutation dependence is under control.

$d(.,.)$	μ_{1234}	μ_{1324}	μ_{1342}	μ_{L^2}	μ_{ex}
μ_{1234}	0	0.2846	0.9622	0.4031	1.0373
μ_{1324}	0.2846	0	0.8049	0.2654	1.0720
μ_{1342}	0.9622	0.8049	0	0.5659	0.6905
μ_{L^2}	0.4031	0.2654	0.5659	0	0.8251
μ_{ex}	1.0373	1.0720	0.6905	0.8251	0

Table 5.1: The angular distances in Degrees between different means for the set $\{x_i\}_{i=1}^4$ in example of Section 5.5.1.

average the more sophisticated means μ_{1234}, μ_{1324}, μ_{1342} and μ_{L^2} are closer to each other than to μ_{ex}, in the sense that the average of the distances between the means in the group is less than the average of the distances between μ_{ex} and the means in the group.

5.5.2 Cyclic Pursuit on the Sphere

Imagine a group of N frogs dispersed uniformly on a globe. If each frog starts chasing its closest group member, what would happen after a while? Since the situation is independent of labels, we assume frog i starts following frog $i+1$ which is its closest group member. Obviously, by x_{N+1} we mean x_1. This is exactly the formulation we introduced in Section 5.3.2. We consider the discrete-time case, and assume at each unit of time frog i hops to the midpoint of the minimal geodesic between itself and from $i+1$. Theorem 5.3.2 tells us that the frogs will either all meet at a point or will converge to circular formations on great circles. We know that if they start in a convex ball, they will converge to a point; however, if they are not initially located in a strongly convex ball either of the two endings can happen. Note that by our

assumption of uniform placement on the globe, if N is large enough the chance for them of not being in a strongly convex ball is low. It seems to be difficult to give a condition to guarantee whether a given initial configuration of the group which is dispersed on entire globe will converge to a point or to great circles. However, as Theorem 5.3.2 suggests a great circle in not a stable formation. We can observe this fact in a numerical simulation. We assume there are $N = 200$ frogs in cyclic pursuit. Figure 5.5 shows the evolution of their positions and P_k the perimeter of the closed geodesic polygon $x_1^k \dots x_N^k$ in terms of the discrete-time index k. As it can be seen, after initial approach to a great circle, due to finite precision errors in computer simulations, P_k ($k \geq 18583$) becomes smaller than 2π and converges to 0, eventually. The conclusion is that unless the frogs commit no errors, they will meet each other at a point. Whether they can avoid this event depends on whether they can employ smarter rules to control their formation!

5.5.3 An Example of GMI Means on the Sphere

In this example, we find the GMI means of types $p = 1.1$ and $p = 3$, and the MI mean for the data points $\{x_i\}_{i=1}^N$ in the example of Subsection 4.7.3. The base case for the GMI means is $K = 3$. The corresponding L^p means are found by the gradient descent algorithm. Figure 5.6 shows the resulted means. For comparison we have shown the $L^{1.1}$, L^3 and L^2 means of the same data set. While it seems very difficult to give any useful quantitative measure of closeness between the L^p mean and its GMI version, this example shows that they are close. However, as mentioned before, the more important issue is their behavior with respect to the outliers. We see that

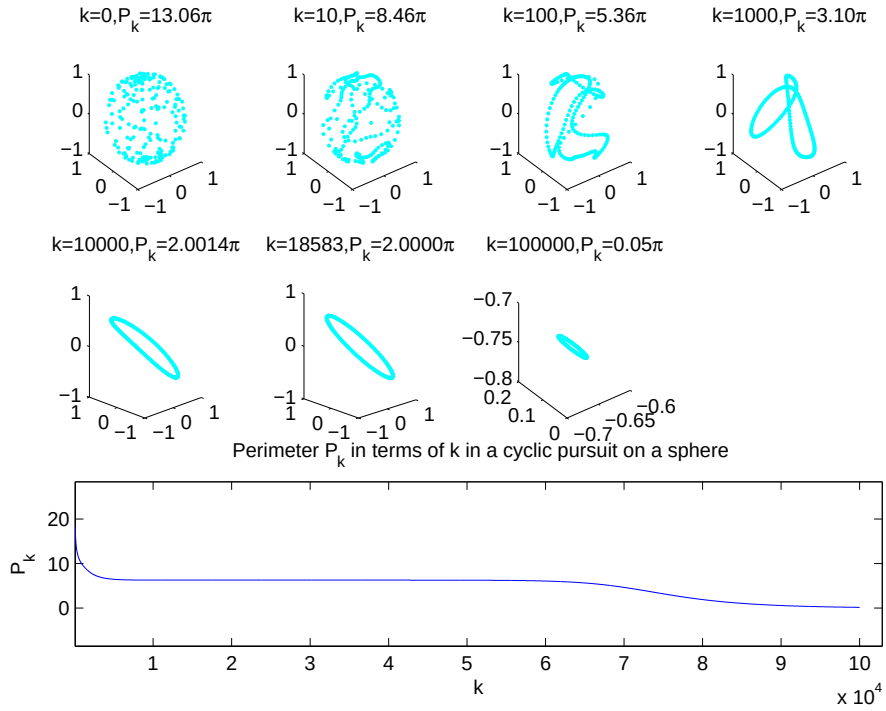


Figure 5.5: Top pictures show the evolution of the positions of a group of $N = 200$ frogs on a globe in discrete-time cyclic pursuit in terms of time k. Convergence is slow and after $k = 18584$ iterations they reach to a formation which is the closest to a great circle in perimeter. However, since this configuration is not stable, due to finite precision errors in calculations they keep getting closer to each other, and after $k = 100,000$ iteration the perimeter will be 0.05π. The bottom graph shows the perimeter P_k in terms of k, which after initial "convergence" to 2π, due to finite precision tends towards zero.

the GMI means of types $p = 1.1$ and $p = 3$ have different behaviors and resemble their L^p versions.

In the simulations for this example we observed two difficulties in computing the GMI means:

1. The first difficulty arises when the scheme is close to convergence, in this situation we need to find the L^p mean of data points which are very close to each other and this can be numerically difficult. Two reasons for this can be found. The first reason is that in such a case the formulas (4.32) and (4.31) for finding the exponential map and its inverse at a point which is close to the data points involves divisions by small numbers and also finding the arccos of small angles. The second reason is that one needs to use very small stepsize or adaptive stepsize in the gradient algorithm to make sure that the new updates are not out of the convex hull of the data points (which is a very small).

2. The second difficulty also is related to finding the L^p mean for values of p close to 1 or very large. We already mentioned the difficulty with large valued of p in Subsection 4.7.3. A look at (3.6), the formula for ∇f_p, shows that if the iterate of the gradient descent is close to one of the data points, then evaluation of ∇f_p requires again division by a small number. This adds some error in finding the L^p mean; however, since in the iterative process we find many L^p means the errors can accumulate. Again, the use of very small stepsize helps in maintaining the error in acceptable levels.

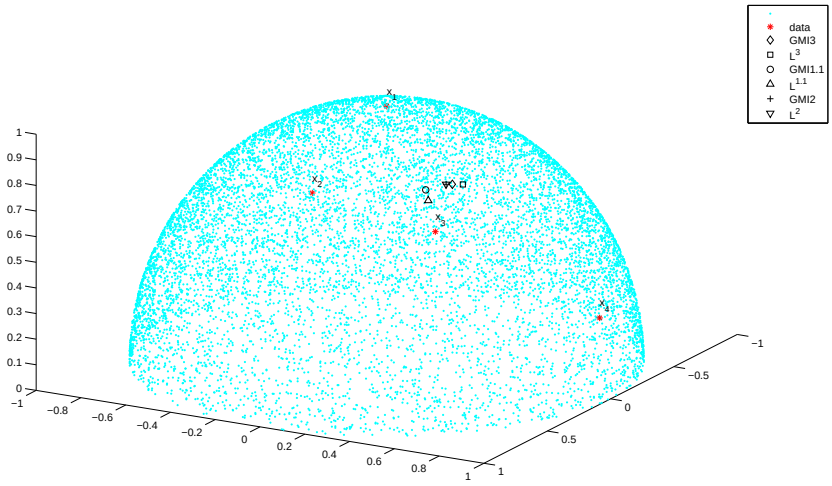


Figure 5.6: GMI mean of types $p = 1.1$ and $p = 3$ together with the MI mean for four points $\{x_i\}_{i=1}^4$ on the unit sphere are shown. For comparison the corresponding L^p means are also shown. The base case for the GMI means is $K = 3$.

Chapter 6

Results Specific to the Orthogonal Group and the Grassmannian

6.1 Introduction

Feasibility of numerical computations related to manifold-valued data depends heavily on our ability in computing the exponential map and geodesics of the underlying manifold. For an arbitrary manifold computing the exponential map can be very difficult. However, for a class of manifolds known as Riemannian symmetric spaces that can be done with relative ease. Fortunately, Riemannian symmetric spaces naturally appear in many applications, and for this reason they deserve special study. In this chapter we study the Riemannian L^p means and recursive-iterative means in two important symmetric spaces: the special orthogonal group $SO(n)$ (comprised of $n \times n$ orthogonal matrices of determinant 1), and the Grassmannian manifold $G_k(n)$ which is the space of k-dimensional subspaces of $\mathbb{R}^n$. Finding geodesics and other related calculations in these two manifolds boil down to matrix computations. We place emphasize on developing efficient methods for calculation of the pairwise and MI means in these manifolds.

6.1.1 Contributions and Outline of the Chapter

It is an interesting fact observed in [36] (see also [16, Chapter 7]) that existence and uniqueness of the local Riemannian L^2 mean could be extended to rather large non-ball domains in $SO(n)$. By large we mean relative to the diameter or volume of $SO(n)$ or in the sense of inclusion. Such a domain can be described via a Finsler distance induced by the matrix 2-norm -instead of the Riemannian norm- in the Lie algebra of $SO(n)$ (see Section (6.4)). In this chapter we try to build our results based on this observation. More concretely, we give large domains for existence and uniqueness of the means. These domains are not Riemannian balls rather Finsler balls, and are larger compared with the Riemannian balls. Not all the results we derive are completely new; however, in most occasions the already known results are derived based on new observations (e.g., Theorem 6.5.1). Two main contributions of this chapter are as follows:

1. In Theorem 6.6.1 we derive the existence and uniqueness results for pairwise, MI and WMI mean of points in Finsler 2-norm balls of appropriate size on $SO(n)$ and $G_k(n)$. Also in Theorem 6.5.1 we give the same result for local L^p means.

2. In Theorems 6.7.1 and 6.7.2 we give efficient methods to find the midpoint of the geodesic between two points in $SO(n)$ (or $G_k(n)$) and the reflection of a point with respect to another point in $SO(n)$ (or $G_k(n)$). The significance of these results is that they allow us to interpolate or extrapolate a geodesic between two points without the explicit use of matrix logarithm and exponen-

tial, both of which have high computational complexity. This will be helpful, especially in finding the pairwise and MI means (see Sections 5.3, 5.4 and 6.8) in these spaces.

We should emphasize that the results in this chapter most likely can still be strengthened. Our efforts to strengthen Theorem 6.5.1 to include the global Riemannian L^p mean as well as larger Finsler domains were not successful.

The outline of this chapter is as follows: In Section 6.2 we give the needed results about the geometry of $SO(n)$ and $G_k(n)$. In Section 6.3 we give the existence and uniqueness results for means of points in a Riemannian ball in $SO(n)$ or $G_k(n)$. In Section 6.4 we describe the Finsler balls based on the matrix 2-norm on the tangent space and study convexity of the Riemannian distance function within these strongly convex Finsler balls in $SO(n)$ and $G_k(n)$. In Section 6.5 existence and uniqueness of the local L^p means in 2-norm Finsler balls is studied. In Section 6.6 we give existence and uniqueness results for pairwise, MI and WMI means in the mentioned 2-norm balls. In Section 6.7 we derive efficient methods to find midpoint of a geodesic between two point in $SO(n)$ and $G_k(n)$. In Section 6.8 we apply the methods of Section 6.7 to implement two algorithms: the Nelder-Mead method of direct search on $SO(n)$ and the K-means clustering algorithm on $G_k(n)$.

6.2 Riemannian Geometry of $SO(n)$ and $G_p(n)$

Here, we briefly give needed results about the Riemannian geometry of $SO(n)$ and $G_k(n)$.

6.2.1 Riemannian Geometry of $SO(n)$

The special orthogonal group $SO(n)$ is a compact Lie group, and we denote its Lie algebra by $\mathfrak{so}(n)$ which is the space of $n \times n$ skew-symmetric matrices with the Lie bracket operation defined by

$$[X, Y] = XY - YX, \quad X, Y \in \mathfrak{so}(n). \tag{6.1}$$

The associated Lie algebra level adjoint operator $ad_X : \mathfrak{so}(n) \to \mathfrak{so}(n)$ is defined by

$$ad_X(Y) = [X, Y] \tag{6.2}$$

for $X, Y \in \mathfrak{so}(n)$. The corresponding group level adjoint representation $Ad_x(Y)$ is defined as

$$Ad_x(Y) = xYx^{-1} = xYx^T \tag{6.3}$$

for $x \in SO(n)$ and $Y \in \mathfrak{so}(n)$. The two operators are related to each other by the well-known relation

$$e^{ad_X}(Y) = Ad_{e^X}(Y) \tag{6.4}$$

for $X, Y \in \mathfrak{so}(n)$, where e^X is the matrix exponential of $n \times n$ matrix X. In the above $e^{ad_X} : \mathfrak{so}(n) \to \mathfrak{so}(n)$ is a linear operator defined as

$$e^{ad_X}(Y) = e^a|_{a=ad_X}(Y) = \sum_{k=0}^{\infty} \frac{ad_X^k}{k!}(Y), \tag{6.5}$$

where $ad_X^k(Y) = [X, ad_X^{k-1}(Y)]$ and $ad_X^0(Y) = Y$.

Remark 6.2.1. We also will use the inverse of the matrix exponential, i.e., the matrix logarithm. It is well-known that if a non-singular $n \times n$ matrix x has no eigenvalues on the negative real line, then its logarithm $X = \log x$ is a unique matrix such that $x = e^X$ and all the eigenvalues of X have imaginary parts between $-\pi$ and π. Note that a skew-symmetric matrix has pure imaginary eigenvalues; hence, if $x \in SO(n)$ has no eigenvalues equal to -1, then there exists a unique $X \in \mathfrak{so}(n)$ such that all the eigenvalues of X are smaller than π in absolute value and that $x = e^X$. See Theorem 6.2.2 which gives the injectivity radius of $SO(n)$ in the standard Riemannian metric.

For $x \in SO(n)$ the left translation $L_x : SO(n) \to SO(n)$ is defined by $L_x(y) = xy$ for $y \in SO(n)$. The right translation can be defined similarly. A tangent vector $Y \in \mathfrak{so}(n)$ can be left translated to $T_xSO(n)$ via the derivative of L_x. This way, by left or right translations, corresponding to any $Y \in \mathfrak{so}(n)$, we can build left or right invariant vector fields Y_L or Y_R. Therefore, we can define the Lie bracket operation for left or right invariant vector fields based on the definition of Lie bracket on $\mathfrak{so}(n)$. This gives an algebraic-differential picture of $SO(n)$. There is a natural way to equip $SO(n)$ and more generally any compact Lie group with a Riemannian structure which matches the group structure, in the sense that both left and right translations are isometries. This follows from the fact that there always exists a Riemannian structure g which is invariant under both left and right translations, i.e., it is bi-invariant. We do not mention the details and only mention the following

theorem [46, 33, 40]:

Theorem 6.2.1. There exists a bi-invariant Riemannian metric g on $SO(n)$ for which

1. $SO(n)$ is a complete Riemannian manifold,

2. The geodesics through I, the identity of $SO(n)$, coincide with the one-parameter subgroups of $SO(n)$, i.e, all such geodesics are of the form e^{tX} where $X \in \mathfrak{so}(n)$. Hence, all geodesics coincide with the one-parameter subgroups and their translations. In this metric inversion is an isometric reflection with respect to I.

3. If ∇ is the Levi-Civita connection for g and R is its corresponding curvature tensor at identity then:

   (a) $\nabla_X Y = \frac{1}{2}[X, Y], \quad X, Y \in \mathfrak{so}(n)$

   (b) $R(X, Y)Z = -\frac{1}{4}[[X, Y], Z], \quad X, Y, Z \in \mathfrak{so}(n)$

   (c) $\langle R(X, Y)Z, U\rangle_I = -\frac{1}{4}\langle [X, Y], [Z, U]\rangle, \quad X, Y, Z, U \in \mathfrak{so}(n)$

The most common such a Riemannian metric is constructed from the inner product induced by the the so-called Killing form on $\mathfrak{so}(n)$. We avoid the details and give the important facts which we need. The mentioned inner product on $\mathfrak{so}(n)$ after a scaling will be

$$\langle X, Y\rangle_I = g_I(X, Y) = \frac{1}{2}\mathrm{tr}(X^T Y), \quad X, Y \in \mathfrak{so}(n). \tag{6.6}$$

For left-invariant vector fields X_L and Y_L we have

$$g_x(X_L(x), Y_L(x)) = g_I(x^{-1}X_L(x), x^{-1}Y_L(x)). \tag{6.7}$$

We call this metric the standard Riemannian metric of $SO(n)$. Note that at $\mathfrak{so}(n)$ we have $g_I(X, X) = \frac{1}{2}\|X\|_F^2$, where $\|.\|_F$ denotes the Frobenius norm. It can be shown that for $X, Y \in \mathfrak{so}(n)$ [13]

$$0 \leq \|[X, Y]\|_F \leq \|X\|_F \|Y\|_F. \tag{6.8}$$

From this and item 3c in Theorem 6.2.1 we conclude that the sectional curvature of $SO(n)$ in the standard Riemannian metric satisfies $0 \leq K \leq \frac{1}{2}$. Also it is quite easy to establish the following Theorem, using the properties of matrix exponential and skew-symmetric matrices:

Theorem 6.2.2. we have the following:

1. The injectivity radius of $SO(n)$ in the standard Riemannian metric is π.

2. Let $n' = [\frac{n}{2}]$, where $[a]$ is the smallest integer not larger than $a \in \mathbb{R}$. Closed simple geodesics in $SO(n)$ are of length $2\sqrt{l_1^2 + \ldots + l_{n'}^2}\pi$, where l_i's are non-negative integers with no common factor other than 1. A shortest (nontrivial) closed geodesic is of length 2π. A longest simple geodesic is of length $2\sqrt{n'}\pi$. Hence, the Riemannian diameter of $SO(n)$ is $\sqrt{n'}\pi$.

3. In addition, for even n the farthest point from the identity I is $-I$ and for

odd n the farthest points from I are of the form

$$\hat{I} = -I + 2VV^T, \tag{6.9}$$

where $v \in \mathbb{R}^n$ and $\|V\| = 1$. ($\|.\|$ is the standard norm of $\mathbb{R}^n$.)

6.2.2 Riemannian Geometry of $G_k(n)$

The following materials can be found in many references, but we mostly follow [26, 91, 46, 33]. The Grassmannian $G_k(n)$ is defined as the manifold of k-dimensional subspaces of $\mathbb{R}^n$, where we assume $k \leq \left[\frac{n}{2}\right]$. Here, $[a]$ is the integer part of real number a. For $k = 1$, $G_k(n)$ is the real projective plant RP^{n-1}, which is found by identifying the antipodal points on the unit sphere S^{n-1}. The more complicated and interesting case is when k is larger than 1; and henceforth we assume $k \geq 2$. $G_k(n)$ is a smooth manifold which can be endowed with a Riemannian structure. There is a natural Riemannian structure which is closely related to the quotient topology associated with $G_k(n)$. We briefly mention these facts. Consider a subgroup of $SO(n)$ defined as

$$K_k(n) = \{\left[\begin{smallmatrix} o_1 & \mathbf{0} \\ \mathbf{0} & o_2 \end{smallmatrix}\right] | o_1 \in O(k), o_2 \in O(n-k), \det(o_1 o_2) = 1\}. \tag{6.10}$$

Here, $O(n)$ denotes the Lie group of orthogonal matrices. $K_k(n)$ is also denoted by $S(O(k) \times O(n-k))$. Now let $q = [U_{n\times k} | U^{\perp}_{n\times(n-k)}]$ denote an element of $SO(n)$. The first k-columns of q denoted by U represent a k-dimensional subspace of $\mathbb{R}^n$, and

the rest of the columns denoted by $U^{\perp}$ represent the orthogonal complement of that subspace. Consider an action of $K_k(n)$ on $SO(n)$ which is defined by multiplication from the right of $q \in SO(n)$. $G_k(n)$ is identified with the set of orbits of this action which coincides with the left coset of the subgroup $K_k(n)$. Moreover, $G_k(n)$ can be considered as a Riemannian manifold whose metric is derived from the standard Riemannian metric of $SO(n)$. Each element $q \in SO(n)$ can be associated to an element $\{qK_k(n)\} \in G_k(n)$ which consists of the orbit of Q under the action of $K_k(n)$:

$$\{qK_k(n)\} = \{qo | o \in K_k(n)\}. \tag{6.11}$$

Note that this is the collection of elements of $SO(n)$ whose first k columns span the same subspace as the first k columns of q. The tangent space to $SO(n)$ at the identity will be decomposed to two orthogonal subspaces according to Cartan decomposition: $\mathfrak{so}(n) = \mathfrak{g}_k(n) \oplus \mathfrak{k}_k(n)$, where

$$\mathfrak{k}_k(n) = \{ \left[\begin{smallmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{smallmatrix} \right] | A \in \mathfrak{so}(k), B \in \mathfrak{so}(n-k) \} \tag{6.12}$$

and

$$\mathfrak{g}_k(n) = \{ \left[\begin{smallmatrix} \mathbf{0} & -A^T \\ A & \mathbf{0} \end{smallmatrix} \right] | A \in \mathbb{R}^{(n-k)\times k} \}. \tag{6.13}$$

$\mathfrak{g}_k(n)$ and $\mathfrak{k}_k(n)$ are called horizontal and vertical spaces, respectively. At any other point $q \in SO(n)$ we have the corresponding decomposition via left translation. Note that $\mathfrak{g}_k(n)$ is identified with the tangent space of $G_k(n)$ at $\{IK_k(n)\}$. Therefore, if $X_i = \left[\begin{smallmatrix} \mathbf{0} & -\bar{X}_i^T \\ \bar{X}_i & \mathbf{0} \end{smallmatrix} \right]$ $(i = 1, 2)$ are two tangent vectors at $\{IK_k(n)\}$, then the Riemannian

metric is nothing but

$$\langle X_1, X_2 \rangle_{\{IK_k(n)\}} = g_I(X_1, X_2) = -\frac{1}{2}\mathrm{tr}(X_1 X_2) = \mathrm{tr}(\bar{X}_1^T \bar{X}_2), \tag{6.14}$$

where g is the standard Riemannian metric of $SO(n)$ (6.6). We denote the corresponding norm at $\mathfrak{g}_k(n)$ by $\|.\|$, as before. A consequence of the above relation is that the geodesic in $G_k(n)$ along $X \in \mathfrak{g}_k(n)$ is noting but the equivalent class of $\{e^{tX} K_k(n)\}$, where e^A is the standard matrix exponential of matrix $A \in \mathbb{R}^{n \times n}$. Let $\bar{X} = U_{(n-k)\times(n-k)} \left[\begin{smallmatrix} \Sigma_{k\times k} \\ 0 \end{smallmatrix} \right] V_{k\times k}^T$ be a Singular Value Decomposition (SVD) of $\bar{X}_{(n-k)\times k}$. We can write $U = \left[\begin{smallmatrix} U_1, U_2 \end{smallmatrix} \right]$, where U_1 is $(n-k) \times k$ and U_2 is $(n-k) \times (n-2k)$. Then $X \in \mathfrak{g}_k(n)$ can be written as:

$$X = \left[\begin{smallmatrix} \mathbf{0} & -\bar{X}^T \\ \bar{X} & \mathbf{0} \end{smallmatrix} \right] = \left[\begin{smallmatrix} V & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & U_1 & U_2 \end{smallmatrix} \right] \begin{bmatrix} \mathbf{0} & -\Sigma & \mathbf{0} \\ \Sigma & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \left[\begin{smallmatrix} V^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & U_1 & U_2 \end{smallmatrix} \right]^T. \tag{6.15}$$

Assume that $\Sigma = \mathrm{diag}(\theta_1, \ldots, \theta_k)$, where $\theta_1 \geq \ldots \geq \theta_k \geq 0$. Here, $\mathrm{diag}(a)$ is a diagonal matrix whose diagonal is the vector a. Note that $\|X\| = (\sum \theta_i^2)^{\frac{1}{2}}$. Now, e^{tX} can be computed easily as

$$e^{tX} = \left[\begin{smallmatrix} V & \mathbf{0} \\ \mathbf{0} & U \end{smallmatrix} \right] \begin{bmatrix} \mathbf{C}(t\theta_1,\ldots,t\theta_k) & -\mathbf{S}(t\theta_1,\ldots,t\theta_k) & \mathbf{0} \\ \mathbf{S}(t\theta_1,\ldots,t\theta_k) & \mathbf{C}(t\theta_1,\ldots,t\theta_k) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_{(n-2k)\times(n-2k)} \end{bmatrix} \left[\begin{smallmatrix} V & \mathbf{0} \\ \mathbf{0} & U \end{smallmatrix} \right]^T, \tag{6.16}$$

where

$$\mathbf{C}(\theta_1, \ldots, \theta_k) = \mathrm{diag}\big(\cos\theta_1, \ldots, \cos\theta_k\big) \text{ and } \mathbf{S}(\theta_1, \ldots, \theta_k) = \mathrm{diag}\big(\sin\theta_1, \ldots, \sin\theta_k\big). \tag{6.17}$$

The decomposition in (6.16) is known as the CS decomposition of the orthogonal matrix e^{tX} [31, 26]. An element $s \in G_k(n)$ can be represented by a subspace that is a member of $\{e^{tX}K_k(n)\}$ as

$$S(tX) = \left[\begin{smallmatrix} V & \mathbf{0} \\ \mathbf{0} & U \end{smallmatrix}\right] \begin{bmatrix} \mathbf{C}(t\theta_1,\ldots,t\theta_k) \\ \mathbf{S}(t\theta_1,\ldots,t\theta_k) \\ \mathbf{0}_{(n-2k)\times k} \end{bmatrix} \tag{6.18}$$

We call this the standard subspace representation of $s = \{e^{tX}K_k(n)\} \in G_k(n)$. For any $v \in SO(k)$, $S(tX)v$ represents the same subspace, but it is in the standard form anymore. Note that $S(tX)$ is found from (6.16) by dropping the right most factor and two columns of the matrix in middle. Also note that the geodesics along $-X$ (or X^T) is given by $\{(e^{tX})^T K_k(n)\}$ which can be found simply by a transposition operation. Therefore, a subspace S' which corresponds to reflection of S with respect to $\{IK_k(n)\}$ is simply

$$S'(tX) = S(-tX) = \left[\begin{smallmatrix} V & \mathbf{0} \\ \mathbf{0} & U \end{smallmatrix}\right] \begin{bmatrix} \mathbf{C}(t\theta_1,\ldots,t\theta_k) \\ -\mathbf{S}(t\theta_1,\ldots,t\theta_k) \\ \mathbf{0}_{(n-2k)\times k} \end{bmatrix}. \tag{6.19}$$

Note that S' can be found by replacing $t\theta_i$ with $-t\theta_i$. The Riemannian distance between two elements of $G_k(n)$ can be found from their subspace representations (standard or non-standard) easily: If $s_1, s_2 \in G_k(n)$ have subspace representations S_1 and S_2, respectively, let $S_1^T S_2 = UCV^T$ be an SVD of $S_1^T S_2$, then $C = \mathrm{diag}(\cos\theta_1, \ldots, \cos\theta_k)$, where $0 \leq \theta_1, \ldots, \theta_k \leq \frac{\pi}{2}$ are the principal angles between

s_1 and s_2 and the Riemannian distance between s_1 and s_2 is

$$d(s_1, s_2) = (\sum_{i=1}^{k} \theta_i^2)^{\frac{1}{2}}. \tag{6.20}$$

Theorem 6.2.3. We have the following

1. The injectivity radius of $G_k(n)$ in the standard Riemannian metric is $\text{inj}G_k(n) = \frac{\pi}{2}$.

2. Closed geodesics in $G_k(n)$ are of length $\sqrt{l_1^2 + \ldots + l_k^2}\pi$, where $l_1, \ldots, l_k$ are nonnegative integers with no common factor other than 1. A shortest closed geodesic is of length π and a longest is of length $\sqrt{k}\pi$. Hence, the distance between two farthest points in $G_k(n)$ (or equivalently the Riemannian diameter of $G_k(n)$) is $\frac{\sqrt{k}}{2}\pi$.

3. Let $R : \mathfrak{g}_k(n) \times \mathfrak{g}_k(n) \times \mathfrak{g}_k(n) \to \mathfrak{g}_k(n)$ denote curvature tensor of $G_k(n)$ at $\{IK_k(n)\}$. We have

$$R(X, Y)Z = -[[X, Y], Z], \quad X, Y, Z \in \mathfrak{g}_k(n), \tag{6.21}$$

where $[.,.]$ is the matrix Lie bracket. Therefore, the sectional curvature along two orthonormal tangent vectors $X, Y \in \mathfrak{g}_k(n)$ is $-\text{tr}([[X, Y], Y]X) = \|[X, Y]\|^2$, where $\|.\|$ is the standard norm in $\mathfrak{so}(n)$. From this, one can conclude that the sectional curvature of $G_k(n)$ satisfies $0 \leq K \leq 2$.

Remark 6.2.2. In Theorem 5.3.2, a condition on the finiteness of the number of

possible values for the lengths of closed geodesic is mentioned. From Theorems 6.2.2 and 6.2.3, this condition is satisfied in $SO(n)$ and $G_k(n)$.

6.2.2.1 Isometric Identification of $G_k(n)$ with the Space of Orthogonal Projections Matrices of Rank k

It is obvious that to any k-dimensional subspace of $\mathbb{R}^n$ we can associate, in a unique way, an $n \times n$ orthogonal projection matrix. Such a matrix has the form SS^T, where $S \in \mathbb{R}^{n\times k}$ is a matrix comprised of an orthogonal basis of that subspace. Let us denote the linear space of $n \times n$ symmetric matrices by $\mathcal{S}_n$ and its subset of orthogonal projection matrices of rank k by $\mathcal{P}_{k,n}$. $\mathcal{P}_{k,n}$ can be isometrically identified with $G_k(n)$ as Riemannian manifolds via an isometric map [89]. This map, simply maps a tangent vector at $\{IK_k(n)\}$ which can be represented as $X = \left[\begin{smallmatrix} \mathbf{0} & -\bar{X}^T \\ \bar{X} & \mathbf{0} \end{smallmatrix}\right]$ to a tangent vector $\tilde{X} = \left[\begin{smallmatrix} \mathbf{0} & \bar{X}^T \\ \bar{X} & \mathbf{0} \end{smallmatrix}\right]$ in the tangent space to $\mathcal{P}_{k,n}$ at $I_k = \left[\begin{smallmatrix} I_{k\times k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{smallmatrix}\right]$. This can be easily extended to other points in $G_k(n)$ via rotating a subspace in $G_k(n)$ and congruence transformation in $\mathcal{P}_{k,n}$. The actual formula for the geodesic in this space can be found e.g., in [41].

6.3 Existence and Uniqueness of L^p and MI Means in Riemannian Balls

The following is a direct consequence of what preceded in Section 6.2:

Theorem 6.3.1. Theorems 3.2.1, 3.2.2, 5.4.1 and 5.4.4 hold true in $M = SO(n)$ equipped with the standard Riemannian metric in which case we have $\Delta = \frac{1}{2}$ and

$\mathrm{inj}M = \pi$, and hence (see (3.5))

$$\rho_{\Delta,p} = \begin{cases} \frac{\sqrt{2}}{4}\pi & 1 \leq p < 2 \\ \\ \frac{\pi}{2}(= r_{\mathrm{cx}}) & 2 \leq p \leq \infty. \end{cases} \tag{6.22}$$

The same theorems hold for $M = G_k(n)(2 \leq k \leq [\frac{n}{2}])$ where have $\Delta = 2$ and $\mathrm{inj}M = \frac{\pi}{2}$, and accordingly

$$\rho_{\Delta,p} = \begin{cases} \frac{\sqrt{2}}{8}\pi & 1 \leq p < 2 \\ \\ \frac{\pi}{4}(= r_{\mathrm{cx}}) & 2 \leq p \leq \infty. \end{cases} \tag{6.23}$$

The main shortcoming of the above theorem is that the balls of radius $\rho < \rho_{\Delta,p}$ become increasingly small compared with the diameter of $SO(n)$ or $G_k(n)$ as n increases (see Theorems 6.2.2 and 6.2.3). Note that $\rho_{\Delta,p}$ remains constant as n increases, while the diameters of $SO(n)$ and $G_k(n)$ increase by n. The crucial observation made in [36] is that one can find larger non-ball strongly convex domains in a compact Lie group in which the local center of mass is unique. This domain is easily described in terms of the matrix 2-norm in the Lie algebra $\mathfrak{so}(n)$ (rather than the norm (6.6) related to the standard Riemannian metric). We shall elaborate on this issue next.

6.4 Large Non-ball Convex Domains in $SO(n)$ and $G_k(n)$ and a Finsler Structure

For our purposes a Finsler manifold is defined as follows [72]:

Definition 6.4.1. A Finsler metric F on a smooth manifold M assigns a norm $F(x,.)$ to each tangent space T_xM in a smooth manner as x varies. The pair (M,F) is called a Finsler manifold.

Remark 6.4.1. A more restricted definition requires a convexity condition on the norm, i.e., the norm must be a Minkowski norm[1]. The main consequence of this restriction is that the geodesics are well behaved in the sense of uniqueness and smoothness. An example clears the issue [4]. In $\mathbb{R}^2$ the standard Euclidean norm has only the straight lines as its length minimizing curves between two points. However, the L^∞ norm which is defined by $\|\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\|_\infty = \max\{|x|,|y|\}$ in addition to straight lines has some other merely piece-wise differentiable length minimizing curves too. The same holds for the L^1 norm defined as $\|\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\|_1 = |x| + |y|$. Neither of these two norms are Minkowski norms. Also note that in these two norms the unit sphere is not smooth, in contrast with the unit spheres in the L^p with $1 < p < \infty$. For our purposes the issue of uniqueness of Finsler geodesics are not important; hence, in what follows we shall use a norm which is not Minkowski.

[1]A Minkowski norm is defined as follows: Let V be a finite dimensional vector space. A function on V is called a Minkowski norm if:

1. $F(u) \geq 0$ for every $u \in V$ and $F(u) = 0$ if and only if $u = 0$,
2. $F(\lambda u) = |\lambda| u$ for every $u \in V$ and $\lambda \in \mathbb{R}$ and
3. F is smooth on $V - \{0\}$ and its Hessian at every point y is positive definite, i.e., the symmetric bi-linear form:

$$F_y(u,v) = \frac{1}{2}\frac{\partial^2}{\partial s \partial t} F^2(y + su + tv)|_{s=t=0} \tag{6.24}$$

is positive definite.

6.4.1 The Matrix 2-Norm as a Finsler Norm

The matrix induced 2-norm of $A \in \mathbb{R}^{n\times n}$ is defined as:

$$\|A\|_2 = \max_{\|u\|_2=1} \|Au\| \tag{6.25}$$

where $\|.\|$ is the standard Euclidean norm in $\mathbb{R}^n$. $\|A\|_2$ is the largest singular value of A and for a normal matrix (hence a skew-symmetric) it is the largest eigenvalue of A in absolute value. Note that this norm is bi-invariant with respect to left and right multiplication by orthogonal matrices, i.e., $\|xAy\|_2 = \|A\|_2$ for $x, y \in O(n)$. We equip $\mathfrak{so}(n)$ with this norm, extend it to any other tangent space via bi-invariance; and we make $SO(n)$ into a Finsler manifold. Now the length of a piecewise smooth curve $c : [0, 1] \to SO(n)$ is

$$\mathcal{L}(c) = \int_{t=0}^{1} \|\frac{\mathrm{d}c}{\mathrm{d}t}\|_2 \mathrm{d}t. \tag{6.26}$$

The 2-norm Finsler distance between two points $x, y \in SO(n)$ is defined as the infimum of the length of all piecewise smooth curves in $SO(n)$ connecting x and y. We denote the 2-norm distance between x and y by $d_2(x, y)$. A curve between x and y is a length minimizing curve or minimal if its length is equal to distance between x and y. We refrain from using the term geodesic hence defining them as stationary points of the length functional requires defining a connection which requires further preparation.

Remark 6.4.2. For $n = 2, 3$ and $A \in \mathfrak{so}(n)$ we have $\|A\|_2 = \frac{1}{\sqrt{2}}\|A\|_F$. Therefore, $\|.\|_2$ is a Minkowski norm in this case. However, it is easy to see that $\|.\|_2$ is not a

Minkowski norm for $n \geq 4$. For example, for $n = 4$, consider three $\mathfrak{so}(n)$ elements Y, U and V. Let $Y = Q\Lambda Q^H$ with $\Lambda = \text{diag}(+\mathbf{i}, -\mathbf{i}, 0, 0)$, where $\mathbf{i} = \sqrt{-1}$, be the eigen decomposition of Y. Here Q^H is the conjugate transpose of the unitary matrix Q. Set $U = QDQ^H$ with $D = \text{diag}(0, 0, +\mathbf{i}, -\mathbf{i})$ and $V = U$. For small enough $|s|$ and $|t|$ $(s, t \in \mathbb{R})$ we have $\|Y + sU + tV\|_2 = \|Y\|_2 = 1$. As a result the bi-linear form in (6.24) is zero with $U = V \neq 0$; and it cannot be positive definite. This automatically extends to every $n > 4$. Therefore, we expect to have length minimizing curves with cranks in $SO(n)$ $(n \geq 4)$ as a Finsler manifold. In fact, consider this curve in $\mathfrak{so}(4)$:

$$\Omega(t) = \begin{cases} tY + tU & 0 \leq t \leq \frac{1}{2} \\ tY + (1-t)U & \frac{1}{2} \leq t \leq 1, \end{cases} \tag{6.27}$$

where Y and U are defined as before. Note that $\|\dot{\Omega}(t)\|_2 = \|Y\|_2$. The piecewise linear curve $\Omega(t)$ connects the origin to Y and its length is $\|Y\|_2$, so it is a length minimizing curve. Obviously, there is the straight line segment $\Psi(t) = tY$ with $0 \leq t \leq 1$ which does the same, i.e, its length is equal to the distance between Y and the origin. Theorem 6.4.1 tells us that the curve e^{tY} is a length minimizing curve in $SO(4)$ connecting I to e^Y; and since the curve $\gamma(t) = e^{\Omega(t)}$ also has the same 2-norm length it too is a length minimizing curve in $SO(4)$. Therefore, in $(SO(4), \|.\|_2)$ the length minimizing curves are not unique, even locally, which is in contrast with the Riemannian case- where the length minimizing curves are locally unique and they are the geodesics.

Remark 6.4.3. The 2-norm is not an admissible norm with respect to the Lie bracket operation. A norm $\|.\|$ on a Lie algebra $\mathfrak{g}$ is admissible if $\|[A,B]\| \leq \|A\|\|B\|$ for $A, B \in \mathfrak{g}$ [60]. As we mentioned before $\|.\|_F$ is an admissible norm on $\mathfrak{so}(n)$. Any induced norm like the 2-norm, can be converted to an admissible norm via scaling. For example, if we define $\|A\|_{2'} = 2\|A\|_2$, then from the triangle inequality and since $\|AB\|_2 \leq \|A\|_2\|B\|_2$ we have $\|[A,B]\|_{2'} \leq \|A\|_{2'}\|B\|_{2'}$ for all $A, B \in \mathfrak{so}(\mathfrak{n})$. However, the unit ball of $\|.\|_{2'}$ is contained in the unit ball of $\|.\|_2$. In this sense we have shrunk the unit ball. Some of our results here, are essentially derived in [36] in the context of a general compact Lie group and with an admissible norm. Applying those results to $SO(n)$ with the $\|.\|_{2'}$ norm, results in too conservative results about the radius of largest convex balls in $SO(n)$.

The following theorem shows that curves of the form e^{tX} and their translates minimize the 2 Finsler distance in addition to the Riemannian distance. Therefore, the Finsler structure is quite simple. The proof of the theorem can be found in [16, Chapter 7] (cf. [36]):

Theorem 6.4.1. Let $X, Y \in \mathfrak{so}(n)$ with $\|X\|_2 + \|Y\|_2 < \pi$. Then there exists a unique $H \in \mathfrak{so}(n)$ with $\|H\|_2 < \pi$ such that $h = e^H = e^X e^Y$. Moreover, $\|H\|_2 \leq \|X\|_2 + \|Y\|_2$. Therefore, the 1-parameter subgroups of $SO(n)$ and their translates locally minimize the 2-norm Finsler distances. Here, locally means on 2-norm lengths not larger than π.

The following corollaries are immediate:

Corollary 6.4.1. If X, Y are as in Theorem 6.4.1 and $x = e^X$ and $y = e^Y$, then

$$d_2(x, y) = \| \log x^T y \|_2 \leq \|X\|_2 + \|Y\|_2. \tag{6.28}$$

Corollary 6.4.2. The tangent space $\mathfrak{g}_k(n)$ can be endowed with the Finsler 2-norm induced from $\mathfrak{so}(n)$. Hence, $G_k(n)$ will be a Finsler manifold in which horizontal geodesics and their translates locally minimize the 2-norm Finsler distance. Here, locally means on 2-norm lengths not larger than $\frac{\pi}{2}$. We denote the 2-norm distance between $x, y \in G_k(n)$ by $d_2(x, y)$.

6.4.1.1 Metric Decreasing Property of the Matrix Exponential in $SO(n)$

The exponential map in a Riemannian manifold of nonnegative curvature has a metric decreasing property. This means that the distances between two points in the tangent spaces is not smaller than the distance between their images under the exponential on the manifold. One can show that on $SO(n)$ this holds for any bi-invariant Finsler distance. This fact was observed in [36] for admissible bi-invariant norms (see also [16, Chapter 7]). Here, we give a rather different proof, where we do not use the admissibility assumption. First, we give a short preparation about the derivative of the matrix exponential (see [45] or [38]). Let $f(t) = e^{A(t)}$ where $A : [0, 1] \to \mathfrak{so}(n)$ is a curve in $\mathfrak{so}(n)$. Then the derivative of $f(t)$ is computed as

$$\frac{\mathrm{d}}{\mathrm{d}t} f(t) = \mathrm{dexp}_{A(t)}(\dot{A}(t)) f(t), \tag{6.29}$$

where $\dot{A}(t) = \frac{\mathrm{d}A(t)}{\mathrm{d}t}$, and $\mathrm{dexp}_X : \mathfrak{so}(n) \to \mathfrak{so}(n)$ is defined as

$$\mathrm{dexp}_X(Y) = \frac{e^a - 1}{a}|_{ad_X}(Y) \tag{6.30}$$

for any $Y \in \mathfrak{so}(n)$. Here, by $\frac{e^a-1}{a}|_{ad_X}(Y)$ we mean that in the Laurant expansion of $u \mapsto \frac{e^a-1}{a}$, $a \in \mathbb{C}$ is replaced by ad_X and then applied to Y (see (6.2)).

Theorem 6.4.2. Let $\|.\|$ be any bi-invariant norm in $\mathfrak{so}(n)$ and d its induced distance function in $SO(n)$ [1]. We have

$$d(e^X, e^Y) \leq \|X - Y\|, \quad \forall X, Y \in \mathfrak{so}(n). \tag{6.31}$$

Proof. Consider the function $f : [0,1] \times [0,1] \to SO(n)$ with $f(t,s) = e^{t(X+s(Y-X))}$. Note that $f(1,0) = e^X$ and $f(1,1) = e^Y$, i.e., $s \mapsto f(1,s)$ is a curve connecting e^X to e^Y, therefore:

$$d(e^X, e^Y) \leq \int_0^1 \|\frac{\partial f}{\partial s}\|\mathrm{d}s. \tag{6.32}$$

But from (6.29) we have

$$\frac{\partial f}{\partial s} = \mathrm{dexp}_{t(X+s(Y-X))}t(Y - X)f(t,s). \tag{6.33}$$

From bi-invariance of $\|.\|$ we have $\|\frac{\partial f}{\partial s}\| = \|\mathrm{dexp}_{t(X+s(Y-X))}t(Y - X)\|$. Now, we

[1] Here, d can be either the Riemannian distance function or the 2-norm Finsler distance.

calculate $\frac{\partial g}{\partial t}$ where $g(t,s) = \mathrm{dexp}_{t(X+s(Y-X))}(t(Y-X))$. Note that

$$g(t,s) = \sum_{i=1}^{\infty} \frac{t^i}{i!} ad^{i-1}_{X+s(Y-X)}(Y-X). \tag{6.34}$$

Hence

$$\begin{aligned} \frac{\partial g}{\partial t} &= \sum_{i=0}^{\infty} \frac{t^i}{i!} ad^i_u|_{u=X+s(Y-X)}(Y-X) = e^{ad_{t(X+s(Y-X))}}(Y-X) \\ &= e^{t(X+s(Y-X))}(Y-X)e^{-t(X+s(Y-X))}. \end{aligned} \tag{6.35}$$

Therefore, $\|\frac{\partial g}{\partial t}\| = \|Y-X\|$ and this immediately implies that $d(e^X, e^Y) \leq \|Y-X\|$.

□

6.4.2 Hessian of the Distance Function in Symmetric Spaces

Both $SO(n)$ and $G_k(n)$ are symmetric spaces. One consequence is that the eigenvalues of the curvature operator do not change along a geodesic, and the Jacobi equation becomes a constant coefficient equation which can be solved explicitly. In the sequel assume M is a symmetric Riemannian manifold of dimension m. Let $c : [0, d(q,x)] \to M$ be a unit speed geodesic connecting q to x. The curvature operator $R_{c(t)} : T_{c(t)} \to T_{c(t)}$ defined by

$$X \mapsto R_{c(t)}(X, c'(t))c'(t) \tag{6.36}$$

is a self-adjoint operator which has a set of m real eigenvalues $\kappa_1(t) \geq \ldots \geq \kappa_m(t)$. Here, $c'(t)$ is the velocity vector of $c(t)$. Due to symmetry of M the eigenvalues

remain constant in time and due to both the symmetry and compactness of M they are all nonnegative, therefore $\kappa_i(t) = \kappa_i \geq 0$ [46]. Denote a set of orthonormal eigenvectors of $R_{c(0)}$ by $\{V_i(0)\}_{i=1}^n$. For every i by parallel translating the vector $V_i(0)$ along $c(t)$, we will have a parallel vector $V_i(t)$ which for each t is an eigenvector of $R_{c(t)}$ associated with κ_i. Hence, $\{V_i(t)\}_{i=1}^m$ will be an orthonormal frame along $c(t)$. Any Jacobi field $J(t)$ along $c(t)$ with initial condition $J(0) = 0$ can be written as $J(t) = \sum_{i=1}^m A_i J_i(t)$, where $J_i(t)$ satisfies $J_i''(t) + \kappa_i J(t) = 0$ and $\|J_i'(0)\| = 1$. We can solve explicitly for J_i as $J_i(t) = \mathrm{sn}_{\kappa_i}(t)V_i(t)$, where $\mathrm{sn}_\kappa(t)$ is defined in (2.6). Now, recalling formula (2.4) for the Hessian of $x \mapsto d(q,x)$ we have $\mathrm{Hess}(d(q,x))(J_i(t), J_j(t)) = \langle J_i(t)^{\perp'}, J_j^{\perp}(t)\rangle \delta_{ij}$, where δ_{ij} is Kronecker's delta. Note that from this we have $\mathrm{Hess}(d(q,x))(J_m(t), J_m(t)) = 0$, since $J_m(t)$ has no normal component along $c(t)$. From that formula we have:

$$\mathrm{Hess}(d(q,x))(V_i(d(q,x)), V_j(d(q,x))) = \begin{cases} 0 & i \text{ or } j = m \\ \frac{\mathrm{cs}_{\kappa_i}(d(q,x))}{\mathrm{sn}_{\kappa_i}(d(q,x))}\delta_{ij} & 1 \leq i,j < m. \end{cases} \tag{6.37}$$

Note that since $\kappa_i \geq 0$, we have $\frac{\mathrm{cs}_{\kappa_i}(d(q,x))}{\mathrm{sn}_{\kappa_i}(d(q,x))} = \sqrt{\kappa_i}\cot(\sqrt{\kappa_i}d(q,x))$, except when $\kappa_i = 0$ in which case the expression is equal to $\frac{1}{d(q,x)}$. It is obvious that the largest eigenvalue of the curvature operator, i.e., κ_1 is what controls the sign of the Hessian. For $SO(n)$ and $G_k(n)$ it is possible to find the κ_i's more explicitly and identify the conditions to guarantee that $x \mapsto d(q,x)$ is positive-definite.

6.4.2.1 Eigenvalues of the Curvature Operator in $SO(n)$ and $G_k(n)$

There are different ways to derive the results we are going to present. One short way, which we shall follow, is to use the notion of reduced commutator introduced in [13]. Let $\{\eta_i\}_{i=1}^m$, where $m = \frac{n(n+1)}{2}$, be an orthonormal basis for $\mathfrak{so}(n)$. Any $Y \in \mathfrak{so}(n)$ can be written as $Y = \sum_{i=1}^m y_i \eta_i$. Denote the vector $\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^m$ by $\nu(Y)$, and hence identify $\mathfrak{so}(n)$ (as a vector space) with $\mathbb{R}^m$. The reduced commutator $C_{ad_X} : \mathbb{R}^m \to \mathbb{R}^m$ is linear operator defined via the identity

$$\nu([X, Y]) = C_{ad_X} \nu(Y). \tag{6.38}$$

Note that C_{ad_X} is a matrix representing the adjoint operator ad_X. We recall the following proposition from [13]:

Proposition 6.4.1. Let $n = 2n'$ be an even number and let $\pm\mathbf{i}\lambda_1, \ldots, \pm\mathbf{i}\lambda_{n'}$ ($\mathbf{i} = \sqrt{-1}$) be the eigenvalues of $X \in \mathfrak{so}(n)$. Then the eigenvalues of C_{ad_X} are $\mathbf{i}(\pm\lambda_i \pm \lambda_j)(1 \leq i < j \leq n')$ as well as zero with multiplicity n'. If $n = 2n' + 1$ and $\pm\mathbf{i}\lambda_1, \ldots, \pm\mathbf{i}\lambda_{n'}, 0$ are the eigenvalues of X, then the eigenvalues of C_{ad_X} are $\pm\mathbf{i}\lambda_i (1 \leq i \leq n')$, $\mathbf{i}(\pm\lambda_i \pm \lambda_j)(1 \leq i < j \leq n')$ as well as zero with multiplicity n'.

Now, back to our application, we want to find the eigenvalues of the curvature operator for $SO(n)$ and $G_k(n)$. From Theorems 6.2.1 and 6.2.3 we see that the curvature operator at I and along X is given by $Y \mapsto -\frac{1}{4} ad_X^2(Y)$ for $SO(n)$ and $Y \mapsto -ad_X^2(Y)$ for $G_k(n)$ with the extra condition of $\|X\| = 1$. The above proposition

enables us to find the eigenvalues of the curvature operator. Note that we can write

$$\lambda_1(-ad_X^2) = (|\lambda_1(X)| + |\lambda_2(X)|)^2, \tag{6.39}$$

where $\lambda_1(A)$ and $\lambda_2(A)$ are the first and second largest eigenvalues in absolute value of the symmetric matrix (or skew-symmetric) operator A, respectively.

6.4.3 2-norm Finsler Balls in $SO(n)$

The following Theorem characterizes strongly convex 2-norm (Finsler) balls in $SO(n)$:

Theorem 6.4.3. For $o \in SO(n)$ define the 2-norm (Finsler) ball

$$B_2(o, \rho) = \{oe^{tX} | X \in \mathfrak{so}(n), \|X\|_2 < \rho\}. \tag{6.40}$$

Then $B_2(o, \rho)$ is a (Riemannian) strongly convex set in $SO(n)$ for every $\rho < \frac{\pi}{2}$. Moreover, the Riemannian distance function $x \mapsto d(o, x)$ is strictly convex inside $B_2(o, \rho)$ along non-radial directions; and it is only convex along radial directions.

Proof. We partially follow [36] to prove the first part. Without loss of generality we can assume that o coincides with I, the identity of $SO(n)$. Let $X, Y \in \mathfrak{so}(n)$ and $\|X\|_2, \|Y\|_2 < \rho$. Set $x = e^X$ and $y = e^Y$. We have $x, y \in B_2(I, \rho)$, and since $d_2(x, y) < \pi$ there is a unique minimizing geodesic connecting x to y . Let z be the unique midpoint of the minimizing geodesic connecting x to y. Note that $z^{-1}x$ and $z^{-1}y$ have I as their midpoint. Therefore, by inversion we have $(z^{-1}x)^{-1} = z^{-1}y$ and $z^2 = zxz^{-1}y$. We write $\hat{x} = zxz^{-1}$. Note that $d_2(I, \hat{x}) = d_2(I, x) < \pi$.

Therefore, there exists $\hat{X} \in \mathfrak{so}(n)$ with $\|\hat{X}\|_2 < \rho$ such that $\hat{x} = e^{\hat{X}}$, and we have $z^2 = e^{\hat{X}}e^{Y}$. From this and Theorem 6.4.1, due to the bi-invariance of d_2, we have $d_2(I, z^2) \leq \|\hat{X}\|_2 + \|Y\|_2 < 2\rho < \pi$, which implies that there exists $Z' \in \mathfrak{so}(n)$ with $\|Z'\|_2 < 2\rho < \pi$ with $z^2 = e^{2Z'}$. Therefore, $z = e^{Z}$ with $Z = \frac{Z'}{2}$ and $\|Z\|_2 < \rho$ or equivalently $d_2(I, z) < \rho$. This means that the midpoint of any minimizing geodesic connecting two points in $B_2(I, \rho)$ belongs to $B_2(I, \rho)$. This proves the strong convexity of $B_2(o, \rho)$. For the second statement, let $X \in \mathfrak{so}(n)$ be such that $\|X\|_2$ and $x = e^{X}$. In order to use (6.37) we normalize X as $\tilde{X} = \frac{X}{\|X\|} = \frac{X}{d(I,x)}$ (note that $\|X\|$ is the Riemannian norm in (6.6)). If $\lambda_1(|\lambda_1| = \rho)$ and λ_2 are the two largest eigenvalues (in absolute value) of X, then the largest eigenvalue of $-ad^2_{\tilde{X}}$ by (6.39) is equal to $\frac{(\rho+|\lambda_1|)^2}{d^2(I,x)}$. Hence, the largest eigenvalue of the curvature operator along X is $\kappa_1 = \frac{1}{4}\frac{(\rho+|\lambda_1|)^2}{d^2(I,x)}$. Therefore, we have $\kappa_1 \leq \frac{\rho^2}{d^2(I,x)}$. Next, from (6.37) we see that Hess$d(I, x)$ will be positive definite along non-radial directions if $\sqrt{\kappa_1}d(I, x) < \frac{\pi}{2}$ or equivalently if $\rho < \frac{\pi}{2}$. Along a radial direction the Hessian is zero. □

6.4.4 2-norm Finsler Balls in $G_k(n)$

The next Theorem is the counterpart of Theorem 6.4.3 in $G_k(n)$:

Theorem 6.4.4. For $o \in SO(n)$ define the 2-norm (Finsler) ball

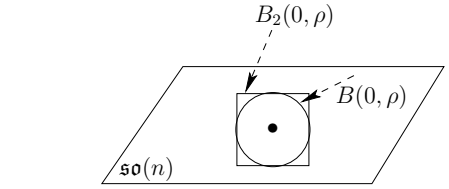
$$B_2(\{oK_k(n)\}, \rho) = \{\{oe^{tX}K_k(n)\}|X \in \mathfrak{g}_k(n), \|X\|_2 < \rho\}. \tag{6.41}$$

Then $B_2(o, \rho)$ is a (Riemannian) strongly convex set in $G_k(n)$ for every $\rho < \frac{\pi}{4}$. Moreover, the Riemannian distance function $x \mapsto d(o, x)$ is strictly convex inside

$B_2(o, \rho)$ along non-radial directions; and it is only convex along radial directions.

Proof. Both parts follow with a very minor modification from the proof of Theorem 3.1 in [47]. The second part also can be shown similar to the previous theorem, by noting that here $\kappa_1 < \frac{4\rho^2}{d^2(\{IK_k(n)\},x)}$. □

Remark 6.4.4. These Finsler balls should not look esoteric. Recall that one can associate $[\frac{n}{2}]$ angles, called angles of rotation, with any matrix in $SO(n)$. $B_2(I, \rho)$ is the set of all elements of $SO(n)$ whose maximum angles of rotation (in absolute value) are less than ρ. A similar interpretation is true for $G_k(n)$ with principal angles instead. However, here we need to allow for both positive and negative principal angles. This means that for any $s \in B_2(\{IK_k(n)\}, \rho)$, its reflection with respect to $\{IK_k(n)\}$ is also in $B_2(\{IK_k(n)\}, \rho)$. Figure 6.1 shows both $B(0, \rho)$ and $B_2(0, \rho)$ in $\mathfrak{so}(n)$, which transform to $B(I, \rho)$ and $B_2(I, \rho)$ in $SO(n)$ via the matrix exponential. The meaning of $B_2(I, \rho)$ being larger than $B(I, \rho)$ is clear in the picture. The relation between $B(I, \rho)$ and $B_2(I, \rho)$ in $SO(n)$ is very similar to the relation between the unit circle and square in the plane.



$B(0,\rho) = \{X \in \mathfrak{so}(n) | \|X\| = \frac{1}{\sqrt{2}}\|X\|_F < \rho\}$
$B_2(0,\rho) = \{X \in \mathfrak{so}(n) | \|X\|_2 < \rho\}$
$B(0,\rho) \subset B_2(0,\rho) \subset \mathfrak{so}(n)$

Figure 6.1: In $\mathfrak{so}(n)$, the 2-norm Finsler ball of radius $\rho \leq \pi$ is larger than the Riemannian ball of radius ρ.

6.5 Existence and Uniqueness of the Local Riemannian L^p Mean in Finsler Balls

The following theorem gives existence and uniqueness of the local L^p in Finsler balls. Recall that the local L^p mean in a set B is defined by solving $\min_{\bar{B}} f_p(x)$ (see (3.2) and Section 3.1.1).

Theorem 6.5.1. Let $\{x_i\}_{i=1}^N$ be a set of data points in $M = SO(n)$ which lies in a Finsler ball $B_2(o, \rho)$ with $\rho < \frac{\pi}{4}$. The local L^p Riemannian mean $(1 < p \leq \infty)$ with respect to the weight vector $\mathbf{w}$ for this set exists is unique and lies inside $B_2(o, \rho)$. For $p = 1$ if the points do not lie on a geodesic, again the same holds, and if the the points lie on a geodesic the mean might not be unique. The same holds for $M = G_k(n)$ with $\rho < \frac{\pi}{8}$.

Proof. We briefly give one proof for both $SO(n)$ and $G_k(n)$. First, note that the local mean exists. Next, from Theorems 6.4.3, 6.4.4 and Corollary 2.4.1 we conclude that for any two points $q, x \in B(o, \rho)$, $x \mapsto d^p(q, x)(p > 1)$ is a strictly convex function; so is $x \mapsto f_p(x)$. For $p = 1$, as before, we have the strict convexity of $f_1(x)$ if not all the data points lie on a geodesic. For $p = \infty$, we might assume that the L^∞ is the minimum of $x \mapsto \hat{f}(x) = \max_i d^2(x_i, x)$, which is again strictly convex[1]. Therefore, except for the possible degenerate case when $p = 1$, if the local L^p mean does not lie on the boundary of $B(o, \rho)$, then it is unique. To see that the mean cannot lie on the boundary of B, first, recall the properties of strongly convex sets

[1]We borrow this trick from [76, p. 168], in which the trick is used to avoid the unpleasant dealing with non-strict convexity of $x \mapsto d(x_i, x)$.

from Theorem 2.3.2. Let x be a point on the boundary of $B_2(o,\rho)$. Then, the gradient of f_p $(1 \leq p < \infty)$ at x is a linear combination of a finite number of vectors pointing inside B (see (3.6)); hence, itself will be pointing inside. This shows that the $L^p(1 \leq p < \infty)$ mean cannot be on the boundary of B. For $p = \infty$, note that there exists a vector in T_xM pointing inside B to which all thee vectors $\exp_x^{-1} x_i$ make acute angle; and by the first variation of arc length moving along this vector reduces f_∞. Therefore, for $1 \leq p \leq \infty$, x cannot be minimizer of f_p. This completes the proof. □

Remark 6.5.1. The result about $SO(n)$ is not completely new, it could be derived with minor modifications from results in [36] or [16]. Ideally, one would like to have the claims of the theorem to hold for the global L^p mean. Our efforts, unfortunately, were not successfully in proving this. The reflection technique used by Buss and Fillmore [17] which we also used to prove the global aspect of Theorem 3.2.1, seems not adequate. One reason is that a 2-norm Finsler ball is not a Riemannian ball (except for $n = 2, 3$). As a result a radial geodesic does not meet the boundary of the Finsler ball perpendicularly.

Remark 6.5.2. Note that the local L^∞ mean of $\{x_i\}_{i=1}^N \subset B_2(o,\rho)$ is the center of the minimal (Riemannian) ball of $\{x_i\}_{i=1}^N$, but this ball is not strongly convex, necessarily.

6.6 Existence and Uniqueness of Mean-Invariance Means in Finsler Balls in $SO(n)$ and $G_k(n)$

The following theorem gives the existence and uniqueness condition for pairwise, MI and WMI means for points in Finsler balls in $SO(n)$ and $G_k(n)$:

Theorem 6.6.1. Let $M = SO(n)$ and assume $\{x_i\}_{i=1}^N \subset B_2(o,\rho) \subset M$ where $\rho < \frac{\pi}{2}$. The pairwise mean, MI and WMI means of this set exist, are unique and lie in $B_2(o,\rho)$ and the results of Theorems 5.3.1 and 5.4.3 hold. The same statements hold true if $M = G_k(n)$ and $\rho < \frac{\pi}{4}$.

Proof. The proof is based on our previous results in Chapter 5. We give one short proof for both parts. Note that since we have finite number of data points we can assume that the points lie in $B_2(o,\rho')$ with $\rho' < \rho$. Also note that all the operations related to MI or WMI means remain valid inside $B_2(o,\rho')$. Now, the claim about the pairwise mean follows from Theorem 5.3.1 with $A = \overline{B_2(o,\rho')}$. The claim for the case of MI and WMI means also follows similarly from Theorems 6.4.3 (and 6.4.4 in case $M = G_k(n)$) and 5.4.3, with $A = B(o,\rho')$ and $A' = B(o,\rho)$. Note that in all cases the mean lies in $B_2(o,\rho)$. □

Remark 6.6.1. This theorem gives the existence and uniqueness of pairwise, MI and WMI means in 2-norm balls which are larger than the balls in Theorem 6.5.1 for local L^p means. While this might be considered as an advantage for recursive-iterative means, we conjecture that the global $L^p (p \geq 2)$ mean should exist uniquely in the same balls as Theorem 6.6.1 gives.

Remark 6.6.2. In Theorem 5.4.4 we proved the existence and uniqueness of GWMI means for data points in a Riemannian ball of specified radius. A generalization of that theorem to Finsler balls, requires accurate analysis of the boundary (which is nonsmooth) of the Finsler ball $B_2(o,\rho)$ in $SO(n)$ or $G_p(n)$. In particular, we need to show that for $\{x_i\}_{i=1}^N \subset \overline{B_2(o,\rho)}$, the L^p mean of the data points belongs to the interior of $B(o,\rho)$. Notice that Theorems 3.2.1 and 6.5.1 have different assumptions about the relative positions of the data points with respect to the boundary of the balls. To show existence and uniqueness of the GWMI means, we have to establish that the $L^p(1 < p \leq \infty)$ mean is a strongly convex mean function with respect to Finsler balls instead of Riemannian balls (see Definition 5.2.1). This seems quite plausible, but our efforts in showing that also were not completely successful.

Remark 6.6.3. The Perimeter Shrinkage (PS) scheme introduced in Section 5.3 remains valid if the data points do not belong to a strongly convex ball provided $d_2(x_i, x_{i+1}) < \pi (1 \leq i \leq N)$. However, as explained in Theorem 5.3.2 the points might converge to equidistance positions on closed geodesics.

6.7 Efficient Geodesic Interpolation and Extrapolation

In numerical computations we need to choose a numerical representation for our manifold-valued data. Usually, this can be accomplished by choosing a suitable embedding of the manifold in some Euclidean space. This leads to another possible definition of mean, which is the so-called extrinsic mean:

Definition 6.7.1. Assume (M,d) is embedded in $(\mathbb{R}^{n_E}, \|.\|)$, where $\|.\|$ is the stan-

dard Euclidean distance. Also denote the embedding of each $x \in M$ with the same letter. Denote by $\text{Proj}_M : \mathbb{R}^{n_E} \to M$ the orthogonal projection operator from the ambient space $\mathbb{R}^{n_E}$ onto the submanifold M. Then the extrinsic mean of $\{x_i\}_{i=1}^N \subset M \subset \mathbb{R}^{n_E}$ with weight vector $\mathbf{w} = [w_1, \ldots, w_N]^T$ is defined by $\mu_{\text{ex}} = \text{Proj}_M(\mu_{\text{EC}})$ where $\mu_{\text{EC}} = \sum_{i=1}^N w_i x_i$ is the standard Euclidean weighted average.

In some occasions it is easy to find the extrinsic mean. For example, as mentioned in Subsection 5.5.1 for unit sphere S^n in $\mathbb{R}^{n+1}$, finding the related projection operator is easy. It can be shown that the set of points in M whose projection onto M is not unique has Lebesque measure zero [11]; hence, the extrinsic mean with respect to a given embedding is almost always unique. On the other hand, since there are many embedding possibilities, a given set of points can have many different extrinsic means. Yet, there are some embeddings which have interesting properties and bring about ease of calculations. The standard embedding of $SO(n)$ in $\mathbb{R}^{n\times n}$ and the embedding of $G_k(n)$ (as orthogonal projection matrices of rank k) in the space of symmetric matrices are so. These two embeddings allow us to perform fast geodesic midpoint calculation, as we shall see in the next two subsections.

6.7.1 Efficient Geodesic Interpolation and Extrapolation in $SO(n)$

The standard embedding of $SO(n)$ in $\mathbb{R}^{n\times n}$ is just the usual representation of elements of $SO(n)$ as $n \times n$ matrices. We also measure the distance between $x_1, x_2 \in SO(n) \subset \mathbb{R}^{n\times n}$ by the Frobenius norm as $\|x_1 - x_2\|_F$. The projection

of $x \in \mathbb{R}^{n\times n}$ onto $SO(n)$ is defined by

$$\text{Proj}_{SO(n)}(x) = \text{argmin}_{y\in SO(n)}\|x-y\|_F^2. \tag{6.42}$$

It is easy to check that, if the determinant of x is positive, then $\text{Proj}_{O(n)}(x) = UV^T$, where $x = U\Sigma V^T$ is an SVD of x [43, p. 432]. Recall that an $n \times n$ matrix can be written as $x = P\Theta$, where P is positive semi-definite and Θ (called the right polar factor) is orthogonal [43, p. 413]. There is also the so-called left polar decomposition for which $x = QR$, where the polar factor Q is orthogonal and R is positive semi-definite. If x is non-singular, then the polar factors are unique and also $Q = \Theta$. Moreover, if x is a normal matrix (i.e., $xx^T = x^Tx$), then $R = P$. Since $x = U\Sigma U^T UV^T$, we have $\Theta = UV^T$, provided x is non-singular. With this brief preparation we state the following theorem:

Theorem 6.7.1. Consider the standard embedding of $SO(n)$ in $\mathbb{R}^{n\times n}$. Let $x_1, x_2 \in SO(n)$. We have the followings:

1. If $\mu_{\text{EC}}(x_1, x_2) = \frac{x_1+x_2}{2} = U\Sigma V^T$ is an SVD of $\mu_{\text{EC}}(x_1, x_2)$, then $\mu_{\frac{1}{2}}(x_1, x_2) = \mu_{\text{ex}}(x_1, x_2) = UV^T$. This operation produces a unique midpoint as long as x_1 is not a cut point of x_2. In particular, if $d_2(x_1, x_2) < \pi$ the operation remains valid.

2. The extrapolated point $\text{Ext}_1(x_1, x_2) = x_1x_2^Tx_1$ is the reflection of x_2 with respect to (w.r.t) x_1 along the minimizing geodesic connecting x_2 to x_1. Unless, x_1 and x_2 are antipodal points (i.e., $d_2(x_1, x_2) = \pi$) the extrapolated point will

be different from x_2.

Proof. Moakher in [67] has shown that $\mu_{\text{ex}}(x_1, x_2) = \text{Proj}_{SO(n)}(\frac{x_1+x_2}{2})$ and $e^{\frac{1}{2}\log(x_2^T x_1)}$ both are equal to the polar factor of $\frac{x_1+x_2}{2}$. This holds if $\frac{x_1+x_2}{2}$ is nonsingular, in which case the polar factor is equal to UV^T, where $U\Sigma V^T$ is an SVD of $\frac{x_1+x_2}{2}$. It is easy to check that $\frac{x_1+x_2}{2}$ is nonsingular, and its determinant is positive if no eigenvalue of $x_2^T x_1$ is -1. This will be true if $d_2(I, x_2^T x_1) < \pi$. This proves item 1 of the theorem. The first part of item 2 follows from the fact that inversion (or transpose) is an reflection isometry w.r.t to the identity. Unless the reflected point and the identity lie on a closed geodesic and in an antipodal positions the reflection will not coincide with the original point. This proves the second part of 2. □

Corollary 6.7.1. The above interpolation and extrapolation operations are well-defined if x_1 and x_2 lie in a strongly convex domain, in particular in a Finsler ball of radius $\rho < \frac{\pi}{2}$.

The above theorem helps us in constructing the geodesic between two points or extrapolating the geodesic between them in an iterative process and without the use of matrix logarithm and exponential. Table 6.1 gives the pseudo-codes for construction of the geodesic between two points and extending the geodesic beyond them based on midpoint calculation and reflection operation. For the interpolation part, at each level, given a set of points on the geodesic, we generate new midpoints and double the number of intermediate points this way. The extrapolation part is nothing but a reflection operation, but it can be complemented by another interpolation between the reflected point and the new point. Note that the direct numerical

Dyadic geodesics interpolation

1. Let $M = SO(n)$ or $M = G_k(n)$. Assume two non-antipodal points $x_0, x_1 \in M$. Set $i = 1$ and fix integer $K \geq 1$.
2. Level i: Given points $\{x_0, \ldots, x_{2^{i-1}}\}$, for $0 \leq j \leq 2^{i-1} - 1$, find $x_{j+\frac{1}{2}} = \mu_{\frac{1}{2}}(x_j, x_{j+1})$ from item 1 of Theorem 6.7.1 or 1 of Theorem 6.7.2.
3. Re-index the $2^i + 1$ points to $0, \ldots, 2^i - 1$, increment i. If $i > K$ stop; otherwise goto step 2.

Dyadic geodesics extrapolation

1. Let $M = SO(n)$ or $M = G_k(n)$. Consider $x_0, x_1 \in M$. Set $i = 1$ and fix integer $J \geq 1$.
2. Level i: Given points $\{x_0, \ldots, x_{i-1}, x_i\}$, find $x_{i+1} = \mathrm{Ext}_1(x_i, x_{i-1})$ from item 2 of Theorem 6.7.1 or item 2 of Theorem 6.7.2.
3. Increment i. If $i > J$ stop; otherwise goto step 2.

Table 6.1: Pseudo-code for dyadic geodesic interpolation and extrapolation.

calculation of the form $\mu_{\frac{1}{2}}(x_1, x_2) = x_1 e^{\frac{1}{2}\log x_1^T x_2}$ compared with the SVD based midpoint calculation in 1 has much higher complexity. Finding the matrix logarithm has the heaviest load, although matrix exponential is also computationally demanding. Moreover, in numerical calculation of $\log x_1^T x_2$, there is no guarantee that the result is skew-symmetric to machine precision and one might need to use special algorithms for this purpose [23]. Without going into the details of the complexity of the related numerical algorithms we give an example to compare the direct and dyadic interpolation methods.

Example 6.7.1. We perform a comparison between the computational times for direct and the dyadic geodesic construction techniques in MATLAB©. Table 6.2 shows the main parts of the used MATLAB code. We try levels $K = 1$ and $K = 2$,

```
[U,S,V] = svd(x1+x2);\%Level 1
M11 = U*V';
[U,S,V] = svd(M11+x1);
M21 = U*V';
[U,S,V] = svd(M11+x2);
M22 = U*V';
```

```
N = 2^K;%K=Level
X = logm(x1'*x2);
M = x1;
E = expm(1/N*X);
for i=1:N-1
M = M * E;% M = x_1 e^{(i/N) log x_1^T x_2}
end
```

Table 6.2: Left: Main part of a MATLAB code to compute the geodesic between $x_1, x_2 \in SO(n)$ at two levels $K = 1$ and $K = 2$ using dyadic efficient interpolation (See Table 6.1. Right: Main part of a MATLAB code to compute the geodesic between $x_1, x_2 \in SO(n)$ at level K, by direct matrix logarithm-exponential calculations.

i.e., 1 and 3 intermediate points. In the direct method, instead of finding $\log x_1^T x_2$ per computed point we just find $E = e^{\frac{1}{2^K} \log x_1^T x_2}$, and in the subsequent calculation use it to make the method more efficient. Table 6.3 shows the results of the median of the computation time for the two methods. It is clear that the dyadic method is much faster. However, note that its efficiency decreases by the level or the number of the generated points geodesic. As explained before, the efficiency of the algorithm mostly stems from avoiding the use of matrix logarithm. At the same time the direct method needs to find the logarithm only once; therefore, as K increases that advantage of the dyadic method decreases. It is interesting to mention that the first step in some of the most powerful methods for computing the logarithm of a matrix A (including the method used in MATLAB) is the so-called "inverse scaling and squaring" [22], in which successive square roots of A, that is $A^{\frac{1}{2^i}} (i = 1, 2, \ldots)$, are found. On the other hand, finding the geodesic midpoint involves finding $e^{\frac{1}{2} \log x_1^T x_2}$, which is nothing but the square root of $x_2^T x_1$. Theretofore, it should not be surprising that the dyadic method is at least 10 times faster than the direct method at Level $K = 1$.

Level $K=1$		
n	dyadic	direct
3	0.0001	0.0020
30	0.0029	0.0410
100	0.0470	0.4586

Level $K=2$		
n	dyadic	direct
3	0.002	0.0020
30	0.0095	0.0596
100	0.1428	0.3963

Table 6.3: Comparison between computation times for two geodesic interpolations MATLAB codes in Table 6.2. We try two levels $K = 1$ and $K = 2$. Times are in Seconds and each is the median of 100 independent runs.

6.7.2 Efficient Geodesic Interpolation and Extrapolation in $G_k(n)$

We mentioned about the isometric identification of $\mathcal{P}_{k,n}$ the space of $n \times n$ orthogonal projection matrices of rank k and $G_k(n)$ in Subsection 6.2.2.1. By standard embedding of $\mathcal{P}_{k,n}$ in $\mathcal{S}_n$, the space of $n \times n$ positive semi-definite symmetric matrices, we can define the extrinsic mean of $\{x_i\}_{i=1}^N \subset \mathcal{P}_{k,n}$ as $\text{Proj}_{\mathcal{P}_{k,n}}(\mu_{EC}(x_1, \ldots, x_N))$, as before. Here, $\text{Proj}_{\mathcal{P}_{k,n}} : \mathcal{S}_n \to \mathcal{P}_{k,n}$ is defined as

$$\text{Proj}_{\mathcal{P}_{k,n}}(x) = \text{argmin}_{y \in \mathcal{S}_n} \|x - y\|_F^2. \tag{6.43}$$

The following lemma gives this projection operator explicitly:

Lemma 6.7.1. Let $A \in \mathbb{R}^{n \times n}$ be symmetric and positive semi-definite. Then we have $\text{Proj}_{\mathcal{P}_{k,n}}(A) = P_A = UU^T$, where $U_{n \times (n-k)}$ is a matrix whose columns are the eigenvectors of A associated to its k largest eigenvalues. If the k largest eigenvalues of A are strictly larger than the rest, then P_A is unique.

Proof. We recall that any $n \times n$ orthogonal projection P of rank k, can be associated to a k-dimensional subspace of $\mathbb{R}^n$, S, and P can written as $P = U_P U_P^T$ where U_P is

an $n \times k$ matrix whose columns are an orthonormal basis of S. Now let $A = U\Sigma U^T$ be the full SVD or EVD of A, where $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$ with $\sigma_1 \geq \ldots \geq \sigma_n \geq 0$. We want to find a P or a U_P which minimizes

$$f(P) = \|U_P U_P^T - A\|_F^2. \tag{6.44}$$

Up to a additive and multiplicative constants $f(P)$ is equal to $-\text{tr}(U_P U_P A)$. So we need to maximize $\max_{U_P} \text{tr}(U_P U_P^T A)$. Which is equivalent to solving

$$\max_{U_P} \text{tr}(U^T U_P (U^T U_P)^T \Sigma). \tag{6.45}$$

Note that $\tilde{P} = U^T U_P (U^T U_P)^T$ again is a rank k orthogonal projection matrix. The above function is equal to $\sum_i \tilde{P}_{ii}\sigma_i$, where $\tilde{P}_{ii}$ is the i^{th} element on the diagonal of $\tilde{P}$. It is easy to check that any entry of an orthogonal projection matrix is not larger than 1 in absolute value. Therefore, the maximum value $\sum_i \tilde{P}_{ii}\sigma_i$ can achieve is $\sum_{i=1}^k \sigma_i$. This can happen if $\tilde{P} = \left[\begin{smallmatrix} I_{k\times k} & 0 \\ 0 & 0 \end{smallmatrix}\right]$. But by choosing U_P equal to the first k columns of U we can achieve this particular $\tilde{P}$ and that maximum. This completes the proof. □

Next, we have the following theorem:

Theorem 6.7.2. Let S_1 and S_2 be subspaces representations of s_1, s_2 in $G_k(n)$, respectively. Also let $P_1 = S_1 S_1^T$ and $P_2 = S_2 S_2^T$ denote the corresponding orthogonal projection matrices associated with s_1 and s_2. We have:

1. The extrinsic mean of s_1 and s_2 (or P_1 and P_2) can be expressed as $\mu_{\text{ex}}(P_1, P_2) =$

$\text{Proj}_{\mathcal{P}_{k,n}}(\mu_{\text{EC}}(P_1, P_2)) = SS^T$, where S is an $n \times k$ matrix of k eigenvectors associated with the k largest eigenvalues of $\mu_{\text{EC}}(P_1, P_2) = \frac{P_1+P_2}{2}$. Moreover, SS^T is the orthogonal projection matrix associated with the midpoint of the minimizing geodesic between s_1 and s_2 and S is a subspace representative of the midpoint. The above projection gives a unique a point if $d_2(s_1, s_2) < \frac{\pi}{2}$.

2. Denote the reflection of s_2 w.r.t s_1 along the minimizing geodesic connecting them by $\text{Ext}_1(s_1, s_2)$. Then $\text{Ext}_1(s_1, s_2)$ has the subspace representation $(2P_1 - I)S_2$.

Proof. Without loss of generality we can assume $S_1 = \left[\begin{smallmatrix} I_{k\times k} \\ \mathbf{0}_{(n-k)\times k} \end{smallmatrix}\right]$ and $S_2 = Q\left[\begin{smallmatrix} \mathbf{C}(\theta_1,\ldots,\theta_k) \\ \mathbf{S}(\theta_1,\ldots,\theta_k) \\ 0_{(n-2k)\times k} \end{smallmatrix}\right]$, where Q is of the form $Q = \left[\begin{smallmatrix} V & 0 \\ 0 & U \end{smallmatrix}\right]$ (see (6.16)). Note that $P_1 = \left[\begin{smallmatrix} I_{k\times k} & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and

$$P_2 = Q\begin{bmatrix} \mathbf{C}^2(\theta_1,\ldots,\theta_k) & \mathbf{C}(\theta_1,\ldots,\theta_k)\mathbf{S}(\theta_1,\ldots,\theta_k) & \mathbf{0} \\ \mathbf{C}(\theta_1,\ldots,\theta_k)\mathbf{S}(\theta_1,\ldots,\theta_k) & \mathbf{S}^2(\theta_1,\ldots,\theta_k) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} Q^T. \tag{6.46}$$

Denote the matrix in the middle by $\tilde{P}_2(\theta_1, \ldots, \theta_k)$. Note that if $\tilde{P} = U_{\tilde{P}_2}\Sigma U_{\tilde{P}_2}^T$ is an SVD or EVD of P, then $P_2 = QU_{\tilde{P}_2}\Sigma(QU_{\tilde{P}_2})^T$ is an the SVD of P_2. Moreover, for SVD's of $\mu_{EC} = \frac{P_1+P_2}{2}$ and $E = \frac{P_1+\tilde{P}_2}{2}$ we have a similar relation, due to the special forms of P_1 and Q. Now, we compute an SVD of E. Note that since E is symmetric and positive semi-definite its EVD and SVD coincide. We rewrite E more explicitly as

$$E = \frac{1}{2}\begin{bmatrix} \text{diag}(\cos^2\theta_1+1,\ldots,\cos^2\theta_k+1) & \text{diag}(\cos\theta_1\sin\theta_1,\ldots,\cos\theta_k\sin\theta_k) & \mathbf{0} \\ \text{diag}(\cos\theta_1\sin\theta_1,\ldots,\cos\theta_k\sin\theta_k) & \text{diag}(\sin^2\theta_1,\ldots,\sin^2\theta_k) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{6.47}$$

Note that θ_i appears in four entries of E at locations (i, i), $(i, i + k)$, $(i + k, i)$ and

$(i+k, i+k)$. Those entries can be put in a matrix E_i of dimension 2×2 as:

$$E_i = \frac{1}{2}\begin{bmatrix} 1+\cos^2\theta_i & \cos\theta_i \sin\theta_i \\ \cos\theta_i \sin\theta_i & \sin^2\theta_i \end{bmatrix}. \tag{6.48}$$

It is easy to check that E_i has $\lambda_{i,1} = \cos^2\frac{\theta_i}{2}$ and $\lambda_{i,2} = \sin^2\frac{\theta_i}{2}$ as eigenvalues. Note that since $\theta_i < \frac{\pi}{2}$, we have $\lambda_{i,1} > \lambda_{i,2}$. In fact, for the same reason we have $\cos^2\frac{\theta_i}{2} > \frac{1}{2}$ and $\sin^2\frac{\theta_i}{2} < \frac{1}{2}$; and hence, $\lambda_{i,1} > \lambda_{j,2}$ for $1 \leq i, j \leq k$. The corresponding eigenvectors can be found as $v_{i,1} = \begin{bmatrix} \cos\frac{\theta_i}{2} \\ \sin\frac{\theta_i}{2} \end{bmatrix}$ and $v_{i,2} = \begin{bmatrix} -\sin\frac{\theta_i}{2} \\ \cos\frac{\theta_i}{2} \end{bmatrix}$. Now one can check that E has $2k$ eigenvalues $\lambda_{k,1} \geq \ldots \geq \lambda_{1,1} > \lambda_{1,2} \geq \ldots \geq \lambda_{k,2}$ $(1 \leq i \leq k)$ and the rest of its $n-k$ eigenvalues are zero. The corresponding eigen matrix can be constructed as $U_E = [\tilde{U}_E, \tilde{U}_E^{\perp}]$, where $\tilde{U}_E$ is of dimension $n \times 2k$ and is defined as

$$\tilde{U}_E = \begin{bmatrix} \mathbf{C}(\frac{\theta_1}{2},\ldots,\frac{\theta_k}{2}) & -\mathbf{S}(\frac{\theta_1}{2},\ldots,\frac{\theta_k}{2}) \\ \mathbf{S}(\frac{\theta_1}{2},\ldots,\frac{\theta_k}{2}) & \mathbf{C}(\frac{\theta_1}{2},\ldots,\frac{\theta_k}{2}) \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{6.49}$$

and $\tilde{U}_E^{\perp}$ is an $n \times (n-2k)$ matrix whose columns span the orthogonal complement of the subspace spanned by the columns of $\tilde{U}_E$. Also define the diagonal matrix Λ as

$$\Lambda = \mathrm{diag}(\cos^2\theta_1, \ldots, \cos^2\theta_1, \sin^2\theta_1, \ldots, \sin^2\theta_k, \underbrace{0, \ldots, 0}_{n-2k}). \tag{6.50}$$

Note that $E = U_E \Lambda U_E^T$ is an SVD (or EVD) of E. By Lemma 6.7.1, the projection of E onto the space of orthogonal projection matrices of dimension $n \times n$ and rank k is $P_E = UU^T$ where U is of dimension $n \times k$ and $U = \begin{bmatrix} \mathbf{C}(\frac{\theta_1}{2},\ldots,\frac{\theta_k}{2}) \\ \mathbf{S}(\frac{\theta_1}{2},\ldots,\frac{\theta_k}{2}) \\ \mathbf{0} \end{bmatrix}$. This proves the first statement. To prove item 2, we recall the reflection formula (6.19). The geodesic reflection of S_2 with respect to S_1 is found by negating $\mathbf{S}(\theta_1, \ldots, \theta_k)$ in

$S_2 = Q \begin{bmatrix} \mathbf{C}(\theta_1,\ldots,\theta_k) \\ \mathbf{S}(\theta_1,\ldots,\theta_k) \\ 0_{(n-2k)\times k} \end{bmatrix}$. Note that

$$(2P_1 - I_{n\times n})S_2 = Q \begin{bmatrix} \mathbf{C}(\theta_1,\ldots,\theta_k) \\ -\mathbf{S}(\theta_1,\ldots,\theta_k) \\ 0_{(n-2k)\times k} \end{bmatrix}. \tag{6.51}$$

Therefore, $(2P_1 - I_{n\times n})S_2$ spans the same subspace as S_2' (the geodesic reflection of S_2 with respect to S_1). □

6.8 Two Applications

There has been an increasing number of real-world applications of averaging or finding the mean for manifold-valued data. Of course, most applications are related to the cases where the underlying manifold is a homogenous space. Here, we consider two less explored applications, where the ideas introduced in this chapter fit and serve very well. The first one is the extension of the famous Nelder-Mead algorithm to $SO(n)$ for maximization of a non-smooth function. The second application is the clustering of subspaces by extension of the K-means algorithm to $G_k(n)$[1]. Our emphasis is on using the pairwise mean (instead of the Riemannian L^2 or the MI mean). Also for finding the pairwise mean and other related calculations we use the efficient methods introduced before.

[1]The author envisioned these two applications in the early stages of this research. However, later he found out that they had already been studied by other researchers. Papers [25, 63] and [37] (and possibly few others) address the two mentioned applications, respectively. Nevertheless, the treatment presented here is developed independently and differs from the ones in the mentioned references.

6.8.1 Nelder-Mead Direct Search Algorithm in $SO(n)$

Direct search methods for optimization try to find a minimum of a function without the use of derivative information [55]. They usually are very useful in cases where the function is non-smooth, its derivative is expensive or impossible to compute or the function is not a even numerical in its nature. A very popular direct search algorithm known as Nelder-Mead simplex algorithm [70] can be easily extended from $\mathbb{R}^n$ to a manifold and in fact it can be efficiently implemented in $SO(n)$ or $G_k(n)$ using the methods introduced, before. Interestingly, the Nelder-Mead algorithm encompasses all the operations we mentioned so far: finding the center of mass, geodesic midpoint assignment and geodesic reflection of a point with respect to another one. We briefly describe the algorithm and give an example of its use for solving a non-differentiable maximization problem with applications in adaptive optics.

In Table 6.4 a pseudo code for the Nelder-Mead algorithm which is adapted to a Riemannian manifold M is given. The adaptation of the original Euclidean version to M is straightforward, and mainly is achieved by the obvious use of the exponential map of M. The algorithm consists of very simple steps. Denote the dimension of M by m. The algorithm will be given initial $m+1$ points considered as vertices of a simplex. Then, through a sequence of function evaluations, numerical comparisons, center of mass calculations and reflections the initial simplex will move, shrink or expand, and typically by getting close to a minimum point the simplex starts to shrink such that it finally will converge to that point. The stopping criterion in Step 3, compares the size of the current simplex with a given threshold to stop. We

should mention that except for special cases there is no proof for convergence of this algorithm. Nevertheless, the Nelder-Mead algorithm has remained a popular tool among many practitioners in many different fields [55]. Here, we choose to use the pairwise mean as the center of mass in Step 4. Note that, interestingly, all the operations in Steps 4-8 can be calculated via pairwise operations. Also note that the operation of expansion in Step 6 which gives $x_e = \exp_{\bar{x}}(-2\exp_{\bar{x}}^{-1} x_{m+1})$ can be viewed as the reflection of $\bar{x}$ with respect to x_r which itself is the reflection of x_{m+1} with respect to $\bar{x}$, and was computed in Step 5. Our concern, of course, should be that the defined operations remain well-defined. For this also there is no guarantee, unless we keep track of the points and force them to remain in valid distances, which can be prohibitively expensive or impractical. Nevertheless, one hopes that in the case of $SO(n)$ or $G_k(n)$ by starting in a large convex 2-norm Finsler ball the iterates will stay there.

6.8.1.1 An Application in Adaptive Optics: Maximization of a Non-differentiable Function in $SO(n)$

In [27], for applications in adaptive optics, maximization of the non-differentiable cost function $f : O(n) \to \mathbb{R}$ defined as

$$f(x) = \sum_{i=1}^{n} \max_{1 \le j \le N} (xC_j x^T)_{ii} \tag{6.52}$$

is proposed. Here, $\{C_i\}_{i=1}^{N} (N \ge 2)$ is a set of symmetric matrices. Without loss of generality we can assume that the domain of f is $SO(n)$. The mentioned max-

Nelder-Mead Algorithm on a Riemannian Manifold

1. Let (M, d) be a complete Riemannian manifold of dimension m with exponential map $\exp_x$ at $x \in M$. Consider $f : M \to \mathbb{R}$. Set $\epsilon > 0$
2. Choose $x_1, \ldots, x_{m+1} \in M$.
3. `if` $\max_{i,j} d(x_i, x_j) < \epsilon$, `then stop`; `else`, reorder the points such that $f(x_1) \leq \ldots \leq f(x_{m+1})$ and `goto` step 4.
4. Find $\bar{x}$ the center of mass of $\{x_i\}_{i=1}^m$.
5. (Reflection) Find x_r the reflection of x_{m+1} w.r.t. $\bar{x}$ as $x_r = \exp_{\bar{x}}(-\exp_{\bar{x}}^{-1} x_{m+1})$
   (a) `if` $f(x_1) \leq f(x_r) < f(x_m)$, `then` $x_{m+1} \leftarrow \bar{x}$ `and goto` step 3.
6. (Expansion) `if` $f(x_r) < f(x_1)$, `then` find x_e the expansion of x_{m+1} w.r.t. $\bar{x}$ as $x_e = \exp_{\bar{x}}(-2 \exp_{\bar{x}}^{-1} x_{m+1})$.
   (a) `if` $f(x_e) < f(x_1)$, `then` set $x_{m+1} \leftarrow x_e$ `and goto` 3, `else` set $x_{m+1} \leftarrow x_r$ `and goto` step 3
7. (Contraction) `else` find x_c the contraction of x_{m+1} w.r.t. $\bar{x}$ as $x_c = \exp_{\bar{x}}(\frac{1}{2} \exp_{\bar{x}}^{-1} x_{x_{m+1}}) = \mu_{\frac{1}{2}}(\bar{x}, x_{m+1})$
   (a) `if` $f(x_c) \leq f(x_{m+1})$, `then` set $x_{m+1} \leftarrow x_c$ `and goto` 3
   (b) `else goto` step 8
8. (Reduction) `For` $i = 2, \ldots, m+1$ set $x_i \leftarrow \mu_{\frac{1}{2}}(x_i, x_1)$ `and goto` step 3

Table 6.4: Pseudo-code of the Nelder-Mead algorithm on a Riemannian manifold M.

imization problem does not have a unique solution [27]. In the special case that $N = 2$, $C_1 = C$ and $C_2 = \mathbf{0}_{n \times n}$, the maximum value of f is equal to sum of the positive eigenvalues of C, and certainly a solution in this case is the transpose of the eigen-matrix of C [27]. For this special case we examine the Nelder-Mead algorithm adapted to $SO(n)$ to minimize $-f$. Let C be as follows (to 4 decimal digits)

$$C = \begin{bmatrix} 1.0133 & 0.7204 & 1.0904 & 0.1530 \\ 0.7204 & -0.4630 & -0.5926 & 1.4637 \\ 1.0904 & -0.5926 & 0.8661 & -0.9719 \\ 0.1530 & 1.4637 & -0.9719 & -0.6416 \end{bmatrix}. \tag{6.53}$$

The eigenvalues of C are $2.2136, 1.6085, -0.9703$ and -2.0769. Therefore, the maximum of f which is equal to the sum of positive eigenvalues of C is 3.8220. To start the Nelder-Mead algorithm, we need to choose a simplex with $m+1$ vertices, where m is the dimension of M. Here, $m = \frac{n(n+1)}{2} = 10$. We generate $m+1 = 11$ random points inside the Finsler ball $B_2(0, \frac{\pi}{2})$ as the initial simplex. We use the pairwise mean as the center of mass in Step 4 of the algorithm. For this step, reflection, expansion, contraction and reduction steps (see Steps 5-8) we use efficient interpolation and extrapolation schemes introduced in Theorem 6.7.1. Note that in the reflection step 5 we have $x_r = \bar{x}x_{m+1}^T\bar{x}$, and in the expansion step x_e is the reflection of $\bar{x}$ with respect to x_r; therefore $x_e = x_r\bar{x}^T x_r$, which can be computed efficiently again. After almost 110 iterations the calculated value for $\max f$ will be equal to 3.8220, in four decimal digits. Denote the solution in four decimal digits by x_*. We have x_* and $x_* C x_*^T$ as follows:

$$x_* = \begin{bmatrix} 0.5325 & -0.5910 & -0.5908 & -0.1346 \\ -0.7519 & -0.0551 & -0.6477 & 0.1102 \\ 0.3817 & 0.6164 & -0.4003 & 0.5605 \\ -0.0734 & -0.5175 & 0.2670 & 0.8097 \end{bmatrix} \text{ and } x_* C x_*^T = \begin{bmatrix} \mathbf{-1.0807} & 0.0000 & 0.0000 & -0.3316 \\ 0.0000 & \mathbf{2.1020} & 0.2347 & -0.0000 \\ 0.0000 & 0.2347 & \mathbf{1.7200} & 0.0000 \\ -0.3316 & -0.0000 & 0.0000 & \mathbf{-1.9665} \end{bmatrix}. \tag{6.54}$$

Note that with $C_1 = C$ as above and $C_2 = \mathbf{0}_{4\times 4}$, we have $f(x_*) = 4.8311$. The convergence is slow in terms of the number of iterations. However, we should note that the computation load for each step is not significant. To see the behavior of the calculated maximizer in terms of iteration index, we have tabulated the approximate maximum in 15 decimal digits in Table 6.5. The last row shows $\max f$ in 15 decimal digits, which is found using the (numerical) eigenvalues of C. We should note, despite their slow convergence, in many practical problems direct search methods

iteration i	computed max f
1	2.99792831533123
10	3.40130355940283
20	3.74564184272491
30	3.77369724540551
40	3.81199282263081
50	3.81796650623473
⋮	⋮
110	3.822008220185086
⋮	⋮
180	3.82201461161657
190	3.82201461242503
200	3.82201461328261
15digit	3.82201461365772

Table 6.5: Computed maximum of the non-differentiable function f in (6.52) in terms of the iteration index in the Nelder-Mead algorithm in $SO(4)$.

including the Nelder-Mead algorithm are the only choices.

6.8.2 Clustering of Subspaces

Data clustering is an important task in many data and signal processing algorithms. The standard clustering problem consists of clustering N data points $x_1, \ldots, x_N$ in $\mathbb{R}^n$ to N_c clusters $\mathcal{C}_1, \ldots, \mathcal{C}_{N_c}$ such that the members in cluster $\mathcal{C}_i$ are closer to the center of $\mathcal{C}_i$ than to the center of any other clusters. By the center of $\mathcal{C}_i$ we mean the Euclidean mean of the data points in $\mathcal{C}_i$. The center of each cluster can be used as a representative of the cluster or its members in subsequent analysis. A very popular algorithm to solve this problem is an iterative algorithm called the K-means algorithm [39]. In occasions, we might encounter a situation where our data points are linear subspaces themselves and we would like to cluster the data. This problem can be named as Grassmannian Clustering [37]. Interestingly, Grassmannian-valued data appear in different applications and clustering of such data has many potential applications. Two interesting application are the pre-coder codebook design [68] and channel quantization in Multi-Input Multi-Output wireless

communication channels [48, 54]. In [37] a Grassmannian K-means algorithm and in [54] a similar K-means algorithm together with the related Vector Quantization algorithm are proposed. These two algorithms use the extrinsic mean (see Section 6.7) for points on $G_k(n)$, mostly due to the ensued ease of calculation. Here, we implement a K-means algorithm with the pairwise mean, which is the easiest mean to compute (among all the means we studied).

K-means Algorithm on a Riemannian Manifold

1. Let (M, d) be a complete Riemannian manifold. Consider a set of data points $\{x_i\}_{i=1}^N$. Fix $N_c \leq N$.
2. Randomly partition $\{x_i\}_{i=1}^N$ into N_c non-empty clusters $\mathcal{C}_1, \ldots, \mathcal{C}_{N_c}$.
3. `for` $i = 1, \ldots, N_c$, find $\bar{x}_{\mathcal{C}_i}$ the center of $\mathcal{C}_i$.
4. `for` $i = 1, \ldots, N$, `if` $d(x_i, \bar{x}_{\mathcal{C}_j}) \leq d(x_i, \bar{x}_{\mathcal{C}_l}) \forall l \neq j$ `then` place x_i in $\mathcal{C}_j$. `if` more than one j satisfies that condition choose one of them randomly.
5. `if` one of the clusters is empty assign one data point to it randomly.
6. `if` in Step 4, none of the clusters have changed, `then stop`; `else goto` Step 3.

Table 6.6: Pseudo-code for the K-means algorithm on a Riemannian Manifold M. It is assumed that finding the center of mass is a well-defined operation.

6.8.2.1 An Example of Clustering of Subspaces With the K-means Algorithm

In our example, we generate some data points in $G_k(n)$ ($k = 3$ and $n = 20$) around $N_c = 3$ centers, and then try to cluster the data points and recover the centers via the K-means algorithm. First, we briefly describe the process to generate these data points. Denote the centers by $c_1, c_2, c_3 \in G_k(n)$ and the associated subspace representation by $S_1, S_2, S_3 \in \mathbb{R}^{n \times k}$. We set $d_2(c_1, c_2) = \frac{\pi}{24} = l_2$ and $d_2(c_1, c_3) = \frac{\pi}{6} =$

l_3 and generate c_2 and c_3 with reference to c_1; hence, we can set $S_1 = \begin{bmatrix} I_{k\times k} \\ \mathbf{0} \end{bmatrix}$. For $i = 2, 3$ we generate

$$C_i = \begin{bmatrix} \mathbf{0} & -\bar{C}_i^T \\ \bar{C}_i & \mathbf{0} \end{bmatrix}, \tag{6.55}$$

where the elements of $\bar{C}_i$ are of uniform distribution on the interval $[0, 1]$. We then normalize C_i as $C_i \leftarrow \frac{C_i}{\|C_i\|_2} l_i$. Now S_i is equal to the first k columns of e^{C_i}. Note that $C_1 = 0$. Next, we generate the actual data points around these centers. We generate $n_1 = 6$, $n_2 = 3$ and $n_3 = 2$ points around c_1, c_2 and c_3, respectively, as follows:

$$X_{i,j} = C_i + \nu \begin{bmatrix} \mathbf{0} & -W_{i,j}^T \\ W_{i,j} & \mathbf{0} \end{bmatrix} \quad 1 \le i \le N_c, \quad 1 \le j \le n_i, \tag{6.56}$$

where $W_{i,j} \in \mathbb{R}^{(n-k)\times k}$ manifests the noise, and is of standard normal elements and ν is a constant controlling the noise power. We set X_{ij} as the first k columns of $e^{X_{i,j}}$. After shuffling all these subspaces we relabel them as $X_1, \ldots, X_N$, where $n = n_1 + n_2 + n_3 = 11$. The parameter ν is a variable that we shall change. First we set $\nu = 0.05$. We implement the K-means algorithm as in Table 6.6 and choose the pairwise mean to find the center of mass of the clusters in Step 3. We perform 20 iterations to compute the pairwise mean. If the algorithm recovers the clusters correctly, then either based on the number of points in each estimated cluster or based on the distances $d(c_i, \hat{c}_j)$ we can recover the label of each of center. We do this relabeling and show the distances between the estimated centers of clusters and the original centers in Table 6.7. The left panel in Table 6.7 shows the median of the distances between the estimated centers and the original centers for $T = 100$

$d(c_i, \hat{c}_i)$	$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$
c_1	0.0416	0.5657	0.4712
c_2	0.5614	0.0392	0.4592
c_3	0.4639	0.4613	0.0406

$d(c_i, \hat{c}_i)$	$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$
c_1	0.1884	0.5268	0.3821
c_2	0.5310	0.1885	0.4356
c_3	0.3911	0.5160	0.1434

Table 6.7: Distances between the centers of the estimated clusters by the K-means algorithm (on $G_3(20)$) and the original centers in the example of Subsection 6.8.2.1. Each number is the median of the results of 100 independent runs of the experiment. The table on the left shows the distances for noise level $\nu = 0.05$, and the one on the right shows the distances for $\nu = 0.20$. The ability of the K-means algorithm to correctly detect the clusters reduces as ν increases.

independent runs. The distances are computed from (6.20). The centers are kept fixed and the data around them are generated in each run. Next, we increase the noise level ν to 0.20, and show the results of $T = 100$ runs in the right panel of Table 6.7. Expectedly, the ability of the K-means algorithm to resolve the clusters decreases by increasing the noise, and as can be seen $d(c_i, \hat{c}_i)$ increases as ν increases.

Chapter 7

Conclusions and Future Research Directions

In this dissertation we studied some mostly theoretical aspects of the problem of averaging on Riemannian manifolds. There are different ways of defining means of data points on a Riemannian manifold, and we studied in detail two of them: the Riemannian L^p mean (in Chapters 3 and 4) and the recursive-iterative means (in Chapter 5). The most immediate question is that of existence and uniqueness of a defined mean. In Theorem 3.2.1 one such a question is answered for Riemannian L^p means. This result improves upon previous results such as the one in [61]. However, there might be room for further improvement, e.g., in Theorem 3.2.1 can we get rid of the specific shape of the domain that includes the support of the probability measure (i.e., a ball)?

In Chapter 4, we studied the L^∞ center of mass, smoothness and convexity properties of the Riemannian L^p mean, and averaging or sensitivities properties of the L^2 mean. With regards to convexity there are still several open questions as explained in Section 4.5 which especially might be of interest to pure geometers, as the convex hull of finite points in a Riemannian manifold is still a mystery. Possible

practical implications of the non-averaging properties whcih we explained in Section 4.6 can be a subject for further research. The fact is that in a manifold of positive curvature, our intuition about the notion of average or mean might turn out to be wrong, and this issue deserves further experimental study. Also we introduced the notion of weak averaging property which is a global notion, and in Theorem 4.6.1 we identified the radius of a ball containing the data points to guarantee the weak averaging. We also conjecture that this radius can be improved (see Conjecture 4.6.1). We did not study in detail the computational methods for finding the L^p mean. However, convergence analysis of gradient descent and other algorithms are widely open questions. See [61] for a result which guarantees convergence of the gradient descent for finding the L^2 mean in symmetric spaces, provided the data points are in a ball which is smaller than the one that is needed to have a unique mean.

In Chapter 5 we studied a large class of recursive-iterative means which are defined based on the notion of mean-invariance. Our analysis generalizes the theories developed in [5] and [59]. The notion of a primitive vectorial mean function and the use of the radius of the minimal ball as a Lyapunov function allowed us to develop a theory, in Section 5.2, which parallels the Perron-Frobenius theory about the infinite power of a primitive stochastic matrix. Of course, an important part of the Perron-Frobenius theorem is the explicit description of the limit mean function, which in a Riemannian manifold is very difficult describe. We also studied the dynamical system point of view of recursive-iterative means. We also studied the simple pairwise mean and its relation with Perimeter Shrinkage scheme and cyclic

pursuit on manifolds. An open problem in this regard is to design an algorithm which brings a set of agents in cyclic pursuit to a closed geodesic in a stable fashion. We also studied weighted MI and Generalized Weighted MI (GWMI) means which can resemble the behavior of the L^p means. A main shortcoming of the recursive-iterative means is that it is not easy to associate statistical notions to them. For example, it is not obvious what the best way is to define a quantity associated with the MI mean which parallels the variance in the case of the L^2 mean.

In Chapter 6, we focused on results specific to $SO(n)$ and $G_k(n)$. These manifolds appear naturally in many applications, and numerical computations on them boil down to standard matrix calculations. Our major focus was on identifying 2-norm Finsler balls of appropriate radius as domains of existence and uniqueness of the L^p and MI means. Our efforts were not completely successful especially with respect to the L^p mean. In particular, we were only able to show that in a 2-norm ball of radius smaller than half of the largest strongly convex 2-norm Finsler ball, the local (and not global) L^p mean exists uniquely (see Theorem 6.5.1). Also while we could show that in the largest convex Finsler balls the MI and WMI means exist uniquely, we were not able to extend this result to GWMI means (see Theorem 6.6.1 and remarks afterwards). As explained in the remarks of Theorem 6.6.1, we conjecture that these results can be improved. We also studied some computational aspects of averaging on $SO(n)$ and $G_k(n)$. In particular, using some standard embeddings of $SO(n)$ and $G_k(n)$ in Euclidean space and in the space of symmetric matrices, respectively, we gave the formulas to find the midpoint of the minimizing geodesic between two points without the use of the matrix exponential and logarithm. This

calculation can be used to efficiently find the MI and pairwise means. We also studied two simple examples: the Nelder-Mead direct search algorithm on $SO(n)$ and the K-means clustering algorithm on $G_k(n)$ both of which have components that can be implemented efficiently via the methods we introduced.

Besides the above mentioned open problems which are directly related to our work, there is a major research direction that can be pursued. An important class of problems arises when we know that a data set belongs to an unknown manifold, e.g., a low dimensional manifold and we still want to process this data set. Therefore, one has to both learn the manifold and its exponential map, and perform data analysis. Learning a manifold has close relations with approximating a manifold with a graph (see e.g., [86, 57, 8]), and it would be very interesting to study existence and uniqueness conditions and other properties of the means based on the quantities associated with the graph of the data.

Bibliography

[1] U. Abresch and W. T. Meyer. Pinching below $\frac{1}{4}$, injectivity radius, and conjugate radius. *Journal of Differential Geometry*, 40(3):643–691, 1994.

[2] P.-A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematicae*, 80(2):199–220, January 2004.

[3] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008.

[4] J. C. Alvarez and C. Duran. An introduction to Finsler geometry. http://www.math.poly.edu/research/finsler/intro/one.html.

[5] T. Ando, C. K. Li, and R. Mathias. Geometric means. *Linear Algebra and its Applications*, 385:305–334, 2004.

[6] V. Arsigny. *Processing Data in Lie Groups: An Algebraic Approach. Application to Non-Linear Registration and Diffusion Tensor MRI*. PhD thesis, École Polytechnique, Nov. 2006.

[7] E. Begelfor and M. Werman. Affine invariance revisited. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2087–2094, 2006.

[8] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):13731396, 2003.

[9] M. Berger. *A Panoramic View of Riemannian Geometry*. Springer, 2007.

[10] A. Bhattacharya. *Nonparametric Statistics On Manifolds With Applications To Shape Spaces*. PhD thesis, The University of Arizona, Nov. 2008.

[11] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *The Annals of Statistics*, 31(1), 1-29 2003.

[12] D. Le Bihan, J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat. Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging*, 13:534–546, 2001.

[13] A. Bloch and I. Iserles. Commutators of skew-symmetric matrices. *International Journal of Bifurcation and Chaos*, 15:793–801, 2005.

[14] A. M. Bruckstein. Why the ant trails look so straight and nice. *The Mathematical Intelligencer*, 15(2):59–62, June 1993.

[15] D. Burago, Y. Burago, and S. Ivanov. *A Course in Metric Geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, 2001.

[16] P. Buser and H. Karcher. *Gromov's almost flat manifolds*. Société Mathématique de France, 1981.

[17] S.R. Buss and J.P. Fillmore. Spherical averages and application to spherical splines and interpolation. *ACM Transcations on Graphics*, 20(2):95–126, April 2001.

[18] I. Chavel. *Riemannian Geometry: A Modern Introduction*. Cambridge University Press, second edition edition, 2006.

[19] J. Cheeger and D. Gromoll. On the structure of complete manifolds of nonnegative curvature. *Annals of Mathematics*, 96(3):413–443, November 1972.

[20] F. H. Clarke, Yu. S. Ledyaev, R.J. Stern, and P. R. Wolenski. *Nonsmooth Analysis and Control Theory*. Graduate Texts in Mathematics. Springer, 1998.

[21] T. H. Colding and W. P. Minicozzi II. Width and mean curvature flow, July 2007. arXiv0705.3827C.

[22] P. I. Davies and N. J. Higham. A Schur-Parlett algorithm for computing matrix functions. *SIAM Journal on Matrix Analysis and Applications*, 25(2):464–485, 2003.

[23] L. Dieci. Considerations on computing real logarithms of matrices, Hamiltonian logarithms, and skew-symmetric logarithms. *Linear Algebra and its Applications*, 244:35–54, 1996.

[24] M. P. do Carmo. *Riemannian Geometry*. Birkhauser, 1992.

[25] D. W. Dreisigmeyer. Direct search algorithms over Riemannian manifolds, December 2006. Available at http://ddma.lanl.gov/Documents/publications /dreisigm-2007-direct.pdf.

[26] A. Edelman, T. A. Arias, and S. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[27] B. L. Ellerbroek, C. Van Loan, N. P. Pitsianis, and R. J. Plemmons. Optimizing closed-loop adaptive-optics performance with use of multiple control bandwidths. *Journal of Optical Society of America A*, 11(11):2871–2886, November 1991.

[28] N. I. Fisher, T. Lewis, and B. J. J. Embleton. *Statistical Analysis of Spherical Data*. Cambridge University Press, 1987.

[29] P.T. Fletcher, L. Conglin, S.M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, Aug 2004.

[30] K. A. Gallivan, A. Srivastava, X. Liu, and P. Van Dooren. Efficient algorithms for inferences on Grassmann manifolds. In *Proc. 12th IEEE Workshop on Statistical Signal Processing*, St. Louis, October 2003.

[31] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 1996.

[32] D. Groisser. Newton's methods, zeros of vector fields, and the Riemannian center of mass. *Advances in Applied Mathematics*, 33, Nov 2004.

[33] K. Grove. *Riemannian Geometry: A Metric Entrance*. Number 65 in Lecture Note Series. Department of Mathematics, University of Aarhus, 2nd edition, Sept. 2002.

[34] K. Grove and H. Karcher. How to conjugate C^1-close group actions? *Mathematische Zeitschrift*, 132(1):11–20, March 1973.

[35] K. Grove, H. Karcher, and E. A. Ruh. Group actions and curvature. *Inventiones math*, 23:31–48, 1974.

[36] K. Grove, H. Karcher, and E.A. Ruh. Jacobi fields and Finsler metrics on compact Lie groups with an application to differentiable pinching problems. *Math. Ann.*, 211:7–21, 1974.

[37] P. Guber and F. J. Theis. Grassmann clustering. In *Proceedings of the European Signal Processing Conference(EUSIPCO)*, 2006.

[38] B. C. Hall. *Lie Groups, Lie Algebras, and Representations*. Graduate Texts in Mathematics. Springer, 2003.

[39] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, 2001.

[40] S. Helgason. *Differential Geometry and Symmetric Spaces*. Academic Press, 1962.

[41] U. Helmke, K. Hper, and J. Trumpf. Newton's method on Graßmann manifolds, 2007. arXiv:0709.2205v2 [math.OC].

[42] U. Helmke and J.B. Moore. *Optimization and Dynamical Systems*. Springer, 2nd edition, 1996.

[43] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[44] M. Humbert, N. Gey, J. Muller, and C. Esling. Determination of a mean orientation from a cloud of orientations. application to electron back-scattering pattern measurements. *Journal of Applied Crystallography*, 29:662–666, 1996.

[45] A. Iserles, H.Z. Munthe-Kaas, S.P. Norsett, and A. Zanna. Lie-group methods. *Acta Numerica*, 9:215–365, 2000.

[46] J. Jost. *Riemannian Geometry and Geometric Analysis*. Universitext. SPringer, 3rd edition, 2002.

[47] J. Jost and Y.L. Xin. Bernstein type theorems for higher codimension. *Journal Calculus of Variations and Partial Differential Equations*, 9(4):277–296, 1999.

[48] R. W. Heath Jr. Communication and signal processing on the Grassmann manifold. 9th IEEE Workshop on Signal Processing for Wireless Communications 2008 Plenary Talk, Available at: http://spawc2008.org/Plenaries/SPAWC2008_Heath.pdf, July 2008.

[49] E.W. Justh and P.S. Krishnaprasad. Equilibria and steering laws for planar formations. *Systems & Control Letters*, 52:25–38, 2004.

[50] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications of Pure and Applied Mathematics*, XXX:509–541, 1977.

[51] H. Karcher. *Global Differential Geometry*, volume 27 of *MAA Studies in Mathematics*, chapter Riemannian Comparison Constructions, pages 170–222. The Mathematical Association of America, 1989.

[52] W. S. Kendall. Probability, convexity, and harmonic maps with small image I: Uniqueness and fine existence. *Proceedings of the London Mathematical Society*, 61(2):371–406, 2 1990.

[53] H. K. Khalil. *Nonlinear Systems*. Prentice Hall, third edition, 2002.

[54] J. H. Kim, W. Zirwas, and M. Haardt. Efficient feedback via subspace-based channel quantization for distributed cooperative antenna systems with temporally correlated channels. EURASIP Journal on Advances in Signal Process, January 2008. Article No. 131.

[55] T. G. Kolda, R. M. Lewis, and Virginia Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, 45(3):385–482, 2003.

[56] K. A. Krakowski. *Geometrical Methods of Inference*. PhD thesis, The University of Western Australia, Aug. 2002.

[57] S. Lafon and A. B. Lee. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1393–1403, Sept. 2006.

[58] S. Lang. *Fundamentals of Differential Geometry*. Graduate Texts in Mathematics. Springer, 1999.

[59] J. Lawson and Y. Lim. A general framework for extending means to higher orders, 2006. arXiv.org:math/0612293.

[60] M. Lazard and J. Tits. Domaines d'injectivite de l'application exponentielle. *Toplogy*, 4:315–322, 1966.

[61] H. Le. Locating Fréchet means with application to shape spaces. *Advances in Applied Probability*, 33(2):324–338, July 2001.

[62] H. Le and D. G. Kendall. The Riemannian structure of Euclidean shape spaces: A novel environment for statistics. *Annals of Statistics*, 21(3):1225–1271, 1993.

[63] S. Li, M. Choi, H. Kim, and F.C. Park. Geometric direct search algorithms for image registration. *IEEE Transactions on Image Processing*, 16(9):2215 – 2224, Sept. 2007.

[64] D. J. Love, R. W. Heath Jr., and T. Strohmer. Grassmannian beamforming for Multiple-Input Multiple-Output wireless systems. *IEEE Transactions on Information Theory*, 49(10):2735–2747, Oct. 2003.

[65] K. V. Mardia and P. E. Jupp. *Directional Statistics*. Wiley, 2000.

[66] W. Meyer. Toponogov's theorem and applications. Lecture notes for the College on Differential Geometry at Trietse, 1989. Available at http://wwwmath.uni-muenster.de/u/meyer/publications/toponogov.html.

[67] M. Moakher. Means and averaging in the group of rotations. *SIAM Journal on Matrix Analysis and Applications*, 24(1):1–16, 2002.

[68] B. Mondal, R. W. Heath Jr., and L. Hanlen. Quantization on the Grassmann manifold: Applications to precoded MIMO wireless systems. *IEEE Transactions on Signal Processing*, 5(8):1025–1028, March 2005.

[69] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, Feb. 2005.

[70] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

[71] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, January 2007.

[72] J.C. Alvarez Paiva and A.C. Thompson. *Volumes on Normed and Finsler Spaces*, volume 50 of *Riemann-Finsler Geometry*. MSRI, 2004.

[73] D. A. Paley. Stabilization of collective motion on a sphere. *Automatica 45*, 45:212–216, 2009.

[74] F. C. Park and B. Ravani. Smooth invariant interpolation of rotations. *ACM Transactions on Graphics*, 16(3):277–295, July 1997.

[75] X. Pennec. *Statistical Computing on Manifolds for Computational Anatomy*. Habilitation à diriger des recherches, Université Nice Sophia-Antipolis, December 2006.

[76] P. Petersen. *Riemannian Geometry*. Springer, 2006.

[77] D. Petz. Means of positive matrices: Geometry and a conjecture. *Annales Mathematicae et Informaticae*, 32:129–139, 2005.

[78] T. Sakai. *Riemannian Geometry*, volume 149. American Mathematical Society, 1996.

[79] A. Sarlette and R. Sepulchre. Consensus optimization on manifolds. *SIAM Journal on Control and Optimization*, 48(1):56–76, 2009.

[80] A. Sarlette, R. Sepulchre, and N.E. Leonard. Autonomous rigid body attitude synchronization. In *Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, LA (USA)*, pages 2566–2571, December 2007.

[81] A. Sarlette, R. Sepulchre, and N.E. Leonard. Cooperative attitude synchronization in satellite swarms: A consensus approach. In *Proceedings of the 17th IFAC Symposium on Automatic Control in Aerospace, Toulouse (France)*, June 2007.

[82] R. Schneider. *Convex Bodies: The Brunn-Minkowski Theory*. Cambridge University Press, 1993.

[83] K. Shoemake. Animating rotation with quaternion curves. In *ACM SIGGRAPH*, volume 19, pages 245–254, 1985.

[84] S. T. Smith. *Geometric optimization methods for adaptive filtering*. PhD thesis, Harvard University, May 1993.

[85] A. Srivastava. A Bayesian approach to geometric subspace estimation. *IEEE Transactions on Signal Processing*, 48(5):1390–1400, May 2000.

[86] J. Tenenbaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:23192323, 2000.

[87] P. Turaga, A. Veeraraghavan, and R. Chellappa. Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision. In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, June 2008.

[88] J. Wallner and N. Dyn. Convergence and C^1 analysis of subdivision schemes on manifolds by proximity. *Computer Aided Geometric Design*, 22(7):593–622, 2005.

[89] A. Weinstein. Almost invariant submanifolds for compact group actions. *Journal of the European Mathematical Society*, 2(1):53–86, March 2000.

[90] J. Wolfowitz. Products of indecomposable, aperiodic, stochastic matrices. *Proceedings of the American Mathematical Society*, 15:733–736, 1963.

[91] Y. C. Wong. Differential geometry of grassmann manifolds. *Proceedings of the National Academy of Sciences of the United States of America*, 57(3):589–594, March 1967.

[92] M. Zefran and V. Kumar. Planning smooth motions on SE(3). In *IEEE International Conference on Robotic Automation*, pages 121–126, Minneapolis, MN, April 1996.