# Geometric Data Analysis: Advancements of the Statistical Methodology and Applications

Dissertation
zur Erlangung des Grades eines Doktors der Naturwissenschaften
*(Dr. rer. nat.)*

am Fachbereich Mathematik und Informatik
der Freien Universität Berlin

vorgelegt von

## Martin Hanik

Berlin 2023

Erstgutachter:

    Professor Dr. Christof Schütte, Freie Universität Berlin, Deutschland

Zweitgutachter:

    Professor Dr. Sarang Joshi, University of Utah, USA

Tag der Disputation: 13. Juli 2023

**Abstract**

Data analysis has become fundamental to our society and comes in multiple facets and approaches. Nevertheless, in research and applications, the focus was primarily on data from Euclidean vector spaces. Consequently, the majority of methods that are applied today are not suited for more general data types. Driven by needs from fields like image processing, (medical) shape analysis, and network analysis, more and more attention has recently been given to data from non-Euclidean spaces—particularly (curved) manifolds. It has led to the field of *geometric data analysis* whose methods explicitly take the structure (for example, the topology and geometry) of the underlying space into account. This thesis contributes to the methodology of geometric data analysis by generalizing several fundamental notions from multivariate statistics to manifolds. We thereby focus on two different viewpoints. First, we use Riemannian structures to derive a novel regression scheme for general manifolds that relies on splines of generalized Bézier curves. It can accurately model non-geodesic relationships, for example, time-dependent trends with saturation effects or cyclic trends. Since Bézier curves can be evaluated with the constructive de Casteljau algorithm, working with data from manifolds of high dimensions (for example, a hundred thousand or more) is feasible. Relying on the regression, we further develop a hierarchical statistical model for an adequate analysis of longitudinal data in manifolds, and a method to control for confounding variables. We secondly focus on data that is not only manifold- but even Lie group-valued, which is frequently the case in applications. We can only achieve this by endowing the group with an affine connection structure that is generally not Riemannian. Utilizing it, we derive generalizations of several well-known dissimilarity measures between data distributions that can be used for various tasks, including hypothesis testing. Invariance under data translations is proven, and a connection to continuous distributions is given for one measure.

A further central contribution of this thesis is that it shows use cases for all notions in *real-world* applications, particularly in problems from shape analysis in medical imaging and archaeology. We can replicate or further quantify several known findings for shape changes of the femur and the right hippocampus under osteoarthritis and Alzheimer's, respectively. Furthermore, in an archaeological application, we obtain new insights into the construction principles of ancient sundials. Last but not least, we use the geometric structure underlying human brain connectomes to predict cognitive scores. Utilizing a sample selection procedure, we obtain state-of-the-art results.

iv

# Acknowledgements

## Publication List

This thesis is based on six (first-author) works of the author, which were published in scientific journals, as part of conference proceedings, or are currently submitted to a journal. Two of them received awards. The paper "Nonlinear Regression on Manifolds for Shape Analysis using Intrinsic Bézier Splines" was valued with a Student Travel Award for the present author at the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2020, while "Bi-invariant Two-Sample Tests in Lie Groups for Shape Analysis" received the Best Paper Award at the "Shape in Medical Imaging" workshop at MICCAI 2020. The full list of works reads:

M. Hanik, M. A. Demirtaş, M. A. Gharsallaoui, and I. Rekik. "Predicting cognitive scores with graph neural networks through sample selection learning". In: *Brain Imaging and Behavior* 16.3 (2022), pp. 1123–1138. DOI: 10.1007/s11682-021-00585-7. © The Author(s) 2021

M. Hanik, B. Ducke, H.-C. Hege, F. Fless, and C. von Tycowicz. "Intrinsic shape analysis in archaeology: A case study on ancient sundials". In: *ACM Journal on Computing and Cultural Heritage*. 2023. DOI: 10.1145/3606698. © The Author(s) 2023

M. Hanik, H.-C. Hege, A. Hennemuth, and C. von Tycowicz. "Nonlinear Regression on Manifolds for Shape Analysis using Intrinsic Bézier Splines". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. 2020, pp. 617–626. DOI: 10.1007/978-3-030-59719-1_60. © Springer Nature Switzerland AG 2020

M. Hanik, H.-C. Hege, and C. von Tycowicz. "Bi-invariant Two-Sample Tests in Lie Groups for Shape Analysis". In: *Shape in Medical Imaging*. 2020, pp. 44–54. DOI: 10.1007/978-3-030-61056-2_4. © Springer Nature Switzerland AG 2020

M. Hanik, H.-C. Hege, and C. von Tycowicz. "Bi-invariant Dissimilarity Measures for Sample Distributions in Lie Groups". In: *SIAM Journal on Mathematics of Data Science*. 4.4 (2022), pp. 1223–1249. DOI: 10.1137/21M1410373. Published under the terms of the Creative Commons 4.0 license

M. Hanik, H.-C. Hege, and C. von Tycowicz. "A Nonlinear Hierarchical Model for Longitudinal Data on Manifolds". In: *2020 IEEE 19th International Symposium on Biomedical Imaging (ISBI 2022)*. 2022, pp. 1–5. DOI: 10.1109/ISBI52829.2022.9761465. © 2022 IEEE

## Declaration of authorship

Name: Hanik
First Name: Martin

I declare to the Freie Universität Berlin that I have completed the submitted dissertation independently and without the use of sources and aids other than those indicated. The present thesis is free of plagiarism. I have marked as such all statements that are taken literally or in content from writings other than those declared on the previous page. This dissertation has not been submitted in the same or similar form in any previous doctoral procedure.

I agree to have my thesis examined by a plagiarism examination software.

Date:                Signature:

# Contents

# Introduction

Mathematical data analysis has become fundamental to many applications, impacting our lives profoundly. Indeed, almost every scientific discipline nowadays uses its tools. This development has come with a paradigm shift away from traditional *modeling* of systems, which usually requires ample a priori knowledge, towards *learning* the underlying laws directly from the data [9, 132, 263]; the ever-increasing popularity of machine—especially deep—learning techniques exemplifies this trend very well [6, 151]. Distilling knowledge purely from (numerical) data works best if there are large amounts available and sufficient realism and variability are included; in this case, it might even be advisable to neglect some of the information to increase computing speed. But when only a few samples are available, it is vital to utilize every bit of information for good results. Indeed, this can be very relevant in practice: Small sets of samples frequently arise, for example, in medical applications (for example, because the use of an intervention is restricted) and archaeology (for example, because findings are scarce). Furthermore, they are also common when we deal with high-dimensional data because the dimension of the underlying space strongly influences how "large" a given sample size is (the ratio of the number of unknown parameters and the number of samples being the deciding factor [42]). Today, data from spaces with several hundred thousand dimensions is not uncommon [42, 73, 135, 189, 245] making sophisticated analysis tools a necessity.

A pillar of the methodology underlying modern data analysis is (mathematical) statistics. Although traditional multivariate statistics plays an undiminished role, the scope has widened considerably over the last decades, with larger and more complex data types being of interest. This development is fuelled by the ongoing increase in computing power, which makes computationally more demanding analysis tools usable in practice. Just 20 years ago, geometric relations between samples differing from the usual Euclidean vector space structure were often overlooked; therefore, most data analysis methods are still formulated exclusively for Euclidean space [51]. Thus, we often do *not* take all the available information into account, as there are many cases in which data comes from other spaces: for example, different manifolds [54, 184, 191, 200], graph [7, 62, 257], and even stratified [85, 128, 142] spaces. One reason why this was not exploited more in the last century is that methods for non-Euclidean spaces can be computationally demanding; today, processing power has increased so much that they can be employed sensibly—also for high dimensional data. Indeed, over the last two decades, several data analysis techniques have been generalized to manifold-valued data with success [108, 189]. Most of them are formulated for *Riemannian* manifolds and exploit their rich structure. For example, the covariance (and

1

higher moments) can be defined [184] using the Frechét mean [95] as the center of the data distribution.[1] Often (but not always) relying on the latter, much work has been done to make *principal component analysis*—one of the most widely used dimension reduction techniques in multivariate statistics—available [93, 129, 185, 229]. Regression analysis has also been investigated, leading, amongst others, to generalizations of linear [89], kernel [69], and Gaussian process regression [161]. Other examples are the Riemannian median [92], dissimilarity measures of sample distributions [126, 172], and several methods from non-parametric statistics [182]. There are also approaches for stratified spaces [85, 129, 182, 185].

A relatively recent research direction investigates the benefits of non-Riemannian structures for statistics in manifolds. For Lie groups, using a non-metric affine connection has fundamental advantages [186, 188]: It allows to define notions like the mean in a way that is compatible with the group operations, resulting in highly desirable invariance/equivariance under data translation.

Albeit the progress, much more is left to be done [128], and there is an ongoing research effort to make the geometry of the data space usable for analysis. This thesis should be understood as part of this process: *It develops new methods for (statistical) geometric data analysis and identifies applications.* It focuses on manifold-valued data, with a particular emphasis on Lie groups in some chapters. While we take the Riemannian viewpoint to derive methods for general manifolds, we adopt the more recent affine viewpoint for Lie groups to benefit from its compatibility with the group structure. Utilizing both, we make the following contributions to the theory of geometric data analysis in this thesis:

- We generalize polynomial (spline) regression to manifolds with Riemannian structure using Bézier curves. A benefit of our approach is that its components can often be computed from explicit formulas, leading to efficient regression algorithms.

- Building upon the new regression scheme, we derive a novel hierarchical statistical model for the analysis of longitudinal manifold-valued data that can capture non-geodesic relations.

- We propose a procedure that allows controlling for continuous confounding variables in manifold-valued data through normalization.

- Relying on the affine viewpoint, we generalize well-known dissimilarity measures for data distributions to Lie groups and prove invariance properties under group operations for them. For one notion—the *bi-invariant Bhattacharyya distance*—we also investigate a natural relation to continuous distributions in the group.

Shape analysis is an important application area for mathematical statistics, in which manifold-valued data often arises. In this field, one wants to infer knowledge about the geometric shape of (usually) two or three-dimensional objects [12, 75, 142]. Its methods are used very often in the field of medical imaging [189] to analyze

---

[1]The Fréchet mean was introduced by Maurice René Fréchet already in the 1940s.

shapes of anatomical structures; apart from this, shape analysis has found applications, for example, in biology [75] and archaeology [142]. We apply our new methods for shape analysis in medical image analysis and archaeology. Thereby, we achieve the following:

- We can accurately capture the development of osteoarthritis (OA) in the knee with the new regression technique; thereby, we observe an acceleration of disease progression which shows that higher-order regression is beneficial in this application.

- Using the novel dissimilarity measures, we perform hypothesis tests for known differences in the shape of the hippocampus under Alzheimer's and for known differences in knee configuration under (OA). We observe statistically significant differences in both cases, verifying the sensitivity of the notions in real-world scenarios.

- Utilizing information we obtain by applying our hierarchical statistical model, we successfully classify distal femur bones according to the development of OA.

- Applying geodesic[2] regression in an archaeological context, we show that ancient Roman craftsmen adapted the bending of the shadow-receiving surface to the installation site when constructing sundials.

Furthermore, we build a deep (graph) neural network that predicts cognitive scores from brain connectome data. By viewing graphs as elements of a manifold, we preselect a subset of samples such that training only on them improves the network's performance yielding better results than state-of-the-art methods.

The thesis consists of three parts: Part II and III contain the results of the author's research, while the necessary theory from differential geometry and an overview of the relevant part of geometric statistics are given in Part I. Part II deals with theoretical and algorithmic findings, whereas Part III showcases their transfer to applications. On a finer level, the structure is as follows.

Part I consists of three chapters: In Chapter 1, we introduce the necessary notions from differential geometry. In Chapter 2, we recall Bézier splines for Riemannian manifolds and summarize the mathematical theory of shape. In Chapter 3, we review existing methods for statistics on manifolds; in doing so, we contrast the Riemannian and the affine viewpoint for Lie groups to highlight the differences between them.

Part II consists of two chapters: In Chapter 4, we focus on Lie group-valued data and generalize dissimilarity measures for data in a bi-invariant way. Changing to the Riemannian viewpoint for data in general manifolds, we introduce a novel generalization of higher-order polynomial regression in Chapter 5. Using it, we propose a method to normalize parameter-dependent data and define a hierarchical model for longitudinal data.

Part III consists of three chapters: In Chapter 6, several applications of the new methods to shape analysis in medicine are shown, while in Chapter 7, we use them

---

[2]Geodesic regression was introduced by Fletcher in [89]. It is an important special case of the more general spline regression we present.

to analyze the shape of certain archaeological artifacts. Finally, in Chapter 8, we use tools from geometric data analysis in manifolds to predict cognitive scores from brain connectomes via graph neural networks.

# Part I

# Background

# Chapter 1

# Foundations of Differential Geometry

In this section, we summarize the necessary background from differential geometry needed for later chapters. Good references on this are, amongst many others, [45], [155], and [197]. The latter, in particular, gives a detailed account of affine connection structures; a standard reference on Lie groups and symmetric spaces is [121]. We also shortly summarize shape analysis—a field with many interesting real-world applications in which data in both Lie groups and general manifolds naturally arises. We will tackle some problems that belong to it in the last part of the thesis. Throughout this work, we use "smooth" synonymously with "infinitely often differentiable

## 1.1 Manifolds, Lie Groups, and their Basic Properties

### 1.1.1 Smooth Manifolds

Informally, a manifold locally "looks" like Euclidean space. It took mathematicians many years to formalize this intuition. Underlying the formal definition is the notion of a chart, which allows obtaining "cartographies" of sets. Let $M$ be a Hausdorff topological space that has a countable basis of topology. (For the foundations of topology, we refer, for example, to [16].) A *chart* on $M$ is a pair $(U, \phi)$, where $U \subseteq M$ is open and $\phi : U \to \mathbb{R}^d$ a homeomorphism. We often also say that $\phi = (x^1, \dots, x^d)$ is a chart of $U$ when we are interested in the component functions of $\phi$, which are called *local coordinates* of $U$. Two charts $(U_1, \phi_1)$ and $(U_2, \phi_2)$ on $M$ are *compatible*, if either $U_1 \cap U_2 = \emptyset$ or $\phi_2 \circ \phi_1^{-1} : \phi_1(U_1 \cap U_2) \to \phi_2(U_1 \cap U_2)$ is a $C^\infty$-diffeomorphism[1]; see Figure 1.1 for a visualization. If there is a collection $\{(U_i, \phi_i)\}_{i \in I}$ of compatible charts such that $\cup_{i \in I} U_i = M$, then we call the collection *atlas*. A *maximal atlas* is then an atlas that contains all possible compatible charts. Finally, if $M$ is endowed with a maximal atlas, then we call it *smooth manifold of dimension d*. There are many examples of smooth manifolds, including (open subsets of) vector spaces, spheres, and sets of matrices like the space of positive definite matrices.

---

[1] It is possible to allow for less restrictive transitions between charts (for example, $C^k$-diffeomorphisms) but we will not need this in this work.

Given two smooth manifolds $M$ and $N$, a *smooth function* $f : M \to N$ can be defined unambiguously by requiring that $\Phi := \psi \circ f \circ \phi^{-1}$ is smooth for every chart $\phi$ of $M$ and $\psi$ of $N$ such that $\Phi$ is well-defined. We will often deal with the case $N = \mathbb{R}$; therefore, the set of smooth functions from $M$ into $\mathbb{R}$ is denoted by $C^\infty(M)$ in the following. Throughout this work, we always assume that manifolds and functions are smooth. Therefore, we omit the word "smooth" and the $C^\infty$-prefix in the following unless we want to emphasize it.
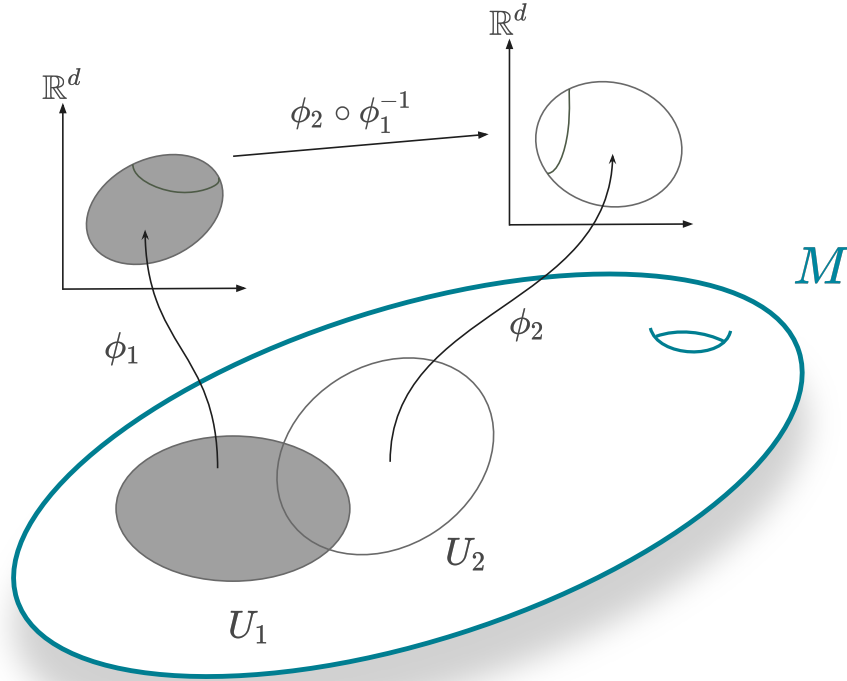


Figure 1.1: Charts $\phi_1$ and $\phi_2$ on a smooth manifold $M$ with domains $U_1$ and $U_2$, respectively. They are compatible if $\phi_2 \circ \phi_1^{-1} : \phi_1(U_1 \cap U_2) \to \phi_2(U_1 \cap U_2)$ is smooth.

Many notions of interest to us are instances of the following general structure:

**Definition 1.1.1** (Vector bundle). *A smooth vector bundle of rank $k$ is a triple $(E, \pi, M)$, where $E$ and $M$ are smooth manifolds; $\pi : E \to M$ is the smooth, surjective projection map; each fiber $E_p := \pi^{-1}(\{p\})$ has the structure of a $k$-dimensional vector space; and the following "triviality condition" is satisfied: For each $p \in M$ there is a neighborhood $U \subseteq M$ and a diffeomorphism*

$$\Phi : \pi^{-1}(U) \to U \times \mathbb{R}^k$$

*such that for every $q \in U$*

$$\Phi\big|_{E_q} : E_q \to \{q\} \times \mathbb{R}^k$$

*is a vector space isomorphism, that is, a bijective linear map.*

Throughout this thesis, it will always be clear which projection $\pi$ is meant and what the rank $k$ is; therefore, we will only say that $E$ is a vector bundle over $M$. Also

important are structure-preserving maps between two bundles over the same base manifold: An isomorphism of vector bundles $(E, \pi_1, M)$ and $(F, \pi_2, M)$ is a diffeomorphism $f : E \to F$ such that $\pi_1 = \pi_2 \circ f$ (that is, $f$ maps fibers to fibers) and $f|_{E_p} : E_p \to F_p$ is a vector space isomorphism for all $p \in M$. Finally, a *section* of a vector bundle $(E, \pi, M)$ is a map $\xi : M \to E$ such that $\pi(\xi(p)) = p$ for all $p \in M$. We denote the set of all sections over $E$ by $\Gamma(E)$.

Many notions from calculus can now be generalized to manifolds. Importantly, we can define a tangent space at every point. Let $f \in C^\infty(M)$ and $\alpha_1, \alpha_2 : (-\varepsilon, \varepsilon) \to M$, $\varepsilon > 0$, be curves through $M$ with $\alpha_1(0) = \alpha_2(0) = p$. We call both equivalent at $p$ if $(f \circ \alpha_1)'(0) = (f \circ \alpha_2)'(0)$ for all $f \in C^\infty(M)$. It can be shown that this defines an equivalence relation. A *tangent vector at $p$* is then a linear map $v : C^\infty(M) \to \mathbb{R}$ corresponding to an equivalence class $[\alpha]$ of curves through $p$ such that $vf := (f \circ \alpha)'(0)$ for all $f \in C^\infty(M)$. We can visualize tangent vectors by imagining that $M$ sits (that is, is embedded) in some higher dimensional $\mathbb{R}^D$. (Thanks to the celebrated Nash embedding theorem, there always is $D \geq 1$ such that this is possible.) A tangent vector corresponds then naturally to a usual "arrow vector" in $\mathbb{R}^D$; see Figure 1.2 for a visualization.
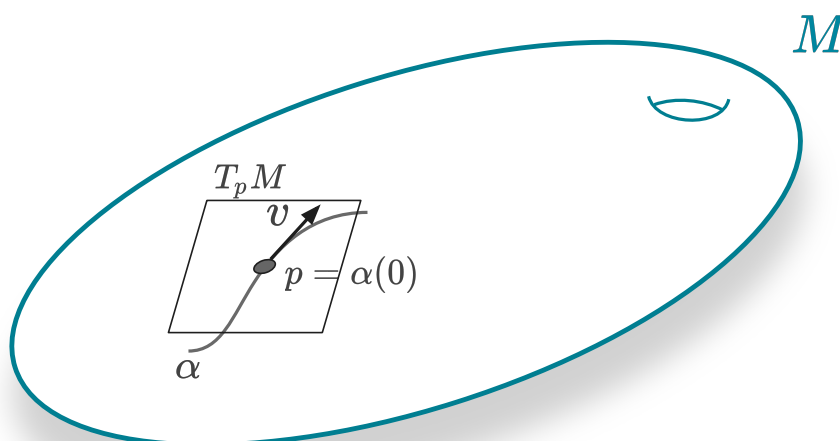


Figure 1.2: Tangent vector $v$ at $p \in M$ defined by a curve $\alpha$.

As in standard calculus, the set of all tangent vectors at $p$ forms a $d$-dimensional vector space: the *tangent space $T_pM$ at $p$*. Since tangent spaces at different points consist of disparate objects, their disjoint union $TM := \sqcup_{p \in M} T_pM$ is of interest; it is called the *tangent bundle of $M$*. It can be shown that $TM$ is not only a manifold. Moreover, it is also a vector bundle over $M$ (the projection mapping each tangent vector to its foot point). A smooth *vector field $X$* on $M$ can thus be defined as a smooth section of the vector bundle $TM$; that is, $X \in \Gamma(TM)$. We often use $X_p$ instead of $X(p)$. Note that every $X \in \Gamma(TM)$ defines a self-map $f \mapsto Xf$ on $C^\infty(M)$ by $Xf(p) := X_pf$; again, we will write $X_pf$ instead of $Xf(p)$. Another important notion is a vector field along a curve $\alpha : [a, b] \to M$, which is a smooth map $X : [a, b] \to TM$ such that $X(t) \in T_{\alpha(t)}M$ for all $t \in [a, b]$.

Corresponding to a tangent space $T_pM$, there also is the *cotangent space*

$$T_p^*M := \{\omega : T_pM \to \mathbb{R} \mid \omega \text{ is linear}\}$$

of linear forms on $T_pM$; elements of $T_p^*M$ are also called *cotangent vectors*. Mirroring the construction of $TM$, the *cotangent bundle* is the disjoint union

$$T^*M := \sqcup_{p \in M} T_p^*M.$$

It also constitutes a vector bundle over $M$ whose sections are called *1-forms*.

Each tangent space has a basis consisting of $d$ vectors. A more interesting question is whether there are $d$ vector fields $(X_1, \ldots, X_n) \in (\Gamma(TM))^n$ such that $((X_1)_p, \ldots, (X_d)_p)$ form a basis *at every point* $p$. The answer, in general, is "No".[2] If a $d$-tuple of vector fields $(X_1, \ldots, X_d)$ with this property exists, we call it (global) *frame field* on $M$. Local frame fields (that is, fields defined only in some coordinate neighborhood $U$ and forming bases at each $p \in U$) exist around every point. Indeed, let $\phi := (x^1, \ldots, x^d)$ be coordinates of $U$, and define vector fields $\mathrm{d}/\mathrm{d}x^1, \ldots, \mathrm{d}/\mathrm{d}x^d \in \Gamma(TU)$ by

$$\frac{d}{dx^i}\Big|_p f := \partial_i(f \circ \phi^{-1})(\phi(p))$$

for all $p \in U$ and $f \in C^\infty(U)$.[3] (Here, $\partial_i$ denotes the partial derivative with respect to the $i$-th canonical coordinate of $\mathbb{R}^d$.) Then, $(\mathrm{d}/\mathrm{d}x^1, \ldots, \mathrm{d}/\mathrm{d}x^d) \in (\Gamma(TU))^d$ constitutes the so-called *coordinate frame field* of $\phi$ over $U$. Importantly, to every coordinate frame $(\mathrm{d}/\mathrm{d}x^1, \ldots, \mathrm{d}/\mathrm{d}x^d)$, there also exists the *coframe* $(\mathrm{d}x^1, \ldots, \mathrm{d}x^d)$. It is the frame field of $T^*U$ dual to the coordinate frame, that is,

$$\mathrm{d}x^i\left(\frac{\mathrm{d}}{\mathrm{d}x^j}\right) = \delta_{ij}$$

for all $i, j = 1, \ldots, d$. (Here, for all $p \in U$, $\delta_{ij}(p) := 1$ if $i = j$ and $\delta_{ij}(p) := 0$ otherwise.) Also worth noting is that to every basis $(v_1, \ldots, v_d)$ of a tangent space $T_pM$, there are coordinates $\phi = (x^1, \ldots, x^d)$ such that $(\mathrm{d}/\mathrm{d}x^1|_p, \ldots, \mathrm{d}/\mathrm{d}x^d|_p) = (v_1, \ldots, v_d)$, that is, any basis of every tangent space is the restriction of some coordinate frame to the given point.

We now turn to the transformation behavior of 1-forms under coordinate changes. Let $\phi_1 = (x^1, \ldots, x^d)$ and $\phi_2 = (\widetilde{x}^1, \ldots, \widetilde{x}^d)$ be charts whose domains of definition have non-empty intersection $U$. If $X \in \Gamma(TU)$, then $X = \sum_{i=1}^d v_i \frac{\mathrm{d}}{\mathrm{d}x^i}$ with $v_i \in C^\infty(U)$ for $i = 1, \ldots, d$. The components of $X$ transform with the Jacobi matrix (function) $J = [\partial \widetilde{x}^l / \partial x^i]$ of the coordinate change, that is,

$$X = \sum_{i=1}^d v_i \frac{\mathrm{d}}{\mathrm{d}x^i} = \sum_{l=1}^d \sum_{i=1}^d v_i \frac{\partial \widetilde{x}^l}{\partial x^i} \frac{\mathrm{d}}{\mathrm{d}\widetilde{x}^l}. \tag{1.1}$$

---

[2] A basic counterexample is the two-dimensional sphere $\mathcal{S}^2$: As a consequence of the famous Hairy Ball Theorem, every smooth vector field on $\mathcal{S}^2$ (indeed on every even-dimensional sphere) is zero at some point.

[3] Note that we could also write $(\frac{d}{dx^i})_p$ instead of $\frac{d}{dx^i}\big|_p$. But it is the convention to use the latter for vector fields that come from partial derivatives in a chart.

For $\omega = \sum_{j=1}^{d} \omega_j dx^j \in \Gamma(T^*U)$ we therefore find

$$\omega = \sum_{j=1}^{d} \omega_j dx^j = \sum_{k=1}^{d} \sum_{j=1}^{d} \omega_j \frac{\partial x^j}{\partial \widetilde{x}^k} d\widetilde{x}^k, \tag{1.2}$$

since

$$\sum_{j=1}^{d} \omega_j dx^j \left( \sum_{i=1}^{d} v_i \frac{d}{dx^i} \right) = \sum_{i=1}^{d} \omega_i v_i = \omega(X)$$

and

$$\sum_{k=1}^{d} \sum_{j=1}^{d} \omega_j \frac{\partial x^j}{\partial \widetilde{x}^k} d\widetilde{x}^k \left( \sum_{l=1}^{d} \sum_{i=1}^{d} v_i \frac{\partial \widetilde{x}^l}{\partial x^i} \frac{d}{d\widetilde{x}^l} \right) = \sum_{i=1}^{d} \sum_{j=1}^{d} \omega_j v_i \sum_{l=1}^{d} \frac{\partial x^j}{\partial \widetilde{x}^l} \frac{\partial \widetilde{x}^l}{\partial x^i}$$

$$= \sum_{i=1}^{d} \omega_i v_i = \omega(X).$$

Thus, a 1-form transforms with the transposed inverse $(J^{-1})^T$ of the Jacobi matrix of the coordinate change.

One of the most important notions we deal with is the derivative of a function. Let $M$ and $N$ be manifolds of possibly different dimensions and $f : M \to N$. Then, the *derivative* $df : \Gamma(TM) \to \Gamma(TN)$ of $f$ is defined by $df(X)h := X(h \circ f)$ for all $h \in C^\infty(N)$. We denote the value of $df(X)$ at $p \in M$ (which is in $T_{f(p)}N$) by $d_p f(X)$. The definition of the derivative is visualized in Figure 1.3.
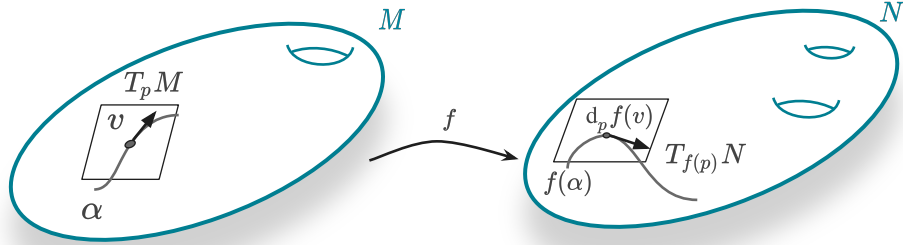


Figure 1.3: Derivative of a smooth map $f$ between two manifolds $M$ and $N$. The vector $v \in T_p M$ is defined through the curve $\alpha$ in $M$. It is mapped to the vector $d_p f(v) \in T_{f(p)}N$ that is induced by the image of $\alpha$ under $f$.

The usual rules from calculus hold for the derivative. In particular, the chain rule applies; that is, if $M, N, O$ are manifolds (of possibly different dimensions) and $f : M \to N, g : N \to O$, then

$$d_p(g \circ f) = d_{f(p)}g \circ d_p f \tag{1.3}$$

for all $p \in M$. Furthermore, when we want to find a local extremum of a function $f \in C^\infty(M)$ in any open neighborhood, the usual necessary condition that $df$ vanishes at the extremum holds. Hence, many optimization strategies known for Euclidean space $\mathbb{R}^d$ can be transferred to manifolds (see, for example, [2]).

Coming back to vector fields, their *integral curves* are defined as in the Euclidean case: Let $p \in M$ and $X \in \Gamma(TM)$. Then, there is $\varepsilon > 0$ and a curve $\alpha : (-\varepsilon, \varepsilon) \to M$, called *integral curve*, with $\alpha(0) = p$ and $\alpha'(t) = X_{\alpha(t)}$ for all $t \in (-\varepsilon, \varepsilon)$. (Note that we write the usual $\alpha'$ instead of $d\alpha(d/dt)$ in the case of curves.) We can interpret $\alpha(t)$ as the point a small ball starting at $p$ would end up at after "flowing" along the vector field $X$ for time $t$.

The so-called *Lie bracket* $[\cdot, \cdot] : \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM)$ of two vector fields $X, Y \in \Gamma(TM)$ quantifies how far $X$ and $Y$ fail to commute. It is defined by $[X, Y]f := XYf - YXf$ for all $f \in C^\infty(M)$. Although it cannot directly be seen, $[X, Y]_p$ is a tangent vector at every $p \in M$.

A final remark, which will become important later in this work, is that the Cartesian product of manifolds is again a manifold (a so-called *product manifold*); its dimension is the sum of the dimensions of the individual manifolds, and all the above operations work component-wise for them.

## 1.1.2 Lie Groups

A Lie group $G$ is a manifold that has a compatible group structure; that is, there is a smooth (not necessarily commutative) group operation $G \times G \ni (f, g) \mapsto fg \in G$ with corresponding identity element $e \in G$ such that the inversion map $g \mapsto g^{-1}$ is also smooth. Vector spaces (with addition) are instances of Lie groups. Another example is the general linear group $GL(d)$; that is, the set of all bijective linear mappings on an $d$-dimensional vector space, where the group operation is the composition of mappings, with $e = \mathrm{Id}$ being the identity map. It is well-known that (closed subgroups of) $GL(d)$ can be identified with (closed subgroups of) the group of invertible $d$-by-$d$ matrices with matrix multiplication. Throughout this thesis, we will regularly make use of this identification without mentioning it every time. Whenever we speak of a matrix group, an arbitrary, closed subgroup of $GL(d)$ is meant.

Each $f \in G$ defines two natural automorphisms on $G$: the *left* and *right translation* $L_f : g \mapsto fg$ and $R_f : g \mapsto gf$. Their derivatives $d_h L_f$ and $d_h R_f$ at $h \in G$ map tangent vectors $v \in T_h G$ bijectively to the tangent spaces $T_{fh} G$ and $T_{hf} G$, respectively. We visualize translations and their derivatives in Figure 1.4. Importantly, it holds that

$$T_g G = \{d_e L_g(v) \mid v \in T_e G\} = \{d_e R_g(w) \mid w \in T_e G\}.$$

Thus, each $v \in T_e G$ determines a smooth vector field $X \in \Gamma(TG)$ by $X_g = d_e L_g(v)$ for all $g \in G$. It is called left invariant because $X_{L_f(g)} = d_g L_f(X_g)$ for all $f, g \in G$, that is, the value at a left translated point is the left translated vector. Together with the Lie bracket, left-invariant vector fields form the so-called Lie algebra $\mathfrak{g} \subset \Gamma(TG)$ of $G$. Furthermore, the converse also holds: Every left invariant vector field is uniquely determined by its value at the identity (or at any other point). Consequently, $T_e G$ and $\mathfrak{g}$ are isomorphic[4].

Since $d_e L_g$ is bijective for every $g \in G$ (and $L_g$ smooth in $g$), any basis of $T_e G$ determines a unique *global* frame consisting of left invariant vector fields. In particular, for each $g \in G$, any basis of $T_e G$ can be smoothly transported to $T_g G$ without dependence

---

[4]To be exact, $T_e G$ with the bracket $[X_e, Y_e] := [X, Y]_e$ is isomorphic to $\mathfrak{g}$ with the usual Lie bracket.
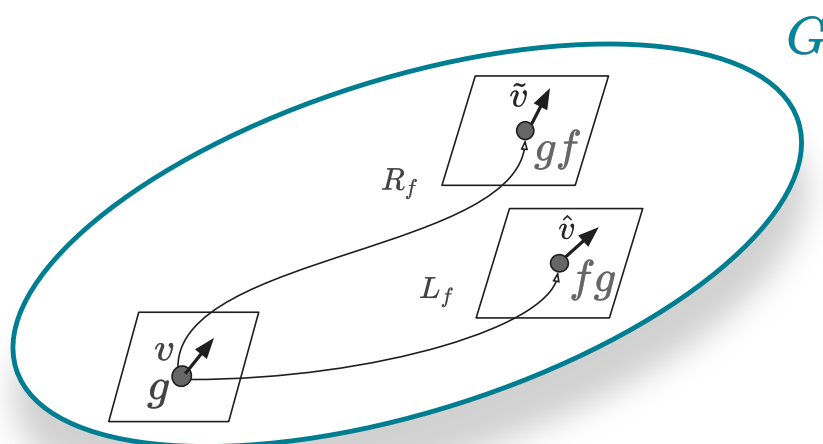
Figure 1.4: Left and right translation and their derivatives in a Lie group $G$. Here, $\widetilde{v} = \mathrm{d}_g R_f(v)$ and $\hat{v} = \mathrm{d}_g L_f(v)$.

on the "path" ("take the basis from the corresponding global frame"). Because of this, we can view $T_e G$ as the *reference tangent space* of $G$.

Right invariant vector fields are defined analogously to left invariant fields: If $v \in G$, then $X := d_e R(v)$ is right invariant, that is, $X_{R_f(g)} = \mathrm{d}_g R_f(X_g)$ for all $f, g \in G$. Left invariant fields work analogously.

The integral curve $\alpha_v : \mathbb{R} \to G$ through $e$ of an invariant (left or right) vector field $X \in \Gamma(TG)$ with $X_e = v$ is defined on the whole of $\mathbb{R}$. It can be shown that $\alpha_v(s+t) = \alpha_v(s)\alpha_v(t) = \alpha_v(t)\alpha_v(s) = \alpha_v(t+s) =$ for all $s, t \in \mathbb{R}$; that is, $\alpha_v$ determines a unique one-parameter subgroup of $G$. This motivates to define the *group exponential* $\exp : T_e G \to G$ by

$$\exp(v) := \alpha_v(1).$$

Note that we also have $\exp(tv) = \alpha_v(t)$ for all $t \in \mathbb{R}$. The group exponential is (also) a diffeomorphism in a neighborhood $V$ of $e$ and, hence, we can define the *group logarithm* log as its inverse there. In the case of matrix groups, they coincide with the matrix exponential and logarithm. Important for us will be the inverse consistency of the group logarithm, that is, for all $g \in V$

$$\log(g^{-1}) = -\log(g). \tag{1.4}$$

Another fundamental automorphism of $G$ is the conjugation $C_f : g \mapsto fgf^{-1}$. We need its differential at the identity, which we call *group adjoint* and denote by $\mathrm{Ad}(f)$. It acts bijectively on vectors $v \in T_e G$ by

$$\mathrm{Ad}(f)v = \mathrm{d}_{f^{-1}} L_f(\mathrm{d}_e R_{f^{-1}}(v)) = \mathrm{d}_f R_{f^{-1}}(\mathrm{d}_e L_f(v)).$$

The group adjoint yields the following crucial relation [186, Thm. 6]: For $f, g \in G$ such that $\log(fg^{-1})$ exists, it links logarithms of left and right translated points according to

$$\log(gf^{-1}) = \mathrm{Ad}(f)\log(f^{-1}g). \tag{1.5}$$

13

A product of Lie groups can be given the product structure, so its group and manifold operations work component-wise.

An important role of Lie groups is that they represent symmetries of geometric spaces. For a manifold $M$, they are formalized by the following notion: Let $G$ be a Lie group. A (smooth) map $\mu : G \times M \to M$, such that $\mu(fg, p) = \mu(f, \mu(g, p))$ and $\mu(e, p) = p$ for all $f, g \in G$, $p \in M$ is called *group action on the left*. (Actions on the right are defined analogously.) Then, for fixed $g \in G$, the map $\mu(g, \cdot) : M \to M$ is an automorphism—a symmetry of $M$.

We finish the section with examples of important Lie groups.

**Example 1.1.** *One of the most important Lie groups is the* general linear group *[252]*

$$\mathrm{GL}(d) := \left\{ A \in \mathbb{R}^{d,d} \mid \det(A) \neq 0 \right\}$$

*with matrix multiplication. It is a manifold of dimension $d^2$ and, at each $A \in G$, the tangent space $T_A\mathrm{GL}(d) = \mathbb{R}^{d,d}$ is the vector space of all d-by-d matrices. (As an open subset of $\mathbb{R}^{d,d}$, charts of $\mathrm{GL}(d)$ are given by the usual coordinate maps of the former.) It is not connected since it consists of two disjoint parts: matrices with positive and negative determinants. The Lie bracket of $\mathrm{GL}(d)$ is the usual matrix commutator. Left and right translations are given by*

$$L_A(B) = AB,$$
$$R_A(B) = BA,$$

*for all $A, B \in \mathrm{GL}(d)$, and*

$$\mathrm{d}_B L_A(C) = AC,$$
$$\mathrm{d}_B R_A(C) = CA,$$

*for all $C \in T_B\mathrm{GL}(d)$. Hence, we arrive at*

$$\mathrm{Ad}(A)C = ACA^{-1}.$$

*The group acts on $\mathbb{R}^d$ on the left by the action $\mu : (A, x) \mapsto Ax$.*

*The group exponential and logarithm are given by the matrix exponential and (principal) matrix logarithm, respectively. That is, denoting the d-by-d identity matrix by I, we have[5]*

$$\exp(C) = \sum_{k=0}^{\infty} \frac{C^k}{k!}$$

*for all $C \in T_I\mathrm{GL}(d)$, while $\log(A)$ is defined for all $A \in \mathrm{GL}(d)$ with no eigenvalues on the (closed) half-line of negative real numbers as the unique solution $C$ of $\exp(C) = A$ whose eigenvalues have imaginary parts in $(-\pi, \pi)$. Both can be computed quite efficiently; see, for example, [124, Secs. 10–11].*

*Crucially, all closed[6] algebraic subgroups of $\mathrm{GL}(d)$ are also Lie groups [121]. For them, left and right translations and actions, the group exponential, and logarithm are the restrictions of*

---

[5]There are several equivalent definitions of both the matrix exponential and logarithm; see [124] for an excellent account.

[6]Here, closed in $\mathrm{GL}(d)$ is meant. A subgroup $G$ is closed in $\mathrm{GL}(d)$ if for any sequence $(A_n)_{n \in \mathbb{N}}$ in $G$ with $\lim_{n \to \infty} = A$, we have that $A \in G$ or $A$ is not invertible.

*those of* $\mathrm{GL}(d)$ *to the respective groups. An example of such a subgroup is the group* $\mathrm{GL}^+(d)$ *of matrices with positive determinants (all linear orientation preserving transformations of* $\mathbb{R}^d$*). Since it is an open subset of* $\mathrm{GL}(d)$*, its dimension is* $d^2$*, and the tangent spaces coincide with those of* $\mathrm{GL}(d)$*.*

*Another important subgroup of* $\mathrm{GL}(d)$ *is the* orthogonal group

$$\mathrm{O}(d) := \{O \in \mathrm{GL}(d) \mid O^T O = I\}.$$

*Its dimension is also* $(d(d-1)/2)$*; its tangent space at* $O$ *is given by all translated* $d$*-by-*$d$ *skew-symmetric matrices*

$$T_O \mathrm{O}(d) = \{OS \in \mathbb{R}^{d,d} \mid S^T = -S\}.$$

*Like* $\mathrm{GL}(d)$*, it has two components of elements with determinants* $\pm 1$*. The part that contains* $I$ *forms the* special orthogonal group

$$\mathrm{SO}(d) := \{R \in \mathrm{GL}(d) \mid R^T R = I, \det(R) = 1\}$$

*consisting of all rotations of* $\mathbb{R}^d$*. Its dimension is* $(d(d-1)/2)$*, too, and its tangent spaces coincide with those of* $\mathrm{O}(d)$*.*

**Example 1.2.** *Another important Lie group is the* Euclidean group *[188], which is the semi-direct[7] product*

$$\mathrm{E}(d) := \mathrm{O}(d) \ltimes \mathbb{R}^d;$$

*Its dimension is* $d(d+1)/2$*. For any two elements* $(O, x), (P, y) \in \mathrm{E}(d)$*, the group operation is given by*

$$(O, x)(P, y) = (OP, x + Oy);$$

*the neutral element of the group is* $(I, 0)$*, and the inverse of each element* $(O, x) \in \mathrm{E}(d)$ *is given by* $(O^T, -O^T x)$*. At each* $(O, x) \in \mathrm{E}(d)$*, the tangent space is the product of the tangent spaces*

$$T_{(O,x)}\mathrm{E}(d) = \{(OS, z) \in \mathbb{R}^{d,d} \times \mathbb{R}^d \mid S^T = -S\}.$$

*The group* $\mathrm{E}(d)$ *acts on* $\mathbb{R}^d$ *on the left by* $\mu : ((O, x), y) \mapsto x + Oy$ *and constitutes the largest class of isometries of* $\mathbb{R}^d$*.*

*We can represent* $\mathrm{E}(d)$ *in homogeneous coordinates, that is, by using the (bijective) mapping*

$$(O, x) \mapsto \begin{bmatrix} O & x \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{d+1,d+1}.$$

*Analogously, tangent vectors* $(S, z) \in T_{(O,x)}$ *are mapped according to*

$$(OS, z) \mapsto \begin{bmatrix} OS & z \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{d+1,d+1}.$$

*In homogeneous coordinates, the group operation coincides with matrix multiplication and, thus,* $\mathrm{E}(d)$ *becomes a (closed) subgroup of* $\mathrm{GL}(d+1)$ *[208], from which it thus inherits the exponential and logarithm maps.*

---

[7]Let $H, N$ be Lie groups and $G = H \times N$. Further, let $G$ be endowed with a group operation. Then, we say that $G$, together with its group operation, is a semi-direct product (denoted in short by $G = H \ltimes N$), if $N$ is a normal subgroup (that is, invariant under conjugation) of $G$ and for every $g \in G$ there are unique $h \in H, n \in N$ such that $g = hn$.

### 1.1.3 Tensors and Tensor Fields

Given a $d$-dimensional manifold $M$, there is the $m$ times contravariant and $n$ times covariant *tensor bundle*

$$\underbrace{TM \otimes \cdots \otimes TM}_{m \text{ times}} \otimes \underbrace{T^*M \otimes \cdots \otimes T^*M}_{n \text{ times}},$$

over $M$ whose fiber at any $p \in M$ is given by

$$\{\beta : \underbrace{T_p^*M \times \cdots \times T_p^*M}_{m \text{ times}} \times \underbrace{T_pM \times \cdots \times T_pM}_{n \text{ times}} \to \mathbb{R} \mid \beta \text{ is multilinear}\}.$$

To shorten the notation, we often write $T_n^m M$ for the tensor bundle. Being a vector bundle of rank $d^{mn}$, its sections are called *m times contravariant and n times covariant tensor fields* over $M$; we usually say $(m,n)$-tensor field instead. In the literature, the word "field" is often omitted; since we are primarily interested in elements of fibers of a tensor bundle, we always call these "tensors" and stick to the phrase "tensor field" when referring to sections of the tensor bundle. Examples of tensors that appear regularly in this thesis are bilinear forms on tangent and cotangent spaces, which are $(0,2)$- and $(2,0)$-tensors, respectively.

Importantly, tensor fields are pointwise objects whose behavior at a point is determined by their restrictions and the values of their inputs at this point [45, p. 101]. This allows us to investigate single tensors but still view them as restrictions of an underlying tensor field to a particular tangent space if the need should arise.

Tensor fields can be added and multiplied pointwise with functions from $C^\infty(M)$. Furthermore, there is the tensor product: Let $S \in \Gamma(T_{n_1}^{m_1}M)$ and $T \in \Gamma(T_{n_2}^{m_2}M)$. Their *tensor product* $S \otimes T$ is an $(m_1 + m_2, n_1 + n_2)$-tensor field that is defined for all $X_1, \ldots, X_{n_1+n_2} \in \Gamma(TM)$ and $\omega_1, \ldots, \omega_{m_1+m_2} \in \Gamma(T^*M)$ by

$$(S \otimes T)(\omega_1, \ldots, \omega_{m_1+m_2}, X_1, \ldots, X_{n_1+n_2}) := S(\omega_1, \ldots, \omega_{m_1}, X_1, \ldots, X_{n_1})$$
$$T(\omega_{m_1+1}, \ldots, \omega_{m_1+m_2}, X_{n_1+1}, \ldots, X_{n_1+n_2}).$$

Importantly, any tensor field in $\Gamma(T_n^m M)$ is a (finite) linear combination of tensor products of $m$ vector fields and $n$ 1-forms.

The tensor product can be used to define chart-induced, local frame fields of the tensor bundle: In a coordinate neighborhood $U$ with chart $\phi = (x^1, \ldots, x^d)$, tensor fields can be written as tensor products of elements of coordinate frames and their coframes; more precisely, an $(m,n)$-tensor field $S$ with *component functions*[8]

$$S^{i_1,\ldots,i_m}_{j_1,\ldots,j_n} := S\left(\mathrm{d}x^{i_1}, \ldots, \mathrm{d}x^{i_m}, \frac{d}{\mathrm{d}x^{j_1}}, \ldots, \frac{d}{\mathrm{d}x^{j_n}}\right) \in C^\infty(U), \tag{1.6}$$

can always be uniquely written as

$$S\big|_U = \sum_{1 \le i_1,\ldots,i_m,j_1,\ldots,j_n \le d} S^{i_1,\ldots,i_m}_{j_1,\ldots,j_n} \frac{d}{\mathrm{d}x^{i_1}} \otimes \cdots \otimes \frac{d}{\mathrm{d}x^{i_m}} \otimes \mathrm{d}x^{j_1} \otimes \cdots \otimes \mathrm{d}x^{j_n}. \tag{1.7}$$

---

[8] In physics, especially in General Relativity, tensor fields are usually identified with their component functions. These are then manipulated according to the algebraic rules that arise from their dependence on the chart. This view of tensors is the foundation of the so-called *Ricci calculus*.

Using Equations (1.1) and (1.2), we can now describe the behavior of tensor fields under coordinate changes: An $m$ times contravariant and $n$ times covariant tensor field is transformed $m$ times by the Jacobi matrix $J$ of the coordinate change and $n$ times by the matrix $(J^{-1})^T$.

Let now $m_1 + n_1 = m_2 + n_2$. It is often of interest to identify tensors of the same rank. To this end, one needs to use a bundle isomorphism between $T^{m_1}_{n_1}(U)$ and $T^{m_2}_{n_2}(U)$. Generally, many isomorphisms could be used, but none of them is distinguished. In contrast, when the manifold is additionally endowed with a Riemannian metric, there is a clear choice of isomorphism. We will come back to this later.

**Remark 1.1.2.** *An isomorphism between $T^{m_1}_{n_1}(U)$ and $T^{m_2}_{n_2}(U)$ that might seem natural at first glance is obtained by fixing coordinates and simply replacing the required amount of tangent/cotangent vector fields in the corresponding decomposition (1.7) of any tensor field in $T^{m_1}_{n_1}(U)$ with their duals. In these coordinates, the component functions of the tensor and those of its image then coincide. However, since coordinate frames and their coframes behave differently under a change of basis, this is generally not true anymore when this isomorphism is written in terms of another coordinate system. In other words, the mathematical form of the component functions depends on the choice of coordinates, so the above isomorphism only seems distinguished when viewed in one (arbitrarily chosen) coordinate system.*

## 1.2 Geometric Structures on Manifolds and Lie Groups

### 1.2.1 Affine Connections and Geodesics

Let $M$ be a manifold. An *affine connection*

$$\nabla : \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM),$$
$$(X, Y) \mapsto \nabla_X Y,$$

on $M$ is a bilinear map such that for all $f \in C^\infty(M), X, Y \in \Gamma(TM)$,

$$(i) \quad \nabla_{fX} Y = f \nabla_X Y,$$
$$(ii) \quad \nabla_X fY = X fY + f \nabla_X Y;$$

that is, it is linear in the first variable and satisfies the product rule in the second. Connections are the natural generalization of the directional derivative along vector fields. (We say that $\nabla_X Y$ is the derivative of $Y$ in direction $X$.) There are many different connections, and without an additional structure on $M$, there is no best choice a priori. If, on the other hand, $M$ is a Lie group or is endowed with a Riemannian metric, then there are canonical choices, as we will see.

With the help of a connection $\nabla$, we can also differentiate vector fields along curves. Let $\alpha : [a, b] \to M$ be a smooth curve and $X$ be a vector field along $\alpha$. Furthermore, let $\widetilde{\alpha}', \widetilde{X} \in \Gamma(TM)$ such that they coincide with $\alpha'$ and $X$ on $\alpha([a, b])$. The *covariant derivative of $X$ along $\alpha$* is then defined by

$$\nabla_{\alpha'} X := \left( \nabla_{\widetilde{\alpha}'} \widetilde{X} \right) \Big|_{\alpha([a,b])};$$

see Figure 1.5 for a visualization. It can be shown that this definition does not depend on the choice of the global vector fields. With this notion, we can introduce parallelism of vectors *along curves*. A vector field $X$ along $\alpha$ is called *parallel* if

$$\nabla_{\alpha'} X = 0; \tag{1.8}$$

we also say that $X_{\alpha(b)}$ is the *parallel transport* of $X_{\alpha(a)}$ to $T_{\alpha(b)}M$. Given $X_{\alpha(a)} = v$, well-known results from the field of differential equations ensure that Equation (1.8) always has a unique solution, that is, the parallel transport of a vector along a curve is well-defined.
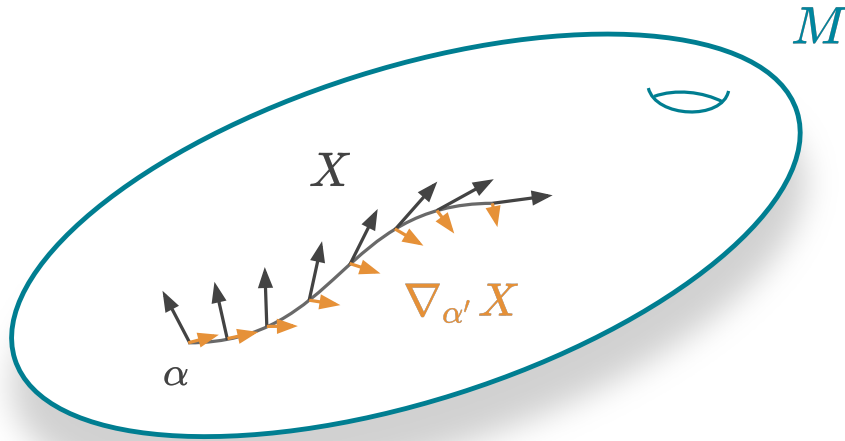


Figure 1.5: Vector field $X$ (grey) along a curve $\alpha$ through $M$. Its covariant derivative $\nabla_{\alpha'} X$ along $\alpha$ is shown in orange.

We can now define a *geodesic* $\gamma : (-\varepsilon, \varepsilon) \to M$ by $\nabla_{\gamma'} \gamma' = 0$ as a curve without acceleration, generalizing the notion of a straight line. Results for differential equations imply that every vector is the tangent vector of a geodesic that is defined on some interval of positive length $2\varepsilon$. Note that affine reparametrizations of $\gamma$ do not change the fact that $\nabla_{\gamma'} \gamma' = 0$; therefore, the interval on which $\gamma$ is defined is arbitrary. Furthermore, we can also define geodesics on closed intervals by taking limits. Thus, through reparametrization, one can then show that for each $p \in M$, there is a neighborhood $V \subseteq T_p M$ such that for all $v \in V$, there is a geodesic $\gamma_v : [0,1] \to M$ with $\gamma_v(0) = p$ and $\gamma_v'(0) = v$. This motivates the definition of the *exponential of the connection* $\nabla$ *at* $p$:

$$\mathrm{Exp}_p : V \to M,$$
$$v \mapsto \gamma_v(1);$$

the map is visualized in Figure 1.6. It turns out that, at every point, the exponential is a local diffeomorphism. We call the local inverse *logarithm of the connection* $\nabla$ and denote it by $\mathrm{Log}_p$. The manifold $M$ is called *geodesically complete* if for every $p \in M$ the exponential $\mathrm{Exp}_p$ is defined on all of $T_p M$. Equivalently, one can require that all geodesic in $M$ extend indefinitely.
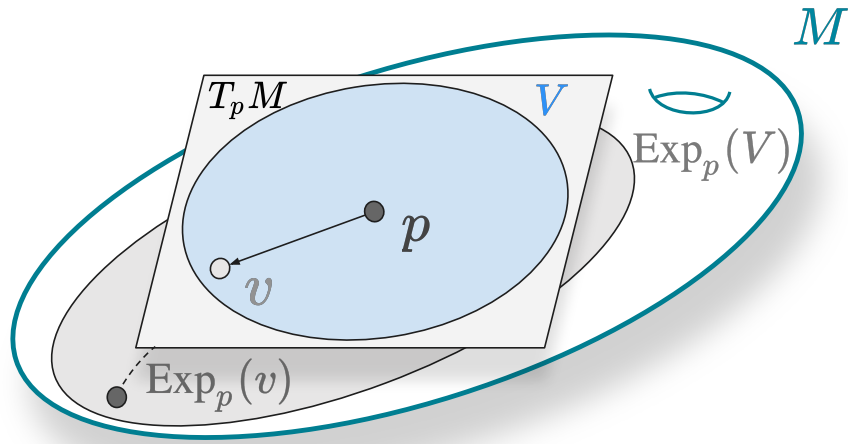
Figure 1.6: Exponential mapping $\mathrm{Exp}_p$ applied to $v \in V \subseteq T_pM$. The broken line indicates the geodesic $\gamma_v$ from $p$ to $\mathrm{Exp}_p(v)$.

A fundamental fact is that every point $p \in M$ has a so-called *normal convex neighborhood*[9] $U$: Each pair $q, r \in U$ can be joined by a unique geodesic $\gamma(\,\cdot\,;q,r) : t \mapsto \gamma(t;q,r)$ with $t \in [0,1]$, which never leaves $U$; see [197, Ch. 1 § 5]. (Throughout this work, we use " $\cdot$ " as a placeholder for a varying parameter.) Furthermore, with $\gamma'(0;p,q) = v$, we find

$$\mathrm{Exp}_p(v) = \gamma(1;p,q)$$

and

$$\mathrm{Log}_p(q) = \gamma'(0;p,q)$$

in $U$.

The connection determines another central object of differential geometry: the *curvature tensor*[10]

$$R : \Gamma(TM) \times \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM),$$
$$(X, Y, Z) \mapsto R(X, Y)Z := \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$

Intuitively, $R$ quantifies the deviation of $M$ from (flat) Euclidean space. It does so by measuring how far partial derivatives $\nabla_X, \nabla_Y$ fail to commute[11]. The last term ensures that a possible non-commutativity of the underlying vector fields themselves does not have an influence.

A core fact of differential geometry is that the parallel transport of vectors depends on the underlying curve whenever $R \neq 0$ (that is, it is not the zero map). Thus, whenever there is curvature, we cannot simply identify different tangent spaces by

---

[9]Other names are also used for such a neighborhood; (at least) "geodesically convex" and "strongly convex" neighborhood can also be found.

[10]There are different sign conventions for the curvature tensor; we follow [197].

[11]They commute in Euclidean space, with the ordinary directional derivative as connection, due to Schwarz's theorem

parallel translating each vector from one to the other (something we always do when working in Euclidean space). Indeed, in general, there is no consistent way of identifying tangent spaces with each other. It only becomes possible when global frame fields exist with which we can consistently relate vectors between tangent spaces (for example, by keeping the coefficients w.r.t. basis constant).

The curvature also determines how fast initially close geodesics deviate from each other (or meet). The notion of the Jacobi field formalizes this observation. Let $\gamma$ be a geodesic in $M$. A *Jacobi field J along $\gamma$* is a solution of the second-order linear ordinary differential equation

$$\nabla_{\gamma'}\nabla_{\gamma'}J + R(J, \gamma')\gamma' = 0; \tag{1.9}$$

an example is shown in Figure 1.7. Most importantly, Jacobi fields describe the spreading of geodesics (globally for geodesically complete manifolds and locally otherwise) in the following sense: Let $\alpha : (-\varepsilon, \varepsilon) \times [0, 1] \to M$ be a geodesic variation, that is, $\alpha$ is a smooth map such that $\gamma_s := \alpha(s, \cdot)$ is a geodesic for each $s \in (-\varepsilon, \varepsilon)$. We denote by $\frac{\partial \alpha}{\partial s}\big|_{(0,t)}$ the vector field along $\gamma_0$ that is defined by

$$\frac{\partial \alpha}{\partial s}\bigg|_{(0,t)} f := \frac{\partial}{\partial s}(f \circ \alpha)(0, t)$$

for all $f \in C^\infty(M)$ and $t \in [0, 1]$. Then

$$J_{\gamma(t)} := \frac{\partial \alpha}{\partial s}\bigg|_{(0,t)}$$

is a Jacobi field along $\gamma_0$.

Now, it is no surprise that Jacobi fields play a central role when one wants to differentiate geodesics with respect to their start and end points. Let $\gamma(\cdot ; p, q) : [0, 1] \to U$ be defined as above with $p, q$ coming from a normal convex neighborhood and $v \in T_pM$. Considering (1.9) with the added constraints $J(0) = v$, $J(1) = 0$ (that is, the boundary value problem with fixed endpoint), we denote its unique solution by $J_v$. Then, the derivative of $\gamma$ with respect to its starting point at $p$ is given by

$$d_p\gamma(t; \cdot, q)(v) = J_v(t) \tag{1.10}$$

for all $t \in [0, 1]$ [29, Sec. 3.1]; see Figure 1.7 for a visualization. Another useful observation is that $\gamma(t; p, q) = \gamma(1 - t; q, p)$ for all $t \in [0, 1]$. Therefore, with $w \in T_qM$, we find

$$d_q\gamma(t; p, \cdot)(w) = d_q\gamma(1 - t; \cdot, p)(w) = J_w(1 - t) \tag{1.11}$$

for all $t \in [0, 1]$. Note that the vector $w$ in $J_w$ encodes the orientation of the underlying geodesic: It is always in the tangent space at the starting point of the geodesic.

In order to compute (1.10) and (1.11), the operator

$$\mathfrak{R} : T_pM \to T_pM,$$

$$v \mapsto R\left(v, \frac{\gamma'(0)}{\|\gamma'(0)\|_p}\right)\frac{\gamma'(0)}{\|\gamma'(0)\|_p}, \tag{1.12}$$

will be important. It is a symmetric linear operator that can thus be diagonalized with real eigenvalues [45, Chs. 4.2 and 5 (Ex. 5)]. Note that $\mathfrak{R}(\gamma'(0)) = 0$, so one eigenvalue is always zero with corresponding (normalized) eigenvector $\gamma'(0)/\|\gamma'(0)\|$.
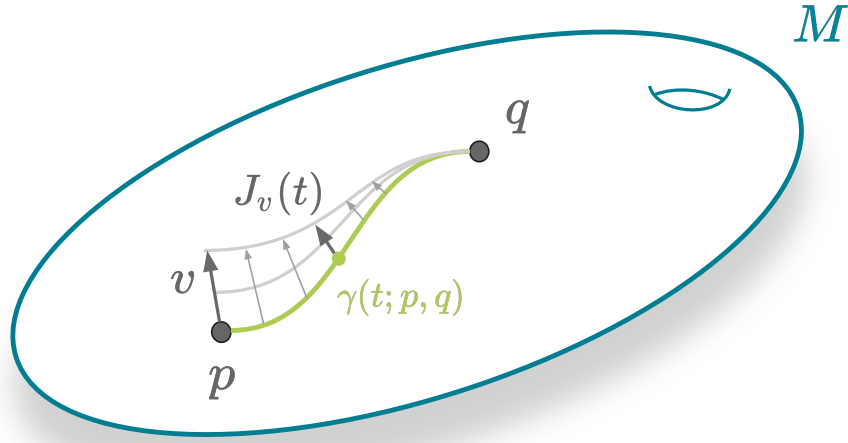
Figure 1.7: Jacobi field $J_v$ along the geodesic $\gamma(t; p, q)$ (green) with boundary values $J_v(0) = v \in T_p M$ and $J_v(1) = 0 \in T_q M$. The grey curves indicate the geodesic variation that corresponds to $J_v$. The vector $J_v(t) \in T_{\gamma(t;p,q)} M$ is also the value of $\mathrm{d}_p \gamma(t; \,\cdot\, , q)$ at $v$, that is, the direction in which $\gamma(t; p, q)$ moves when $p$ is varied in the direction of $v$.

### 1.2.2 Riemannian Metrics

Let $M$ be a manifold. A *Riemannian metric* on $M$ is a smooth map $\langle \,\cdot\, , \,\cdot\, \rangle : \Gamma(TM) \times \Gamma(TM) \to C^\infty(M)$ such that for each $p \in M$ its restriction to $T_p M$ is a Euclidean inner product; that is, for each $p \in M$ the map $\langle \,\cdot\, , \,\cdot\, \rangle_p := \langle \,\cdot\, , \,\cdot\, \rangle|_{T_p M \times T_p M} : T_p M \times T_p M \to \mathbb{R}$ is a symmetric positive definite bilinear form, and $\langle \,\cdot\, , \,\cdot\, \rangle_p$ depends smoothly on $p$. In the following, we will also call the restricted map $\langle \,\cdot\, , \,\cdot\, \rangle_p$ Riemannian metric (or just metric) as it will be clear from the context what is meant. Importantly, every manifold can be endowed with a Riemannian metric.

As in Euclidean space, the metric allows to measure *angles* between and *lengths* of tangent vectors; for $v, w \in T_p M$, they are defined as usual by

$$\|v\|_p := \sqrt{\langle v, v \rangle_p}$$

and

$$\measuredangle(v, w) := \arccos \left( \frac{\langle v, w \rangle_p}{\|v\|_p, \|w\|_p} \right),$$

respectively. Furthermore, if $\alpha : [a, b] \to M$, $[a, b] \subset \mathbb{R}$, is a smooth curve, its *length* can be defined (independently from the parametrization of $\alpha$) by

$$\text{length}(\alpha) := \int_a^b \|\alpha'(t)\|_{\alpha(t)} \, \mathrm{d}t.$$

For $p, q \in M$ this induces the distance (function)

$$\text{dist}(p, q) = \inf\{\text{length}(\alpha) \mid \alpha : [0, 1] \to M \text{ smooth}, \alpha(0) = p, \alpha(1) = q\}.$$

Using the metric we can define the *gradient* $\mathrm{grad} f \in \Gamma(TM)$ of a function $f \in C^\infty(M)$ implicitly by requiring

$$\mathrm{d}f(X) = \langle \mathrm{grad} f, X \rangle$$

for all $X \in \Gamma(TM)$, generalizing the ordinary Euclidean gradient. Importantly, just like in the Euclidean case, $\mathrm{d}_p f = 0$ if and only if $\mathrm{grad}_p f = 0$. Furthermore, the gradient still points in the direction of the steepest ascent, so optimization algorithms relying on gradients can be transferred to Riemannian manifolds [2].

Once we have a metric, the adjoint $A^*$ of a linear operator $A : T_p M \to T_q M$ is given, as usual, by the linear operator from $T_q M$ to $T_p M$ that conserves the scalar product, that is, $\langle Av, w \rangle_q = \langle v, A^* w \rangle_p$ for all $v \in T_p M$, $w \in T_q M$. Further, if $B : T_r M \to T_p M$ is another linear operator, then $(A \circ B)^* = B^* \circ A^*$. An adjoint operator appears, for example, when one wants to calculate the gradient of a composition of functions. Let $f : M \to M$ and $g : M \to \mathbb{R}$. Then, the chain rule for gradients reads

$$\mathrm{grad}_p(g \circ f) = d_p f^* \left( \mathrm{grad}_{f(p)} g \right), \tag{1.13}$$

that is, the gradient of $g$ at $f(p)$ is "transported" to the tangent space at $p$ by the adjoint differential of $f$.

An important class of maps keeps the metric unchanged. Let $M$ and $N$ be manifolds (necessarily of the same dimension) with metrics $\langle \, \cdot \, , \, \cdot \, \rangle^M$ and $\langle \, \cdot \, , \, \cdot \, \rangle^N$, respectively. A diffeomorphism $f : M \to N$ is called *isometry* if

$$\langle v, w \rangle_p^M = \langle \mathrm{d}_p f(v), \mathrm{d}_p f(w) \rangle_{f(p)}^N$$

for all $p \in M$ and $v, w \in T_p M$.

The metric also allows for a different view of the curvature of $M$. To this end, let $p \in M$ and $v, w \in T_p M$ be orthonormal (w.r.t. $\langle \, \cdot \, , \, \cdot \, \rangle_p$). The *sectional curvature* of the plane that is spanned by $v$ and $w$ in $T_p M$ is then defined by

$$K(v, w) := \langle R(v, w)w, v \rangle_p.$$

Being a real-valued measure of how much a manifold deviates from flat Euclidean space, the sectional curvature plays a vital role in many bounds on the maximal sizes of regions with interesting geometric properties; one of those will become important later on.

If $M$ is endowed with a Riemannian metric $\langle \, \cdot \, , \, \cdot \, \rangle$, then there is a canonical choice of affine connection: The *Levi-Civita connection* $\nabla$ is the unique affine connection that satisfies

$$\begin{aligned} &(i) && \nabla_X Y = \nabla_Y X + [X, Y], && \text{(torsion-free)} \\ &(ii) && X\langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle, && \text{(metric)} \end{aligned}$$

for all $X, Y, Z \in \Gamma(TM)$. The Levi-Civita connection conserves angles between parallel transported vectors; its parallel transport operation is thus an isometry of vector spaces. Furthermore, the geodesics of the Levi-Civita connection are locally shortest paths of the Riemannian metric. Therefore, for it we find

$$\mathrm{dist}(p, q) = \|\mathrm{Log}_p(q)\|_p = \|\mathrm{Log}_q(p)\|_q$$

for all $p, q \in M$ such that $\mathrm{Log}_p(q)$ is defined. In particular, if $U \subseteq M$ is a normal convex neighbourhood and $p, q \in U$, then the geodesic $\gamma(\,\cdot\,; p, q)$ is also the unique shortest path between $p$ and $q$.

As is standard in the literature, we call a manifold together with a Riemannian metric and its Levi-Civita connection *Riemannian manifold* and the corresponding curvature tensor *Riemannian curvature tensor*.

When $M$ is endowed with a Riemannian metric, there is a canonical vector bundle isomorphism between certain types of tensor fields. Let $\flat$ denote the metric-induced vector bundle homomorphism between $TM$ and $T^*M$ (the "flat operator") that maps each vector field $X \in \Gamma(TM)$ to the 1-form $X^\flat \in \Gamma(T^*M)$ defined by

$$ X^\flat(Y) = \langle X, Y \rangle $$

for all $Y \in \Gamma(TM)$. It has an inverse (the "sharp operator") denoted by $\sharp$ that maps any $\omega \in \Gamma(T^*M)$ to the vector field $\omega^\sharp \in \Gamma(TM)$ that fulfills

$$ \omega(Y) = \langle \omega^\sharp, Y \rangle $$

for all $Y \in \Gamma(TM)$. Both flat and sharp constitute the *musical isomorphism*. When $m, n \in \mathbb{N}$, $c \in \mathbb{Z}$ such that $m - c \geq 0$ and $n + c \geq 0$, the musical isomorphism induces a bundle isomorphism between $T_n^m M$ and $T_{n+c}^{m-c} M$ through entry-wise application of $\flat$ or $\sharp$. The process of applying $\flat$ and $\sharp$ is also known as *lowering* and *raising of indices*, respectively. The reason for these names will become visible soon.

First, we state how the musical isomorphism works in a coordinate neighborhood $U$. Let $g_{ij} \in C^\infty(U)$ be the component functions of the Riemannian metric in a coordinate neighborhood $U$, that is,

$$ \langle\,\cdot\,,\,\cdot\,\rangle\big|_U = \sum_{i,j=1}^{d} g_{ij}\, \mathrm{d}x^i \otimes \mathrm{d}x^j. $$

Because the metric is symmetric, we have $g_{ij} = g_{ji}$ for all $i, j = 1, \ldots, d$. Let further $g^{ij} \in C^\infty(U)$ be the component functions of the inverse of the matrix $[g_{ij}]$. Then

$$ \left( \frac{\mathrm{d}}{\mathrm{d}x^i} \right)^\flat = \sum_{j=1}^{d} g_{ij}\, \mathrm{d}x^j, \quad i = 1, \ldots, d, \tag{1.14} $$

and

$$ \left( \mathrm{d}x^i \right)^\sharp = \sum_{j=1}^{d} g^{ij}\, \frac{\mathrm{d}}{\mathrm{d}x^j}, \quad i = 1, \ldots, d $$

see, for example, [190, pp. 48–49] for more details. A bundle isomorphism maps an $(m, n)$-tensor field $S$ to an $(m - c, n + c)$-tensor field $\widetilde{S}$, when the latter can be obtained from $S$ by replacing $c$ vector fields in the decomposition (1.7) by their corresponding 1-forms under $\flat$, or vice versa (applying $\sharp$) if $c$ is negative. Thereby, it is important to which position the musical isomorphism is applied; different choices yield different bundle isomorphisms when $m - c > 0$ and $n + c > 0$ (that is, for "mixed" tensors). The phrases "lowering/raising of indices" now come from the fact that, in the notation of Equation (1.7), certain superscripts of the component functions become subscripts,

and the other way round, when a musical isomorphism is applied. (An example is shown below.) Tensor fields on a Riemannian manifold that correspond to each other under *any* such isomorphism are commonly seen as versions of the same tensor that differ in the number of contra- and covariant components and their ordering.

**Example 1.3.** *We now show how "indices are lowered" for (2,0)-tensors fields since this is relevant at several places in this thesis. As above, let $M$ be a Riemannian manifold and $\phi = (x^1, \dots, x^d)$ coordinates of a neighborhood $U$. Let $S \in T_0^2 M$; then*

$$S\big|_U = \sum_{i,j=1}^d S^{ij} \frac{d}{dx^i} \otimes \frac{d}{dx^j}.$$

*Applying Equation (1.14) to both entries results in*

$$\begin{aligned}
\widehat{S}\big|_U &= \sum_{i,j=1}^d S^{ij} \left(\frac{d}{dx^i}\right)^\flat \otimes \left(\frac{d}{dx^j}\right)^\flat \\
&= \sum_{i,j=1}^d S^{ij} \left(\sum_{k=1}^d g_{ik}\, dx^k\right) \otimes \left(\sum_{l=1}^d g_{jl}\, dx^l\right) \\
&= \sum_{k,l=1}^d \sum_{i,j=1}^d g_{ik} g_{jl} S^{ij} dx^k \otimes dx^l \\
&= \sum_{i,j=1}^d \sum_{k,l=1}^d g_{ki} g_{lj} S^{kl} dx^i \otimes dx^j.
\end{aligned}$$

*In the last step we simply re-labeled indices: $i \leftrightarrow k$ and $j \leftrightarrow l$. Using the fact that the metric is symmetric and changing the order of summation yield (omitting the "hat" as is standard)*

$$S_{ij} = \sum_{k,l=1}^d g_{ik} g_{lj} S^{kl}, \quad i,j = 1, \dots, d. \tag{1.15}$$

*Note how the indices of $S^{kl}$ are "lowered" by multiplying with the metric.*

   *Interpreting all functions as entries of matrices (more precisely, matrix-valued functions) $[S_{ij}]$, $[g_{ij}] = [g_{ji}]$, and $[S^{ij}]$, we can then see that Equation (1.15) is equivalent to*

$$[S_{ij}] = [g_{ij}] [S^{ij}] [g_{ij}].$$

*In local coordinates, one can thus describe lowering indices of (2,0)-tensors with matrix algebra. "Raising" the indices of $\widehat{S}_{ij}$ works analogously with the inverse $[g^{ij}]$.*

**Example 1.4.** *Let $M$ be a Riemannian manifold and $\phi = (x^1, \dots, x^d)$ be coordinates of a neighborhood $U$. If the coordinate frame field $(d/dx^1, \dots, d/dx^d)$ is orthonormal, that is,*

$$\left(\frac{d}{dx^1}\bigg|_p, \dots, \frac{d}{dx^d}\bigg|_p\right)$$

*is an orthonormal basis of $T_p M$ for each $p \in U$, then $g_{ij} = \delta_{ij}$. Thus, raising and lowering indices are null operations in this coordinate frame. Note, however, that such frames often do not exist [155, p. 330 ff.].*

   *An important example where such coordinates exist globally is $\mathbb{R}^d$ with the Euclidean metric: The canonical coordinate system has the (partial derivatives in the directions of the) standard basis vectors as coordinate vectors.*

**Example 1.5.** *Let M be a Riemannian manifold and $p \in M$. There always exists a coordinate system $\phi = (x^1, \ldots, x^d)$ on a neighborhood $U$ of $p$ such that $[g_{ij}(p)] = I$ is the identity matrix. Indeed, let $(v_1, \ldots, v_d)$ be an orthonormal basis of $T_pM$ and $E : T_pM \to \mathbb{R}^d$ the corresponding vector space isomorphism that maps the basis $(v_1, \ldots, v_d)$ to the canonical basis of $\mathbb{R}^d$. Let further $U \subseteq M$ be the maximal domain of $\mathrm{Log}_p$. Geodesic normal coordinates with center $p$ are then given by*

$$\phi : U \to \mathbb{R}^d$$
$$q \mapsto E(\mathrm{Log}_p(q)).$$

*With $\phi = (x^1, \ldots, x^d)$ we find [137, p. 27]*

$$\frac{\mathrm{d}}{\mathrm{d}x^i}\bigg|_p = v_i$$

*for $i = 1, \ldots, d$. Thus, when we are only interested in raising and lowering indices of a tensor $S$ at a fixed point $p \in M$, it suffices to represent $S$ in geodesic normal coordinates (that is, w.r.t. any orthonormal basis of $T_pM$). Since the representing matrix of the metric at $p$ is then the identity matrix, raising and lowering indices are null operations. We want to stress that the above is only valid when $(v_1, \ldots, v_d)$ is orthonormal, as otherwise $[g_{ij}(p)] \neq I$.*

### 1.2.3 Riemannian Metrics on Lie Groups

Turning to Riemannian metrics on Lie groups $G$, there are two classes of metrics that "respect the group structure": left and right invariant metrics. The former is defined as follows: We call a metric $\langle \, \cdot \, , \, \cdot \, \rangle$ on $G$ *left invariant*, if left translations are isometries, that is,

$$\langle v, w \rangle_g = \langle d_g L_f(v), d_g L_f(w) \rangle_{L_f(g)} \tag{1.16}$$

for all $f, g \in G$. *Right invariant metrics* are defined analogously by replacing $L_f$ with $R_f$ in Equation (1.16). Equivalent conditions are that $\langle X, Y \rangle$ be constant functions for left/right invariant vector fields $X, Y$ that are determined by $v$ and $w$, respectively. Any Euclidean inner product on $T_eG$ can be extended to a left or right invariant metric on $G$. For example, if $\langle \, \cdot \, , \, \cdot \, \rangle_e$ is a Euclidean inner product on $T_eG$, then the metric on $G$ defined by

$$\langle v, w \rangle_g := \left\langle d_g L_{g^{-1}}(v), d_g L_{g^{-1}}(w) \right\rangle_e$$

for all $v, w \in T_gG$ and all $g \in G$ is left invariant.

An obvious question is whether there exists a *bi-invariant metric* that is both left *and* right invariant. Unfortunately, there are many Lie groups where this is not the case. Indeed, any Lie group that is not a direct product of compact and commutative groups does not have a bi-invariant metric [165], including $\mathrm{GL}(d)$, $\mathrm{GL}^+(d)$, $\mathrm{E}(d)$, $\mathrm{SE}(d)$, the special linear group $\mathrm{SL}(d)$ (that is, all linear, orientation and volume preserving transformations of $\mathbb{R}^d$), and the Heisenberg group [188] (all for $d > 1$). On the other hand, compact Lie groups always possess a bi-invariant metric. Below are two examples of groups with a bi-invariant metric. We will encounter both of them in later chapters.

**Example 1.6.** *The special orthogonal group* $\mathrm{SO}(d)$ *is compact and, therefore, can be endowed with a bi-invariant metric. Indeed, the restriction of the standard inner product on* $\mathbb{R}^{d,d}$ *to* $\mathrm{SO}(d)$ *turns out to have this property. It is given by*

$$\langle RS, RT \rangle_R := \mathrm{trace}\left( (RS)^T RT \right) = \mathrm{trace}\left( S^T T \right)$$

*for all* $RS, RT \in T_R\mathrm{SO}(d)$ *and all* $R \in \mathrm{SO}(d)$. *We then obtain*

$$\mathrm{dist}(R, U) = \| \log(R^T U) \|_{\mathrm{F}},$$

*where F denotes the Frobenius norm.*

**Example 1.7.** *The space of d-by-d symmetric positive definite matrices*

$$\mathrm{SPD}(d) = \{ P \in \mathbb{R}^{d,d} : P^T = P, \text{ all eigenvalues of } P \text{ are positive} \}$$

*is a* $((d+1)d/2)$*-dimensional manifold. Denoting the vector space of symmetric d-by-d matrices by* $\mathrm{Sym}(d)$, *it can be shown that* $\mathrm{SPD}(d)$ *constitutes an open half cone of* $\mathrm{Sym}(d)$ *[18, Prop. 2.7]. Furthermore, the tangent space* $T_P\mathrm{SPD}(d)$ *at each* $P \in \mathrm{SPD}(d)$ *can be identified with* $\mathrm{Sym}(d)$. *An important property now is that the matrix exponential* $\exp : \mathrm{Sym}(d) \to \mathrm{SPD}(d)$ *is a bijection. (We will see that the matrix exponential is also the group exponential, so the notation is consistent.) In [18, 19], Arsigny et al. utilized this property to give* $\mathrm{SPD}(d)$ *the structure of a vector space by defining addition and multiplication by*

$$P \oplus Q := \exp(\log(P) + \log(Q)),$$
$$x \otimes P := \exp(x \log(P)),$$

*for all* $P, Q \in \mathrm{SPD}(d)$ *and* $x \in \mathbb{R}$. *Consequently, with this structure,* $\mathrm{SPD}(d)$ *is a Lie group with commutative group operation. Since its one-parameter subgroups are given by* $\mathbb{R} \ni t \mapsto \exp(tS)$, *the group exponential and logarithm coincide with the matrix exponential and logarithm.*

*Analogously to the vector space construction, Arsigny et al. also proposed to transfer the usual Euclidean metric from* $\mathrm{Sym}(d)$ *to* $\mathrm{SPD}(d)$ *via the matrix exponential. This leads to the Log-Euclidean metric* $\langle \, \cdot \, , \, \cdot \, \rangle$ *that is defined by*

$$\langle S, T \rangle_P := \mathrm{trace}\left( \mathrm{d}_P \log(S)^T \mathrm{d}_P \log(T) \right)$$

*for all* $P \in \mathrm{SPD}(d)$ *and* $S, T \in T_P\mathrm{SPD}(d)$. *Since* $\mathrm{SPD}(d)$ *with the Log-Euclidean metric is isometric to* $\mathrm{Sym}(d)$ *with the standard Euclidean metric, it is flat (that is, has vanishing curvature tensor everywhere) as a Riemannian manifold. Furthermore, since the group operation is commutative, the Log-Euclidean metric is bi-invariant.*

*We obtain*
$$\mathrm{dist}(P, Q) = \| \log(P) - \log(Q) \|_F.$$

*The Riemannian exponential and logarithm can be expressed in terms of the matrix exponential and logarithm; they are given by*

$$\mathrm{Exp}_P(S) = \exp\left( \log(P) + \mathrm{d}_P \log(S) \right), \qquad P \in \mathrm{SPD}(d), S \in T_P\mathrm{SPD}(d),$$
$$\mathrm{Log}_P(Q) = \mathrm{d}_{\log(P)} \exp\left( \log(Q) - \log(P) \right), \quad P, Q \in \mathrm{SPD}(d).$$

*Both can be computed efficiently as there are algorithms for the derivative of the matrix exponential and logarithm [168, 169].*

Last but not least, we have an explicit formula for parallel transport (which must be path-independent since there is no curvature). Let $P, Q \in \text{SPD}(d)$ and $S \in T_P\text{SPD}(d)$. Taking advantage of the fact that the exponential is an isometry (and that parallel vector fields in $\text{Sym}(d)$ are constant), the parallel transport $T$ of $S$ from $P$ to $Q$ is given by

$$T = \left( d_{\log(Q)} \exp \circ \, d_P \log \right) (S).$$

### 1.2.4 A Canonical Affine Structure for Lie Groups

We now turn to the choice of affine connection $\nabla$ on a Lie group $G$. If $G$ is endowed with a Riemannian metric, then its Levi-Civita connection is the clear candidate. Nevertheless, a metric might not be needed, problematic to find, or its Levi-Civita connection lacks valuable properties. In this section, we will see that the latter is the case for several relevant Lie groups, but they also offer another canonical connection, which is better suited.

Since $G$ has a group structure, it is natural to require that $\nabla$ "respects" it. The following two conditions make this precise. First, $\nabla$ should be equivariant under both left and right translations when applied to left or right invariant vector fields; that is, for left invariant fields

$$\nabla_{dL_g(X)} dL_g(Y) = dL_g(\nabla_X Y)$$

for all $X, Y \in \mathfrak{g}$ and $g \in G$, while the analog should hold for right invariant ones. Second, geodesics through the identity $e$ should coincide with one-parameter subgroups. Together with the equivariance property, this implies that any geodesic is a left and right translated one-parameter subgroup, that is, for any geodesic $\gamma$ in $G$ with $\gamma(0) = g$ both $g^{-1}\gamma : t \mapsto g^{-1}\gamma(t)$ and $\gamma g^{-1} : t \mapsto \gamma(t)g^{-1}$ are (generally different) one-parameter subgroups of $G$; this is visualized in Figure 1.8. Because connections with the above properties were first studied by Cartan and Shouten in [46], they are called bi-invariant *Cartan-Shouten connections*;[12] for more on them; see [197, Ch. 6]. Amongst them, we find the *canonical Cartan-Shouten (CCS) connection* $\nabla$ of $G$. It is the unique torsion-free Cartan-Shouten connection. On $\mathfrak{g}$, it is defined by

$$\nabla_X Y := \frac{1}{2}[X, Y], \quad X, Y \in \mathfrak{g}; \tag{1.17}$$

see [197, Ch. 6]. Using Equation (1.17), one can show [197, p. 72] that the curvature tensor of a Lie group with CCS connection is given by

$$R(X, Y)Z = \frac{1}{4}[Z, [X, Y]], \quad X, Y, Z \in \mathfrak{g}. \tag{1.18}$$

We can extend both (1.17) and (1.18) to general vector fields using the multilinearity of $\nabla$ and $R$ and a (global) frame field (which is provided by $\mathfrak{g}$).

---

[12]In general, Cartan-Shouten connections need only be left invariant.

Since geodesics are translated one-parameter subgroups, the local exponential and logarithm of the CCS connection are given by

$$\operatorname{Exp}_g(w) = g \exp\left(\mathrm{d}_g L_{g^{-1}}(w)\right) = \exp\left(\mathrm{d}_g R_{g^{-1}}(w)\right) g, \quad w \in T_g G, \qquad (1.19)$$

$$\operatorname{Log}_g(h) = \mathrm{d}_e L_g\left(\log(g^{-1}h)\right) = \mathrm{d}_e R_g\left(\log(hg^{-1})\right), \quad h \in U, \qquad (1.20)$$

where $U \subseteq G$ is the maximal neighborhood of $g$ such that $\log(g^{-1}h)$ is defined; see [188, Cor. 5.1]. In particular, $G$ is geodesically complete.
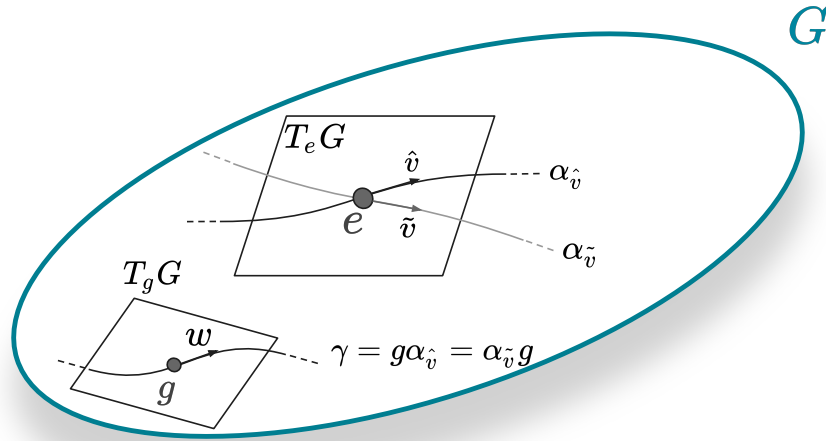


Figure 1.8: Geodesics in a Lie group $G$ with CCS connection. The geodesic $\gamma$ through $g \in G$ with tangent vector $w \in T_g G$ coincides with the left and right translated one-parameter subgroups $g\alpha_{\hat{v}}$ and $\alpha_{\tilde{v}}g$ defined by the tangent vectors $\hat{v} := \mathrm{d}_g L_{g^{-1}}(w) \in T_e G$ and $\tilde{v} := \mathrm{d}_g R_{g^{-1}}(w) \in T_e G$, respectively. (Note that if the group operation is commutative, then $\hat{v} = \tilde{v}$.)

Crucially, the CCS connection is the Levi-Civita connection of a Riemannian metric if and only if the latter is bi-invariant, that is, invariant under left *and* right translations. This is the only case in which $\operatorname{Exp}_e = \exp$ and $\operatorname{Log}_e = \log$ on a Lie group with Riemannian structure; only then, geodesics and translated one-parameter subgroups coincide. Whenever the group structure is of interest, the CCS connection is thus the first choice.

We finish with one example of a Lie group with a bi-invariant metric and two where the CCS connection differs from all Levi-Civita connections.

**Example 1.8.** *As shown in Example 1.6, the group SO(d) has a bi-invariant metric, and its Levi-Civita connection thus corresponds to the CCS connection. Consequently, geodesics are given by 1-parameter subgroups, and the exponential and logarithm of the connection are given by translations of the matrix exponential and logarithm, respectively. There also is an explicit formula to compute the parallel transport along a geodesic, which will be helpful later. Let $\gamma$ be a geodesic in SO(d), $\gamma(0) = R$, and $\gamma'(0) = RA$. Further, let $RS \in T_R SO(d)$ and $\gamma(1) = U$. The parallel transport $UT \in T_U SO(d)$ of $RS$ from $R$ to $U$ is then given by [78, p.*

$$UT = U \exp\left(-\frac{A}{2}\right) S \exp\left(\frac{A}{2}\right).$$

**Example 1.9.** *In contrast to* SO(*d*)*, there is no bi-invariant metric on* GL(*d*) *and, therefore, on its component* GL$^+$(*d*)*. Therefore, no Levi-Civita connection on either of them has translated 1-parameter subgroups as geodesics. The latter only holds when the group's CCS connection is used.*

**Example 1.10.** *The Euclidean group* E(*d*) *and its component* SE(*d*) *also do not possess a bi-invariant metric, implying that geodesics of a Levi-Civita connection and 1-parameter subgroups never coincide. In homogeneous coordinates, the exponential and logarithm of the group's CCS connection can also be computed with the corresponding matrix maps (c.f. Example 1.2).*

## 1.3  Riemannian Symmetric Spaces

A Riemannian symmetric space is a Riemannian manifold $M$ with a certain regularity. Let $p \in M$ and Id the identity map. We say that an isometry $s_p : U \to U$ of some neighbourhood $U$ of $p$ is a geodesic symmetry about $p$, if $s_p(p) = p$ and $\mathrm{d}_p s_p = -\mathrm{Id}$. The name originates from the fact that $s_p$ mirrors all geodesics through $p$ in $p$ without changing distances. A connected manifold $M$ is called *Riemannian symmetric space* (symmetric space in short) if for every $p$ there is a geodesic symmetry $s_p : M \to M$ (that is, every $s_p$ is a *global* isometry). Examples are Euclidean space, the sphere (see below), and Lie groups with bi-invariant metric (where concatenations of the inversion map with left or right translations are the symmetries). A direct consequence of the existence of symmetries is that symmetric spaces are geodesically complete; because if a geodesic $\gamma$ is defined on $[0, s)$, then we can always reflect it in some $t \in (s/2, s)$ to extend it beyond $s$.

Symmetric spaces are of interest to us because of three reasons: First, they appear in many applications; second, they allow for *explicit* formulas for Jacobi fields, which enables us to formulate specialized algorithms with less iterative subroutines; and third, essential results concerning the relationship of maximum likelihood and least squares estimation can be shown for them.

A fundamental property of symmetric spaces is that their Riemannian curvature tensor $R$ is parallel, that is, $\nabla R = 0$. (Here, $\nabla$ denotes the *generalization* of the connection to tensor fields on $M$; see [45, Ch. 4] for details.) This fact is crucial in proving the following lemma from [29, Lem. 6] (see also [20, Sec. 3]), which opens up the possibility to compute Jacobi fields (and thus (1.10) and (1.11)) explicitly in symmetric spaces. Note that a parallel translated orthonormal basis (ONB) is again an ONB since angles are preserved.

**Lemma 1.3.1.** *Let $M$ be a d-dimensional symmetric space, $p, q \in M$ and $\gamma(\,\cdot\,; p, q) : [0, 1] \to M$ be a geodesic between $p$ and $q$. Further, let $(v_1, \ldots, v_d)$ be an ONB of $T_p M$ that diagonalizes the operator $\mathfrak{R}$ (1.12) at $p$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_d$ and $(X_1, \ldots, X_d)$ the parallel transported frame field of $(v_1, \ldots, v_d)$ along $\gamma$. Set $l_\gamma := \mathrm{length}(\gamma)$.*

*Then, for $i = 1, \ldots, d$, the value of the Jacobi field $J_{v_i}$ along $\gamma$ at $t \in [0,1]$ is given by*

$$J_{v_i}(t) = \begin{cases} \dfrac{\sinh\left((1-t)l_\gamma\sqrt{-\lambda_i}\right)}{\sinh\left(l_\gamma\sqrt{-\lambda_i}\right)} X_i(t), & \lambda_i < 0, \\[3mm] \dfrac{\sin\left((1-t)l_\gamma\sqrt{\lambda_i}\right)}{\sin\left(l_\gamma\sqrt{\lambda_i}\right)} X_i(t), & \lambda_i > 0, \\[3mm] (1-t)X_i(t), & \lambda_i = 0. \end{cases}$$

*Furthermore, for an arbitrary vector $v = \sum_{i=1}^{d}\langle v, v_i\rangle_p\, v_i \in T_pM$ and all $t \in [0,1]$ we have*

$$J_v(t) = \sum_{i=1}^{d}\langle v, v_i\rangle_p\, J_{v_i}(t).$$

Note that $l_\gamma = \mathrm{dist}(p, q)$ in a normal convex neighborhood $U \subseteq M$ with $p, q \in U$ (see Section 1.2.2). Furthermore, we emphasize that the operator $\mathfrak{R}$ is defined using the tangent vector of $\gamma$ *scaled to unit-length*. This is equivalent to formulating $\mathfrak{R}$ without the normalization of $\gamma'(0)$ (that is, $v \mapsto R(v, \gamma'(0))\gamma'(0)$) but requiring that $\gamma$ is *arc-length parametrized* (that is, defined on $[0, l_\gamma]$, necessarily with unit speed); because then $\gamma'(0)$ already has norm one. The formulation of Lemma 1.3.1 in [29] can be misunderstood to mean that $\mathfrak{R}$ is defined without the normalization of $\gamma'(0)$ but for $\gamma$ being parametrized in $[0, 1]$ (as, for example, in [206, Sec. 4.3], where $v \mapsto R(v, \gamma'(0))\gamma'(0)$ is called "Jacobi operator"[13]). This interpretation would lead to different results since linear reparametrizations of $\gamma$ (which do not change the fact that it is a geodesic) change the norm of $\gamma'(0)$ and thus the eigenvalues of $v \mapsto R(v, \gamma'(0))\gamma'(0)$. However, deriving the lemma from [20, Prop. 3.5] reveals that the formulation (1.12) is correct.

We provide examples.

**Example 1.11.** *The d-dimensional unit sphere $\mathcal{S}^d$ is one of the most important manifolds. It can be defined as the set of unit vectors in $\mathbb{R}^{d+1}$ (with Euclidean 2-norm $\|\cdot\|$), that is,*

$$\mathcal{S}^d := \left\{ p \in \mathbb{R}^{d+1} \mid \|p\| = 1 \right\}.$$

*An atlas of $\mathcal{S}^d$ can, for example, be obtained by taking all orthogonal projections of open hemispheres onto the corresponding planes that divide the sphere [45, p. 21]. At each $p \in \mathcal{S}^d$ the tangent space is the orthogonal complement of $p$ in $\mathbb{R}^{d+1}$, that is,*

$$T_p\mathcal{S}^d = \left\{ v \in \mathbb{R}^{d+1} \mid p^T v = 0 \right\}.$$

*We can restrict the standard Euclidean scalar product of the surrounding space to the tangent spaces to obtain a Riemannian metric on $\mathcal{S}^d$. Using the corresponding Levi-Civita connection, the geodesics in $\mathcal{S}^d$ are (segments of) great circles and, thus, open hemispheres are (maximal) normal convex neighbourhoods. (Throughout this work, without mentioning it again, we will always assume that $\mathcal{S}^d$ is endowed with the above structure.) Importantly, we have*

$$K \equiv 1,$$

---

[13]We do not use this name because there might be confusion as other works call $\nabla_{\gamma'}^2 + \mathfrak{R}$ "Jacobi operator".
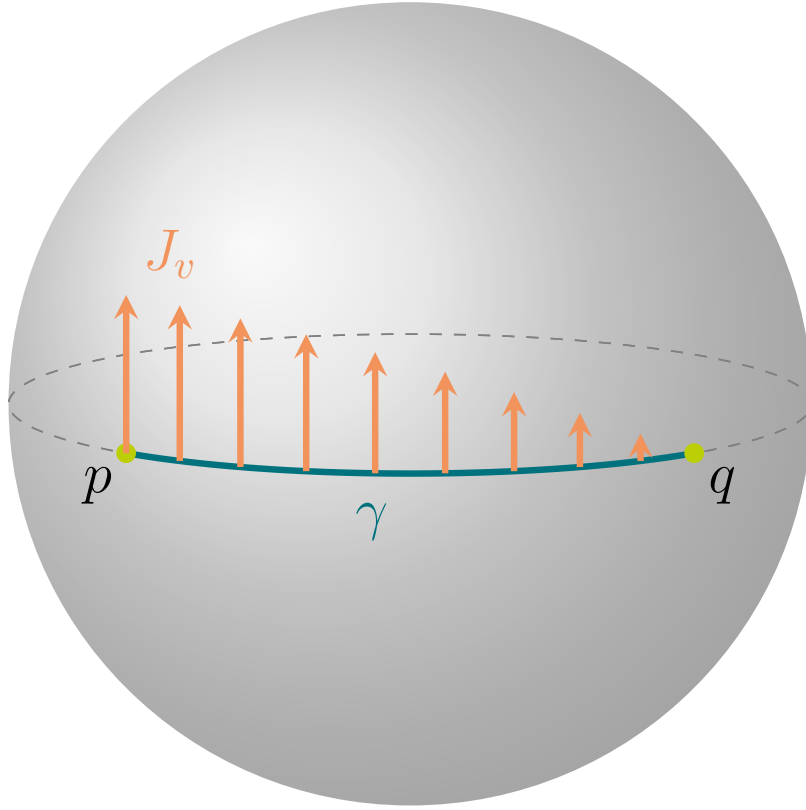
Figure 1.9: Jacobi field $J_v$ along the geodesic $\gamma$ between $p = [1, 0, 0]^T$ and $q = [0, 1, 0]^T$ in the sphere $\mathcal{S}^2$ with $v = [0, 0, 0.4]^T$.

*that is, $\mathcal{S}^d$ has constant sectional curvature equal to one (see, for example, [197, Ex. 22.3]).*

*The Riemannian exponential at $p \in \mathcal{S}^d$ is given for all $v \in T_p \mathcal{S}^2 \setminus \{0\}$ by*

$$\mathrm{Exp}_p(v) = p \cos(\|v\|) + \frac{v}{\|v\|} \sin(\|v\|),$$

*while the Riemannian logarithm is defined for all $q \in \mathcal{S}^d \setminus \{-q\}$ by*

$$\mathrm{Log}_p(q) = \frac{q - (p^T q)p}{\|q - (p^T q)p\|} \arccos(p^T q).$$

*The distance in $\mathcal{S}^d$ is*

$$\mathrm{dist}(p, q) = \arccos(p^T q).$$

*Parallel transport along geodesics in $\mathcal{S}^d$ can also be calculated explicitly. Let $p, q \in \mathcal{S}^d$ be from the same hemisphere (which is the largest normal convex neighborhood that contains them). The parallel transport $w$ of $v \in T_p \mathcal{S}^d$ along $\gamma(\cdot\,; p, q)$ to $T_q \mathcal{S}^d$ is (not only) then given by [148, p. 116]*

$$w = v - \frac{\mathrm{Log}_p(q)^T v}{\|\mathrm{Log}_p(q)\|^2} \left( \mathrm{Log}_p(q) + \mathrm{Log}_q(p) \right).$$

*The sphere $\mathcal{S}^d$ is also a symmetric space. It is easily verified that the symmetry at each $p \in \mathcal{S}^d$ is the reflection at the line in $\mathbb{R}^{d+1}$ that goes through the origin and $p$; that is,*

$s_p(q) = -q + 2\langle q, p \rangle p$. *Furthermore, if $\gamma : [0,1] \to \mathcal{S}^d$ is a geodesic, then it follows from $K \equiv 1$ that*

$$R\left(X, \frac{\gamma'}{\|\gamma'\|}\right) \frac{\gamma'}{\|\gamma'\|} = X$$

*for vector fields $X$ along $\gamma$ with $\langle X, \gamma' \rangle = 0$ (see [45, Ex. 2.3 in Chap. 5 ]). Consequently, all eigenvalues but the trivial of the operator $\mathfrak{R}$ (1.12) are equal to 1. A set of corresponding orthonormal eigenvectors can be chosen as any ONB of the orthogonal complement of $\mathrm{span}\{\gamma'(0)\}$ in $T_{\gamma(0)}\mathcal{S}^d$. Using Theorem 1.3.1 we can then calculate Jacobi fields; an example for $d = 2$ is shown in Figure 1.9.*

**Example 1.12.** *Being a connected Lie group, the special orthogonal group $\mathrm{SO}(d)$ with the Riemannian structure from Example 1.6 is a symmetric space. Since the Riemannian metric is bi-invariant, the Levi-Civita connection coincides with the CCS connection (1.17), and the Riemannian curvature tensor is given by Equation (1.18). The operator $\mathfrak{R}$ along a geodesic $\gamma$ between $R, U \in \mathrm{SO}(d)$ with $\gamma'(0) = RS$ is thus*

$$\mathfrak{R}(RT) = \frac{1}{4}R(-S^2T + 2STS - TS^2).$$

*Herefrom, we can compute the eigenvalues and eigenvectors we need for Jacobi fields. For convenience, we only give the formulas for $d = 3$ from [206, pp. 77–78], which we will need in later chapters. We denote the $i$-th canonical basis vector of $\mathbb{R}^3$ by $e_i$, $i = 1, 2, 3$. Let $T = V b(e_3 e_2^T - e_2 e_3^T) V^T$ be the Schur decomposition of the skew-symmetric matrix $T \in T_I \mathrm{SO}(3)$. The non-trivial eigenvalues of $\mathfrak{R}$ are*

$$\lambda_{2,3} = \frac{b^2}{4};$$

*corresponding orthonormal eigenvectors are*

$$S_1 = V \left(e_2 e_1^T - e_1 e_2^T\right) V^T,$$
$$S_2 = V \left(e_3 e_1^T - e_1 e_3^T\right) V^T.$$

**Example 1.13.** *The space of symmetric positive definite matrices $\mathrm{SPD}(d)$ together with the Log-Euclidean structure from Example 1.7 also constitutes a symmetric space (since it is connected). It is a flat space. Hence, all orthonormal bases diagonalize $\mathfrak{R}$. Therefore, the Jacobi fields in Theorem 1.3.1 are given by scaled parallel translations.*

# Chapter 2

# Bézier Splines in Riemannian Manifolds and Shape Spaces

Here, we give the necessary background on Bézier splines and shape analysis. The former form the basis for Chapter 5, while the latter is an important application area of geometric data analysis of which we will investigate several problems in Chapter 6.

## 2.1   Bézier Splines in Riemannian Manifolds

Fundamental to the constructions in Chapter 5 will be splines built from generalized Bézier curves [105, 195]. Therefore, we summarize the relevant facts about them here. We restrict the domain of a (generalized) Bézier curve to $[0, 1]$ for clarity. This does not influence generality since reparametrizations are always possible.

Let $M$ be a Riemannian manifold and $U \subseteq M$ be a normal convex neighborhood. A family[1] of $k + 1 \geq 2$ *control points* $(p_0, \ldots, p_k) \in U^{k+1}$ defines a *Bézier curve* $\beta : [0, 1] \to U$ *of degree $k$* according to the *generalized de Casteljau algorithm*

$$
\begin{aligned}
\beta_i^0(t) &:= p_i, \\
\beta_i^l(t) &:= \gamma(t; \beta_i^{l-1}(t), \beta_{i+1}^{l-1}(t)), \quad l = 1, \ldots, k, \quad i = 0, \ldots, k - l,
\end{aligned}
\tag{2.1}
$$

by

$$
\beta(t) := \beta_0^k(t).
$$

The algorithm is visualized in Figure 2.1 for the sphere. In Euclidean space $\mathbb{R}^d$, it reduces to the familiar de Casteljau algorithm (because there the geodesic $\gamma(\,\cdot\,; x, y)$ is the straight line between $x, y \in \mathbb{R}^d$) and it is a well-known fact [84] that $\beta$ is then a curve with polynomials of order at most $k$ as entries. Hence, $\beta$ is a generalization of a polynomial curve of degree $k$ to Riemannian manifolds. Whenever of interest, we will make the dependence of $\beta$ on its control points explicit by writing $\beta(t; p_0, \ldots, p_k)$. Note that if there are only two control points $p_0, p_1$, then $\beta$ is just the geodesic from $p_0$ to $p_1$.

An important property of a Bézier curve $\beta$ is that

$$
\beta(0) = p_0 \text{ and } \beta(1) = p_k.
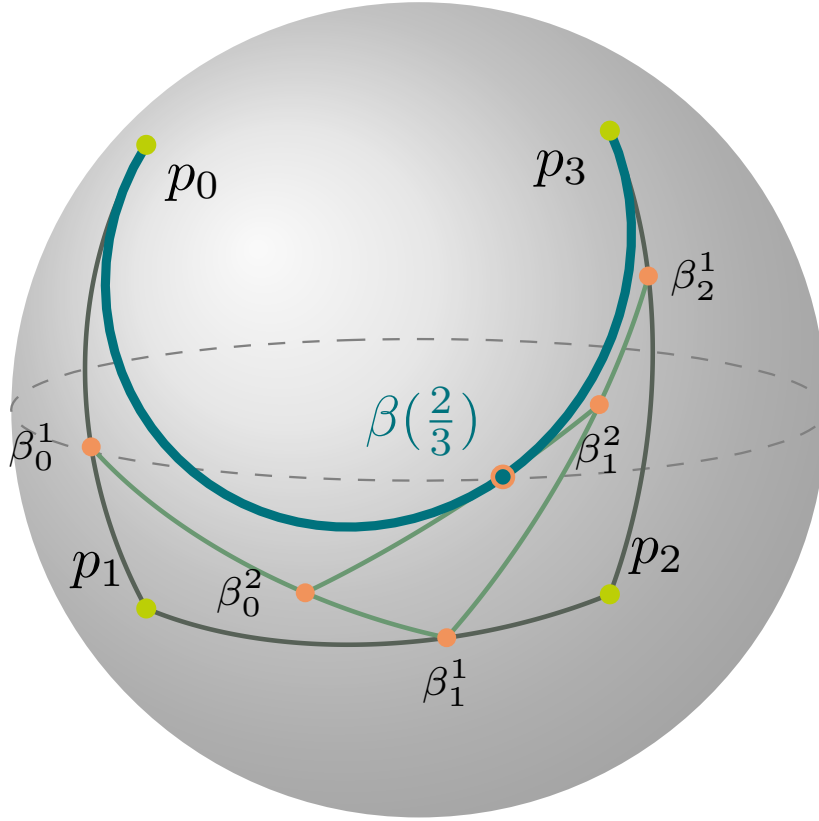\tag{2.2}
$$

---

[1] Copies of the same point are allowed.

Figure 2.1: Cubic Bézier curve $\beta$ on the sphere $\mathcal{S}^2$ and the construction of $\beta(2/3)$ by the de Casteljau algorithm.

Furthermore, the velocities at these points are

$$\beta'(0) = k\operatorname{Log}_{p_0}(p_1) \text{ and } \beta'(1) = -k\operatorname{Log}_{p_k}(p_{k-1}); \tag{2.3}$$

see [195, Thm. 1]. Popiel and Noakes also derived the accelerations at $t = 0$ and $t = 1$; their result is given in the following theorem.

**Theorem 2.1.1** ([195]). *A Bézier curve $t \mapsto \beta(t; p_0, \ldots, p_k)$ satisfies*
*(i) $\nabla_{\beta'}\beta'(0) = k(k-1)\widetilde{u}_0$, where*

$$\widetilde{u}_0 := \begin{cases} \gamma'(0; p_1, p_2), & \text{if } p_0 = p_1, \\ (\mathrm{dExp}_{p_0})^{-1}_{\gamma'(0;p_0,p_1)}\big(\gamma'(0; p_1, p_2) - \gamma'(1; p_0, p_1)\big), & \text{if } p_0 \neq p_1; \end{cases}$$

*(ii) $\nabla_{\beta'}\beta'(1) = k(k-1)\widetilde{u}_k$, where*

$$\widetilde{u}_k := \begin{cases} -\gamma'(1; p_{k-2}, p_{k-1}), & \text{if } p_{k-1} = p_k, \\ (\mathrm{dExp}_{p_k})^{-1}_{-\gamma'(1;p_{k-1},p_k)}\big(\gamma'(0; p_{k-1}, p_k) - \gamma'(1; p_{k-2}, p_{k-1})\big), & \text{if } p_{k-1} \neq p_k. \end{cases}$$

Using (2.2) and (2.3), we obtain the following corollary of Theorem 2.1.1.

34

**Corollary 2.1.2** ([195]). *A Bézier curve $t \mapsto \beta(t; p_0, \ldots, p_k)$ satisfies*

$$p_0 = \beta(0; p_0, \ldots, p_k),$$
$$p_k = \beta(1; p_0, \ldots, p_k),$$
$$p_1 = \mathrm{Exp}_{p_0}\left(\frac{1}{k}v_0\right),$$
$$p_{k-1} = \mathrm{Exp}_{p_k}\left(-\frac{1}{k}v_k\right),$$
$$p_2 = \mathrm{Exp}_{p_1}\left(\frac{1}{k(k-1)}w_0\right),$$
$$p_{k-2} = \mathrm{Exp}_{p_{k-1}}\left(\frac{1}{k(k-1)}w_k\right),$$

*where $v_0 := \beta'(0; p_0, \ldots, p_k)$, $v_k := \beta'(1; p_0, \ldots, p_k)$, $u_0 := \nabla_{\beta'}\beta'(0)$, $u_k := \nabla_{\beta'}\beta'(1)$, and*

$$w_0 := \begin{cases} u_0, & \text{if } p_0 = p_1, \\ (\mathrm{dExp}_{p_0})_{\gamma'(0; p_0, p_1)}\big(u_0 + k(k-1)\gamma'(0; p_0, p_1)\big), & \text{if } p_0 \neq p_1, \end{cases}$$

$$w_k := \begin{cases} u_k, & \text{if } p_{k-1} = p_k, \\ (\mathrm{dExp}_{p_k})_{-\gamma'(1; p_{k-1}, p_k)}\big(u_k - k(k-1)\gamma'(1; p_{k-1}, p_k)\big), & \text{if } p_{k-1} \neq p_k. \end{cases}$$

Property (2.3) allows us to fit Bézier curves of possibly different orders together to a differentiable spline. For $i = 0, \ldots, L-1$ let $\big(p_0^{(i)}, \ldots, p_{k_i}^{(i)}\big)$ be the control points of $L \geq 2$ Bézier curves $\beta^{(0)}, \ldots, \beta^{(L-1)}$ such that

$$p_{k_i}^{(i)} = p_0^{(i+1)} \quad \text{and} \quad \gamma\left(\frac{k_i + k_{i+1}}{k_{i+1}}; p_{k_i-1}^{(i)}, p_{k_i}^{(i)}\right) = p_1^{(i+1)} \tag{2.4}$$

for all $i = 0, \ldots, L-2$. Then, we define the *Bézier spline B* with control points $\big(p_0^{(i)}, \ldots, p_{k_i}^{(i)}\big)$, by

$$B(t) := \begin{cases} \beta^{(0)}\big(t; p_0^{(0)}, \ldots, p_{k_0}^{(0)}\big), & t \in [0, 1], \\ \beta^{(i)}\big(t - i; p_0^{(i)}, \ldots, p_{k_i}^{(i)}\big), & t \in (i, i+1], \quad i = 1, \ldots, L-1. \end{cases} \tag{2.5}$$

We also view Bézier curves as Bézier splines with $L = 1$ segments. From Equation (2.3), it follows that $B$ is $C^1$. Hence, we can make $B$ differentiable by aligning the four control points at the connections, thereby removing two degrees of freedom (see [105, Sec. 2.3] for more details). Corollary 2.1.2 can further be used to ensure that $B$ is $C^2$, but we restrict to $C^1$ splines in this work.

The spline $B$ is a *closed $C^1$ curve*, if (2.4) extends cyclically, that is,

$$p_{k_{L-1}}^{(L-1)} = p_0^{(0)} \quad \text{and} \quad \gamma\left(\frac{k_{L-1} + k_0}{k_0}; p_{k_{L-1}-1}^{(L-1)}, p_0^{(0)}\right) = p_1^{(0)} \tag{2.6}$$

also holds. Closed $C^1$ splines can readily be extended to cyclic curves defined on the whole of $\mathbb{R}$ when the input parameter is viewed modulo $L$.

Because of Equations (2.4) and (2.6), not all control points are free parameters of a Bézier spline. Indeed, if $p_{k_i}^{(i)} = p_0^{(i+1)}$ are *connecting control points*, joining the *i*-th
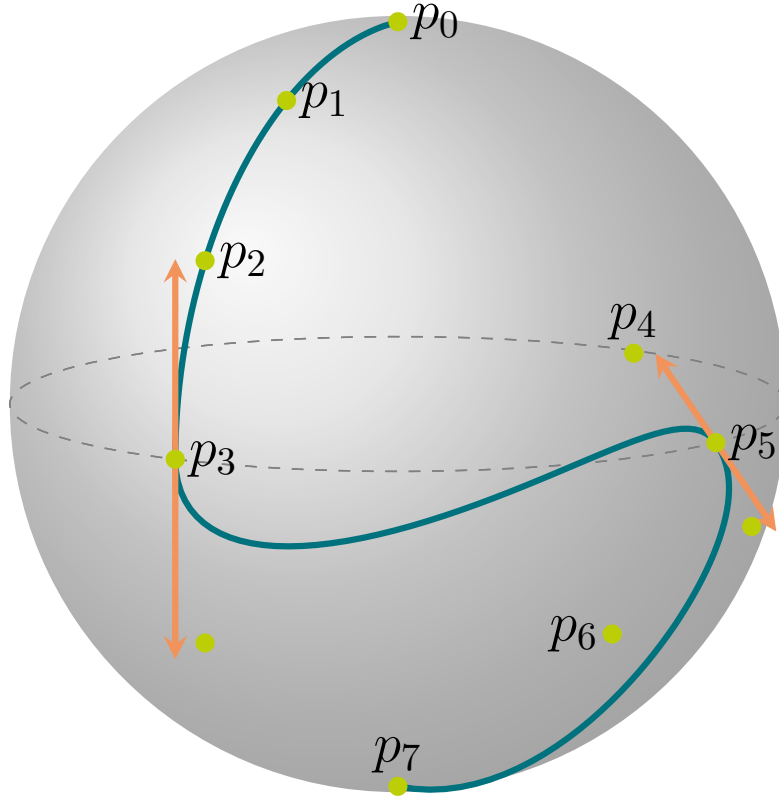
Figure 2.2: Bézier spline with 3 cubic segments on $\mathcal{S}^2$. The unlabeled green points are dependent control points.

to the $(i+1)$-th segment[2], we omit $p_0^{(i+1)}$ as a variable. Furthermore, we *choose* that $p_{k_i}^{(i)}$ and its *predecessor* $p_{k_i-1}^{(i)}$ are free variables, while the *successor* $p_1^{(i+1)}$ shall also be eliminated as a variable. With $\bar{k} := (k_{i-1} + k_i)/k_i$ we therefore have the additional condition $p_1^{(i+1)} = \gamma(\bar{k}; p_{k_i-1}^{(i)}, p_{k_i}^{(i)})$ at the connection. Setting

$$K := \begin{cases} k_0 + k_1 + \cdots + k_{L-2} + k_{L-1} - L + 1, & \text{if } B \text{ is not closed,} \\ k_0 + k_1 + \cdots + k_{L-2} + k_{L-1} - L - 1, & \text{if } B \text{ is closed,} \end{cases} \tag{2.7}$$

it then follows that the number of *independent* control points of a $C^1$ Bézier spline $B$ is $K + 1$.

We denote the set of $K + 1$ independent control points of $B$ by $p_0, \ldots, p_K$. In the non-closed case, this means

$$(p_0, \ldots, p_K) := \left( p_0^{(0)}, \ldots, p_{k_0}^{(0)}, p_2^{(1)}, \ldots, p_{k_1}^{(1)}, \ldots, p_2^{(L-1)}, \ldots, p_{k_{L-1}}^{(L-1)} \right) \in U^{K+1},$$

while $p_0^{(0)}$ and $p_1^{(0)}$ are left out for closed $B$. An example of a non-closed $C^1$ spline with three cubic segments and nine independent control points is shown in Figure 2.2, while a closed $C^1$ spline with two cubic segments and four independent control points

---

[2]For closed splines the segment numbers are meant modulo $L$, so the connection between the first and last segment is included.
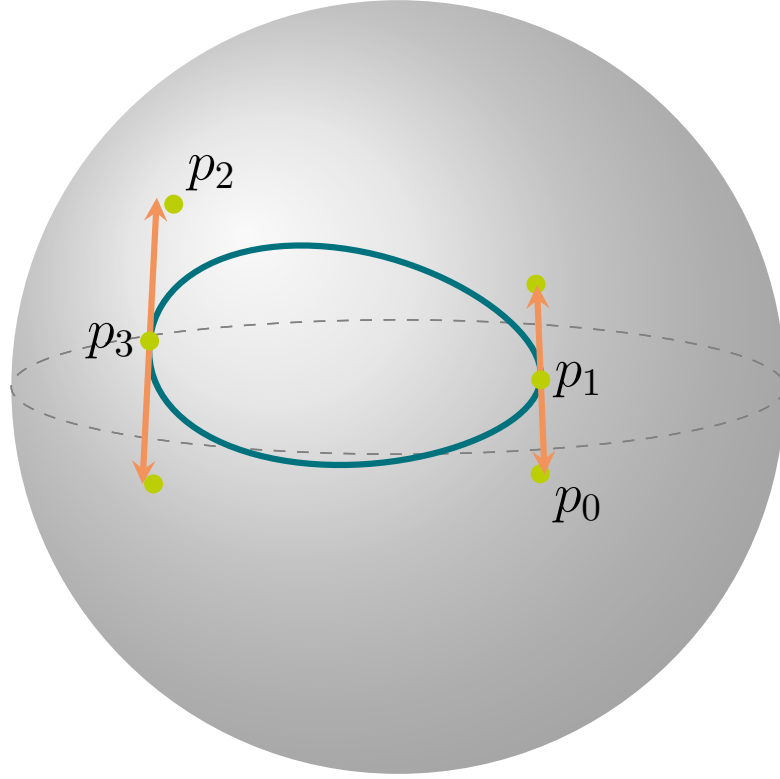
Figure 2.3: Closed spline with two cubic segments on $\mathcal{S}^2$. The unlabeled green points are dependent control points.

is depicted in Figure 2.3. When it is important, we make the dependence of $B$ on its control points clear by writing $B(t; p_0, \dots, p_K)$.

We introduce a notation for the set of Bézier splines

$$\mathcal{B}^L_{k_0,\dots,k_{L-1}}(U) := \{B : [0, L] \to U \mid B \text{ is a } C^1 \text{ Bézier spline} \\ \text{with } L \text{ segments of degrees } k_0, \dots, k_{L-1}\} \tag{2.8}$$

and its subset of closed splines

$$\mathring{\mathcal{B}}^L_{k_0,\dots,k_{L-1}}(U) := \{B : [0, L] \to U \mid B \text{ is a closed } C^1 \text{ Bézier spline} \\ \text{with } L \text{ segments of degrees } k_0, \dots, k_{L-1}\}. \tag{2.9}$$

*Except for clearly marked cases, we assume in the following that $L, k_0, \dots, k_{L-1}$, and the type of the curve are arbitrary but fixed. Therefore, we only write $\mathcal{B}(U)$ unless we want to emphasize the parameters.*

Later, we need to compute the adjoint of the derivative of a Bézier spline with respect to its control points; in symmetric spaces, this can be done explicitly, as we shall see in the now. (In the following, we summarize [29, Secs. 3–4], which we recommend for a more detailed treatment.

Before we can investigate the adjoint, we must turn to the derivative itself. We start with single curves. Fixing $j \in \{0, \dots, k\}$ we denote the value

$$\mathrm{d}_{p_j} \beta^l_i(t; p_i, \dots, p_{j-1}, \cdot, p_{j+1}, \dots, p_{i+l})(v) \in T_{\beta^l_i(t)} M$$

of the derivative of $\beta_i^l$ with respect to the $j$-th control point at $v \in T_{p_j}M$ by $w_i^{(l)}$.[3] (We set $w_i^{(l)} := 0 \in T_{\beta_i^l(t)}M$, if $j < i$ or $j > i + l$.) Then, using the chain rule (1.3) and denoting the Kronecker delta by $\delta_{ij}$, we obtain from (2.1) that

$$w_i^{(0)} = \delta_{ij}v,$$

$$w_i^{(l)} = \mathrm{d}_{\beta_i^{l-1}(t)}\gamma\left(t;\,\cdot\,,\beta_{i+1}^{l-1}(t)\right)\left(w_i^{(l-1)}\right) + \mathrm{d}_{\beta_{i+1}^{l-1}(t)}\gamma\left(t;\beta_i^{l-1}(t),\,\cdot\,\right)\left(w_{i+1}^{(l-1)}\right), \quad (2.10)$$

$$l = 1,\ldots,k, \quad i = 0,\ldots,k-l;$$

in particular,

$$\mathrm{d}_{p_j}\beta(t;p_0,\ldots,p_{j-1},\,\cdot\,,p_{j+1},\ldots,p_k) = w_0^{(k)}.$$

Using Equations (1.10) and (1.11) and omitting the zero derivatives leads to the following theorem.

**Theorem 2.1.3** (Derivative of a Bézier curve [29])**.** *Let $\beta$ be a Bézier curve of degree $k$ and $i \le j \le i + l$. Then,*

$$w_i^{(l)} = \begin{cases} J_{w_i^{(l-1)}}(t), & \text{if } j = i, \\ J_{w_i^{(l-1)}}(t) + J_{w_{i+1}^{(l-1)}}(1-t), & \text{if } i < j < i + l, \\ J_{w_{i+1}^{(l-1)}}(1-t), & \text{if } j = i + k, \end{cases}$$

*for $l = 1,\ldots,k$ and $i = 0,\ldots,k-l$, where $J_{w_i^{(l-1)}}$ is defined along $s \mapsto \gamma\left(s;\beta_i^{l-1}(t),\beta_{i+1}^{l-1}(t)\right)$ while $J_{w_{i+1}^{(l-1)}}$ is defined along the reversed geodesic $s \mapsto \gamma\left(s;\beta_{i+1}^{l-1}(t),\beta_i^{l-1}(t)\right)$.*

The theorem shows that the derivative acts by transporting the vector $v$ via Jacobi fields along the (relevant part of the) "tree of geodesics" as spanned by the de Castelau algorithm; this is visualized in Figure 2.4 for a cubic curve in $\mathcal{S}^2$. Note that whenever separate paths meet, their contributions are added.

We now turn to $C^1$ Bézier splines. In this case, Equations (2.4) and (2.6) must be accounted for, so differentiating with respect to a control point at a connection is different from the case of a single curve. To this end, note that if $p_j$ is a connecting control point, joining the $(i-1)$-th to the $i$-th segment (again, modulo $L$ for closed splines), then for $v \in T_{p_{j-1}}M$ and $w \in T_{p_j}M$

$$\mathrm{d}_{p_{j-1}}\gamma(\overline{k};\,\cdot\,,p_j)(v) = J_v(\overline{k}) \quad \text{and} \quad \mathrm{d}_{p_j}\gamma(\overline{k};p_{j-1},\,\cdot\,)(w) = J_w(1-\overline{k})$$

are the (extrapolated) values at $t = \overline{k}$ of the Jacobi fields (1.10) and (1.11), respectively [29, Sec. 4.2]. With this, the following theorem extends Theorem 2.1.3 to $C^1$ Bézier splines. Bergmann and Gousenbourger proved it for non-closed splines in [29, Lem. 10]. We formulate it so that closed splines are included; the proof trivially extends to this case.

---

[3]Note that the function $\beta_i^l$ in the de Casteljau algorithm (2.1) depends only on the control points $p_i,\ldots,p_{i+l}$.
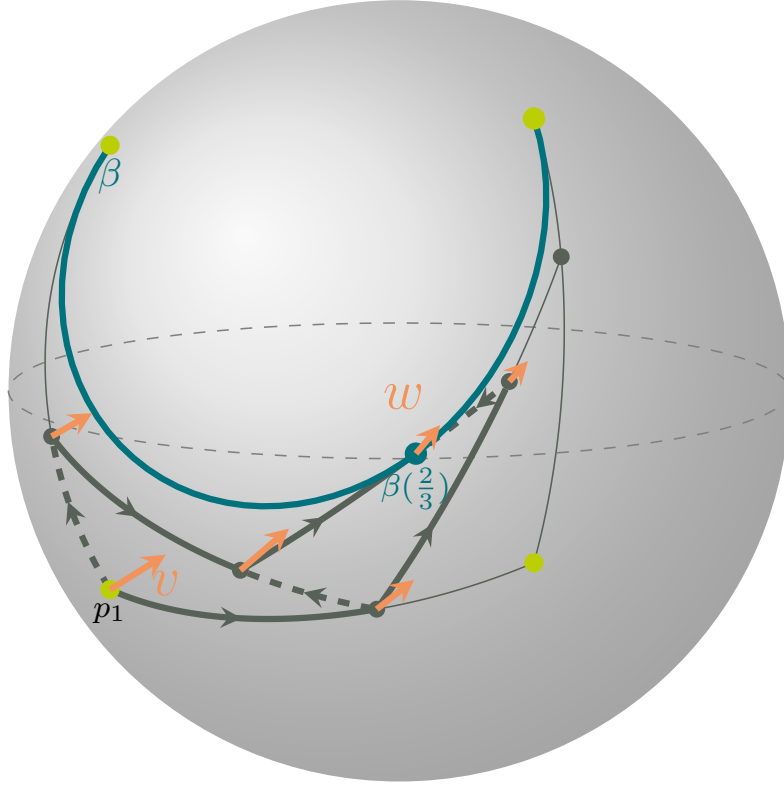
Figure 2.4: Derivative of a cubic Bézier curve $\beta$ in $\mathcal{S}^2$ w.r.t. the control point $p_1$ with $t = 2/3$. We visualize how, for $v \in T_{p_1}\mathcal{S}^2$, the vector $w := \mathrm{d}_{p_1}\beta(2/3; p_0, \cdot, p_2, p_3))(v) \in T_{\beta(2/3)}\mathcal{S}^2$ is generated by transporting $v$ through the tree of geodesics (along the part with thick lines). A solid line indicates that the vector is transported via the derivative of the geodesic w.r.t. its starting point (1.10), while a dashed line implies that it is transported via an endpoint derivative (1.11). Jacobi fields are not shown to avoid cluttering.

**Theorem 2.1.4** (Derivative of a $C^1$ Bézier spline). *Let $B \in \mathcal{B}(U)$ be defined by (2.5), $p_j$ be the j-th control point of B, and $v \in T_{p_j}M$. We set*

$$w_j := \mathrm{d}_{p_j}B(t; p_0, \ldots, p_{j-1}, \cdot, p_{j+1}, \ldots, p_K)(v).$$

*Further, let*

$$i \in \begin{cases} \{1, \ldots, L-1\}, & \text{if } B \text{ is not closed,} \\ \{1, \ldots, L\}, & \text{if } B \text{ is closed.} \end{cases}$$

*In the closed case, $i$, $i-1$, and $i+1$ are to be understood modulo L.*
  *If $p_j$ connects the segments $\beta^{(i-1)}$ and $\beta^{(i)}$, then*

$$w_j = \begin{cases} \mathrm{d}_{p_j}\beta\left(t-i+1; p_0^{(i)}, \ldots, p_{k_i-1}^{(i)}, \cdot\right)(v), & t \in (i-1, i], \\ \mathrm{d}_{p_j}\beta\left(t-i; \cdot, p_1^{(i+1)}, \ldots, p_{k_{i+1}}^{(i+1)}\right)(v) \\ \quad + \mathrm{d}_{\gamma(\bar{k}; p_{j-1}, p_j)}\beta\left(t-i; p_j, \cdot, p_2^{(i+1)}, \ldots, p_{k_{i+1}}^{(i+1)}\right)\left(\mathrm{d}_{p_j}\gamma(\bar{k}; p_{j-1}, \cdot)(v)\right), & t \in (i, i+1]. \end{cases}$$

*If the successor $p_{j+1}$ connects the segments $\beta^{(i-1)}$ and $\beta^{(i)}$,*

$$
w_j = \begin{cases}
\mathrm{d}_{p_j}\beta\left(t-i+1; p_0^{(i)}, \ldots, \cdot, p_{k_i}^{(i)}\right)(v), & t \in (i-1, i], \\
\mathrm{d}_{\gamma(\bar{k}; p_j, p_{j+1})}\beta\left(t-i; p_{j+1}, \cdot, p_2^{(i+1)}, \ldots, p_{k_{i+1}}^{(i+1)}\right)\left(\mathrm{d}_{p_j}\gamma(\bar{k}; \cdot, p_{j+1})(v)\right), & t \in (i, i+1].
\end{cases}
$$

*If $i = 1$, then $i - 1 = 0$ is included in the first interval in both cases.*
*In any other case, $w_j$ can be computed segment-wise using Theorem 2.1.3.*

The theorem shows that as long as we can calculate Jacobi fields, we can also compute the derivative of Bézier splines with respect to their control points. Crucially, in combination with Theorem 1.3.1 and Equation (1.10) we get explicit formulas in symmetric spaces that fundamentally only require parallel transport along geodesics: Let $M$ be a $d$-dimensional symmetric space and $U \subseteq M$ be a normal convex neighborhood. Further, let $p, q \in U$ and $v \in T_pM$. Then, there are coefficients $a_1, \ldots, a_d \in \mathbb{R}$, an ONB $(v_1, \ldots, v_d)$ of $T_pM$, and its parallel translated frame field $(X_1, \ldots, X_d)$ along $\gamma(\,\cdot\,; p, q)$ such that

$$
\mathrm{d}_p\gamma(t; \cdot, q)(v) = J_v(t) = \sum_{i=1}^{d} \langle v, v_i \rangle_p \, a_i X_i(t). \tag{2.11}
$$

Repeated applications of this equation yield the explicit formula for the derivative of a Bézier spline in a symmetric space.

We are now ready to investigate the adjoint derivative and start again with single Bézier curves. Remember that taking the adjoint reverses the order of concatenated operators; in our context, this means that to compute

$$
\mathrm{d}_{p_j}\beta(t; p_0, \ldots, p_{j-1}, \cdot, p_{j+1}, \ldots, p_k)^*(w)
$$

for $w \in T_{\beta(t)}M$, we must transport $w$ backward along all valid paths in the tree of geodesics (that is, along the geodesics whose start or endpoints depend on $p_j$) via adjoint differentials of geodesics and sum up the contributions at $p_j$ (cf. Figure 2.4).[4]

**Example 2.1.** *For a quadratic Bézier curve $\beta$, it holds that*

$$
\mathrm{d}_{p_1}\beta(t; p_0, \cdot, p_2)^*(w) = \mathrm{d}_{p_1}\gamma\left(t; p_0, \cdot\right)^*\left(\mathrm{d}_{\gamma(t; p_0, p_1)}\gamma\left(t; \cdot, \gamma(t; p_1, p_2)\right)^*(w)\right)
$$
$$
+ \mathrm{d}_{p_1}\gamma\left(t; \cdot, p_2\right)^*\left(\mathrm{d}_{\gamma(t; p_1, p_2)}\gamma\left(t; \gamma(t; p_0, p_1), \cdot\right)^*(w)\right).
$$

To compute the adjoint, one must evaluate the operator $\mathrm{d}_p\gamma(t; \cdot, q)^*$ for $p, q \in U$, which is generally difficult. However, if $M$ is a symmetric space, then there is an explicit formula, just like for the derivative itself. Indeed, for $v \in T_pM$, and $w \in$

---

[4]When computing, it is better to add contributions that pass through the same point along the tree as soon as possible.

$T_{\gamma(t;p,q)}M$, Equation (2.11) then implies

$$\left\langle \mathrm{d}_p\gamma(t;\,\cdot\,,q)(v), w \right\rangle_{\gamma(t;p,q)} = \left\langle \sum_{i=1}^{d} \langle v, v_i \rangle_p\, a_i X_i(t), w \right\rangle_{\gamma(t;p,q)}$$

$$= \sum_{i=1}^{d} a_i \left\langle X_i(t), w \right\rangle_{\gamma(t;p,q)} \left\langle v, v_i \right\rangle_p$$

$$= \left\langle v, \sum_{i=1}^{d} \langle X_i(t), w \rangle_{\gamma(t;p,q)}\, a_i v_i \right\rangle_p .$$

Hence,

$$\mathrm{d}_p\gamma(t;\,\cdot\,,q)^*(w) = \sum_{i=1}^{d} \langle X_i(t), w \rangle_{\gamma(t;p,q)}\, a_i v_i.$$

Since Bézier splines are concatenated geodesics, we can also compute

$$\mathrm{d}_{p_j} B(t; p_0, \ldots, p_{j-1}, \,\cdot\,, p_{j+1}, \ldots, p_K)^*(w).$$

Indeed, as a consequence of Theorem 2.1.4 there is (depending on the value of *t*) only one more possible concatenation if $p_j$ connects two segments, or if it is the predecessor of such a control point.

## 2.2 Shape and Shape Spaces

In mathematics, a "shape" is the set of all the geometric properties of a (usually two- or three-dimensional) object that are invariant under similarity transformations, that is, under translations, rotations, and scalings. Depending on the application, different strategies to encode this information are employed. Consequently, there are several "shape spaces", which can be used to model object shapes. A classic example is Kendall shape space [142], which relies on landmark configurations; an excellent textbook on shape spaces based on landmarks is also [75]. An overview of shape spaces whose elements are diffeomorphisms can be found in [24]—with more in-depth discussions, for example, in [231, 265]. Skeletal models are discussed in [194, 224]. Finally, physics-based spaces were investigated in [11, 118, 245].

Not all of these spaces implement full invariance under similarity transformations; it is not done, for example, because of methodological reasons (for example, when size information is deemed decisive), to keep computational costs manageable, or because the representation does not allow for all three (but is useful in applications). Nevertheless, some degree of invariance is always achieved. What almost all of the above shape spaces have (at least locally) in common is an inherent manifold structure that is non-Euclidean. Therefore, tools from geometric statistics are needed for data analysis in them.

Quite often, shape spaces are also Lie groups. This connection dates back at least 100 years when D'arcy Thompson [239] had the idea to model shapes of homogeneous objects as *deformations* of a common *reference*; see Figure 2.5 for a visualization. This concept is present in several modern approaches to shape analysis, where the deformations are usually elements of Lie groups. Examples following this idea are [3,
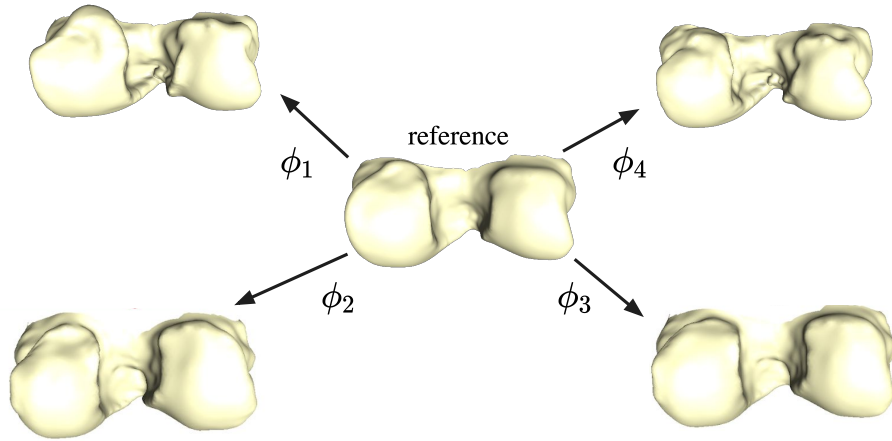
Figure 2.5: Shape via deformations visualized for distal femora. Given samples (here four), the maps $\phi_i$ are the deformations that the reference has to undergo to coincide with the samples. (Different types of deformation are used; for example, they can be diffeomorphisms of the surrounding space or bijections between triangle meshes.) The shape of each object is encoded in terms of the corresponding deformation $\phi_i$ (for example, as the deformation itself or some derived quantity like its derivative).

36], where configurations of the human spine are represented in a product group consisting of translations and rotations. We show two further examples of shape spaces that rely on deformations; both will be used later in this work.

**Example 2.2. $\mathrm{GL}^+(3)$-based shape space ([11]).** *Suppose that n homogeneous objects are given as triangulated surfaces $\mathcal{T}_i \subset \mathbb{R}^3$ that are in correspondence, scaled to the same size, and Procrustes aligned[5]. Then, we can describe their shapes with the help of the Lie group $\mathrm{GL}^+(3)$. To obtain this representation, we view, for each $i = 1, \ldots, n$, the mesh $\mathcal{T}_i$ as the result of a deformation $\phi_i$ of a shared reference mesh $\overline{\mathcal{T}} \subset \mathbb{R}^3$; that is, each $\phi_i : \overline{\mathcal{T}} \to \mathcal{T}_i$ is an orientation-preserving isomorphism that yields a semantic correspondence.*

*The idea now is that the derivatives of these deformations comprise the shape information because they provide a local characterization of the deformation without being influenced by the location in space. Thus, the Jacobian $d\phi_i$ constitutes the shape representation of the i-th object. More formally, let m be the number of triangles of each mesh. (It must be the same for all.) Since each derivative $d\phi_i$ is constant on each face $\overline{F}_j$ of $\overline{\mathcal{T}}$ and orientation-preserving, there are 3-by-3 matrices $G_j^{(i)} \in \mathrm{GL}^+(3)$ such that*

$$d\phi_i\big|_{\overline{F}_j} = G_j^{(i)} \tag{2.12}$$

*for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$. The i-th shape is thus given by*

$$(G_1^{(i)}, \ldots, G_m^{(i)}) \in \mathrm{GL}^+(3)^m,$$

*so $\mathrm{GL}^+(3)^m$ constitutes the shape space. We visualize the encoding for two meshes of distal femora in Figure 2.6. Note that this model stems from a continuous formulation [245, Sec. 2] that admits consistent and convergent discretizations.*

---

[5]Generalized (partial) Procrustes alignment [102] superimposes the meshes such that translational and rotational differences are minimized.
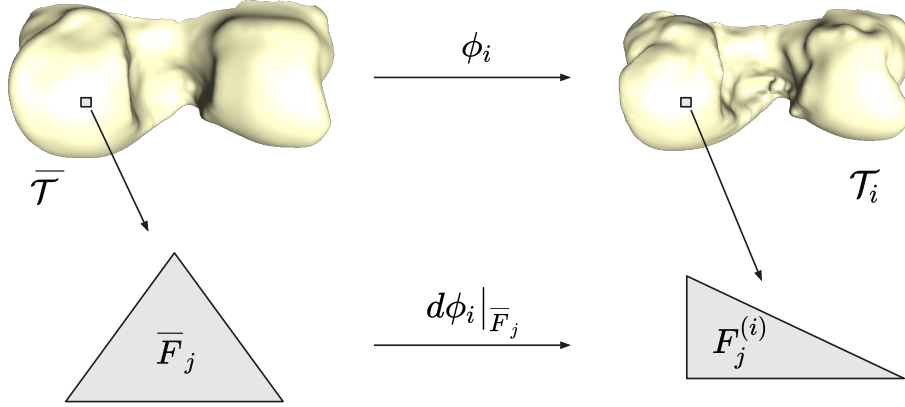
Figure 2.6: Construction of the $\mathrm{GL}^+(3)$-based shape space for femora. The mesh $\overline{\mathcal{T}}$ representing the reference femur is deformed into the mesh $\mathcal{T}_i$ of the $i$-th observation by $\phi_i$. The Jacobian $d\phi_i\big|_{\overline{F}_j} = G_j^{(i)}$ thereby maps the $j$-th triangle $\overline{F}_j$ of $\overline{\mathcal{T}}$ into the $j$-th triangle $F_j^{(i)}$ of $\mathcal{T}_i$.

*The task of obtaining a triangle mesh whose Jacobian is closest to given differential coordinates leads to a variational problem. Its minimizer is given by the solution of the well-known Poisson equation for which fast numerical solvers exist. Furthermore, as a global variational approach, the minimizer given by the Poisson equation tends to distribute errors uniformly such that local gradient field inconsistencies are attenuated.*

*In this work, we always endow the shape space $\mathrm{GL}^+(3)^m$ with its product Lie group structure obtained from Example 1.1. As part of this thesis, we develop new tools that will make the choice of $\overline{\mathcal{T}}$ unimportant: They will be invariant under translations, and selecting another reference $\widetilde{\mathcal{T}}$ only amounts to right-translating all shapes $(G_1^{(i)}, \ldots, G_m^{(i)})$ with the same element from $\mathrm{GL}^+(3)^m$ (viz., the derivative of the simplicial deformation between $\widetilde{\mathcal{T}}$ and $\overline{\mathcal{T}}$).*

We do not specify a Riemannian structure on this shape space as we will not need one (but see [11] for a well-founded choice). It should be emphasized that no bi-invariant metric exists on $\mathrm{GL}^+(3)$, and thus also on the above shape space. Therefore, it does not have a Levi-Civita connection that coincides with its CCS connection.

Whenever we want to analyze shapes in a manifold with a Riemannian structure, we rely on a slightly different space, which we introduce in the following example.

**Example 2.3. Shape space of differential coordinates ([245]).** *Suppose again that n homogeneous objects are given as triangle meshes $\mathcal{T}_i \subset \mathbb{R}^3$ that are in correspondence, scaled to the same size, and Procrustes aligned. To encode shapes as differential coordinates, one initially proceeds as in Example 2.2 and computes the matrices given by Equation (2.12). By using the polar decomposition, we can then find rotation matrices $R_j^{(i)} \in \mathrm{SO}(3)$ and symmetric positive definite matrices $P_j^{(i)} \in \mathrm{SPD}(3)$ such that*

$$G_j^{(i)} = R_j^{(i)} P_j^{(i)}$$

*for all $i = 1, \ldots, n$, and $j = 1, \ldots, m$. Both have a physical meaning: $R_j^{(i)}$ and $P_j^{(i)}$ can be viewed as the rotation and stretch, respectively, that $\overline{F}_j$ undergoes when $\overline{\mathcal{T}}$ is deformed by $\phi_i$. We call*

$$\left( \left( R_1^{(i)}, U_1^{(i)} \right), \ldots, \left( R_m^{(i)}, U_m^{(i)} \right) \right) \in \left( \mathrm{SO}(3) \times \mathrm{SPD}(3) \right)^m$$

*the differential coordinates of $\mathcal{T}_i$. The space of differential coordinates for homogeneous triangle meshes with $m$ faces then is*

$$\Delta := \left( \mathrm{SO}(3) \times \mathrm{SPD}(3) \right)^m.$$

*We endow $\Delta$ with the Riemannian product structure it inherits from $\mathrm{SO}(3)$ and $\mathrm{SPD}(3)$ (as given in Examples 1.6 and 1.7). [6] Hence, all operations work component-wise.*

*Given differential coordinates $q \in \Delta$ and a reference $\overline{\mathcal{T}}$, the triangular mesh corresponding to $q$ can be found as in Example 2.2. (We obtain the matrices (2.12) by multiplying the rotations from the left to the corresponding stretches.)*

*When Riemannian methods are used, it is essential to minimize the bias that comes with a particular choice of reference. We, therefore, compute $\overline{\mathcal{T}}$ in an iterative process as the mesh representation of the data's Fréchet mean as follows. To initialize, we choose one of the given geometries as the reference, say $\mathcal{T}_1$, and encode the shapes of each $\mathcal{T}_i$ in $\Delta$ for $i = 1, \ldots, n$. Then, we compute their Fréchet mean with Algorithm 1 and use its mesh representation as the new reference. This procedure is repeated until it converges or the residual norm does not decrease any further.*

Later in this work, we often visualize the results of computations in one of the above shape spaces. The depicted objects are the triangle meshes that correspond to the computed shapes (which are the solutions to the Poisson problem described above).

---

[6]Note that we have endowed $\Delta$ with a bi-invariant metric (since it is a product of bi-invariant metrics). Thus, the Levi-Civita connection coincides with the CCS connection given by the group structure of $\Delta$. Because, to us, the group structure of $\mathrm{GL}^+(3)^m$ feels more natural and there are promising results in real-world application of the $\mathrm{GL}^+(3)$-based space, we rely on the latter when there exist bi-invariant tools for the task at hand.

# Chapter 3

# Statistics in Riemannian Manifolds and Lie Groups

In this chapter, we repeat several concepts from statistics in manifolds that will be important later. As our focus in this thesis lies mainly on the analysis of *finite* sets of data points, we concentrate on sample statistics. Still, there are often analogous notions for random variables with general probability distributions that motivate them, and we often mention touch on the continuous case to give a broader picture.

When transferring statistical methods to manifolds, it has been very fruitful to utilize the geometry of the underlying space [189]. Connections and metrics provide the structure that makes it possible to define many notions in ways that are natural and often consistent (the latter meaning that they coincide with their multivariate counterparts in Euclidean space). The Riemannian framework has proven particularly useful. We will utilize it in Chapter 5 to extend regression theory for general manifolds.

In the case of Lie group–valued data, one further differentiates between notions that presuppose a Riemannian structure and notions that only rely on the CCS connection. These represent parallel approaches that concentrate on the compatibility of the theory with different (geometric and group-algebraic) properties of the underlying space. Because a good understanding of their differences will be necessary, both approaches are contrasted in this chapter. First, we repeat the multivariate concepts and then explain how these are transferred to the different settings.

As part of our exposition, we pay special attention to the differentiation and possible identification of contra- and covariant formulations of covariance tensor fields. A fundamental understanding of this topic is highly relevant; nevertheless, it has not received much space in works where the Riemannian framework is used.

## 3.1   Notions of Interest from Multivariate Statistics

We start by recalling central notions from multivariate statistics that will be of interest later on. There is a plethora of literature on the subject; standard references are [157] and [209]. We silently assume an underlying probability space whenever we speak of a random variable. When we consider an i.i.d. data set, say $(y_1, \ldots, y_d)$, consisting of independent realizations of a random variable $\mathcal{Y}$, we write $(y_1, \ldots, y_d) \overset{i.i.d.}{\sim} \mathcal{Y}$.

### 3.1.1  Mean

The mean, ubiquitous in multivariate statistics and probability theory, is probably the most fundamental characteristic we can assign to a (data) distribution. Let $\mathcal{Y}$ be a random variable on $\mathbb{R}^d$ with probability density function (pdf) $\nu$. The *mean* (or expectation) of $\mathcal{Y}$ is defined by

$$\mathbb{E}(\mathcal{Y}) := \int_{\mathbb{R}^d} x\nu(x)\,\mathrm{d}x; \tag{3.1}$$

here $\mathrm{d}x$ denotes the standard volume form on $\mathbb{R}^d$. (As is well known, there is a more general definition of the mean involving probability measures. Nevertheless, the above one suffices for us.) The *sample mean* $\overline{y}$ of a data set $(y_1,\dots,y_n) \overset{i.i.d.}{\sim} \mathcal{Y}$ in $\mathbb{R}^d$ is then defined by

$$\overline{y} := \frac{1}{n}\sum_{i=1}^{n} y_i. \tag{3.2}$$

It is a consistent and unbiased estimator of (3.1) if the given samples are independent realizations of a random variable $\mathcal{Y}$.

   The above definitions cannot be directly generalized to a manifold $M$; the reasons are that the latter lacks the necessary analogs to the vector space operations (addition and scalar multiplication) and a notion of an integral of $M$-valued functions. But Fréchet noticed in his seminal work [95] that, with $\mathrm{dist}(x,y) := \|x-y\|$ denoting the Euclidean distance function on $\mathbb{R}^d$, Equations (3.1) and (3.2) are equivalent to

$$\mathbb{E}(\mathcal{Y}) = \arg\min_{y\in\mathbb{R}^d} \int_{\mathbb{R}^d} \mathrm{dist}(x,y)^2\,\nu(x)\,\mathrm{d}x \tag{3.3}$$

and

$$\overline{y} = \arg\min_{y\in\mathbb{R}^d} \sum_{i=1}^{n} \mathrm{dist}(y_i,y)^2, \tag{3.4}$$

respectively. One can further show [139] that

$$\int_{\mathbb{R}^d} (x-y)\,\nu(x)\,\mathrm{d}x = 0 \quad \text{if and only if} \quad y = \mathbb{E}(\mathcal{Y}), \tag{3.5}$$

and

$$\sum_{i=1}^{n} (y_i - y) = 0 \quad \text{if and only if} \quad y = \overline{y}. \tag{3.6}$$

   We will see below that Equations (3.3) to (3.6) can be used to generalize the (sample) mean to different settings on manifolds.

### 3.1.2  Covariance

Let $\mathcal{X}$ and $\mathcal{Z}$ be two $\mathbb{R}$-valued random variables with joint probability density function $\nu_{\mathcal{X},\mathcal{Z}}$ on $\mathbb{R}^2$. Existence assumed, their *covariance* is defined by

$$\begin{aligned}
\mathrm{Cov}(\mathcal{X},\mathcal{Z}) &:= \mathbb{E}\Big(\big(\mathcal{X} - \mathbb{E}(\mathcal{X})\big)\big(\mathcal{Z} - \mathbb{E}(\mathcal{Z})\big)\Big) \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} \big(x - \mathbb{E}(\mathcal{X})\big)\big(z - \mathbb{E}(\mathcal{Z})\big)\nu_{\mathcal{X},\mathcal{Z}}(x,z)\,\mathrm{d}z\,\mathrm{d}x.
\end{aligned}$$

The *variance* of $\mathcal{X}$ is

$$\text{Var}(\mathcal{X}) := \text{Cov}(\mathcal{X}, \mathcal{X}).$$

Let $\mathcal{Y}$ be a random variable on $\mathbb{R}^d$ with pdf $\nu$. Since $\mathcal{Y} = \begin{bmatrix} \mathcal{Y}_1 & \cdots & \mathcal{Y}_d \end{bmatrix}^T$ is a vector with scalar-valued random variables as components, one finds (assuming existence)

$$\Sigma_{\mathcal{Y}} := \text{Cov}(\mathcal{Y}, \mathcal{Y}) = \begin{bmatrix} \text{Var}(\mathcal{Y}_1) & \cdots & \text{Cov}(\mathcal{Y}_1, \mathcal{Y}_d) \\ \vdots & \ddots & \vdots \\ \text{Cov}(\mathcal{Y}_d, \mathcal{Y}_1) & \cdots & \text{Var}(\mathcal{Y}_d) \end{bmatrix} \in \mathbb{R}^{d,d},$$

which is called *covariance matrix* of $\mathcal{Y}$. It is symmetric and positive semi-definite (positive-definite under very reasonable assumptions on $\mathcal{Y}$). Furthermore, there is the additional characterization

$$\Sigma_{\mathcal{Y}} = \int_{\mathbb{R}^d} (x - \mathbb{E}(\mathcal{Y}))(x - \mathbb{E}(\mathcal{Y}))^T \nu(x) \, dx.$$

Since $\text{Log}_x(y) = y - x$ for all $x, y \in \mathbb{R}^d$, the equation can also be written as

$$\Sigma_{\mathcal{Y}} = \int_{\mathbb{R}^d} \text{Log}_{\mathbb{E}(\mathcal{Y})}(x) \text{Log}_{\mathbb{E}(\mathcal{Y})}(x)^T \nu(x) \, dx. \tag{3.7}$$

Hence, letting an integral sign for tensors denote component-wise integration, $\Sigma_{\mathcal{Y}}$ is the representing matrix of the (2,0)-tensor

$$\Sigma_{\mathcal{Y}} = \int_{\mathbb{R}^d} \text{Log}_{\mathbb{E}(\mathcal{Y})}(x) \otimes \text{Log}_{\mathbb{E}(\mathcal{Y})}(x) \nu(x) \, dx \in T_{\mathbb{E}(\mathcal{Y})}\mathbb{R}^d \otimes T_{\mathbb{E}(\mathcal{Y})}\mathbb{R}^d. \tag{3.8}$$

Note that we use the same symbol for the tensor and its representing matrix. This is justified by the fact that, for $\mathbb{R}^d$, there is the canonical identification of bilinear maps on a tangent space $T_x\mathbb{R}^d$ with its representing matrix. It is obtained by identifying the $i$-th partial derivative with the canonical basis vector $e_i$ and computing components of the matrix representation in this basis. (Indeed, we have used this identification of tangent vectors with elements of $\mathbb{R}^d$ already in Equation 3.7.)

We want to emphasize that, first of all, $\Sigma_{\mathcal{Y}}$ is a bilinear map on the *cotangent* space $T^*_{\mathbb{E}(\mathcal{Y})}\mathbb{R}^d$. But since the canonical basis (of partial derivatives) is orthonormal, the identity matrix naturally represents the metric. Hence, as seen in Example 1.5, lowering indices does not change the components of the tensor. Therefore, $\Sigma_{\mathcal{Y}}$ is also the representing matrix of the (0,2)-tensor

$$\Sigma^{\mathcal{Y}} = \int_{\mathbb{R}^d} \text{Log}_{\mathbb{E}(\mathcal{Y})}(x)^\flat \otimes \text{Log}_{\mathbb{E}(\mathcal{Y})}(x)^\flat \nu(x) \, dx \in T^*_{\mathbb{E}(\mathcal{Y})}\mathbb{R}^d \otimes T^*_{\mathbb{E}(\mathcal{Y})}\mathbb{R}^d \tag{3.9}$$

given by

$$(T_{\mathbb{E}(\mathcal{Y})}\mathbb{R}^d)^2 \cong (\mathbb{R}^d)^2 \ni (r, z) \mapsto r^T \Sigma_{\mathcal{Y}} z. \tag{3.10}$$

Denoting the standard Euclidean metric by $\langle \, \cdot \, , \, \cdot \, \rangle$, the map (3.10) encodes the covariance of *arbitrary* linear combinations $r^T \mathcal{Y}$ and $z^T \mathcal{Y}$ via

$$r^T \Sigma_{\mathcal{Y}} z = \int_{\mathbb{R}^d} \langle r, \text{Log}_{\mathbb{E}(\mathcal{Y})}(x) \rangle \langle \text{Log}_{\mathbb{E}(\mathcal{Y})}(x), z \rangle \nu(x) \, dx. \tag{3.11}$$

In particular, we find that the entry at $(i, j)$ of $\Sigma_\mathcal{Y}$ is given by

$$e_i^T \Sigma_\mathcal{Y} e_j = \mathbb{E}[\langle e_i, \text{Log}_{\mathbb{E}[Y]}(\mathcal{Y}) \rangle \langle \text{Log}_{\mathbb{E}[Y]}(\mathcal{Y}), e_j \rangle]. \tag{3.12}$$

When $\nu$ has compact support, one can further define the (2,0)- and (0,2)-*covariance tensor fields* $\mathfrak{M}_\mathcal{Y} \in \Gamma(T_0^2 \mathbb{R}^d)$ and $\mathfrak{M}^\mathcal{Y} \in \Gamma(T_2^0 \mathbb{R}^d)$ of $\mathcal{Y}$. At each $y \in \mathbb{R}^d$ they are given by

$$\mathfrak{M}_\mathcal{Y}\big|_y := \int_{\mathbb{R}^d} \text{Log}_y(x) \otimes \text{Log}_y(x) \nu(x) \, dx \tag{3.13}$$

and

$$\mathfrak{M}^\mathcal{Y}\big|_y := \int_{\mathbb{R}^d} \text{Log}_y(x)^\flat \otimes \text{Log}_y(x)^\flat \nu(x) \, dx, \tag{3.14}$$

respectively. The tensors (3.8) and (3.9) are the restrictions of (3.13) and (3.14) to $y = \mathbb{E}(\mathcal{Y})$. As discussed in Example 1.4, the standard coordinate frame is orthonormal so that $\mathfrak{M}_\mathcal{Y}$ and $\mathfrak{M}^\mathcal{Y}$ everywhere have the same representing matrix

$$\mathfrak{M}_\mathcal{Y}\big|_y = \int_{\mathbb{R}^d} \text{Log}_y(x) \text{Log}_y(x)^T \nu(x) \, dx$$

in this frame.

Analogously, the (2,0)- and (0,2)-*sample covariance tensor fields* $\mathfrak{M}_{y_i} \in \Gamma(T_0^2 \mathbb{R}^d)$ and $\mathfrak{M}^{y_i} \in \Gamma(T_2^0 \mathbb{R}^d)$ of a data set $(y_1, \ldots, y_n) \overset{i.i.d.}{\sim} \mathcal{Y}$ in $\mathbb{R}^d$ are point-wise defined by[1]

$$\mathfrak{M}_{y_i}\big|_y := \frac{1}{n} \sum_{i=1}^n \text{Log}_y(y_i) \otimes \text{Log}_y(y_i)$$

and

$$\mathfrak{M}^{y_i}\big|_y := \frac{1}{n} \sum_{i=1}^n \text{Log}_y(y_i)^\flat \otimes \text{Log}_y(y_i)^\flat,$$

respectively. They also share the representing matrix

$$\mathfrak{M}_{y_i}\big|_y := \frac{1}{n} \sum_{i=1}^n \text{Log}_y(y_i) \text{Log}_y(y_i)^T = \frac{1}{n} \sum_{i=1}^n (y_i - \overline{y})(y_i - \overline{y})^T$$

in canonical coordinates. The *sample covariance (matrix)* of the data is the special case

$$\Sigma_{y_i} := \mathfrak{M}_{y_i}\big|_{\overline{y}} = \frac{1}{n} \sum_{i=1}^n \text{Log}_{\overline{y}}(y_i) \text{Log}_{\overline{y}}(y_i)^T = \frac{1}{n} \sum_{i=1}^n (y_i - \overline{y})(y_i - \overline{y})^T. \tag{3.15}$$

The latter is always a symmetric, positive semi-definite matrix and usually positive definite when enough samples are available. Analogously to the continuous case, we obtain

$$r^T \Sigma z = \sum_{i=1}^n \langle r, \text{Log}_{\overline{y}}(y_i) \rangle \langle \text{Log}_{\overline{y}}(y_i), z \rangle. \tag{3.16}$$

For $r = e_i$ and $z = e_j$, $i, j = 1, \ldots, n$ we obtain an estimator to the expectation (3.12) of the underlying random variable.

We will soon see that the tensor view of the covariance turned is natural for generalizations to manifolds.

---

[1] Another possibility when defining the sample covariance is to use the factor $\frac{1}{n-1}$ instead of $\frac{1}{n}$. Both can be found in the literature on statistics in manifolds. They are connected to different settings and quality criteria for estimators [157]. In Euclidean space, the factor $\frac{1}{n-1}$ turns the sample covariance into an unbiased estimator (but only if one considers the sample covariance as an element of $\mathbb{R}^{d,d}$ [228]).

### 3.1.3 Mahalanobis Distance

The Mahalanobis distance is widely used to determine how far a point is away from a (sample) distribution. We only summarize the sample case; the random variable case is analogous.

Let data $(y_1, \ldots, y_n) \overset{i.i.d.}{\sim} \mathcal{Y}$ in $\mathbb{R}^d$ with mean $\overline{y}$ and invertible sample covariance $\Sigma_{y_i}$ be given. The *squared Mahalanobis distance* of a point $x \in \mathbb{R}^d$ to the distribution of the data is defined by

$$\mu_{(\overline{y}, \Sigma_{y_i})}(x) := \mathrm{Log}_{\overline{y}}(x)^T \Sigma_{y_i}^{-1} \mathrm{Log}_{\overline{y}}(x) \tag{3.17}$$
$$= (x - \overline{y})^T \Sigma_{y_i}^{-1} (x - \overline{y}).$$

It is a (0,2)-tensor on $T_{\overline{y}} \mathbb{R}^d$. Since $\Sigma_{y_i}^{-1}$ is positive definite, its root is well-defined. The Mahalanobis distance measures how many standard deviations the observation is away from the distribution's mean. Directions in which high variability was observed are weighted less than those with little. Note that for $\Sigma_{y_i}$ to be invertible, the family $(y_1, \ldots, y_n) \overset{i.i.d.}{\sim} \mathcal{Y}$ must contain a basis of $\mathbb{R}^d$; in particular, $n \geq d$ must hold. We can further see that

$$\mu_{(\overline{y}, \Sigma_{y_i})}(x) = \mu_{(0, \Sigma_{y_i})}(x - \overline{y}); \tag{3.18}$$

this formulation will become helpful when we work with Mahalanobis distances in Lie groups

### 3.1.4 Hotelling $T^2$ Statistic

Let $(x_1, \ldots, x_m) \overset{i.i.d.}{\sim} \mathcal{X}$ and $(y_1, \ldots, y_n) \overset{i.i.d.}{\sim} \mathcal{Y}$ be two data sets in $\mathbb{R}^d$, each with independent realizations of normally distributed random variables $\mathcal{X}, \mathcal{Y}$. We further assume homoscedasticity, that is, $\mathcal{X}$ and $\mathcal{Y}$ share the same covariance matrix $\Sigma$. The data's *pooled sample covariance* is then given (after trivially lowering indices) by the (0,2)-tensor

$$\widehat{\Sigma}_{x_i, y_i} := \frac{1}{m+n-2} \left( \sum_{i=1}^{m} \mathrm{Log}_{\overline{x}}(x_i) \mathrm{Log}_{\overline{x}}(x_i)^T + \sum_{j=1}^{n} \mathrm{Log}_{\overline{y}}(y_i) \mathrm{Log}_{\overline{y}}(y_i)^T \right) \tag{3.19}$$

$$= \frac{1}{m+n-2} \left( \sum_{i=1}^{m} (x_i - \overline{x})(x_i - \overline{x})^T + \sum_{j=1}^{n} (y_j - \overline{y})(y_j - \overline{y})^T \right).$$

It is an (unbiased) estimator of $\Sigma$.

We can use the pooled sample covariance to define the Hotelling $T^2$ statistic. The latter is a fundamental tool from multivariate statistics that measures the difference between the means of the two sets of samples. It is defined by

$$t^2\big((x_i), (y_i)\big) := \frac{mn}{m+n} \mu^2_{(0, \widehat{\Sigma}_{x_i, y_i})}(\overline{x} - \overline{y}) = \frac{mn}{m+n} (\overline{x} - \overline{y})^T \widehat{\Sigma}_{x_i, y_i}^{-1} (\overline{x} - \overline{y}). \tag{3.20}$$

Its most prominent use is in two-sample hypothesis testing; see, for example, [97, 116]. Other applications are in (statistical) fault detection [134, 163, 217, 260, 262].

49

### 3.1.5 Bhattacharyya Distance

Another index to measure the dissimilarity of two multivariate probability distributions is the Bhattacharyya distance [138]. It has found numerous applications, for example, in statistical hypothesis testing [50, 180, 236], feature selection and extraction [55, 110, 259], image processing [104], and uncertainty quantification [34, 35]. In its most general form, it is defined for distributions with probability density function (pdf). Let $\rho, \nu$ be pdfs on $\mathbb{R}^d$. Then, their Bhattacharyya distance is given by

$$D_B^{\text{int}}(\rho, \nu) := -\ln\left(\int_{\mathbb{R}^d} \sqrt{\rho(x)\nu(x)}\, dx\right). \tag{3.21}$$

Let now $\rho$ and $\nu$ be two multivariate normal distributions with means $\overline{x}, \overline{y}$ and covariance matrices $\Sigma_x, \Sigma_y$, respectively. We define the *averaged covariance* by $\overline{\Sigma}_{x,y} := 1/2\,(\Sigma_x + \Sigma_y)$. Then, it can be shown [138, Eqn. (58)] that (3.21) becomes

$$D_B^{\text{int}}(\rho, \nu) = \frac{1}{8}\mu_{(0,\overline{\Sigma}_{x,y})}^2(\overline{x} - \overline{y}) + \frac{1}{2}\ln\left(\frac{\det(\overline{\Sigma}_{x,y})}{\sqrt{\det(\Sigma_x)\det(\Sigma_y)}}\right). \tag{3.22}$$

Note that the second summand vanishes if $\Sigma_x = \Sigma_y$; that is, the Bhattacharyya distance is the (scaled) square of the Mahalanobis distance for two normal distributions with the same covariance matrix.

We can directly translate Equation (3.22) to distributions of samples $(x_1, \ldots, x_m) \overset{i.i.d.}{\sim} \mathcal{X}$, $(y_1, \ldots, y_n) \overset{i.i.d.}{\sim} \mathcal{Y}$ in $\mathbb{R}^d$ when it can be assumed that each set consists of independent samples drawn from two normal distributions, respectively.[2] Then, we can replace means with sample means and covariance matrices with sample covariance matrices to arrive at

$$D_B\big((x_i), (y_i)\big) := \frac{1}{8}\mu_{(0,\overline{\Sigma}_{x_i,y_i})}^2(\overline{x} - \overline{y}) + \frac{1}{2}\ln\left(\frac{\det(\overline{\Sigma}_{x_i,y_i})}{\sqrt{\det(\Sigma_{x_i})\det(\Sigma_{y_i})}}\right). \tag{3.23}$$

Note that the first summand (without the constant factor) differs from Hotelling's $T^2$ statistic only because the averaged covariance matrix is used instead of the pooled one. An advantage of the Bhattacharyya distance over Hotelling's $T^2$ statistic is that it is also sensitive to differences in the covariance structure: The Hotelling $T^2$ statistic is zero if (and only if) the means coincide.

## 3.2 Statistics in Riemannian Manifolds

We now recall fundamental notions from statistics in Riemannian manifolds. Excellent references are [90] and [184].

---

[2]Technically, Equation (3.23) is well-defined independent from the underlying distributions as long as the involved covariance matrices are invertible. On the other hand, if the distributions differ from normal distributions, then (3.23) is not directly linked to (3.21) anymore.

### 3.2.1 Fréchet Mean

Fréchet derived Equations (3.3) and (3.4) as characterizations of the Euclidean mean. He noticed that they can be directly transferred to metric spaces, including Riemannian manifolds, by using the distance function. He obtained the following generalization of the mean (see Appendix A.1 for a summary of integration on manifolds).

**Definition 3.2.1** (Fréchet mean). *Let M be a Riemannian manifold with distance function* dist *and $\mathcal{Q}$ be a random variable on M with pdf $\rho$. A* Fréchet mean $\mathbb{E}(\mathcal{Q})$ *of $\mathcal{Q}$ is any global minimizer of the* Fréchet variance

$$\widetilde{F}(q) := \int_M \mathrm{dist}(q,p)^2 \, \rho(p) \, \mathrm{d}p.$$

*Accordingly, for data $(q_1, \ldots, q_n) \overset{i.i.d.}{\sim} \mathcal{Q}$ in M, any global minimizer $\overline{q}$ of the data's* sample Fréchet variance

$$F(q) := \sum_{i=1}^n \mathrm{dist}(q, q_i)^2 \tag{3.24}$$

*is called* sample Fréchet mean *of $q_1, \ldots, q_n$.*

We will often omit the word "sample"; it will be clear from the context which mean is meant.

As we deal with data sets in this thesis, we focus on the sample Fréchet mean in the following. Nevertheless, almost all of the following results can also be formulated in the continuous case.

If $M$ together with its distance function is a complete metric space, then the existence of a Fréchet mean is ensured [189, Thm. 2.1]. Let further $U \subseteq M$ be a normal convex neighborhood and $p \in U$. Define the map

$$\tau : U \to \mathbb{R}_{\geq 0},$$
$$q \mapsto \mathrm{dist}(q,p)^2.$$

It can be shown (see, for example, [190, Ch. 5]) that

$$\mathrm{grad}_q \tau = -2 \, \mathrm{Log}_q(p) \tag{3.25}$$

at all $q \in U$. Thus, we obtain the optimality condition

$$\sum_{i=1}^n \mathrm{Log}_{\overline{q}}(q_i) = 0 \tag{3.26}$$

at $\overline{q}$ from the gradient of the Fréchet variance $F$ (3.24); that is, in the tangent space of $\overline{q}$ the sum of "difference vectors" to the data points must vanish. We, therefore, say that $\overline{p}$ is an *exponential barycenter*. This condition is a direct generalization of (3.6); we visualize it in Figure 3.1. In general manifolds, however, $F$ can have local minima (and maxima), so solving (3.26) for $\overline{q}$ is not equivalent to minimizing (3.24) anymore. A solution of (3.26) that is a local minimum of $F$ is therefore usually called *Riemannian center of mass* [4, 139] (see [140] and the references therein for more on the history of the naming). We will later see that a Riemannian center of mass always exists if the
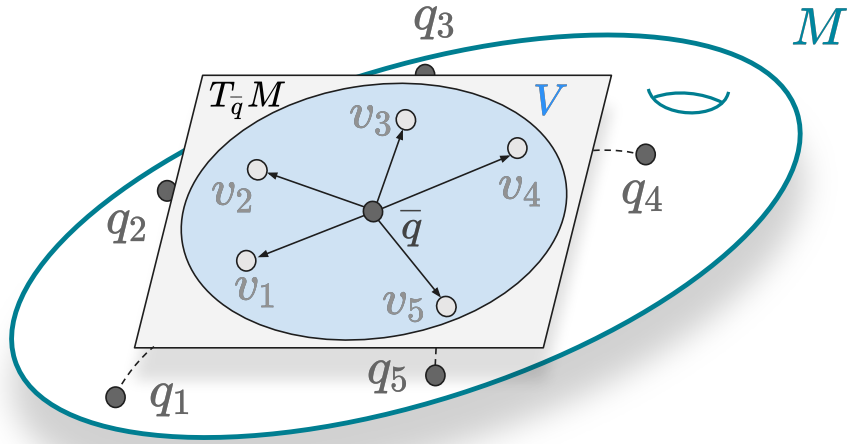
Figure 3.1: Fréchet mean/Riemannian center of mass of $q_1, \ldots, q_5$ in a Riemannian manifold $M$. After mapping the points into the tangent space $T_{\bar{q}}M$ at the Fréchet mean $\bar{q}$ with $\mathrm{Log}_{\bar{q}}$, the resulting tangent vectors $v_i := \mathrm{Log}_{\bar{q}}(q_i)$, $i = 1, \ldots, 5$, sum to zero. The broken lines depict the geodesics from the mean to the data points. Furthermore, $V \subset T_{\bar{q}}M$ indicates a neighborhood that is small enough such that $\mathrm{Exp}_{\bar{q}}(V) \subseteq B_r(\bar{q})$ with $r$ as in (3.27) such that the mean is unique.

data is contained in a normal convex neighborhood. (Notice already that the metric does not appear in Equation (3.26).) Fréchet means are also Riemannian centers of mass.

Importantly, if the data is contained in a sufficiently small neighborhood, then there is a unique Fréchet mean in its interior. Conditions on the size of this neighborhood were found by Karcher [139] and Kendall [143]. The most recent result we know is due to Afsari [4]. To state it, we need several notions. We define a *geodesic ball with center $o \in M$ and radius $r > 0$* by $B_r(o) := \{p \in M \mid \mathrm{dist}(p, o) < r\}$. Furthermore, let $\mathrm{inj}(p) := \sup\{r > 0 \mid \mathrm{Exp}_p \text{ is injective on } B_r(p)\}$ and $\mathrm{inj}(M) := \inf\{\mathrm{inj}(p) \mid p \in M\}$. Denoting the supremum of all sectional curvatures by $\kappa$, then Afsari showed that if the data is contained in a geodesic ball of radius at most

$$r = \begin{cases} \frac{1}{2}\min\left(\mathrm{inj}(M), \frac{\pi}{\sqrt{\kappa}}\right), & \kappa > 0, \\ \frac{1}{2}\mathrm{inj}(M), & \kappa \leq 0, \end{cases} \tag{3.27}$$

then there is a unique Fréchet mean inside the ball.

From now on, we assume that the data is concentrated enough to have a unique Fréchet mean.

The asymptotic behavior of the Fréchet mean has been studied intensively over the last 20 years [32, 33, 79, 123], with results on consistency as well as central limit theorems.

Fréchet means can be computed using Riemannian gradient descent (see [2] for a comprehensive account of optimization in Riemannian manifolds). From Equation (3.25), we thus obtain the following algorithm [184, Sec. 4.6].

---

**Algorithm 1** Fréchet mean

**Input:** data $(q_1, \ldots, q_n)$ within a normal convex neighbourhood of $M$, tolerance $\epsilon$, stepsize $\delta \leq 1$

**Output:** approximation $\widehat{q}$ of the Fréchet mean

$\widehat{q} \leftarrow q_1$

$v \leftarrow \delta \sum_{i=1}^{n} \text{Log}_{\widehat{q}}(q_i)$

**while** $\|v\|_{\widehat{q}} \geq \epsilon$ **do**

   $\widehat{q} \leftarrow \text{Exp}_{\widehat{q}}(v)$

   $v \leftarrow \delta \sum_{i=1}^{n} \text{Log}_{\widehat{q}}(q_i)$

**end while**

---

The convergence properties of this algorithm, which were found to be excellent in practice (for example, in [153, 184]), were studied in [5, 153, 154].

### 3.2.2 Riemannian Covariance

We now discuss how covariance is defined in Riemannian manifolds. Since, in general, tangent spaces of manifolds do not have canonical bases, we make the dependence of coordinate representations on the chosen basis clear by adjusting the notation. Namely, here and later, we denote the vector of coordinates of a tangent vector as well as the representing matrix of a (2,0)- or (0,2)-tensor (field) with brackets $[\,\cdot\,]$. It is always assumed that *arbitrary but fixed (coordinate) bases* are used to obtain the representation.

The tensor view of the covariance from Section 3.1.2 is extremely helpful in manifolds. Indeed, Equations (3.13) and (3.15) generalize directly [184, 188].

**Definition 3.2.2** (Riemannian (2,0)-covariance). *Let $M$ be a Riemannian manifold and $\mathcal{Q}$ a random variable on $M$ with pdf $\nu$ that has compact support in a normal convex neighborhood $U \subseteq M$. The (2,0)-covariance tensor field $\mathfrak{M}_{\mathcal{Q}} \in \Gamma(T_0^2 U)$ of $\mathcal{Q}$ is defined at each $q \in U$ by*

$$\mathfrak{M}_{\mathcal{Q}}\big|_q := \int_U \text{Log}_q(p) \otimes \text{Log}_q(p)\nu(p)\,\mathrm{d}p,$$

*with representing matrix*

$$[\mathfrak{M}_{\mathcal{Q}}\big|_q] = \int_U [\text{Log}_q(p)][\text{Log}_q(p)]^T \nu(p)\,\mathrm{d}p.$$

*Its restriction to $\mathbb{E}(\mathcal{Q})$ is called* (2,0)-covariance tensor *of $\mathcal{Q}$.*

*Let further $(q_1, \ldots, q_n) \overset{i.i.d.}{\sim} \mathcal{Q}$ be a data set in $U$ with Fréchet mean $\bar{q}$. The data's (2,0)-sample covariance tensor field $\mathfrak{M}_{q_i} \in \Gamma(T_0^2 U)$ is at each $q \in U$ defined by*

$$\mathfrak{M}_{q_i}\big|_q := \frac{1}{n} \sum_{i=1}^{n} \text{Log}_q(q_i) \otimes \text{Log}_q(q_i)$$

*with representing matrix*

$$[\mathfrak{M}_{q_i}\big|_q] = \frac{1}{n} \sum_{i=1}^{n} [\text{Log}_q(q_i)][\text{Log}_q(q_i)]^T. \tag{3.28}$$

*Its restriction*

$$\Sigma_{q_i} := \mathfrak{M}_{q_i}\big|_{\bar{q}}$$

*to the sample mean $\bar{q}$ is the data's (2,0)-sample covariance tensor.*

Let $[g_{ij}(q)]$ denote the representing matrix of the metric at $q$ in the chosen basis of $T_q M$. We obtain the (0,2)-sample covariance by lowering indices (see Example 1.3).

**Definition 3.2.3** (Riemannian (0,2)-covariance)**.** *Let $M$ be a Riemannian manifold and $\mathcal{Q}$ a random variable on $M$ with pdf $v$ that has compact support in a normal convex neighborhood $U \subseteq M$. The (0,2)-covariance tensor field $\mathfrak{M}^{\mathcal{Q}} \in \Gamma(T_2^0 U)$ of $\mathcal{Q}$ is defined at each $q \in U$ by*

$$\mathfrak{M}^{\mathcal{Q}}\big|_q := \int_U \mathrm{Log}_q(p)^\flat \otimes \mathrm{Log}_q(p)^\flat v(p)\, \mathrm{d}p \in \Gamma(T_2^0 U),$$

*with representing matrix*

$$[\mathfrak{M}^{\mathcal{Q}}\big|_q] = \int_U [g_{ij}(q)][\mathrm{Log}_q(p)][\mathrm{Log}_q(p)]^T [g_{ij}(q)] v(p)\, \mathrm{d}p.$$

*Its restriction to $\mathbb{E}(\mathcal{Q})$ is called (0,2)-covariance tensor of $\mathcal{Q}$.*

Let further $(q_1, \ldots, q_n) \overset{i.i.d.}{\sim} \mathcal{Q}$ be a data in $U$ with Fréchet mean $\bar{q}$. The data's (0,2)-sample covariance tensor field $\mathfrak{M}^{q_i} \in \Gamma(T_2^0 U)$ is at each $q \in U$ defined by

$$\mathfrak{M}^{q_i}\big|_q := \frac{1}{n}\sum_{i=1}^{n} \mathrm{Log}_q(q_i)^\flat \otimes \mathrm{Log}_q(q_i)^\flat$$

*with representing matrix*

$$[\mathfrak{M}^{q_i}\big|_q] = \frac{1}{n}\sum_{i=1}^{n}[g_{ij}(q)][\mathrm{Log}_q(q_i)][\mathrm{Log}_q(q_i)]^T[g_{ij}(q)] = [g_{ij}(q)]\,[\Sigma_{q_i}]\,[g_{ij}(q)]. \qquad (3.29)$$

*Its restriction*

$$\Sigma^{q_i} := \mathfrak{M}^{q_i}\big|_{\bar{q}}$$

*to the sample mean $\bar{q}$ is the data's (0,2)-sample covariance tensor.*

When the (0,2)-version of the covariance is of interest, one can thus either determine an orthonormal basis and compute the corresponding (2,0)-covariance—as discussed in Example 1.5, lowering indices then is a null operation—or calculate the components explicitly from Equation (3.29) in an arbitrary basis.

The above distinction of tensor fields is crucial since Equations (3.11), (3.12) and (3.16) generalize to the (0,2)- but not the (2,0)-covariance tensor. For example, if $(q_1, \ldots, q_n) \overset{i.i.d.}{\sim} \mathcal{Q}$ are independent realizations of an $M$-valued random variable $Q$ and $v_1, \ldots, v_d$ the basis of $T_{\bar{q}}M$ for which we obtain coordinate representations, we can interpret the $(i, j)$-entry of $[\Sigma^{q_i}]$—but not $[\Sigma_{q_i}]$—as an estimator of

$$\mathbb{E}[\langle \mathrm{Log}_{\mathbb{E}[Q]}(Q), v_i\rangle_{\mathbb{E}[Q]} \langle \mathrm{Log}_{\mathbb{E}[Q]}(Q), v_j\rangle_{\mathbb{E}[Q]}].$$

(See also [186, p. 39] for a discussion of this.)

### 3.2.3 Riemannian Mahalanobis Distance

The tools we gathered so far are can be used to generalize the Mahalanobis distance (3.17) to Riemannian manifolds. One uses the fact that, in Euclidean space, its square is the (0,2)-tensor on the tangent space at the mean whose representing matrix is the inverse covariance matrix.

**Definition 3.2.4** (Riemannian Mahalanobis distance [184])**.** *Let M be a Riemannian manifold. Further, let* $(q_1, \ldots, q_n) \overset{i.i.d.}{\sim} \mathcal{Q}$ *be data with unique Fréchet mean* $\overline{q}$ *and invertible (0,2)-sample covariance* $\Sigma_{q_i}$*. Assume that* $p \in M$ *such that* $\mathrm{Log}_{\overline{q}}(p)$ *exists. The* squared Riemannian Mahalanobis distance *from p to the distribution of the data is then given by*

$$\mu^2_{(\overline{q}, \Sigma_{q_i})}(p) := \left[\mathrm{Log}_{\overline{q}}(p)\right]^T \left[\Sigma_{q_i}\right]^{-1} \left[\mathrm{Log}_{\overline{q}}(p)\right]. \tag{3.30}$$

It can be verified that this definition does not depend on the chosen basis of $T_{\overline{q}}M$ as, indeed, any change of basis cancels out in (3.30). Note that since (2,0)-covariance tensor is used, the Riemannian metric does not appear in the definition. (Indeed, the (0,2)-covariance tensor cannot be used in the above definition since then changes of bases do not cancel out.) Still, the Riemannian Mahalanobis distance indirectly depends on the metric through the Levi-Civita connection, as the latter determines the logarithm.

### 3.2.4 Hotelling $T^2$ Statistic for Riemannian Manifolds

In [172, Sec. 3.3], Muralidharan and Fletcher introduce a generalization of the Hotelling $T^2$ statistic (3.20) to Riemannian manifolds $M$. Let $(p_1, \ldots, p_m) \overset{i.i.d.}{\sim} \mathcal{P}$, $(q_1, \ldots, q_n) \overset{i.i.d.}{\sim} \mathcal{Q}$ be data sets in $M$ with unique Fréchet means $\overline{p}, \overline{q}$ and sample covariance matrices $\Sigma_{p_i}, \Sigma_{q_i}$, respectively. To generalize the statistic, the difference between the means in (3.20) can be replaced by *either* $\mathrm{Log}_{\overline{p}}(\overline{q}) \in T_{\overline{p}}M$ *or* $\mathrm{Log}_{\overline{q}}(\overline{p}) \in T_{\overline{q}}M$. Depending on the choice, the resulting statistic will be different. Furthermore, since vectors from different tangent spaces cannot be canonically identified, there is no clear way to generalize the pooled covariance. Muralidharan and Fletcher, therefore, propose to calculate a generalized $T^2$ statistic at both means and average the results. This leads to the following definition. (Here and in the following, we use the abbreviation $(a_i)$ for a data set $(a_1, \ldots, a_n)$, when the set is the input of some function.)

**Definition 3.2.5** (Riemannian Hotelling $T^2$ statistic [172])**.** *Let M be a Riemannian manifold. Further, let* $(p_1, \ldots, p_m) \overset{i.i.d.}{\sim} \mathcal{P}$, $(q_1, \ldots, q_n) \overset{i.i.d.}{\sim} \mathcal{Q}$ *be data sets in M with unique Fréchet means* $\overline{p}, \overline{q} \in M$ *and sample covariances* $\Sigma_{p_i}, \Sigma_{q_i}$*, respectively. The* Riemannian Hotelling $T^2$ statistic *is then defined by*

$$\mathrm{t}^2\big((p_i), (q_i)\big) := \tfrac{1}{2}\left(\mu^2_{(\overline{q}, \Sigma_{q_i})}(\overline{p}) + \mu^2_{(\overline{p}, \Sigma_{p_i})}(\overline{q})\right). \tag{3.31}$$

**Remark 3.2.6.** *Note that Definition 3.2.5 does not reduce to (3.20) in the Euclidean case, that is, for data sets* $(x_1, \ldots, x_m) \overset{i.i.d.}{\sim} \mathcal{X}, (y_1, \ldots, y_n) \overset{i.i.d.}{\sim} \mathcal{Y}$ *in* $\mathbb{R}^d$*. To see this, observe first that, in this case,* $\mathrm{Log}_{\overline{x}}(\overline{y}) = -\mathrm{Log}_{\overline{y}}(\overline{x}) = \overline{y} - \overline{x}$*. Hence, we get from Equation (3.31) that*

$$\mathrm{t}^2\big((x_i), (y_i)\big) = (\overline{x} - \overline{y})^T \tfrac{1}{2}\left(\Sigma_{x_i}^{-1} + \Sigma_{y_i}^{-1}\right)(\overline{x} - \overline{y}).$$

*Thus, it is enough to show that $\widehat{\Sigma}_{x_i, y_i}^{-1} \neq 1/2 \left( \Sigma_{p_i}^{-1} + \Sigma_{q_i}^{-1} \right)$ in general. But this is true—not only because of scaling—since for symmetric positive definite matrices $P_1, P_2$ (of the same size), we have $P_1^{-1} + P_2^{-1} \neq (P_1 + P_2)^{-1}$ in general (and any such matrix can be a pooled sample covariance as can be seen with the help of the singular value decomposition).*

### 3.2.5 Riemannian Bhattacharyya Distance

A generalization of the Bhattacharyya distance (3.23) to Riemannian manifolds is due to Hong et al.; see [126, Sec. 3]. As with the pooled covariance, we cannot simply average the two covariance matrices since they are defined on different tangent spaces. Therefore, Hong et al. propose to average scalar terms, which leads to the following definition.[3]

**Definition 3.2.7** (Riemannian Bhattacharyya distance [126]). *Let $M$ be a Riemannian manifold. Further, let $(p_1, \ldots, p_m) \overset{i.i.d.}{\sim} \mathcal{P}$, $(q_1, \ldots, q_n) \overset{i.i.d.}{\sim} \mathcal{Q}$ be data sets in $M$ with unique Fréchet means $\overline{p}, \overline{q} \in M$ and sample covariances $\Sigma_{p_i}, \Sigma_{q_i}$, respectively. Then, the* Riemannian Bhattacharyya distance *is defined by*

$$\mathfrak{D}_B\big((p_i), (q_i)\big) := \frac{1}{16} \left( \mu^2_{(\overline{p}, \Sigma_{p_i})}(\overline{q}) + \mu^2_{(\overline{q}, \Sigma_{q_i})}(\overline{p}) \right) + \frac{1}{2} \ln \left( \frac{\det\left([\Sigma_{p_i}]\right) + \det\left([\Sigma_{q_i}]\right)}{2\sqrt{\det\left([\Sigma_{p_i}]\right) \det\left([\Sigma_{q_i}]\right)}} \right).$$

This definition is independent of the chosen bases of $T_{\overline{p}}M$ and $T_{\overline{q}}M$ as this is true for both the squared Mahalanobis distance and the determinant.

**Remark 3.2.8.** *Definition 3.2.7 also is not consistent with its multivariate analog (3.23) because $\det(P) + \det(Q) \neq \det(P + Q)$ for general symmetric positive definite matrices $P, Q$ of the same size.*

## 3.3 Non-metric Statistics in Lie Groups

We now turn to the case of Lie group-valued data. Of course, being manifolds, Lie groups can be endowed with a Riemannian metric, and then, the tools from the previous section can be used for data analysis. In many cases, this has the drawback that central group properties are ignored. Most importantly, the fact that bi-invariant metrics generally do not exist on Lie groups makes the construction of (statistical) notions that are invariant/equivariant under translations from *both* left and right very difficult in the Riemannian setting. On the other hand, Lie groups that are endowed with their CCS connections possess the natural prerequisites that allow us to construct bi-invariant statistical notions. This was first noted by Pennec and Arsigny in [186] and advanced to general affine connection spaces by Pennec and Lorenzi in [188]. In this section, we summarize many of their findings. The definitions in the following all look very similar to the Riemannian setting in Section 3.2. However, they always use the group's CCS connection, which, as mentioned before, is *not* the Levi-Civita connection of a Riemannian metric for general Lie groups. We will see that this alteration is decisive in obtaining notions consistent with the group structure.

---

[3]The formulation in [126] is more involved because the authors suppose that the sample covariance matrices are singular.

### 3.3.1 Group Mean

Let $G$ be a Lie group. We do not fix any Riemannian metric on $G$. Consequently, we do not have a notion of distance, so that Equation (3.4) cannot be generalized (that is, one cannot define a mean as the minimizer of a sum of distances in this setting). However, the local optimality conditions (3.5) and (3.6) only rely on the affine structure. Combined with the CCS connection, they can thus be used to define group means in Lie groups—without a metric—as exponential barycenters. A detail that changes is that we work with an (arbitrary but fixed) left or right Haar measure d$g$ instead of a Riemannian volume form when integrating; for a summary of the corresponding theory see Appendix A.1. With this, we make the following definition.

**Definition 3.3.1** (Group mean [186]). *Let G be a Lie group with CCS connection and $\mathcal{G}$ a random variable on G with pdf $\rho$ that has compact support in a normal convex neighborhood $U \subseteq G$. A group mean $\mathbb{E}(\mathcal{G})$ of $\mathcal{G}$ is implicitly defined by*

$$\int_U \mathrm{Log}_{\mathbb{E}(\mathcal{G})}(g)\rho(g)\,\mathrm{d}g = 0.$$

*Analogously, for data $(g_1, \ldots, g_n) \overset{i.i.d.}{\sim} \mathcal{G}$ in U we call $\overline{g} \in U$ sample group mean of $(g_1, \ldots, g_n)$, if*

$$\sum_{i=1}^n \mathrm{Log}_{\overline{g}}(g_i) = 0. \tag{3.32}$$

We focus on the sample group mean. Because of (1.20), Equation (3.32) is equivalent to both

$$\sum_{i=1}^n \log(\overline{g}^{-1}g_i) = 0 \quad \text{and} \quad \sum_{i=1}^n \log(g_i\overline{g}^{-1}) = 0.$$

Hence, $\overline{g}$ is the group mean of $(g_1, \ldots, g_n)$ if and only if $e$ is the group mean of the left and right translated data sets $(\overline{g}^{-1}g_1, \ldots, \overline{g}^{-1}g_n)$ and $(g_1\overline{g}^{-1}, \ldots, g_n\overline{g}^{-1})$. We can also see that the elements $g_1, \ldots, g_n$ must be "sufficiently localized" to obtain a mean value: if they are too far apart, the logarithm may not be defined. The most general result known so far on existence and uniqueness is the following: If $U \subseteq G$ is a CSLCG (convex with semilocal convex geometry) neighborhood and $g_1, \ldots, g_n \in U$, then their group mean exists and is unique. The details can be found in Appendix A.2.

Central properties of the sample group mean are summarized in the following theorem [188, Thm. 5.13].

**Theorem 3.3.2** (Equivariance of the sample group mean). *Let G be a Lie group endowed with its CCS connection and $\overline{g}$ be a sample group mean of a data set $(g_1, \ldots, g_n)$ as in Definition 3.3.1. Then, for any $f \in G$, the group means of the left translated data $(fg_1, \ldots, fg_n)$, right translated data $(g_1f, \ldots, g_nf)$ and inverted data $(g_1^{-1}, \ldots, g_n^{-1})$ are $f\overline{g}$, $\overline{g}f$ and $\overline{g}^{-1}$, respectively.*

According to the theorem, group means are equivariant under the group operations. These properties make (statistical) data analysis with the mean more robust; we visualize them for translations in Figure 3.3.

Finally, group means can be computed efficiently with a fixed point iteration when the data is sufficiently localized (for example, in a CSLCG neighborhood). Then, the
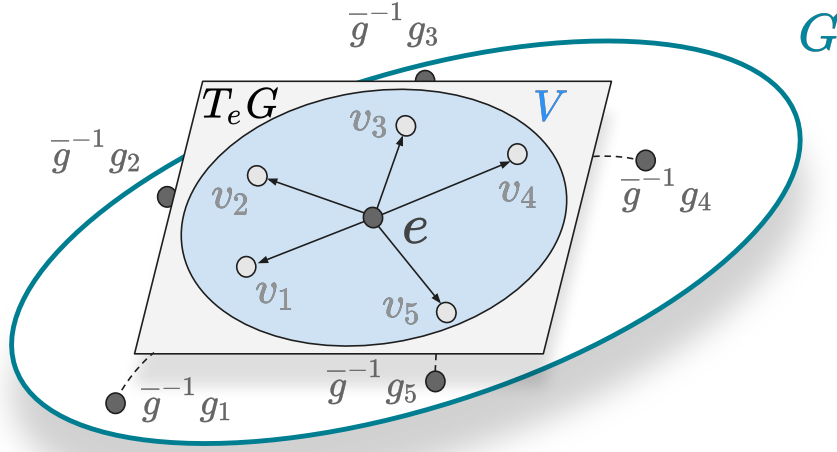
Figure 3.2: Group mean of a data set $(g_1, \ldots, g_5)$ in a Lie group $G$. The element $\overline{g}$ is the group mean if and only if $e$ is the mean of the translated data $(\overline{g}^{-1} g_1, \ldots, \overline{g}^{-1} g_n)$. The set $V$ indicates the image of a CSLCG neighborhood under the group logarithm log. After mapping the points into the tangent space $T_e G$, the resulting tangent vectors $\log(\overline{g}^{-1} g_i)$, $i = 1, \ldots, 5$, sum to zero. The broken lines are the geodesics (one-parameter subgroups) from the mean to the data points.

following algorithm converges at least with a linear rate to the (unique) solution [186, Cor. 5].

---

**Algorithm 2** Group mean

---

**Input:** data $(g_1, \ldots, g_n)$ within a sufficiently small CSLCG neighbourhood of $G$, tolerance $\epsilon$, auxiliary norm $\| \cdot \|$ on $T_e G$

**Output:** approximation $\widehat{g}$ of the group mean

$\widehat{g} \leftarrow g_1$

$v \leftarrow \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} \log(\widehat{g}^{-1} g_i)$

**while** $\|v\| \geq \epsilon$ **do**

$\quad \widehat{g} \leftarrow \widehat{g} \exp(v)$

$\quad v \leftarrow \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} \log(\widehat{g}^{-1} g_i)$

**end while**

---

This algorithm is very similar to Algorithm 1 but only coincides with it when $G$ is endowed with a bi-invariant metric. Note also that we do all vector computations in $T_e G$ instead of $T_{\overline{g}} G$ because it avoids unnecessary translations.

### 3.3.2 Sample Covariance for Lie Groups

As for Riemannian manifolds, the multivariate (0,2)-sample covariance can be generalized directly to Lie groups with affine connection when the group mean is used.
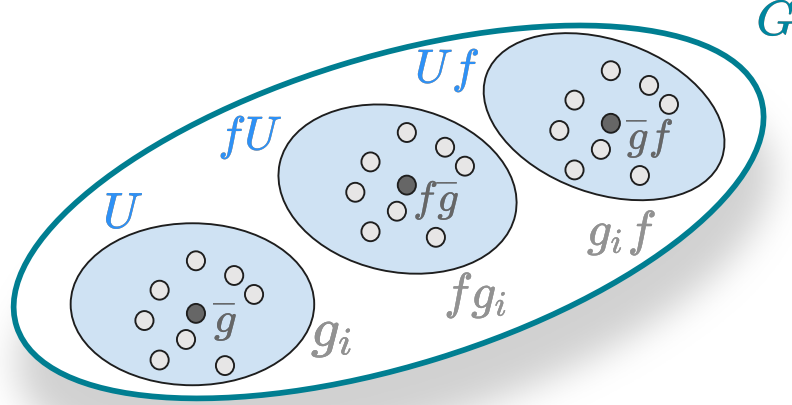
Figure 3.3: Visualization of the equivariance of the group mean under translations for a data set $(g_1, \ldots, g_n)$ in a CSLCG neighborhood $U$ of a Lie group $G$.

**Definition 3.3.3** (Group (2,0)-covariance [188])**.** *Let $G$ be a Lie group with CCS connection and $\mathcal{G}$ a random variable on $G$ with pdf $\nu$ that has compact support in a CSLCG neighborhood $U \subseteq G$. The (group) (2,0)-covariance tensor field $\mathfrak{M}_{\mathcal{G}} \in \Gamma(T_0^2 U)$ of $\mathcal{G}$ is defined at each $f \in U$ by*

$$\mathfrak{M}_{\mathcal{G}}\big|_f = \int_U \mathrm{Log}_f(g) \otimes \mathrm{Log}_f(g)\nu(g)\,\mathrm{d}g,$$

*with representing matrix*

$$[\mathfrak{M}_{\mathcal{G}}\big|_f] = \int_U [\mathrm{Log}_f(g)][\mathrm{Log}_f(g)]^T \nu(g)\,\mathrm{d}g.$$

*Its restriction to $\mathbb{E}(\mathcal{G})$ is called (2,0)-covariance (tensor) of $\mathcal{G}$.*

*Let further $(g_1, \ldots, g_n) \overset{i.i.d.}{\sim} \mathcal{G}$ be a data set in a CSLCG neighborhood $U \subseteq G$ with group mean $\overline{g}$. The data's (2,0)-sample covariance tensor field $\mathfrak{M}_{g_i} \in \Gamma(T_0^2 U)$ is at each $f \in U$ defined by*

$$\mathfrak{M}_{g_i}\big|_f := \frac{1}{n} \sum_{i=1}^n \mathrm{Log}_f(g_i) \otimes \mathrm{Log}_f(g_i)$$

*with representing matrix*

$$[\mathfrak{M}_{g_i}\big|_f] = \frac{1}{n} \sum_{i=1}^n [\mathrm{Log}_f(g_i)][\mathrm{Log}_f(g_i)]^T.$$

*Its restriction*

$$\Sigma_{g_i} := \mathfrak{M}_{g_i}\big|_{\overline{g}}$$

*to the sample mean $\overline{g}$ is the data's (2,0)-sample covariance (tensor).*

As discussed in Section 3.2.2, Equations (3.11), (3.12) and (3.16) do not generalize to this notion of covariance. However, its matrix *inverse* transforms like a (0,2)-tensor and thus yields a bilinear map on the tangent space.

### 3.3.3 Bi-invariant Mahalanobis Distance
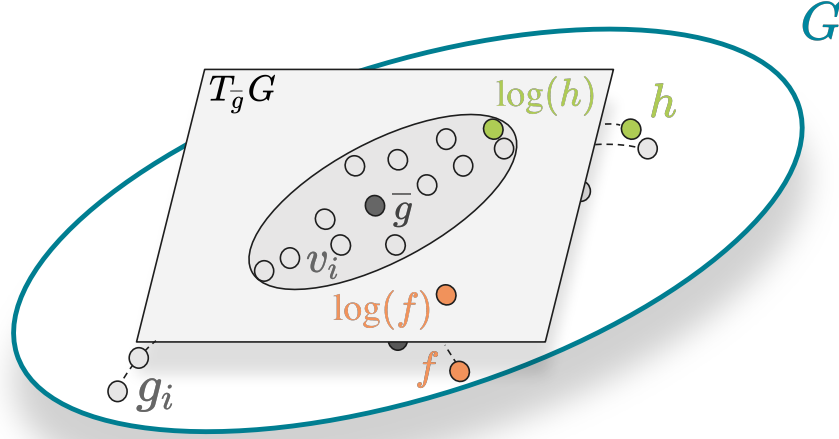


Figure 3.4: Bi-invariant Mahalanobis distance for a data set $(g_1, \ldots, g_n)$ with group mean $\bar{g}$ (dark grey) in a Lie group $G$. Here, $v_i := \mathrm{Log}_{\bar{g}}(g_i)$ are the corresponding vectors in the tangent space (light grey). The point $f \in G$ (orange) has a greater Mahalanobis distance to the distribution of the $g_i$ than the point $h$ (green) since the data has more variability in the direction of $h$ than in the direction of $f$ (as indicated by the ellipse in the tangent space whose boundary is an isoline of the bi-invariant Mahalanobis distance).

Pennec, Arsigny, and Lorenzi noticed that the Mahalanobis distance can be transferred to Lie groups with CCS connection just as for Riemannian manifolds as it only requires the (2,0)-covariance.

**Definition 3.3.4** (Bi-invariant Mahalanobis distance [186])**.** *Let $G$ be a Lie group with CCS connection, and $(g_1, \ldots, g_n) \overset{i.i.d.}{\sim} \mathcal{G}$ data within a CSLCG neighborhood in $G$ with group mean $\bar{g}$ and invertible sample covariance $\Sigma_{g_i}$. Let further $f \in G$ such that $\mathrm{Log}_{\bar{g}}(f)$ exists. Then, the squared bi-invariant Mahalanobis distance of $f$ to the distribution of the $g_i$ is defined by*

$$\mu^2_{(\bar{g}, \Sigma_{g_i})}(f) := \left[\mathrm{Log}_{\bar{g}}(f)\right]^T \left[\Sigma^{g_i}\right]^{-1} \left[\mathrm{Log}_{\bar{g}}(f)\right]. \tag{3.33}$$

One again verifies that this definition does not depend on the chosen basis. Just like its Euclidean counterpart, the bi-invariant Mahalanobis distance measures how far the point $f$ is away from the sample mean *in terms of the observed variability in the given direction*; this is visualized in Figure 3.4.

The following proposition shows that the bi-invariant Mahalanobis distance is indeed invariant under left and right translations—a property that, in general, does not hold when the Riemannian framework is used.

**Proposition 3.3.5.** *Let G be a Lie group with CCS connection, and $(g_1, \ldots, g_n) \stackrel{i.i.d.}{\sim} \mathcal{G}$ data within a CSLCG neighborhood in G with group mean $\overline{g}$. Let further $f \in G$ such that $\mathrm{Log}_{\overline{g}}(f)$ exists, and $h \in G$ be arbitrary. Then,*

$$\mu^2_{(h\overline{g}, \Sigma_{hg_i})}(hf) = \mu^2_{(\overline{g}, \Sigma_{g_i})}(f) = \mu^2_{(\overline{g}h, \Sigma_{g_ih})}(fh).$$

We refer to [188, p. 181] for the proof.

This finishes the overview of the literature. In the following chapters, we use presented notions to develop new statistical tools for manifolds.

# Part II

# Extension of the Theory, Novel Algorithms

# Chapter 4

# Bi-invariant Dissimilarity Measures for Non-metric Statistics in Lie Groups

Manifold-valued data occurs in many applications [54, 184, 191, 200], and it can often be conceived as elements of a Lie group. Examples include representations of skeletal systems (for example, in robotics [181]), inferences of anatomical structures by evaluating diffusion tensor fields in medicine [18, 19, 20, 189], position- and motion-independent recognition of objects in computer vision [127, 247, 248, 249], or analysis of covariance matrices (for example, in feature-based image analysis [244]). The algorithmic tasks involved in Lie group-based computations are correspondingly diverse: They range from the calculation of geometric means [166, 167], to function approximation and regression [115, 211, 251], numerical solution of differential equations [131], numerical minimization [238], signal processing [21, 22, 47, 88], image analysis [25, 77], as well as computer vision [43, 176, 178, 196, 241, 242].

Shape analysis is another area where Lie group–valued measurements are regularly performed. The idea is to represent shapes of objects as deformations of a shared reference, as introduced by D'Arcy Thompson over 100 years ago [239]; the deformations are usually elements of a Lie group. For example, in [3, 36] configurations of the human spine are encoded in a product group consisting of translations and rotations. Classical matrix groups are also used in physically motivated shape spaces [13, 245] as well as in the characterization of volume [255] and surface [11] deformations. Important algorithmic tasks are matching, analysis, and statistics of shapes [91, 129, 231, 234, 243].

For data analysis in Lie groups, it is necessary to generalize statistical methods to the Lie group setting. Keeping in mind that Lie groups possess symmetries, as they act on themselves via translations, it is desirable to look for generalizations that respect them. The task, then, is to derive methods that are invariant/equivariant under in-group translations (from both left and right). This property is not only an established theoretical criterion for the selection of statistical methods [113] but also has practical advantages: In shape analysis, for example, it avoids a bias due to the choice of reference and differing data configurations [3, 11, 255]. In machine learning applications, equivariant convolutional networks have been shown to perform very

well on Lie groups [58, 87].

We build upon the work of Pennec, Arsigny, and Lorenzi on statistics in Lie groups with CCS connection to derive bi-invariant (that is, invariant under translation from both left and right) generalizations of the Hotelling $T^2$ statistic and Bhattacharyya distance. In contrast to the generalizations based on Riemannian structures, the proposed quantities are not only bi-invariant (see Figure 4.1) but also reduce to the multivariate formulations in the case of Euclidean spaces.

Applications of the new notions are shown in Sections 6.1 and 6.2.
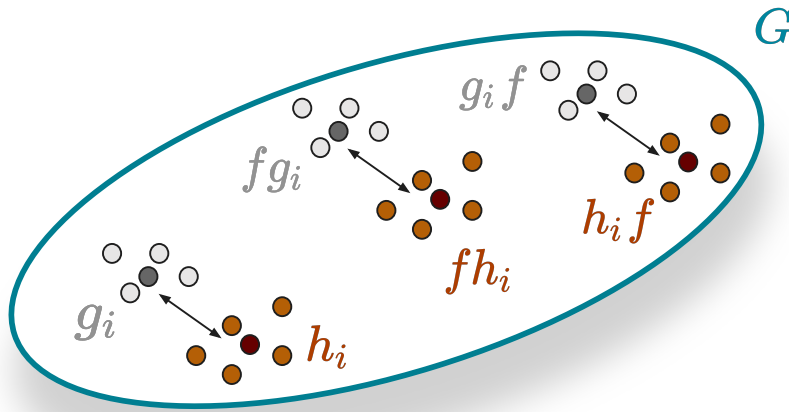


Figure 4.1: Visualization of invariance of dissimilarity measures for sample distributions under translations. The dissimilarity of two data sets $(g_1, \ldots, g_n)$ and $(h_1, \ldots, h_m)$ in a Lie group $G$ (the darker points indicating the group means) should be invariant under translations with any element $f \in G$ from left *and* right.

## 4.1 Extending Dissimilarity Measures

In this section, we derive extensions of the Hotelling $T^2$ statistic and Bhattacharyya distance for Lie group–valued data that are invariant under left and right translations. In the multivariate case, these fundamental properties seem almost trivial as, obviously,

$$t^2\big((x_i + v), (y_i + v)\big) = t^2\big((v + x_i), (v + y_i)\big) = t^2\big((x_i), (y_i)\big)$$

and

$$D_B\big((x_i + v), (y_i + v)\big) = D_B\big((v + x_i), (v + y_i)\big) = D_B\big((x_i), (y_i)\big)$$

for data sets $(x_1, \ldots, x_m)$, $(y_1, \ldots, y_n)$ and vectors $v$ in $\mathbb{R}^d$. Generalizing these properties is our goal.

### 4.1.1 Bi-invariant Hotelling $T^2$ Statistic

We start with defining a slightly modified sample covariance: We translate the data to a neighborhood of the identity and compute the covariance there. This allows for comparing the within-sample variability of data sets, even when their support lies in different parts of the group. The definition only covers the sample case but can easily be transferred to random variables.

**Definition 4.1.1** (Centralized sample covariance). *Let $G$ be a Lie group and $(g_1, \ldots, g_n) \overset{i.i.d.}{\sim} \mathcal{G}$ a data set in a CSLCG neighborhood $U \subseteq G$ with group mean $\overline{g}$. The data's left-centralized sample covariance tensor field $\mathfrak{M}_{g_i}^c \in \Gamma(T_0^2 L_{\overline{g}^{-1}}(U))$ is at each $f \in L_{\overline{g}^{-1}}(U)$ defined by*

$$\mathfrak{M}_{g_i}^c\big|_f := \frac{1}{n} \sum_{i=1}^{n} \mathrm{d}_e L_f\big(\log(\overline{g}^{-1} g_i)\big) \otimes \mathrm{d}_e L_f\big(\log(\overline{g}^{-1} g_i)\big)$$

*with representing matrix*

$$[\mathfrak{M}_{g_i}^c\big|_f] = \frac{1}{n} \sum_{i=1}^{n} \Big[\mathrm{d}_e L_f\big(\log(\overline{g}^{-1} g_i)\big)\Big] \Big[\mathrm{d}_e L_f\big(\log(\overline{g}^{-1} g_i)\big)\Big]^T.$$

*Its restriction*

$$\Sigma_{g_i}^c := \mathfrak{M}_{g_i}^c\big|_e$$

*to the identity[1] is the data's* (group) left-centralized sample covariance (tensor).

In the following, we use the left-centralized covariance tensor to compare second-order moments of data sets. This is motivated by the fact that $T_e G$ can be seen as the reference tangent space for the whole group. Since we never explicitly use the right-centralized covariance (that can be defined in an obvious way), we will drop the word "left" in the following; we will proceed similarly for notions derived from the centralized covariance. The centralized covariance will be fundamental for our constructions because it allows comparing the variability of data sets with different means in a single tangent space. This, in turn, makes it possible to transfer notions from multivariate statistics that involve covariance information from two or more data sets—even if their group means are different. In particular, we can now define the pooled covariance at the identity for two data sets with the same centralized covariance.[2]

**Definition 4.1.2** (Pooled sample covariance). *Let $(g_1, \ldots, g_n) \overset{i.i.d.}{\sim} \mathcal{G}$ and $(h_1, \ldots, h_m) \overset{i.i.d.}{\sim} \mathcal{H}$ be data sets in CSLCG neighborhoods $U_1, U_2 \subseteq G$, respectively, and let $\overline{g}$ and $\overline{h}$ be their group means. Then, their* left-pooled sample covariance *is defined by*

$$\widehat{\Sigma}_{g_i, h_i} := \frac{1}{n + m - 2}\left(n \Sigma_{g_i}^c + m \Sigma_{h_i}^c\right).$$

Because there is no danger of confusion, we only write $\widehat{\Sigma}$ from now on. The principle behind the pooled covariance is visualized in Figure 4.2.

---

[1] Note that $\mathrm{d}_e L_e$ is simply the identity on $T_e G$.

[2] More precisely, we assume that the samples are drawn independently from two distributions such that the limits of their centralized covariance matrices are equal for $n, m \to \infty$.
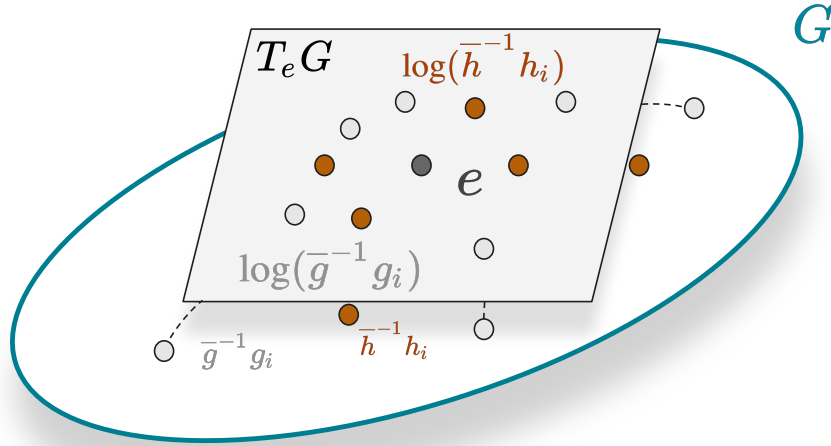
Figure 4.2: Pooling of data. To compute the pooled sample covariance of two data sets $(g_1, \ldots, g_n)$ and $(h_1, \ldots, h_m)$ both are translated via the inverses $\overline{g}^{-1}$ and $\overline{h}^{-1}$ of their means, respectively; the data then lies in a neighborhood of $e$, which is also the new mean.

We need to understand how they transform under joint data translations to obtain bi-invariant indices from the newly-defined notions. For this, we denote the centralized covariance of data that was jointly left or right translated with $f \in G$ by $f \bullet \Sigma^c_{g_i}$ and $\Sigma^c_{g_i} \bullet f$, respectively, (that is, $f \bullet \Sigma^c_{g_i} = \Sigma^c_{fg_i}$ and $\Sigma^c_{g_i} \bullet f = \Sigma^c_{g_if}$); furthermore, $\mathrm{inv} \bullet \Sigma^c_{g_i} = \Sigma^c_{g_i^{-1}}$ denotes the centralized covariance of the inverted data points. We also extend these definitions linearly to weighted sums of covariances. Then, we have the following lemma.

**Lemma 4.1.3.** *Given data $(g_1, \ldots, g_n) \overset{i.i.d.}{\sim} \mathcal{G}$ in a CSLCG neighborhood $U \subseteq G$, its centralized covariance is invariant under left translations, while under right translations with $f \in G$, it transforms according to*

$$\left[ \Sigma^c_{g_i} \bullet f \right] = \left[ \mathrm{Ad}(f^{-1}) \right] \left[ \Sigma^c_{g_i} \right] \left[ \mathrm{Ad}(f^{-1}) \right]^T.$$

*Furthermore, if the data points are inverted, then*

$$\left[ \mathrm{inv} \bullet \Sigma^c_{g_i} \right] = \left[ \mathrm{Ad}(\overline{g}) \right] \left[ \Sigma^c_{g_i} \right] \left[ \mathrm{Ad}(\overline{g}) \right]^T.$$

*Proof.* Using the equivariance of the mean gives

$$\left[ f \bullet \Sigma^c_{g_i} \right] = \frac{1}{n} \sum_{i=1}^{n} \left[ \log \left( \overline{g}^{-1} f^{-1} f g_i \right) \right] \left[ \log \left( \overline{g}^{-1} f^{-1} f g_i \right) \right]^T = \left[ \Sigma^c_{g_i} \right],$$

and applying Equation (1.5) yields

$$
\begin{aligned}
\left[\Sigma_{g_i}^c \bullet f\right] &= \frac{1}{n}\sum_{i=1}^{n}\left[\log\left(f^{-1}\overline{g}^{-1}g_i f\right)\right]\left[\log\left(f^{-1}\overline{g}^{-1}g_i f\right)\right]^T \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\mathrm{Ad}\left(f^{-1}\right)\right]\left[\log\left(\overline{g}^{-1}g_i\right)\right]\left[\log\left(\overline{g}^{-1}g_i\right)\right]^T\left[\mathrm{Ad}\left(f^{-1}\right)\right]^T \\
&= \left[\mathrm{Ad}\left(f^{-1}\right)\right]\left[\Sigma_{g_i}^c\right]\left[\mathrm{Ad}\left(f^{-1}\right)\right]^T.
\end{aligned}
$$

After inverting the data points we find, using Theorem 3.3.2, Equation (1.4), and Equation (1.5), that

$$
\begin{aligned}
\left[\mathrm{inv}\bullet\Sigma_{g_i}^c\right] &= \frac{1}{n}\sum_{i=1}^{n}\left[\log\left(\overline{g}g_i^{-1}\right)\right]\left[\log\left(\overline{g}g_i^{-1}\right)\right]^T \\
&= \frac{1}{n}\sum_{i=1}^{n}(-1)^2\left[\log\left(g_i\overline{g}^{-1}\right)\right]\left[\log\left(g_i\overline{g}^{-1}\right)\right]^T \\
&= \left[\mathrm{Ad}(\overline{g})\right]\left[\Sigma_{g_i}^c\right]\left[\mathrm{Ad}(\overline{g})\right]^T.
\end{aligned}
$$

$\square$

We can translate this result directly to the pooled covariance:

**Corollary 4.1.4.** *Let $(g_1,\ldots,g_n)\overset{i.i.d.}{\sim}\mathcal{G}$ and $(h_1,\ldots,h_m)\overset{i.i.d.}{\sim}\mathcal{H}$ be data sets in CSLCG neighborhoods $U_1,U_2\subseteq G$, respectively, and let $\overline{g}$ and $\overline{h}$ be their group means. The pooled covariance is invariant under left translations of the data and transforms under right translations with $f\in G$ according to*

$$
\left[\widehat{\Sigma}\bullet f\right] = \left[\mathrm{Ad}\left(f^{-1}\right)\right]\left[\widehat{\Sigma}\right]\left[\mathrm{Ad}\left(f^{-1}\right)\right]^T.
$$

*If the data is inverted, then we have*

$$
\left[\mathrm{inv}\bullet\widehat{\Sigma}\right] = \frac{1}{n+m-2}\left(n\left[\mathrm{Ad}(\overline{g})\right]\left[\Sigma_{g_i}\right]\left[\mathrm{Ad}(\overline{g})\right]^T + m\left[\mathrm{Ad}\left(\overline{h}\right)\right]\left[\Sigma_{h_i}\right]\left[\mathrm{Ad}\left(\overline{h}\right)\right]^T\right).
$$

*In particular, if $\overline{g}=\overline{h}$, then*

$$
\left[\mathrm{inv}\bullet\widehat{\Sigma}\right] = \left[\mathrm{Ad}(\overline{g})\right]\left[\widehat{\Sigma}\right]\left[\mathrm{Ad}(\overline{g})\right]^T.
$$

Taking advantage of Proposition 3.3.5 we also have

$$
\mu^2_{(\overline{g},\Sigma_{g_i})}(f) = \mu^2_{(e,\Sigma_{g_i}^c)}(\overline{g}^{-1}f);
$$

that is, we can always evaluate Mahalanobis distances at the identity, analogously to the Euclidean case (3.18); see Figure 4.3 for a visualization.

With this, we propose the following generalization of the Hotelling $T^2$ statistic for Lie groups, using the abbreviated notation $((g_i),(h_i))$ for $((g_1,\ldots,g_n),(h_1,\ldots,h_m))$:
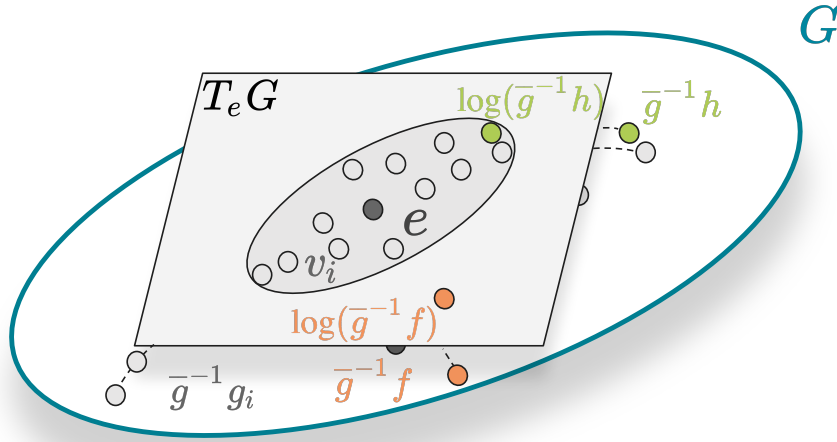
Figure 4.3: Bi-invariant Mahalanobis distance for a data set $(g_1, \dots, g_n)$ in a Lie group $G$ (dark grey) after left translation with $\overline{g}^{-1}$. The operation leaves the distance invariant. Note that right translated points $g_1\overline{g}^{-1}, \dots, g_n\overline{g}^{-1}, f\overline{g}^{-1}, h\overline{g}^{-1}$ would be at different locations in general, but the Mahalanobis distance would be the same by Proposition 3.3.5.

**Definition 4.1.5** (Bi-invariant Hotelling $T^2$ statistic)**.** *Let two data sets* $(g_1, \dots, g_n) \overset{i.i.d.}{\sim} \mathcal{G}$ *and* $(h_1, \dots, h_m) \overset{i.i.d.}{\sim} \mathcal{H}$ *in CSLCG neighborhoods* $U_1, U_2 \subseteq G$ *be given. Let further* $\overline{g}$ *and* $\overline{h}$ *be their group means,* $\widehat{\Sigma}$ *their pooled covariance and assume that* $\log(\overline{g}^{-1}\overline{h})$ *exists. The* left *bi-invariant Hotelling* $T^2$ *statistic of both data sets is then defined by*

$$t^2\big((g_i), (h_i)\big) := \frac{nm}{n+m} \mu^2_{(e, \widehat{\Sigma})}\left(\overline{g}^{-1}\overline{h}\right).$$

A visualization of the bi-invariant Hotelling $T^2$ statistic is shown in Figure 4.4. The definition is independent of the chosen basis of $T_eG$ (because this is true for the Mahalanobis distance). Furthermore, it reduces to (3.20) when $G = \mathbb{R}^n$. This is a consequence of the common reference tangent space; as the latter is missing in Riemannian manifolds, reduction to the multivariate case is problematic using the Riemannian approach. Indeed, as was shown in Section 3.2.4, the only generalization of Hotelling's $T^2$ statistic to Riemannian manifolds known to the authors does not have this property.

While we can apply the bi-invariant Hotelling $T^2$ statistic to any two data sets that fulfill the requirements in its definition, it should be used when we expect that both have the same covariance structure. (Again, that is, when we suppose that the limits of the centralized covariance matrices are equal for $n, m \to \infty$.) The reasons are that only differences between the means are detected (the statistic is zero if and only if both are equal), and there is little reason to pool the covariances otherwise. If we cannot expect homoscedasticity, the bi-invariant Bhattacharyya distance (see Section 4.1.2) should be used instead.
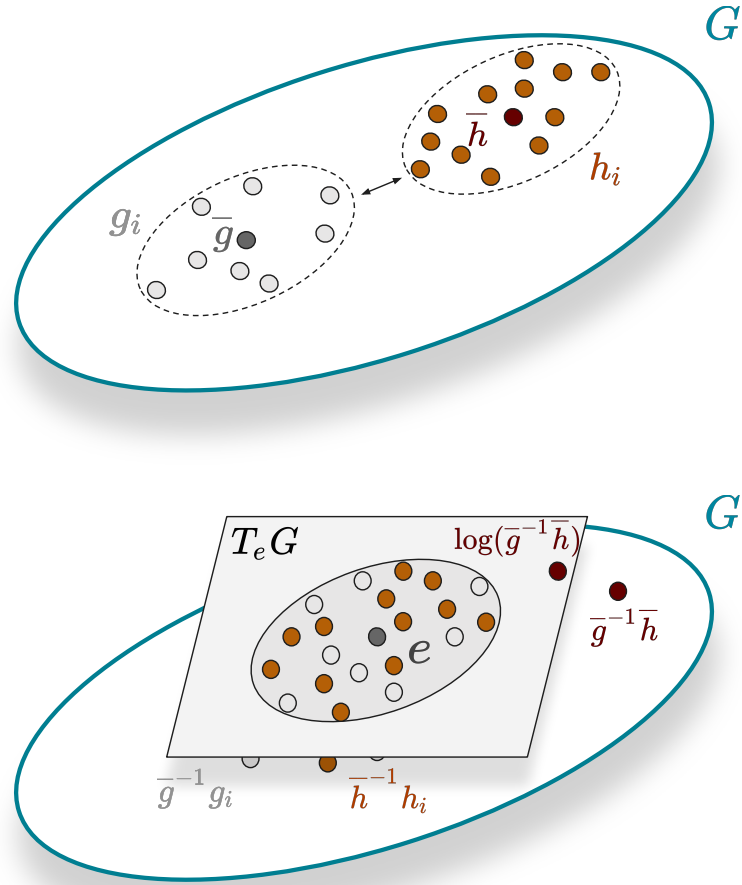
Figure 4.4: Top: Two data sets $(g_1, \ldots, g_n)$ and $(h_1, \ldots, h_m)$ with the same covariance structure (indicated by the broken ellipses). Bottom: After left translating the sets to a neighborhood of $e$ with the inverses of their means, the bi-invariant Hotelling $T^2$ statistic is the (scaled) squared Mahalanobis distance of $\log(\overline{g}^{-1}\overline{h})$ under the pooled covariance.

The following theorem shows that the newly defined notion has the invariance properties we aimed for and does not depend on the order of the data sets.

**Theorem 4.1.6** (Properties of the bi-invariant Hotelling $T^2$ statistic). *Let* $(g_1, \ldots, g_n) \overset{i.i.d.}{\sim}$ *$\mathcal{G}$ and* $(h_1, \ldots, h_m) \overset{i.i.d.}{\sim} \mathcal{H}$ *be data sets in CSLCG neighborhoods* $U_1, U_2 \subseteq G$, *respectively. Let further* $\overline{g}$ *and* $\overline{h}$ *be their group means and assume that* $\log(\overline{g}^{-1}\overline{h})$ *exists. The bi-invariant Hotelling $T^2$ statistic has the following properties:*
*(i)   it is symmetric,*
*(ii)  it is invariant under left and right translations,*
*(iii) if* $\overline{g} = \overline{h}$, *then it is invariant under inversion.*

*Proof.* Note first that the pooled covariance is independent of the order of the data

groups. Thus, using Equation (1.4), property (i) follows from

$$t^2\big((g_i),(h_i)\big) = \frac{nm}{n+m}(-1)^2\Big[\log\big(\overline{h}^{-1}\overline{g}\big)\Big]^T\Big[\widehat{\Sigma}\Big]^{-1}\Big[\log\big(\overline{h}^{-1}\overline{g}\big)\Big] = t^2\big((h_i),(g_i)\big).$$

Next, we show (ii). Invariance under left translations follows directly since $f$ and $f^{-1}$ cancel out as in the proof of Lemma 4.1.3. Let $f \in G$. Then, using Theorem 3.3.2, Equation (1.5), and Corollary 4.1.4, we find

$$\begin{aligned}
t^2\big((g_if),(h_if)\big) &= \frac{nm}{n+m}\Big[\log\big(f^{-1}\overline{g}^{-1}\overline{h}f\big)\Big]^T\Big[\widehat{\Sigma}\bullet f\Big]^{-1}\Big[\log\big(f^{-1}\overline{g}^{-1}\overline{h}f\big)\Big] \\
&= \frac{nm}{n+m}\Big[\log\big(\overline{g}^{-1}\overline{h}\big)\Big]^T\Big[\mathrm{Ad}(f^{-1})\Big]^T\Big[\mathrm{Ad}(f^{-1})\Big]^{-T}\Big[\widehat{\Sigma}\Big]^{-1} \\
&\qquad\qquad \Big[\mathrm{Ad}\big(f^{-1}\big)\Big]^{-1}\Big[\mathrm{Ad}\big(f^{-1}\big)\Big]\Big[\log\big(\overline{g}^{-1}\overline{h}\big)\Big] \\
&= t^2\big((g_i),(h_i)\big).
\end{aligned}$$

Finally, (iii) follows immediately from $\log(e) = 0$. $\qquad\qquad\square$

Note that although the statistic is not invariant under inversion in general, the fact that equality of means is invariantly detected makes it also interesting (for example, to perform hypothesis tests for equality of means) when invariance under inversion is of interest. (Nevertheless, the authors are not aware of any such application.)

**Remark 4.1.7.** *Observe that if we replace* all *left translations with right translations in* all *definitions of this section, the resulting* right bi-invariant Hotelling $T^2$ statistic *will be different from the left one in general but will have the same invariance properties as the statistic from Definition 4.1.5. We can use the average of the left- and right- variant (which is also bi-invariant) as a* central bi-invariant $T^2$ statistic.

The procedure to compute the bi-invariant Hotelling $T^2$ statistic is given in Algorithm 3. Therein, we assume that a way to compute the group exponential and logarithm is known, which is the case for many Lie groups (for example, for those from Examples 1.1 and 1.2).

---

**Algorithm 3** (Left) bi-invariant Hotelling $T^2$ statistic

---

**Input:** data sets $(g_1, \ldots, g_n)$ and $(h_1, \ldots, h_m)$ in a $d$-dimensional Lie group $G$ as in Definition 4.1.5 with group means $\overline{g}$ and $\overline{h}$, vector space isomorphism $E : T_e G \to \mathbb{R}^d$ ("coordinate map")

**Output:** (left) bi-invariant Hotelling $T^2$ statistic $t$

**for** $i = 1, \ldots, n$ **do**
  $x_i \leftarrow E\left(\log(\overline{g}^{-1} g_i)\right)$
**end for**
**for** $i = 1, \ldots, m$ **do**
  $y_i \leftarrow E\left(\log(\overline{h}^{-1} h_i)\right)$
**end for**
$z \leftarrow E\left(\log(\overline{g}^{-1}\overline{h})\right)$
$S \leftarrow \dfrac{1}{n+m-2}\left(\displaystyle\sum_{i=1}^{n} x_i x_i^T + \sum_{i=1}^{m} y_i y_i^T\right)$
$t \leftarrow \dfrac{nm}{(n+m)} z^T S z$

---

### 4.1.2 Bi-invariant Bhattacharyya Distance

Since the first summand of the multivariate Bhattacharyya distance (3.23) coincides with a Hotelling $T^2$ statistic, there is hope that we can generalize the former to Lie groups in a bi-invariant way using the bi-invariant Hotelling $T^2$ statistic; and indeed, it is possible. We begin by generalizing the averaged sample covariance:

**Definition 4.1.8** (Averaged sample covariance). *Let $(g_1, \ldots, g_n) \overset{i.i.d.}{\sim} \mathcal{G}$ and $(h_1, \ldots, h_m) \overset{i.i.d.}{\sim} \mathcal{H}$ be data sets in CSLCG neighborhoods $U_1, U_2 \subseteq G$, respectively, and let $\overline{g}$ and $\overline{h}$ be their group means. Then, their* left-averaged sample covariance *is defined by*

$$\overline{\Sigma}_{g_i, h_i} = \frac{1}{2}(\Sigma_{g_i}^c + \Sigma_{h_i}^c).$$

Again, in the following, we drop the subscripts. Since it only differs in weighting, the averaged covariance has the same properties as the pooled covariance.

**Corollary 4.1.9.** *The averaged covariance is invariant under left translations of the data and transforms under right translations with $f \in G$ according to*

$$\left[\overline{\Sigma} \bullet f\right] = \left[\mathrm{Ad}\left(f^{-1}\right)\right]\left[\overline{\Sigma}\right]\left[\mathrm{Ad}\left(f^{-1}\right)\right]^T.$$

*If the data is inverted, then we have*

$$\left[\mathrm{inv} \bullet \overline{\Sigma}\right] = \frac{1}{2}\left(\left[\mathrm{Ad}(\overline{g})\right]\left[\Sigma_{g_i}^c\right]\left[\mathrm{Ad}(\overline{g})\right] + \left[\mathrm{Ad}\left(\overline{h}\right)\right]\left[\Sigma_{h_i}^c\right]\left[\mathrm{Ad}\left(\overline{h}\right)\right]\right).$$

*In particular, if $\overline{g} = \overline{h}$, then*

$$\left[\mathrm{inv} \bullet \overline{\Sigma}\right] = \left[\mathrm{Ad}(\overline{g})\right]\left[\overline{\Sigma}\right]\left[\mathrm{Ad}(\overline{g})\right]^T.$$
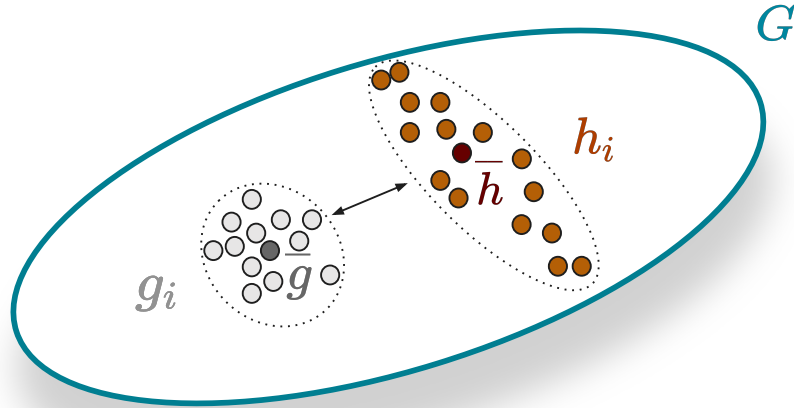
Figure 4.5: Two data sets $(g_1, \ldots, g_n)$ and $(h_1, \ldots, h_m)$ with different covariance structures. The bi-invariant Bhattacharyya also captures the latter dissimilarity in addition to the difference between the means $\overline{g}$ and $\overline{h}$.

With this, we can generalize the Bhattacharyya distance to Lie groups:

**Definition 4.1.10** (Bi-invariant Bhattacharyya distance). *Let* $(g_1, \ldots, g_n) \overset{i.i.d.}{\sim} \mathcal{G}$ *and* $(h_1, \ldots, h_m) \overset{i.i.d.}{\sim} \mathcal{H}$ *be data sets in CSLCG neighborhoods* $U_1, U_2 \subseteq G$, *respectively. Let further* $\overline{g}$ *and* $\overline{h}$ *be their group means,* $\overline{\Sigma}$ *their averaged covariance and assume that* $\log(\overline{g}^{-1}\overline{h})$ *exists. The* left bi-invariant Bhattacharyya distance *between both data sets is then defined by*

$$D_B\big((g_i), (h_i)\big) := \frac{1}{8}\mu_{(e,\overline{\Sigma})}^2\left(\overline{g}^{-1}\overline{h}\right) + \frac{1}{2}\ln\left(\frac{\det\left(\left[\overline{\Sigma}\right]\right)}{\sqrt{\det\left(\left[\Sigma_{g_i}^c\right]\right)\det\left(\left[\Sigma_{h_i}^c\right]\right)}}\right).$$

In the Euclidean case, Definition 4.1.10 reduces to (3.23). The bi-invariant Bhattacharyya can also detect differences in the covariance structure of the sample sets; this scenario is visualized in Figure 4.5. Furthermore, the following theorem shows that the attribute "bi-invariant" is justified.

**Theorem 4.1.11** (Properties of the bi-invariant Bhattacharyya distance). *Let* $(g_1, \ldots, g_n) \overset{i.i.d.}{\sim} \mathcal{G}$ *and* $(h_1, \ldots, h_m) \overset{i.i.d.}{\sim} \mathcal{H}$ *be data sets in CSLCG neighborhoods* $U_1, U_2 \subseteq G$, *respectively. Let further* $\overline{g}$ *and* $\overline{h}$ *be their group means and assume that* $\log(\overline{g}^{-1}\overline{h})$ *exists. The bi-invariant Bhattacharyya distance has the following properties:*
*(i)   it is symmetric,*
*(ii)  it does not depend on the chosen basis,*
*(iii) it is invariant under left and right translations,*
*(iv)  if* $\overline{g} = \overline{h}$, *then it is invariant under inversion.*

*Proof.* Note that we can focus on the second summand since for the first all results can be shown as in the proof of Theorem 4.1.6 (replacing the pooled by the averaged

covariance). Thus, (i) follows immediately, since $\overline{\Sigma}$ is invariant under an exchange of the data sets.

Next, we show (iii). Let $f \in G$ and $a_f = \det(\mathrm{Ad}(f^{-1}))$; note that $a_f \neq 0$. Invariance of the second summand under left translations follows from Lemma 4.1.3 and Corollary 4.1.9. They further imply that

$$\frac{\det\left(\left[\overline{\Sigma} \bullet f\right]\right)}{\sqrt{\det\left(\left[\Sigma_{g_i}^c \bullet f\right]\right)\det\left(\left[\Sigma_{h_i}^c \bullet f\right]\right)}} = \frac{a_f^2 \det\left(\left[\overline{\Sigma}\right]\right)}{a_f^2\sqrt{\det\left(\left[\Sigma_{g_i}^c\right]\right)\det\left(\left[\Sigma_{h_i}^c\right]\right)}},$$

which yields invariance under right translations. The proofs of (ii) and (iv) work analogously (using Corollary 4.1.9 for (iv) and replacing group adjoints with basis transformations for (ii)). $\qquad\square$

Note that in Definition 4.1.10 we take determinants of matrix representations of (2,0)-tensors instead of (0,2)-tensors like in the Euclidean case (for background information see, for example, [45, Ch. 4]). This does not make a difference since lowering indices with any (auxiliary) metric does not influence the *ratio* of the determinants in Definition 4.1.10; the additional terms (that is, the matrix representation of the metric) cancel out analogously to those introduced by translations (that is, matrix representations of group adjoints) in the proofs of Theorems 4.1.6 and 4.1.11.

**Remark 4.1.12.** *Again, if we replace* all *left translations with right translations in* all *definitions (also in those of the bi-invariant Hotelling $T^2$ statistic), we obtain the* right bi-invariant Bhattacharyya distance *with the same invariance properties. We can also average the left- and right- variants to obtain the* central bi-invariant Bhattacharyya distance*.*

**Remark 4.1.13.** *A notion from multivariate statistics that is intimately related to the Bhattacharyya distance is the Hellinger distance [122] (see also [180, p. 51])*

$$H\big((x_i),(y_i)\big) = \sqrt{1 - 1/\exp\big(\mathrm{D}_B\big((x_i),(y_i)\big)\big)} \tag{4.1}$$

*for data sets $(x_1,\dots,x_m)$ and $(y_1,\dots,y_n)$ in $\mathbb{R}^d$. It has found various applications, for example, in decision trees [57], internet telephony [218], and data visualization [202].*

*By using the bi-invariant Bhattacharyya distance in Equation* (4.1)*, we directly obtain a* bi-invariant Hellinger distance *on Lie Groups with the same invariance properties.*

Algorithm 4 summarizes how the bi-invariant Bhattacharyya distance is computed. (We again assume that the group exponential and logarithm can be calculated.)

---

**Algorithm 4** (Left) bi-invariant Bhattacharyya distance

---

**Input:** data sets $(g_1, \ldots, g_n)$ and $(h_1, \ldots, h_m)$ in a $d$-dimensional Lie group $G$ as in Definition 4.1.10 with group means $\overline{g}$ and $\overline{h}$, vector space isomorphism $E : T_e G \to \mathbb{R}^d$ ("coordinate map")

  **Output:** (left) bi-invariant Bhattacharyya distance $d$

  **for** $i = 1, \ldots, n$ **do**
    $x_i \leftarrow E\left(\log(\overline{g}^{-1} g_i)\right)$
  **end for**
  **for** $i = 1, \ldots, m$ **do**
    $y_i \leftarrow E\left(\log(\overline{h}^{-1} h_i)\right)$
  **end for**
  $z \leftarrow E\left(\log(\overline{g}^{-1}\overline{h})\right)$
  $R \leftarrow \dfrac{1}{n} \sum\limits_{i=1}^{n} x_i x_i^T$
  $T \leftarrow \dfrac{1}{m} \sum\limits_{i=1}^{m} y_i y_i^T$
  $S \leftarrow \dfrac{1}{2}(R + T)$
  $d \leftarrow \dfrac{1}{8} z^T S z + \dfrac{1}{2} \ln\left(\dfrac{\det(S)}{\sqrt{\det(R)\det(T)}}\right)$

---

## 4.2 Bi-invariant Bhattacharyya Distance and Densities

In the multivariate setting, the Bhattacharyya distance has the more general integral definition (3.21) for distributions with probability density function (pdf). We show in this section that the bi-invariant Bhattacharyya distance is compatible with this view. Like in Euclidean space, we obtain compatibility for Gaussian-like distributions from the generalized integral definition. On the other hand, in contrast to our definition, the integral version has the drawback that it is not bi-invariant in general.

Let $d$ be the dimension of the Lie group $G$. Integrals of functions are defined using a left Haar measure, as summarized in Appendix A.1. For pdfs $\theta, \nu$ on $G$ we define the integral version of the Bhattacharyya distance on Lie groups

$$D_B^{\text{int}}(\theta, \nu) := -\ln\left(\int_G \sqrt{\theta(g)\nu(g)}\, dg\right) \tag{4.2}$$

Clearly, (4.2) reduces to (3.21) for $G = \mathbb{R}^d$ when we use the usual volume form. Since

$$D_B^{\text{int}}(\theta \circ L_f, \nu \circ L_f) = D_B^{\text{int}}(\theta, \nu)$$

for each $f \in G$, the integral version of the Bhattacharyya distance is left invariant. However, because $dg$ is usually not right invariant this does not hold for right translations. Thus, in contrast to Definition 4.1.10, the integral version (4.2) cannot generally be bi-invariant. (A similar problem arises when we start with a right invariant Haar measure.) For this to hold, we need a *unimodular* Lie group, that is, a Lie group with

bi-invariant volume form. (A well-known example of a non-unimodular Lie group is the group of invertible affine transformations of $\mathbb{R}^n$.) Therefore, we will only be interested in the unimodular case. Since, then, distributions can be "moved around via translations" without changing integrals, we can consider w.l.o.g. only those centered at $e$.

In the following, let $E : \mathbb{R}^d \to T_eG$ be any orientation-preserving vector space isomorphism (which we obtain from choosing a basis of $T_eG$). Further, let $\mathfrak{e} \subseteq T_eG$ be a neighborhood of $0 \in T_eG$ such that, *first*, $\exp|_{\mathfrak{e}}$ is an diffeomorphism onto the neighborhood $W = \exp(\mathfrak{e})$ of $e$ and, *second*, there is an orientation-preserving diffeomorphism $F : T_eG \to \mathfrak{e}$ between the whole of $T_eG$ and $\mathfrak{e}$. Lemma A.2.5 shows that we can always choose $\mathfrak{e} := \log(W)$ with $W$ being any CSLCG neighborhood of $e$. Now, we can define $\widetilde{\exp} = \exp \circ F \circ E$ and $\widetilde{\log} = E^{-1} \circ F^{-1} \circ \log$. Note that the latter is a chart of $W$ with image $\mathbb{R}^d$. To identify densities on $G$ that connect both versions of the Bhattacharyya distance we first need the following lemma.

**Lemma 4.2.1.** *Let $G$ be a d-dimensional Lie group and $g \in G$. Further, let $\rho, \sigma$ be pdfs on $\mathbb{R}^d$ and $\widetilde{\exp}^*(dg|_W) = \varphi \, dx$ with $0 < \varphi \in C^\infty(\mathbb{R}^d)$. Define*

$$\theta(g) = \begin{cases} (\rho/\varphi \circ \widetilde{\log})(g), & g \in W, \\ 0, & g \in G \setminus W, \end{cases} \tag{4.3}$$

*and*

$$\nu(g) = \begin{cases} (\sigma/\varphi \circ \widetilde{\log})(g), & g \in W, \\ 0, & g \in G \setminus W. \end{cases} \tag{4.4}$$

*Then, $\theta$ and $\nu$ are pdfs on $G$ with*

$$\int_G \sqrt{\theta(g)\nu(g)} \, dg = \int_{\mathbb{R}^d} \sqrt{\rho(x)\sigma(x)} \, dx.$$

*Proof.* Clearly, $\theta, \nu \geq 0$. Furthermore, since $\theta$ and $\nu$ are only supported in $W$ and $\widetilde{\log} \circ \widetilde{\exp} = \mathrm{Id}_{\mathbb{R}^d}$, we find $\int_G \theta(g) \, dg = \int_G \nu(g) \, dg = 1$ and

$$\int_G \sqrt{\theta(g)\nu(g)} \, dg = \int_W \sqrt{\theta(g)\nu(g)} \, dg$$

$$= \int_{\mathbb{R}^d} \widetilde{\exp}^* \left( \sqrt{\theta(g)\nu(g)} \, dg \right)$$

$$= \int_{\mathbb{R}^d} (\sqrt{\theta\nu} \circ \widetilde{\exp})(x)\varphi(x) \, dx$$

$$= \int_{\mathbb{R}^d} \left( \frac{\sqrt{\rho\sigma}}{\varphi} \circ \mathrm{Id}_{\mathbb{R}^d} \right)(x)\varphi(x) \, dx$$

$$= \int_{\mathbb{R}^d} \sqrt{\rho(x)\sigma(x)} \, dx.$$

$\square$

We can write the function $\varphi$ also more explicitly. In the domain of a chart, any form of maximal degree $d$ is given as the wedge product of $d$ coordinate 1-forms multiplied by a smooth function (see Appendix A.1). Taking the canonical coordinates

$(x_1, \ldots, x_d)$ on $\mathbb{R}^d$, there these 1-forms are $dx^1, \ldots, dx^d$. (We also have $dx = dx^1 \wedge \cdots \wedge dx^d$.) Setting $dy^i := \widetilde{\log}^*(dx^i)$, $i = 1, \ldots, d$, there is $0 < \psi \in C^\infty(W)$ such that $dg|_W = \psi \, dy^1 \wedge \cdots \wedge dy^d$. Hence, the transformation rule for forms of maximal degree [8, Ch. XII Thm. 2.3] implies

$$\widetilde{\exp}^*(dg|_W) = \det(d\,\widetilde{\exp})(\psi \circ \widetilde{\exp})\, dx^1 \wedge \cdots \wedge dx^d \tag{4.5}$$

and, thus, $\varphi = \det(d\,\widetilde{\exp})(\psi \circ \widetilde{\exp})$.

In the following, $\mathcal{N}(x, [\Sigma])$ denotes the multivariate normal distribution on $\mathbb{R}^d$ with mean $x$ and covariance matrix $[\Sigma]$. The following proposition establishes a connection between the bi-invariant Bhattacharyya distance and the integral version.

**Proposition 4.2.2.** *Let G be a d-dimensional unimodular Lie group with CCS connection, and let $\rho$ be the pdf of $\mathcal{N}(0, [\Sigma^c_{g_i}])$ and $\sigma$ the pdf of $\mathcal{N}([\log(\overline{g}^{-1}\overline{h})], [\Sigma^c_{h_i}])$ on $\mathbb{R}^d$. From $\rho$ and $\sigma$, define pdfs $\theta$ and $\nu$ on G by (4.3) and (4.4), respectively. Further, let $(g_1, \ldots, g_n)$ and $(h_1, \ldots, h_m)$ be data sets in a CSLCG neighborhood in $U \subseteq W$ with group means $\overline{g}$ and $\overline{h}$, respectively. Then,*

$$\mathrm{D}_B^{\mathrm{int}}(\theta, \nu) = \mathrm{D}_B\big((g_i), (h_i)\big).$$

*Proof.* The assertion follows from Lemma 4.2.1 and the fact that the Bhattacharyya distance between two multivariate normal distributions is of the form (3.23). □

According to the proposition, "scaled push-forwards" of the multivariate normal distribution provide a connection between both versions of the Bhattacharyya distance. Densities connected to these were investigated for SE$(n)$ in [52]—also with an emphasis on bi-invariance—and for affine locally symmetric spaces in [53]. Note that the requirement $U \subseteq W$ is not needed in the proof but is introduced to ensure that the data samples have non-vanishing probabilities under $\rho$ and $\sigma$. If one wants to choose $\rho$ and $\sigma$ (e.g, to simulate an underlying model) setting $\mathfrak{e} := \log(U)$ seems sensible because thereby the locality condition that we impose on the data for $\mathrm{D}_B$ is met. On the other hand, larger domains could be used if "outliers" should be possible.

Interesting special cases are Lie groups whose exponentials are *global* diffeomorphisms. Examples are simply connected, nilpotent Lie groups (like the Heisenberg group, which allows for a unique mean in the entire group [189, pp. 198–199]); they are unimodular and the pullback of the Lebesgue measure under the logarithm is a bi-invariant Haar measure [56]. Thus, in a simply-connected, nilpotent Lie group, $dg$ can be chosen such that $\psi \equiv 1$ in Equation (4.5).

# Chapter 5

# Regression with Bézier Splines in Riemannian Manifolds

Regression methods are central to modern statistics. They are used when the dependence of a measured variable on one or more explanatory variables is of interest. Usual multivariate regression methods are not applicable when the observed variable takes values in general manifolds; therefore, there is widespread research for regression in nonlinear spaces—a process fuelled in particular by the need to analyze nonlinear data (for example, images and shapes) from an ever-growing number of large-scale longitudinal studies in medical applications [99]. Consequently, non-parametric [69, 161, 266], semi-parametric [223, 270], and parametric [64, 125, 146] regression models have been studied using Riemannian tools; some of these are applicable to data from general manifolds [69, 125, 146, 161, 223] while others are formulated for specific ones [64, 266, 270].

Parametric models have the advantages that they are usually faster to compute and that their results are easier to interpret than those of their non-parametric counterparts. They should therefore be considered whenever we are confident that the relationship between the measured and explanatory variables is of a known type. The simplest parametric regression model in Euclidean space is linear regression, and geodesic regression [89, 174] was introduced as its generalization to Riemannian manifolds. Geodesic regression allows approximating given instances in a Riemannian manifold by a generalized straight line. However, some processes cannot be accurately described by a geodesic, for example, periodic motion or processes with saturation that slow down after some time.

Riemannian polynomials [125] have been considered for higher-order parametric regression. In these works, they are defined (by employing variational principles) as solutions to differential equations involving curvature terms. Since closed-form solutions of the equations are usually not available, evaluation and optimization are complicated and numerically expensive.

We propose to use manifold-valued Bézier splines as an alternative. They allow for explicit formulas, which can be utilized to improve computational speed without sacrificing accuracy. The fact that they can easily be concatenated additionally stands out in comparison to Riemannian polynomials: While variational spline models allow for piecewise composition, there is no clear way to define them for even degrees [125].

Therefore, introducing flexible, intrinsic splines for regression is a core contribution of this work. Our ansatz leads to closed-form, numerically stable, and efficient expressions for the gradient of the regression objective in symmetric spaces.

In the second part of the chapter, it is shown how the new regression scheme can be extended into a hierarchical model that allows for an adequate, highly flexible analysis of longitudinal manifold-valued data; and how we can use it to control for confounding variables in given data sets that introduce unwanted variance.

## 5.1 Regression with Bézier Splines

Next comes the novel regression model. While we introduce the theory in this section, two applications to shape analysis in medical imaging are shown in Sections 6.3 and 6.4.

### 5.1.1 The Model

For *k*-th order polynomial regression in Euclidean space $\mathbb{R}^d$ one assumes that the relationship between an $\mathbb{R}^d$-valued random variable $Y$ and a scalar variable $t$ is given by the model

$$Y(t) = P(t) + \epsilon(t), \tag{5.1}$$

with $P$ being a (deterministic) polynomial of degree at most $k$ and $\epsilon$ a vector-valued, random variable representing the error. In the following, we generalize this model to Riemannian manifolds using Bézier splines; furthermore, we extend it conceptually by including closed curves.

Let $M$ be a Riemannian manifold and $U \subseteq M$ be a normal convex neighborhood. Further, let $n$ data points $(t_1, q_1), \ldots, (t_n, q_n) \in [0,1] \times U$ with corresponding scalar parameters (for example, points in time) be given. We suppose that there is a Bézier spline $B \in \mathcal{B}(U)$ such that the data points $q_1, \ldots, q_n$ are independent realizations of an $M$-valued random variable $\mathcal{Q}$ that is connected to a deterministic variable $t \in \mathbb{R}$ according to the model

$$\mathcal{Q}(t) = \exp_{B(t;p_0,\ldots,p_K)}\big(\epsilon(t)\big). \tag{5.2}$$

Here, for each $t \in [0,1]$ the random error variable $\epsilon$ takes values in $T_{B(t)}M$ whose realizations are (generally non-continuous) random vector fields along $B$.[1] The control points $p_0, \ldots, p_K$ are the unknown parameters. A visualization of the model for cyclic data on the sphere is shown in Figure 2.3.

As intended, in Euclidean space (5.2) reduces to (5.1) when restricting to Bézier curves. Furthermore, our model is a generalization of geodesic regression [89, Eqn. 3], which it reduces to when $B$ consists of a single segment with two control points.

In the following, we assume that the type of $B$ is known. In practice, this must often be determined first. As for polynomial regression, knowledge-driven (first) guesses of the degrees should be tested, for example, by holding back a testing set from the data to validate the model.

---

[1]The assumption that the explanatory variable $t$ is in $[0,1]$ is purely for reasons of consistency. For a finite set of samples, the realizations of $t$ must be in an interval of finite length, which can always be linearly re-scaled to $[0,1]$. We do this in all later applications of our method.
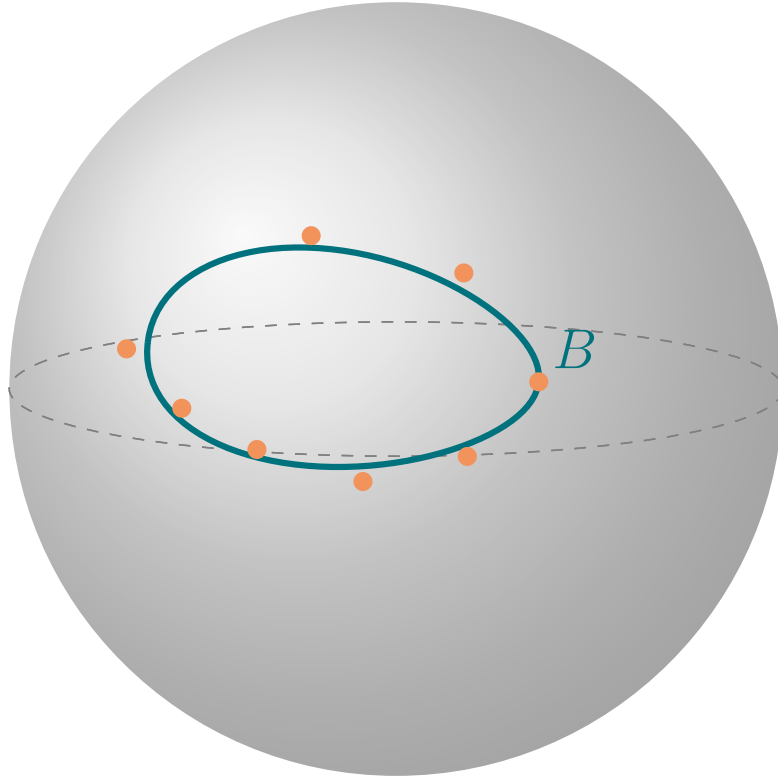
Figure 5.1: Sketch of regression with closed Bézier splines in $\mathcal{S}^2$. The orange dots are data points, while $B$ indicates the regressed spline.

### 5.1.2 Least Squares Estimation

We now discuss how the Bézier spline $B$ in (5.2) should be estimated. Since (noise) distributions on manifolds that allow for tractable computations are rare, using maximum likelihood estimation or empirical Bayes is often impossible. Therefore, geometric least squares estimation is commonly employed. Nevertheless, we will show in the next section that the former coincides with maximum likelihood estimation for symmetric spaces.

To motivate our method, we again look at the multivariate case first. A standard way to estimate the polynomial $P$ in (5.1) is *ordinary least squares*. Let $N$ measurements $(t_1, y_1), \ldots, (t_n, y_n) \in [0,1] \times \mathbb{R}^d$ be given. Denoting the set of polynomial curves over $[0,1]$ of degree at most $k$ by $\mathcal{P}_k$, we can approximate $P$ by the *least squares estimator*

$$\widehat{P} := \arg\min_{P \in \mathcal{P}_k} \frac{1}{2} \sum_{i=1}^{n} \|P(t_i) - y_i\|^2.$$

We can directly transfer this idea to Riemannian manifolds. Let $\mathcal{C} \subset U^{K+1}$ be the largest (that is, containing all others with the same property) *compact*, simply-connected subset such that each point in $\mathcal{C}$ defines an element of $\mathcal{B}(U)$. Then, we can define the least squares estimator of (5.2) as follows.

**Definition 5.1.1.** *Let $n$ realizations $(t_1, q_1), \ldots, (t_n, q_n) \in [0,1] \times U$ of (5.2) be given. Then,*

*the* sum-of-squared error *of a Bézier spline $B \in \mathcal{B}(U)$ is defined by*

$$\mathcal{E}(p_0, \ldots, p_K) := \frac{1}{2} \sum_{i=1}^{n} \mathrm{dist}\Big(B(t_i; p_0, \ldots, p_K), q_i\Big)^2. \tag{5.3}$$

*Furthermore, the* least squares estimator *$\widehat{B} \in \mathcal{B}(U)$ of the spline $B$ in (5.2) is determined by its control points*

$$(\widehat{p}_0, \ldots, \widehat{p}_K) \in \mathcal{C}$$

*according to*

$$(\widehat{p}_0, \ldots, \widehat{p}_K) := \underset{(p_0, \ldots, p_K) \in \mathcal{C}}{\arg\min} \ \mathcal{E}(p_0, \ldots, p_K). \tag{5.4}$$

Note that, for two control points, this definition also boils down to Fletcher's definition [89, Eqn. 5] so that in the geodesic case, the least squares estimates from geodesic regression and our method coincide. We also remark that, generally, none of the control points coincides with a data point, as with spline *interpolation* [29, 105].

Before going on, a comment on our use of $\mathcal{C}$ seems in order. Like Fletcher in [89, Thms. 2 and 3], we introduce it to ensure the existence of a least squares estimator. The existence question in an open set is very involved and, as far as we can see, depends strongly on the "layout" of the data, the geometry of the space, and the degree of the underlying curve. For example, if we want to use a polynomial of high degree for sparse data with similar values of $t$ but highly varying realizations of $\mathcal{Q}$, this will often lead to strongly oscillating least squares estimators (a well-known effect for polynomials of a higher degree in Euclidean space), the "extrema" eventually "trying" to leave $U$. We leave this question for future work and use a compact set instead.

In general, the minimizer (5.4) is not known analytically, which makes iterative schemes necessary. Therefore, we apply Riemannian gradient descent. (For optimization on manifolds see [2].) The gradient of $\mathcal{E}$ can be computed for each control point individually.[2] We write $B_t(p_j)$ for the map $p_j \mapsto B_t(p_j) := B(t; p_0, \ldots, p_j, \ldots, p_K)$ and define the functions $p \mapsto \tau_i(p) := \mathrm{dist}(p, q_i)^2$. Equation (3.25) yields $\mathrm{grad}_p \tau_i = -2 \mathrm{Log}_p(q_i)$ for each $p \in U$. When we consider the $j$-th summand on the right-hand side of (5.3), the chain rule (1.13) implies that its gradient with respect to the $i$-th control point is given by

$$\mathrm{grad}_{p_j}(\tau_i \circ B_{t_i}) = \mathrm{d}_{p_j} B_{t_i}^* \Big(\mathrm{grad}_{B_{t_i}(p_j)} \tau_i\Big) = -2 \, \mathrm{d}_{p_j} B_{t_i}^* \Big(\mathrm{Log}_{B_{t_i}(p_j)}(q_i)\Big). \tag{5.5}$$

Using Equation (5.3) we thus obtain

$$\mathrm{grad}_{p_j} \mathcal{E} = - \sum_{i=1}^{n} \mathrm{d}_{p_j} B_{t_i}^* \Big(\mathrm{Log}_{B_{t_i}(p_j)}(q_i)\Big) \tag{5.6}$$

for each $j = 0, \ldots, K$.

In symmetric spaces, we can thus compute $\mathrm{grad}_{p_j} \mathcal{E}$ explicitly as discussed in Section 2.1. A visualization of the gradient of a cubic Bézier curve in $\mathcal{S}^2$ is shown in Figure 5.2.

---

[2]In fact, here we assume that the product manifold $M^{K+1}$ is endowed with the so-called product structure (see, for example, [45])—a very natural choice since we have no reason to assume that there are dependencies between the control points.
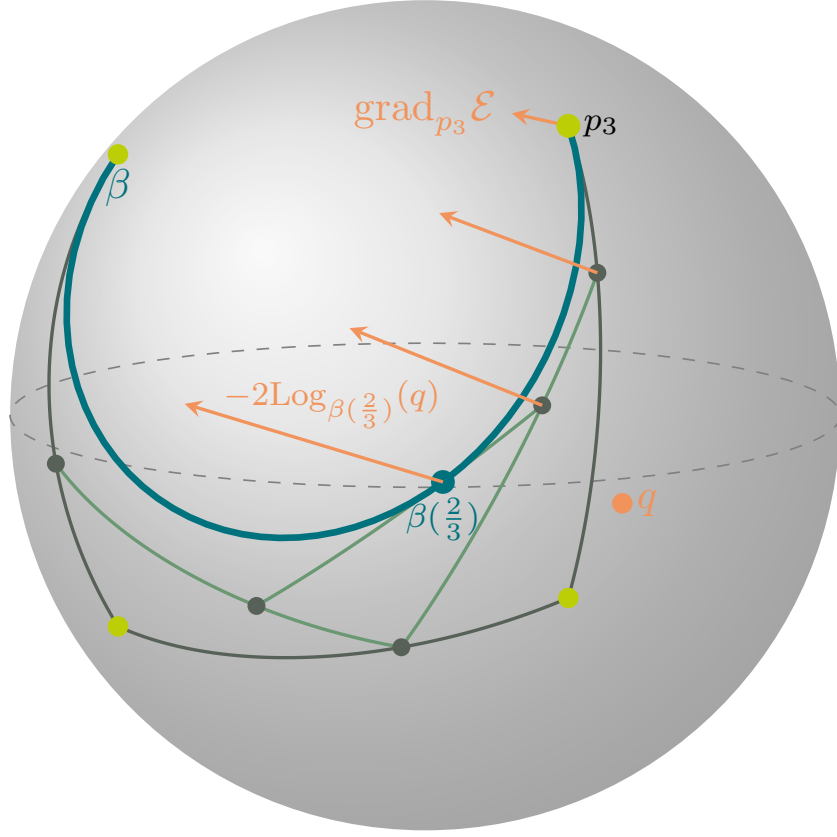
Figure 5.2: Gradient w.r.t. $p_3$ of a cubic Bézier curve $\beta$ on the sphere $\mathcal{S}^2$. In this example, there is only a single data point $p$. Since the gradient is taken w.r.t. the last control point, there is only one path through the tree of geodesics along which $\mathrm{d}_{p_3} B^*_{2/3}$ transports $\mathrm{grad}_{\beta(2/3)} \mathrm{dist}(\,\cdot\,, q)^2 = -2\,\mathrm{Log}_{\beta(2/3)}(q)$ backward.

The gradient descent algorithm to compute $(\widehat{p}_0, \ldots, \widehat{p}_K)$ (not only) in symmetric spaces is given in Algorithm 5. Therein, the exponential and gradient of the product manifold $M^{K+1}$ are used. They are computed component-wise (because we work with the product structure). That is,

$$\mathrm{Exp}_{(p_0,\ldots,p_K)} \left( s\,\mathrm{grad}_{(p_0,\ldots,p_K)} \mathcal{E} \right) = \left( \mathrm{Exp}_{p_0}\left( s\,\mathrm{grad}_{p_0} \right), \ldots, \mathrm{Exp}_{p_K}\left( s\,\mathrm{grad}_{p_K} \right) \right)$$

for $p_0, \ldots, p_k \in U$ and $s > 0$.

As an initial guess for the gradient descent, choosing $(p_0, \ldots, p_K)$ along the geodesic polygon whose corners interpolate the data points closest to knot points (in terms of $t$) has worked well. Note that gradient descent methods converge to *local* minima. Thus, we implicitly assume that there is no "badly behaved" local minimum "far away" from the least squares estimator.

### 5.1.3 Maximum Likelihood Estimation

In Euclidean space, the least-squares estimation coincides with maximum likelihood estimation under the assumption that the distribution of the errors is Gaussian. In [89,

---

**Algorithm 5** Gradient descent algorithm for the least-squares estimator

---

**Input:** Riemannian manifold $M$ with normal convex neighborhood $U$, a compact set $\mathcal{C} \subset U^{K+1}$, data $(t_1, q_1), \ldots, (t_n, q_n) \in [0, 1] \times U$, initial guess $(p_0, \ldots, p_K) \in \mathcal{C}$ of the control points, step size function $s : T\mathcal{C} \to \mathbb{R}_{>0}$ (for example, Armijo step sizes [2, Def. 4.2.2]), tolerances $\epsilon, \delta > 0$

**Output:** approximation $(\widetilde{p}_0, \ldots, \widetilde{p}_K) \in \mathcal{C}$ of the least-squares estimator (5.4)

> **while** $\|\mathrm{grad}_{(p_0,\ldots,p_K)} \mathcal{E}\| > \epsilon$ and $s(\mathrm{grad}_{(p_0,\ldots,p_K)} \mathcal{E}) > \delta$ **do**
>
> $\quad (p_0, \ldots, p_K) \leftarrow \mathrm{Exp}_{(p_0,\ldots,p_K)} \left( -s(\mathrm{grad}_{(p_0,\ldots,p_K)} \mathcal{E}) \, \mathrm{grad}_{(p_0,\ldots,p_K)} \mathcal{E} \right)$
>
> **end while**
>
> $(\widetilde{p}_0, \ldots, \widetilde{p}_K) \leftarrow (p_0, \ldots, p_K)$

---

Sec. 4.4], Fletcher showed that this is also true for geodesic regression in symmetric spaces under a generalized isotropic normal distribution. In the following, we extend his result to regression with Bézier splines. To this end, we deal with integrals of functions on manifolds; a summary of the underlying theory is given in Appendix A.1.

The maximum likelihood method is another probabilistically motivated method to obtain an estimator for the parameters of a statistical model from given data. The idea is to view the parameters (the control points of $B$ in our case) as variables of the underlying probability density function (pdf). The *maximium likelihood estimator* is the set of parameters that maximizes the so-called *likelihood function* under the given data. In practice, because it has the same extrema but is often easier to analyze, one usually applies the logarithm to the likelihood function and works with the *log-likelihood function* instead.

An assumption on the pdf underlying the model is needed to apply the maximum likelihood method. Let $M$ be a connected symmetric space, $\bar{q} \in M$, and define

$$C(\sigma) := \int_M \exp\left( -\frac{\mathrm{dist}(p, \bar{q})^2}{2\sigma^2} \right) \mathrm{d}p.$$

It is important to note that $C$ does not depend on $\bar{q}$: The distance function and thus $C$ are invariant under isometries, and for any two points in a symmetric space, there always is an isometry of the space mapping one into the other [121]. Given $\bar{q}$ and $\sigma > 0$, Fletcher's generalized normal distribution is now defined via the pdf

$$\rho : M \to \mathbb{R}_{\geq 0}$$
$$p \mapsto \rho(p; \bar{q}, \sigma) := \frac{1}{C(\sigma)} \exp\left( -\frac{\mathrm{dist}(p, \bar{q})^2}{2\sigma^2} \right). \tag{5.7}$$

The distance function in the exponential leads to an isotropic assignment of probability mass. Two advantageous properties of $\rho$ are (apart from being an isotropic normal distribution in Euclidean space) that it is smooth and nonzero everywhere; both need not hold for other generalizations of the normal distribution. (They are not given, for example, for the maximum entropy distribution introduced in [184], which, on the other hand, has the advantage that anisotropy can be modeled.)

We say that the model (5.2) has *Gaussian errors*, if the random variable $\mathcal{Q}$ is conditionally distributed according to (5.7), that is, with pdf

$$\nu(q|t=t_0) = \rho\big(q; B(t_0; p_0, \ldots, p_K), \sigma\big) \tag{5.8}$$

for some $\sigma > 0$. Under this assumption, the *log-likelihood function* of our model is given by

$$l\,(p_0, \ldots, p_K; \sigma) := N \log\big(C(\sigma)\big) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \mathrm{dist}\big(B(t_i; p_0, \ldots, p_K), q_i\big)^2$$

for given data $(t_1, q_1), \ldots, (t_n, q_n) \in [0,1] \times U$; the *maximum likelihood estimator* of the model then is the spline whose control points maximize $l$ (with control points in $\mathcal{C} \subseteq U^{K+1}$).

The following theorem shows that, in symmetric spaces, maximum likelihood and least squares estimators coincide under the above assumption.

**Theorem 5.1.2.** *Let $M$ be a symmetric space, $U \subseteq M$ a normal convex neighborhood, and $(t_1, q_1), \ldots, (t_n, q_n) \in [0,1] \times U$ realizations of the model (5.2). Then, least squares optimization and maximum likelihood approximation are equivalent for (5.2) with Gaussian errors.*

*Proof.* Using Equations (5.5) and (5.6) we find for each $j = 1, \ldots, K$ and $p \in U$

$$\mathrm{grad}_p l = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \mathrm{grad}_p \mathrm{dist}\big(B(t_i; p_0, \ldots, p_{j-1}, \,\cdot\,, p_{j+1}, \ldots, p_K), q_i\big)^2$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n} \mathrm{d}_p B_{t_i}^* \Big(\mathrm{Log}_{B(t_i; p_0, \ldots, p_{j-1}, p, p_{j+1}, \ldots, p_K)}(q_i)\Big)$$

$$= -\frac{1}{\sigma^2} \mathrm{grad}_p \mathcal{E}.$$

Hence, in the interior of $\mathcal{C}$, local minimizers of the sum-of-squared error $\mathcal{E}$ are local maximizers of the log-likelihood function $l$, and vice versa. We now show that this also holds for global extrema in $\mathcal{C}$. Let $(\widehat{p}_0, \ldots, \widehat{p}_K)$ be the global minimizer of $\mathcal{E}$, $(p_0, \ldots, p_K) \in \mathcal{C}$ arbitrary, and $\alpha$ a smooth curve in $\mathcal{C}$ that connects them, while ending in the minimizer. Using the definition of the gradient and the relationship above, we see that $\mathrm{d}l = -1/\sigma^2 \, \mathrm{d}\mathcal{E}$. Hence, applying Stoke's Theorem (Theorem A.1.2) yields

$$l(\widehat{p}_0, \ldots, \widehat{p}_K) - l(p_0, \ldots, p_K) = \int_\alpha \mathrm{d}l$$

$$= -\frac{1}{\sigma^2} \int_\alpha \mathrm{d}\mathcal{E}$$

$$= \frac{1}{\sigma^2}\big(\mathcal{E}(p_0, \ldots, p_K) - \mathcal{E}(\widehat{p}_0, \ldots, \widehat{p}_K)\big)$$

$$\geq 0.$$

Since $(p_0, \ldots, p_K)$ is arbitrary, this shows that $(\widehat{p}_0, \ldots, \widehat{p}_K)$ is a global maximum of $l$. Analogously, one can show that global maxima of $l$ are global minima of $\mathcal{E}$. $\qquad\square$

## 5.2   A Nonlinear Hierarchical Model

Longitudinal studies are crucial for medical sciences. They provide much-needed information on developing phenomena, which can be vital for improving prognosis. To faithfully extract this information, the analysis method must handle the characteristics inherent to longitudinal data. Most prominently, it must be able to deal with the problem of having strongly correlated measurements (from the same individual at differing times) and independent measurements (from different individuals) as part of the data set. Standard statistical tools like regression only allow the study of effects across individuals or within a single one; they cannot be applied to mixed data since their fundamental independence assumptions are not fulfilled. Another frequent problem in longitudinal studies is missing data (for example, patients dropping out of a study) or irregular intervals between measurements. To deal with these problems, multivariate mixed-effects models (in particular, hierarchical models) were developed for data from Euclidean spaces [156]. To each individual, they assign an independent functional relation between the dependent variable and the time-dependent explanatory variable. The coefficients that determine these trends are assumed to be random within the population and are called "random effects". The primary object of interest is usually a "fixed effect" that depends on random effects. It is usually a "fixed" (non-random) coefficient that determines the average spatiotemporal behavior of the whole group.

Recently, the considerable potential of longitudinal studies for shape analysis has come into focus [99]. To obtain as much information as possible from the resulting data, it is necessary to leave the realm of Euclidean vector spaces and turn to methods from Riemannian geometry, as curved manifolds are their natural domains. Therefore, standard methods from multivariate statistics to analyze longitudinal data must be transferred to these more general spaces. Until now, this has only partly been done with hierarchical models. Current parametric hierarchical models for manifold-valued data are either based on geodesics [37, 172, 225, 226] or general trajectories [48, 231], while an approach based on non-parametric curves is presented in [44]. As an alternative geodesic model, the authors of [173] proposed to use a different Riemannian structure that allows for faster algorithms in case of high dimensional data. However, to the best of our knowledge, there is no *in-between* hierarchical model for parametric trends of a higher order that (a) allows modeling more complicated trends and (b) is efficient enough to handle (high-dimensional) data sets with a large number of observations. Since many phenomena are only poorly characterized by geodesic models (for example, the cyclic motion of the cardiac structure), there are numerous possible applications.

Thus, we propose a novel *higher order hierarchical model* for analyzing longitudinal manifold-valued data. By modeling, in the first step, subject-wise trends as splines consisting of generalized Bézier curves, we can correctly capture a vast range of phenomena—not least periodic ones. We use the developed regression method to estimate these trends. In the second step, the obtained trajectories are considered perturbations of a common mean (curve). To this end, we adapt the functional-based metric from [173] to compare the obtained subject trajectories within the corresponding *space of Bézier splines*. Thereby, we can analyze the inter-subject variability cor-

rectly without interference from correlating measurements. For efficient calculations, we rely on the geodesic calculus introduced in [120, 210] and also utilized in [173].

We apply the novel model to longitudinal shape data of the human distal femur in Section 6.5.

### 5.2.1 Spaces of Bézier Splines

The hierarchical model relies on the fact that at least lower dimensional sets of Bézier splines are smooth manifolds. To the best of our knowledge, this has not been shown before; therefore, we prove it here. We can then endow these spaces with a known metric for manifolds of curves and use it as the basis for our model.

**Theorem 5.2.1.** *Let $M$ be a $d$-dimensional Riemannian manifold and $U \subseteq M$ a normal convex neighborhood. Furthermore, let $\mathcal{B}_k^1(U)$ and $\mathring{\mathcal{B}}_k^1(U)$ be defined by (2.8) and (2.9), respectively. Then, $\mathcal{B}_k^1(U)$ can be given the structure of a $(k+1)d$-dimensional smooth manifold for $k \in \{0, 1, 2, 3, 4, 5\}$. Furthermore, with this structure, $\mathring{\mathcal{B}}_k^1(U)$ is a smooth $(k-1)d$-dimensional embedded submanifold of $B_k^1(U)$ for $k \in \{2, 3, 4, 5\}$.*

*Proof.* Let $1 \leq k \leq 5$. We define the map

$$F : U^{k+1} \to \mathcal{B}_k^1(U),$$
$$(p_0, \ldots, p_k) \mapsto \beta(\,\cdot\,; p_0, \ldots, p_k).$$

Corollary 2.1.2 implies that $F$ is injective and, thus, bijective. We can then push the product-manifold structure of $U^{k+1}$ forward along $F$ to $B_k^1(U)$ to turn it into a smooth manifold, thereby securing that $F$ is a smooth diffeomorphism [197, Ch. 30 § 9].[3] Analogously, we can push the structure of $U^{k-1}$ to $\mathring{\mathcal{B}}_k^1(U)$ turning it into an embedded submanifold. □

This structure allows us to identify Bézier splines of low degrees and their control points. Application-wise, the spaces covered in Theorem 5.2.1 are the most important ones as, traditionally, the degree of the underlying polynomial curve in regression is kept as low as possible to avoid overfitting, and this should not be different for regression in manifolds. Also, when curves of low degrees are not adequate, it can be advisable to use splines (built from them) instead of curves of higher degrees [162].

From a theoretical perspective, it is nevertheless interesting to advance the result to higher degrees. Indeed, it seems very likely that $\mathcal{B}_k^1(U)$ and $\mathring{\mathcal{B}}_k^1(U)$ can always be given a manifold structure in the above way. We, therefore, state the following conjecture.

**Conjecture.** *Theorem 5.2.1 also holds for all $k \geq 6$.*

We are unaware of explicit formulas for further control points, which would be necessary to extend our proof. What is known is that the $\ell$-th covariant derivative of a Bézier curve at $t = 0$ and $t = 1$ is determined by the first and last $\ell + 1$ control points, respectively [195, p. 112]; but not that a subset of them never suffices.

---

[3]More precisely, we obtain a topology on $\mathcal{B}_k^1(U)$ by requiring that a subset is open if and only if its preimage under $F$ is open. A maximal atlas is then given analogously: If $(V, \phi)$ is a chart of $U^{k+1}$, then $(F(U), \phi \circ F^{-1})$ is a chart of $B_k^1(U)$ (and all charts are constructed this way).

Of course, one could also try to use the inverse function theorem to show that $F$ is *locally* invertible. Then, by restricting to a subset $\widetilde{U} \subseteq U$, we would obtain that $\mathcal{B}_k^1(\widetilde{U})$ is a manifold. To this end, delicate investigations of sums of Jacobi fields along different curves are necessary. We leave this for future work.

Turning to Bézier splines, we can extend Theorem 5.2.1.

**Corollary 5.2.2.** *Let $M$ be a $d$-dimensional Riemannian manifold and $U \subseteq M$ a normal convex neighborhood. Furthermore, let $\mathcal{B}_{k_0,\dots,k_{L-1}}^L(U)$ and $\mathring{\mathcal{B}}_{k_0,\dots,k_{L-1}}^L(U)$ be defined by (2.8) and (2.9), respectively. Then, $\mathcal{B}_{k_0,\dots,k_{L-1}}^L(U)$ can be given the structure of a smooth manifold of dimension*

$$\dim\left(\mathcal{B}_{k_0,\dots,k_{L-1}}^L(U)\right) = d\left(\sum_{i=0}^{L-1} k_i - L + 2\right)$$

*if $k_i \in \{1,2,3,4,5\}$ for all $i = 0,\dots,L-1$. Furthermore, if $k_i \in \{1,2,3,4,5\}$ for all $i = 1,\dots,L-2$ such that $\sum_{i=0}^{L-1} k_i > L$, then $\mathring{\mathcal{B}}_{k_0,\dots,k_{L-1}}^L(U)$ is a smooth embedded submanifold in this structure. The dimension of $\mathring{\mathcal{B}}_{k_0,\dots,k_{L-1}}^L(U)$ is*

$$\dim\left(\mathring{\mathcal{B}}_{k_0,\dots,k_{L-1}}^L(U)\right) = d\left(\sum_{i=0}^{L-1} k_i - L\right).$$

*Proof.* We extend $F$ so that

$$\begin{aligned} F : U^{K+1} &\to \mathcal{B}_{k_0,\dots,k_{L-1}}^L(U), \\ (p_0,\dots,p_K) &\mapsto B(\,\cdot\,; p_0,\dots,p_K), \end{aligned} \tag{5.9}$$

where $K$ is given by Equation (2.7). The assertion now follows using the arguments from the proof of Theorem 5.2.1 for every segment. (Thereby, $F$ is also turned into diffeomorphism.) The construction for $\mathring{\mathcal{B}}_{k_0,\dots,k_{L-1}}^L(U)$ works analogously. $\qquad\square$

Our last results motivate the following definition.

**Definition 5.2.3.** *Let $M$ be a $d$-dimensional Riemannian manifold and $U \subseteq M$ a normal convex neighborhood. Furthermore, let $\mathcal{B}_{k_0,\dots,k_{L-1}}^L(U)$ and $\mathring{\mathcal{B}}_{k_0,\dots,k_{L-1}}^L(U)$ as in (2.8) and (2.9) be endowed with the smooth manifold structure introduced above. We call $\mathcal{B}_{k_0,\dots,k_{L-1}}^L(U)$ Bezierfold over $U$ of degrees $(k_0,\dots,k_{L-1})$ Bézier splines. We further call $\mathring{\mathcal{B}}_{k_0,\dots,k_{L-1}}^L(U)$ Bezierfold over $U$ of closed degrees $(k_0,\dots,k_{L-1})$ Bézier splines.*

Since derivatives of diffeomorphisms map tangent spaces bijectively into each other, we can characterize the tangent spaces of a Bézierfold $\mathcal{B}(U)$ with the help of the map (5.9): For each $B \in \mathcal{B}(U)$, we find

$$T_B\mathcal{B}(U) = \{X : [0,L] \to TM \mid \exists\, v_0 \in T_{p_0}M,\dots,v_K \in T_{p_K}M \ \forall\, t \in [0,L] :$$
$$X(t) = \sum_{j=0}^K \mathrm{d}_{p_j}B(t; p_0,\dots,p_{j-1},\,\cdot\,,p_{j+1},\dots,p_K)(v_j)\}.$$

As a manifold, $\mathcal{B}(U)$ can be endowed with a Riemannian metric. Let $\langle\,\cdot\,,\,\cdot\,\rangle$ be the metric of $M$. We propose to endow $\mathcal{B}(U)$ with the metric of Srivastava and Klassen from [231, Sec. 3.3]. At each $B \in T_B\mathcal{B}(U)$, it is given by

$$\langle\langle X, Y\rangle\rangle_B := \int_0^L \langle X(t), Y(t)\rangle_{B(t)}\mathrm{d}t$$

for all $X, Y \in T_B \mathcal{B}(U)$. Note that it is smooth (w.r.t. the used structure) because Bézier splines depend smoothly on their control points. The metric induces a natural distance between two Bézier splines. Although we do not have explicit formulas for geodesics or the Riemannian exponential $\text{Exp}_B^{\mathcal{B}}$ at $B \in \mathcal{B}(U)$ at hand, we can approximate them efficiently using variational time-discretization [210] as described in Section 5.2.4.

**Remark 5.2.4.** *Several Bézierfolds coincide with sets of constant maps into $U$. Whereas for spaces that contain non-closed curves, this is (only) true for $\mathcal{B}_0^1(U)$, there are infinitely many spaces of closed curves, including $\mathring{\mathcal{B}}_2^1(U)$, $\mathring{\mathcal{B}}_{1,2}^2(U)$, $\mathring{\mathcal{B}}_{1,1,2}^3(U)$, etc. It is then clear that these spaces are all diffeomorphic to $U$. Furthermore, since tangent vectors are constant maps to $TU$ in these cases, they are also isometric to $U$.*

### 5.2.2 The Model

We now introduce the nonlinear hierarchical model. Since its main application is data analysis from medical studies, we speak of measurements taken from "subjects" in the following. Nevertheless, the model is not restricted to data from this field.

Let $M$ be a Riemannian manifold and $U \subseteq M$ a normal convex neighborhood. Further, let $\mathcal{B}(U)$ be a Bézierfold over $U$. Consider that $S$ subjects are given and that, for each, there is (a possibly different number of) $n_s$, $s = 1, \ldots, S$ measurements of the pair of an independent, deterministic scalar variable $t$ (the same for every subject) and a manifold-valued dependent variable $\mathcal{Q}^{(s)}$. The data thus reads

$$\left( t_i^{(s)}, q_i^{(s)} \right) \in \mathbb{R} \times U, \quad i = 1, \ldots, n_s, \quad s = 1, \ldots, S. \tag{5.10}$$

Such data arises, for example, in longitudinal studies that observe shape developments [99], brain networks [219], or human gait [248]. Importantly, in these applications, the intra-subject measurements (that is, samples $q_i^{(s)}$ with the same $s$) are usually correlated, whereas cross-sectional measurements (that is, points $q_i^{(s)}$ with mutually different $s$) are independent of each other. Therefore, intra-subject trends must be treated ahead of the group-wise analysis, as fundamental independence assumptions are violated otherwise [172].

To accurately describe longitudinal data, we propose a hierarchical two-stage model. Thereby, we extend the works of Muralidharan and Fletcher [172], as well as Nava-Yazdani, Hege, and von Tycowicz [173], which both rely on geodesic trends. The model consists of the following two consecutive levels.

**Individual level**: On the lower, individual level, we assume that (5.2) underlies each subject's data independently from the others; the type and the degrees of all Bézier splines are supposed to be the same. That is, it is assumed that the data (5.10) is drawn from random variables

$$\mathcal{Q}^{(s)}(t) := \text{Exp}_{B^{(s)}(t; p_0^{(s)}, \ldots, p_K^{(s)})} \left( \epsilon^{(s)}(t) \right), \quad s = 1, \ldots, S.$$

**Group level:** On the upper group level, we view each trajectory as a perturbation of a common mean trajectory $\overline{B} := \overline{B}( \, \cdot \, ; \overline{p}_0, \ldots, \overline{p}_K) \in \mathcal{B}(U)$ in the Bézierfold $\mathcal{B}(U)$ according to

$$B^{(s)} = \text{Exp}_{\overline{B}}^{\mathcal{B}} \left( X^{(s)} \right), \quad s = 1, \ldots, S,$$
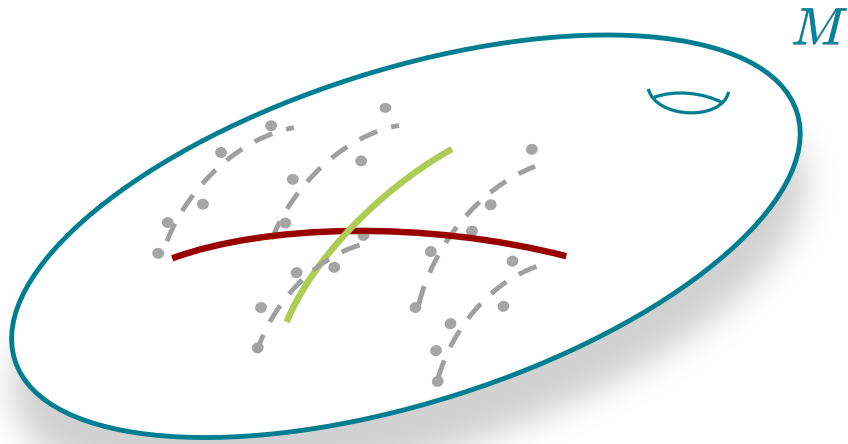
Figure 5.3: Conceptual comparison of a hierarchical model and regression analysis on a longitudinal data set. The measurements are shown as grey dots (without time). Each subject's (correlated) measurements follow its trend (broken lines). Spline regression on the full data set yields the red curve; its direction deviates strongly from all individual trends. The mean of the individual trends (green curve) is a better estimator of the average longitudinal trend in the cohort.

with random tangent vector fields $X^{(s)} \in T_{\overline{B}}\mathcal{B}$. This model first deals with the correlated samples, so, on the group level, independence can be assumed again. Its advantage for longitudinal data is conceptually visualized in Figure 5.3.

### 5.2.3   Parameter Estimation

We propose a two-step least squares estimation for the control points of the individual developments and those of the mean trend. For the control points of the individual trends, we can use the estimator (5.4). To approximate the control points of $\overline{\mathcal{B}}$, we suggest using the control points of the Fréchet mean in $\mathcal{B}(U)$ of the estimated individual trends.

### 5.2.4   Computation

We can compute the estimated control points of each subject as discussed in Section 5.1.2. The following discussion is, therefore, only concerned with approximating the control points of the mean spline. We rely on methods from discrete geodesic calculus by Rumpf and Wirth [210]; see Appendix A.3 for details.

Let $B_1, B_2 \in \mathcal{B}(U)$. A path between $B_1$ and $B_2$ through $\mathcal{B}(U)$ (and defined in $[0,1]$) may be represented as a parametrized surface in $U$, because it induces a map $H : [0,1] \times [0,L] \to U, (r,t) \mapsto H(r,t)$ with $H(0, \cdot) = B_1$ and $H(1, \cdot) = B_2$. A

geodesic between $B_1$ and $B_2$ is a minimizer of the *path energy*[4]

$$E(H) := \int_0^1 \int_0^L \left\langle \frac{dH}{dr}(r,t), \frac{dH}{dr}(r,t) \right\rangle dt \, dr.$$

Discretizing in $\mathcal{B}(U)$ and identifying splines with their control points, we obtain a *discrete $\ell$-geodesic* $(p_0^j, \ldots, p_K^j)_{j=0,\ldots,\ell} \in (U^{K+1})^{\ell+1}$ between $B_1$ and $B_2$ as the minimizer of the *discrete path energy*

$$E_\ell \left( \left( p_0^j, \ldots, p_K^j \right)_{j=0,\ldots,\ell} \right) := \ell \sum_{j=1}^{\ell-1} \int_0^L \text{dist} \left( B\left(t; p_0^j, \ldots, p_K^j\right), B\left(t; p_0^{j+1}, \ldots, p_K^{j+1}\right) \right)^2 dt.$$

When distances on $M$ can be computed, the integral can be evaluated using a suitable quadrature rule.

To approximate minimizers of $E_\ell$, we extend the iterative procedure from [173] to our setting. The algorithm is motivated by the general characteristic of shortest paths that are induced by Srivastava's metric and the corresponding connection in the space of curves: If $\alpha_1, \alpha_2 : [0, L] \to U$ are two smooth curves, then the map $H(\cdot, t) : [0, 1] \to U$ that is induced by the shortest curve between $\alpha_1$ and $\alpha_2$ is a geodesic in $M$ for all $t \in [0, L]$; see [231]. Approximating the integrals in $E_\ell$ with a quadrature and applying an alternating optimization scheme (that is, block coordinate descent) lets us compute the discrete $\ell$-geodesics between two curves by iteratively performing spline regression. First, we initialize the control points of the inner curves equidistantly along the geodesics that connect the corresponding control points of $B_1$ and $B_2$. The inner trajectories are then updated so that they lie "in the middle" of their neighbors; to this end, we replace them with the result of a spline regression on the $K + 1$ data points given by (equidistant) evaluations of the neighboring curves. The procedure is summarized in Algorithm 6.

Next, we discuss the computation of the *discrete $\ell$-mean* of $S$ curves $B_1, \ldots, B_S \in \mathcal{B}(U)$, with which we approximate the common mean (curve). It is the spline $\overline{B} \in \mathcal{B}(U)$ minimizing

$$G_\ell(p_0, \ldots, p_K) := \sum_{s=1}^S E_\ell \left( \left( p_0^j, \ldots, p_K^j \right)_{j=0,\ldots,\ell}^{(s)} \right),$$

$$\text{s.t. } (p_0^\ell, \ldots, p_K^\ell)^{(s)} = (p_0, \ldots, p_K)^{(s)}, \quad s = 1, \ldots, S,$$

where $(p_0^j, \ldots, p_K^j)_{j=0,\ldots,\ell}^{(s)}$ denotes the control points of the discrete $\ell$-geodesic between $B_s$ and $B(\cdot; p_0, \ldots, p_K)$. It can be computed with an alternating optimization scheme. We initialize the control points of $\overline{B}$ with the means of the corresponding control points of the data curves. Then, we compute discrete geodesics toward the mean and update the latter by a spline regression. This process is repeated in an alternating fashion. The whole procedure is summarized in Algorithm 7.

---

[4]Indeed, a curve minimizes the path energy *if and only if* it is a geodesic; see, for example, [45, p. 196].

---

**Algorithm 6** Discrete $\ell$-geodesic in $\mathcal{B}(U)$. Solving spline regression by computing (5.4) with Algorithm 5 is denoted by `reg`.

---

**Input:** $B_1, B_2 \in \mathcal{B}(U)$ with control points $(p_0^0, \ldots, p_K^0)$, $(p_0^\ell, \ldots, p_K^\ell)$, respectively

**Output:** approximations $(\widehat{p}_0^j, \ldots, \widehat{p}_K^j)_{j=0,\ldots,\ell}$ of the control points of the discrete $\ell$-geodesic from $B_1$ to $B_2$

**for** $j = 1, \ldots, \ell - 1$ **do**

$$(\widehat{p}_0^j, \ldots, \widehat{p}_K^j) \leftarrow \left( \gamma \left( \tfrac{j}{\ell}; p_0^0, p_0^\ell \right), \ldots, \gamma \left( \tfrac{j}{\ell}; p_K^0, p_K^\ell \right) \right)$$

**end for**

$(\widehat{p}_0^0, \ldots, \widehat{p}_K^0) \leftarrow (p_0^0, \ldots, p_K^0)$
$(\widehat{p}_0^\ell, \ldots, \widehat{p}_K^\ell) \leftarrow (p_0^\ell, \ldots, p_K^\ell)$

**repeat**

  **for** $j = 1, \ldots, \ell - 1$ **do**

    **for** $i = 0, \ldots, K$ **do**

$$\left( t_i^{j-1}, q_i^{j-1} \right) \leftarrow \left( \tfrac{iL}{K}, B \left( \tfrac{iL}{K}; p_0^{j-1}, \ldots, p_K^{j-1} \right) \right)$$

$$\left( t_i^{j+1}, q_i^{j+1} \right) \leftarrow \left( \tfrac{iL}{K}, B \left( \tfrac{iL}{K}; p_0^{j+1}, \ldots, p_K^{j+1} \right) \right)$$

    **end for**

$$(\widehat{p}_0^j, \ldots, \widehat{p}_K^j) \leftarrow \texttt{reg}\left( \left( t_i^{j-1}, q_i^{j-1} \right)_{i=1,\ldots,K} \cup \left( t_i^{j+1}, q_i^{j+1} \right)_{i=1\ldots,K} \right)$$

  **end for**

**until** convergence

---

**Algorithm 7** Mean trajectory. Computing the Fréchet mean (with Algorithm 1) and $\ell$-geodesic (with Algorithm 6) are denoted `mean` and `ℓ-geo`, respectively.

---

**Input:** $B_1, \ldots B_S \in \mathcal{B}(U)$ with control points $(p_0^{(s)}, \ldots, p_K^{(s)})_{s=1,\ldots,S}$, discretization parameter $\ell \in \mathbb{N}$

**Output:** approximation $(\widehat{p}_0, \ldots, \widehat{p}_K)$ of the control points of the discrete $\ell$-mean curve

$$(\widehat{p}_0, \ldots, \widehat{p}_K) \leftarrow \left( \texttt{mean}\left( p_0^{(1)}, \ldots, p_0^{(S)} \right), \ldots, \texttt{mean}\left( p_K^{(1)}, \ldots, p_K^{(S)} \right) \right)$$

**repeat**

  **for** $s = 1, \ldots, S$ **do**

$$\left( p_0^j, \ldots, p_K^j \right)_{j=0,\ldots,\ell}^{(s)} \leftarrow \ell\texttt{-geo}\big( B \left( \cdot\; ; \widehat{p}_0, \ldots, \widehat{p}_K \right), B_s \big)$$

  **end for**

  **for** $s = 1, \ldots, S$ **do**

    **for** $i = 0, \ldots, K$ **do**

$$\left( t_i^{(s)}, q_i^{(s)} \right) \leftarrow \left( \tfrac{iL}{K}, B \left( \tfrac{iL}{K}; p_0^{(s)}, \ldots, p_K^{(s)} \right) \right)$$

    **end for**

  **end for**

$$(\widehat{p}_0, \ldots, \widehat{p}_K) \leftarrow \texttt{reg}\left( \left( t_i^{(s)}, q_i^{(s)} \right)_{i=1,\ldots,K}^{s=1,\ldots,S} \right)$$

**until** convergence

---

## 5.3 Normalization via Regression

In statistics, one almost always needs to control confounding and other extraneous variables that are not of interest to a study but might influence the outcome. If the pool of samples is large enough, the controlling can be done through sample selection, enforcing, for example, that the variable is constant or follows a particular distribution. Another possibility, which is also applicable in the case of a small data pool, is *normalization*, that is, shifting/scaling the data's statistics such that the influence of the confounding variables is minimized. Such procedures can also be necessary when analyzing manifold-valued data. In this section, we develop a normalization method that helps to reduce the interference of continuous influence parameters. While motivated by the specific application shown in Chapter 7, it can be applied when several sets of manifold-valued data shall be compared that are influenced by a continuous confounding variable. Conceptually similar procedures are used in Euclidean space, for example, when analyzing molecules [111, 221, 250].

### 5.3.1 The Model

Let $M$ be a Riemannian manifold, $U \subseteq M$ a normal convex neighborhood, and $I_s \subset \mathbb{R}$, $s = 1, \ldots, S$, closed intervals with non-empty intersection, that is, $\cap_{s=1}^{S} I_s \neq \emptyset$. (The latter assumption guarantees a *shared* parameter value that can serve as a reference point for the normalization.) Let further $s$ groups of $U$-valued data points be given, each element coming with a parameter:

$$\left( t_j^{(s)}, q_j^{(s)} \right) \in I_s \times U, \quad s = 1, \ldots, S, \quad j = 1, \ldots, n_s.$$

We assume that the data from $s$-th group behaves according to the model (5.2), that is,

$$\mathcal{Q}^{(s)}(t) := \mathrm{Exp}_{B^{(s)}(t; p_0, \ldots, p_K)} \left( \mathcal{R}^{(s)}(t) \right) \quad s = 1, \ldots, S. \tag{5.11}$$

In contrast to the previous sections, here it is supposed that we are interested in the $TM$-valued variables $\mathcal{R}^{(s)}$ (thus the change of notation). We further assume that the "drifts" in the data given by the Bézier splines $B^{(s)}$, $i = 1, \ldots, S$, are *caused by the single confounding but deterministic variable $t$*.

### 5.3.2 Normalization

In this section, we discuss how the data can be normalized at some $t_0 \in \cap_{i=1}^{S} I_i$ so that further analysis is not influenced by the variability caused by $t$. The procedure is as follows. First calculating the least squares estimators (5.4) for each group yields best fitting splines $\widehat{B}^{(s)} : I_i \to M$, $i = 1, \ldots, S$. It is possible that the regressed curves already provide valuable information about the data, as we will see in Chapter 7. Approximations $\widehat{R}^{(s)}(t_i) \in TM$ of the realizations of $\mathcal{R}^{(s)}(t_i) \in TM$ of (5.11) are then given by the logarithms

$$\widehat{R}^{(s)}(t_i) := \mathrm{Log}_{\widehat{B}^{(s)}(t_j^{(s)})} \left( q_j^{(s)} \right) \in T_{\widehat{B}^{(s)}(t_j^{(s)})} M, \quad i = 1, \ldots, S, \quad j = 1, \ldots, n_i.$$

To minimize the influence caused by the variable $t$, we want to normalize the data at some chosen point $t_0$. Since, for each group, $\widehat{B}^{(s)}$ approximates the trend that $t$ enforces, we propose to parallel translate the vectors $\widehat{R}_j^{(s)}$ along $\widehat{B}^{(s)}$ to $\widehat{B}^{(s)}(t_0)$. This results in vectors

$$w_j^{(s)} \in T_{\widehat{B}^{(s)}(t_0)}M, \quad i = 1, \ldots, S, \quad j = 1, \ldots, n_i. \tag{5.12}$$

They represent the differences of the data points to the Bézier splines *normalized at $t_0$*. Mapping them back to the manifold gives the *normalized data points*:

$$\widetilde{q}_j^{(s)} := \mathrm{Exp}_{\widehat{B}^{(s)}(t_0)}\left( w_j^{(s)} \right), \quad i = 1, \ldots, S, \quad j = 1, \ldots, n_i. \tag{5.13}$$

---

**Algorithm 8** Data normalization w.r.t. a continuous parameter. Parallel translation of a vector $v$ from $p$ to $q$ along a Bézier spline $B$ is denoted by $\mathtt{transp}_B(v; p, q)$. Solving spline regression by computing (5.4) with Algorithm 5 is denoted by $\mathtt{reg}$.

---

**Input:** Realizations $(t_j^{(s)}, q_j^{(s)}) \in I_s \times U$ of (5.11) with $s = 1, \ldots, S$, and $j = 1, \ldots, n_s$; a common parameter $t_0 \in \cap_{i=1}^{S} I_i$

**Output:** Normalized data $\widetilde{q}_j^{(s)} \in M$

**for** $s = 1, \ldots, S$ **do**
$\quad B \leftarrow \mathtt{reg}\left( \left( t_i^{(s)}, q_i^{(s)} \right)_{i=1,\ldots,K} \right)$
$\quad$**for** $j = 1, \ldots, n_s$ **do**
$\quad\quad w_j^{(s)} \leftarrow \mathtt{transp}_B\left( \mathrm{Log}_{B\left(t_j^{(s)}\right)}\left( q_j^{(s)} \right); B\left( t_j^{(s)} \right), B\left( t_0 \right) \right)$
$\quad\quad \widetilde{q}_j^{(s)} \leftarrow \mathrm{Exp}_{B(t_0)}\left( w_j^{(s)} \right)$
$\quad$**end for**
**end for**

---

Now, we can perform *inter-group comparison* with the normalized data (5.13). For example, we can compute the Fréchet means and perform group tests for equality as in [172, Section 3.3]. If the curvature near the points $\widehat{B}^{(s)}(t_0)$ is small (that is, all sectional curvatures are close to 0), then we can also use the linearized data (5.12) and apply methods from multivariate statistics; the higher the curvature, though, the stronger will be the introduced distortion. The normalization process and its result for data from two groups are visualized in Figures 5.4 and 5.5. It is important to note that the choice of $t_0$ will influence the results, just like in multivariate statistics. (Imagine, for example, that we measure the height of a group of people. Only considering children at age ten will give results different from 20-years-old adults.)

Computation-wise, for geodesics as underlying trends, explicit formulas for parallel transport are known for many manifolds appearing in applications (see, for example, those discussed in Chapter 1). Otherwise, several numerical schemes can provide approximations of parallel transport along curves [109, 158, 145]. The normalization algorithm is given in Algorithm 8. (We again assume that Riemannian exponentials and logarithms can be computed in the manifold at hand.)
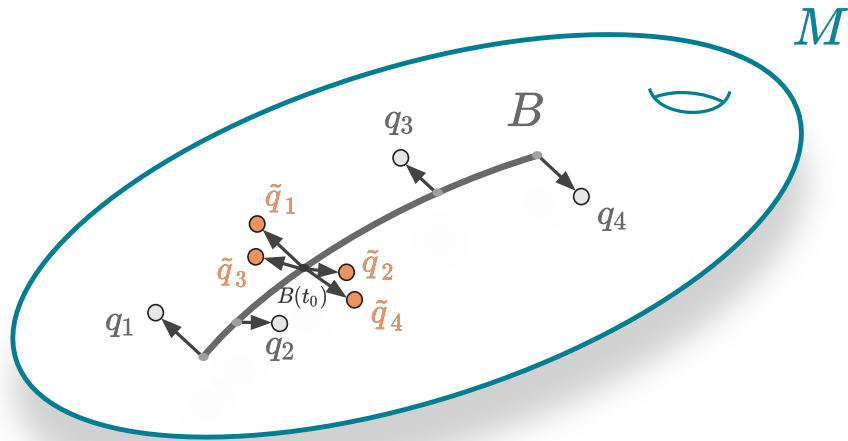
Figure 5.4: Normalization w.r.t. some parameter $t$ for a single data group $(q_j, t_j)$, $j = 1, 2, 3, 4$, in a Riemannian manifold $M$. The curve $B$ is the result of spline regression w.r.t. $t$. The points $B(t_j)$ are depicted in light grey, while the tangent vectors $v_j$ are the black arrows attached to them. Finally, the parallel translation $w_j$ of each $v_j$ is a tangent vector at $B(t_0)$ (also black); it yields the normalized data $\widetilde{q}_j$ shown in orange.
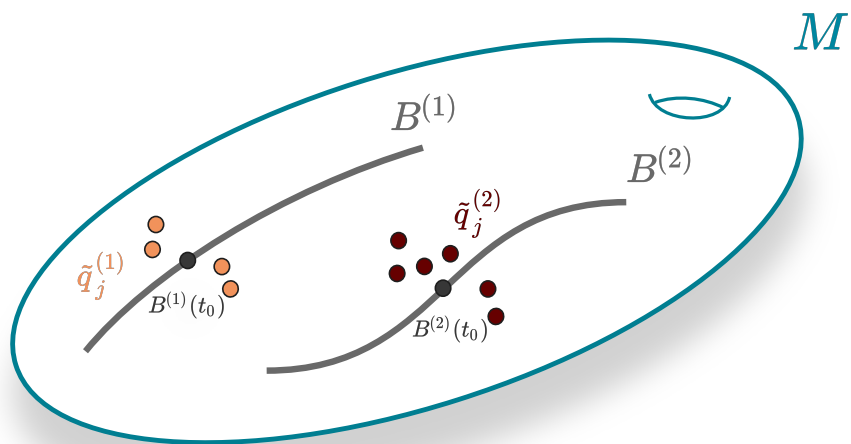


Figure 5.5: Normalization results for two groups. Both splines $B_1$ and $B_2$ are defined for the value $t = t_0$, so the two data sets can be normalized at $B^{(1)}(t_0)$ and $B^{(2)}(t_0)$, respectively. The groups (orange and red) can now be compared without bias caused by the influence of $t$.

# Part III

# Transfer to Applications

# Chapter 6

# Shape Analysis in Medicine

In this chapter, we show several applications of the newly-developed methods to shape analysis in medicine; the importance of the underlying geometry for data analysis is well-established in this field [189]. We start with a short overview of the diseases that motivate research in the applications we investigate. While we do not answer any of the questions that researchers in these areas ask right now, we demonstrate that our tools can be helpful when experts in the field apply them to their problems. We hope the following are good "guiding examples" to show what is possible. As is standard in these applications, we do not check for the existence of means, etc. beforehand (but we would if results called for it).

### Osteoarthritis

Osteoarthritis (OA) is a degenerative disease of the joints that affects millions of people worldwide [15]. It develops when the protective cartilage that cushions the ends of the meeting bones degenerates. While the joint pathology is diverse, the most prominent features are the loss of articular cartilage and remodeling of the adjacent bones; the former also leads to the so-called joint space narrowing as less cartilage is there to keep the bones apart. Although any joint can be affected, the population impact is highest for OA of the hip and knee [15]. While OA treatment traditionally consists of pain management and joint replacements for patients with severe symptoms, an improved understanding of the pathogenesis is shifting the focus to disease prevention [100].

### Alzheimer's

Alzheimer's disease (AD) is a neurodegenerative disorder that is diagnosed more and more often worldwide. The most important risk factor for developing AD is age [170]. People affected by AD often show an increasing lack of episodic memory, word-finding problems, loss of orientation in familiar neighborhoods, and other behavioral changes. The earliest signs of AD are so fine that there is also the category of "mild cognitive impairment" (MCI) for people with only slight memory impairments. However, most people with MCI also progress to AD [170]. Several studies showed that AD is characterized by an atrophy pattern of the brain, particularly of the hippocampi [171, 201, 220]. Their location in the brain is shown in Figure 6.1.

Figure 6.1: Position of the left and right hippocampus in the brain as seen from the underside ventral view. The picture was taken from Wikipedia.

**Mitral Valve Disease**

Diseases of the mitral valve are the second-most common form of valve disease in adults needing surgery [159]. *Mitral valve regurgitation* is an important example thereof. It can have different causes and is characterized by a (backward) blood flow from the left ventricle into the left atrium during systole. This is possible, for example, when the valve's leaflets do not close fully or prolapse into the left atrium during systole [80]. *Mitral valve stenosis* is typically caused by scarring after rheumatic fever and is more common in developing countries [49]. It is characterized, amongst others, by a thickening of the leaflets, which results in a narrowing of the valve opening. As a result, there is reduced blood flow from the left ventricle into the left atrium. Both mitral valve regurgitation and stenosis are visualized in Figure 6.2. Mitral valve diseases are often characterized by specific movement patterns. The corresponding shape anomalies can be observed (at least) at some point in the cardiac cycle. Early detection and assessment of mitral valve regurgitation are necessary for the best short-term and long-term results of treatment [80].

## 6.1 Bi-invariant Two-Sample Tests for Shape Analysis of Knee Configurations

We start by investigating configurations of the human knee joint under OA. The relative position of the femur and tibia (the constituting bones) can be described by a rigid-body transformation in the Lie group SE(3) (which we endow with the group structure from Example 1.2). Using two-sample tests, we infer a significant difference between knee configurations of people with severe OA and healthy controls, applying the bi-invariant Hotelling $T^2$ statistic and Bhattacharyya distance; we thus detect the well-known joint space narrowing that is indicative of OA.
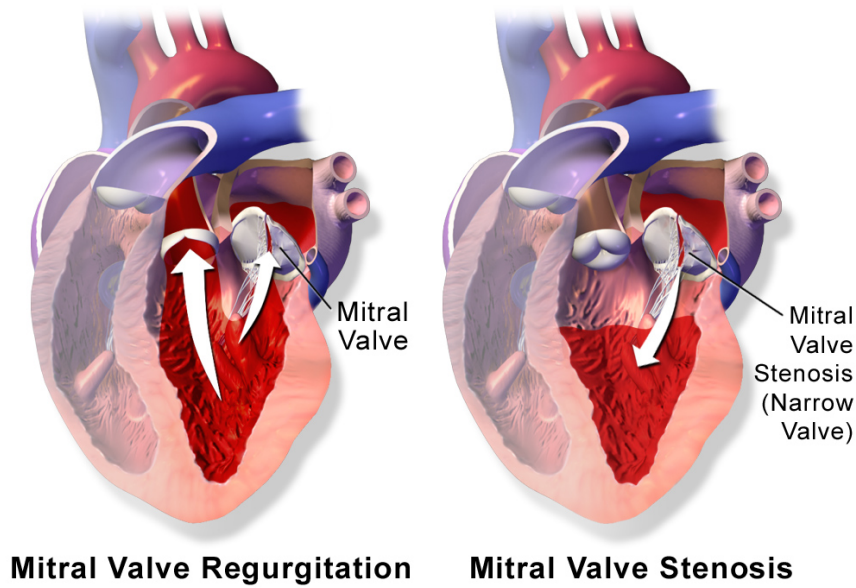
Figure 6.2: Mitral valve regurgitation and stenosis in the human heart. The image was taken from [232].

**Data Description**

The data set, which was also used in [245], is derived from the Osteoarthritis Initiative[1] (OAI). The OAI is a longitudinal study of knee osteoarthritis that provides (among others) publicly accessible clinical evaluation data and radiological images from 4,796 men and women between the ages of 45 and 79. Through a special support plate, it was made sure that the patients' legs were constrained to the same posture. From the baseline data set (that is, the images from the initial visits), we chose 58 severely diseased subjects and 58 healthy subjects according to their Kellgren-Lawrence score [141]—an ordinal scale from 0 to 4 based on radiographic features—for which segmentations of the respective magnetic resonance images are publicly available [14].[2] (We used samples with scores 0 and 1 for the control group and score 4 for the OA group.) The sets were balanced to maximize the statistical power of our hypothesis test. For the 116 subjects, surfaces of the distal femora and proximal tibiae were extracted from the respective 3D weDESS MR images (0.37×0.37 mm matrix, 0.7 mm slice thickness). A supervised post-processing procedure was used to ensure the quality of the segmentations and the correspondence of the resulting meshes (8,988 vertices for each femur and 8,320 for each tibia). During the process, the relative positions of the femur and tibia, which are present in the raw data through the global coordinate system of the scanner, are maintained.
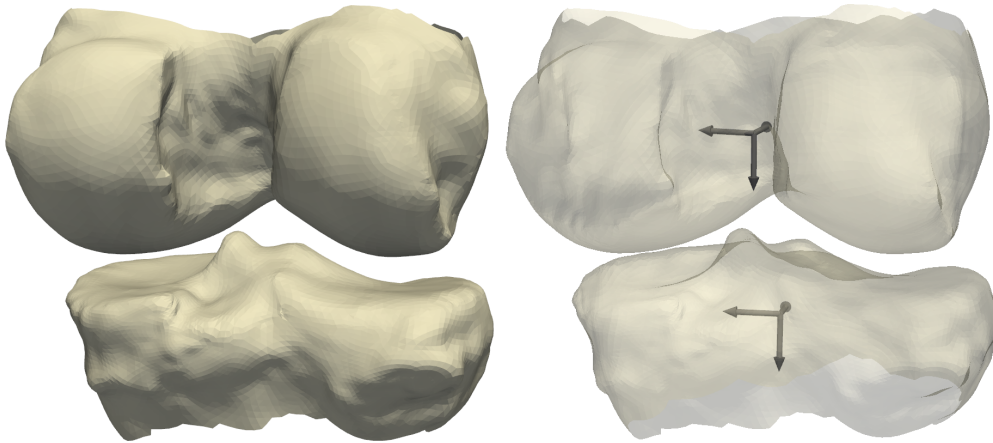
Figure 6.3: Visualization of the knee bones and the computed orthonormal frames. On the left, the distal femur (top) and the proximal tibia (bottom) are shown; the meshes (and their relative position) were reconstructed from MRI data. The space between both (the joint space) is filled with cartilage, ligaments, and menisci (all not shown). On the right, we depict the orthonormal frames of both meshes.

**Encoding Relative Positions in SE(3)**

For each femur-tibia pair, we can capture the position of the tibia relative to the femur by calculating the rigid body transformation that moves the latter onto the former. (If we chose the femur as the reference, we would get the inverse transformation.) More precisely, let $(O_f, x_f), (O_t, x_t) \in \mathrm{E}(3)$ be local orthonormal reference frames for the femur and tibia, respectively. (The vector part yields the origin, while the matrix columns represent the coordinate axes.) By applying $(0, -x_f) \in \mathrm{SE}(3)$ to both frames, we translate the meshes so that the femoral frame is always centered at the origin. The rotation-translation pair $P := (R, v) \in \mathrm{SE}(3)$ that maps the reference frame $(O_f, 0)$ of the femur to the reference frame $(O_t, x_t - x_f)$ of the tibia encodes the relative position of the bones; it follows from

$$(O_t O_f^T, x_t - x_f)(O_f, 0) = (O_t, x_t - x_f)$$

that

$$P = (R, v) = (O_t O_f^T, x_t - x_f).$$

To compute these translations, we must pick the initial frames for the femora and tibiae. We determine them from the principal component analysis of the vertices of the corresponding triangle meshes: While we pick the center of gravity as the origin, the axes are chosen as unit vectors pointing along the principal directions; the latter is done consistently throughout the population for both bone types. Note that left and right translations of $P$ by elements from $\mathrm{SE}(3)$ correspond to other choices of reference frames for the femora and tibiae, respectively (that is, all that are achieved by applying *the same* transformations to each set of frames). Our bi-invariant notions are thus invariant under such changes of reference.

---

Finally, we want to point out that this approach of representing relative positions by elements of the special Euclidean group is also used in many other applications, for example, medical image analysis [36], robotics [181], human action recognition [248], radar detection [22, 47], and state description of molecules [28]. Thus, the permutation test outlined below can be directly applied to data from these domains.

**Bi-invariant Permutation Test**

We use a non-parametric permutation setup to test whether there is a difference between the knee configurations of people with OA and healthy controls. We prefer this setup over a parametric test because we do not want to make assumptions about the distributions of both groups but rather learn this from the data. After the 116 femur-tibia pairs have been processed as described in Section 6.1, we obtain a set $(P_1^{(H)}, \ldots, P_{58}^{(H)})$ in SE(3) of transformations derived from healthy controls and a set $(P_1^{(OA)}, \ldots, P_{58}^{(OA)})$ coming from patients with OA. We consider the null hypothesis $H_0$ of equal distributions, that is, $P^{(H)} \overset{H_0}{\sim} P^{(OA)}$. As test statistics, both the bi-invariant Hotelling $T^2$ statistic $t^2$ and bi-invariant Bhattacharyya distance $D_B$ are used (and thus compared).

We initialize the test by computing the baseline

$$T_0 := T\left( \left(P_i^{(H)}\right), \left(P_i^{(OA)}\right) \right).$$

Then, we perform 10,000 random permutations of the joint set

$$(Z_1, \ldots, Z_{116}) := \left( P_1^{(H)}, \ldots, P_{58}^{(H)}, P_1^{(OA)}, \ldots, P_{58}^{(OA)} \right);$$

that is, denoting the $l$-th permutation by $\sigma_l$, we compute the values

$$T_l := T\left( (Z_{\sigma_l(1)}, \ldots, Z_{\sigma_l(58)}), (Z_{\sigma_l(59)}, \ldots, Z_{\sigma_l(116)}) \right), \quad l = 1, \ldots, 10000.$$

With $\mathbf{1}_{a \geq b}$ being 1 if $a \geq b$ and 0 else, the $p$-value for the statistic $T$ is then given by

$$p_T = \frac{1}{10000} \sum_{l=1}^{10000} \mathbf{1}_{T_l \geq T_0};$$

it is the proportion of test statistics greater than that computed for the original (unpermuted) groups. A standard level to reject the null hypothesis is $p_T < 0.05$; in this case, we call the difference between the distributions *significant* and reject $H_0$. Performing the test with each statistic, we obtain

$$p_{t^2} \approx 0.00019 \quad \text{and} \quad p_{D_B} < 10^{-5}.$$

Hence, the differences are significant with very small $p$-values for both statistics. We thus clearly detect the well-known observation of joint space narrowing under OA. Note that we can deduce a posteriori that the sample size was sufficient to detect significant differences with an error probability (much) smaller than 0.05.
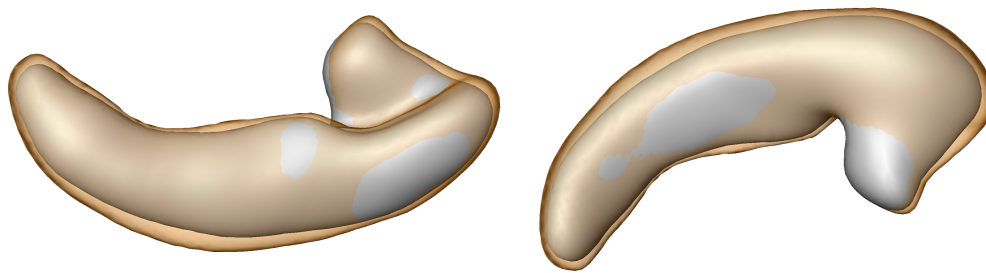
Figure 6.4: Group mean shapes of right hippocampi of the cognitive normal (orange, transparent) and impaired (grey) subjects overlaid.

## 6.2 Bi-invariant Two-Sample Tests for Shape Analysis of the Right Hippocampus

In our second experiment, we analyze hippocampal atrophy patterns due to mild cognitive impairment (MCI) by applying (and again comparing) the bi-invariant Hotelling $T^2$ statistic and Bhattacharyya distance. As is consistently reported in neuroimaging studies, atrophy of the hippocampal formation is a characteristic early sign of MCI. Using a local and global two-sample test, we infer significant differences in the distribution of shapes of right hippocampi between a cognitive normal and the MCI group, in agreement with the literature. This is the case for both of our dissimilarity measures. Moreover, the local test suggests that the shrinkage is more pronounced in some subregions of the hippocampus.

**Data Description**

For our experiment, we prepared a data set consisting of 26 subjects showing MCI and 26 cognitive normal (CN) controls from the open access Alzheimer's Disease Neuroimaging Initiative[3] (ADNI) database. Among others, ADNI provides 1632 magnetic resonance images of brains collected at four different time points with segmented hippocampi. We established surface correspondence (2280 vertices, 4556 triangles) in a fully automatic manner by applying the "deblurring and denoising" approach of functional maps [83] to isosurfaces that were extracted from the available segmentations. The data set was then randomly assembled from all resulting meshes whose segmentations were simply connected and whose surfaces approximated the corresponding isosurfaces well ($\leq 10^{-5}$ mm root mean square surface distance to the isosurface).

**Hippocampal Atrophy Patterns in CN vs. MCI**

To test for a difference in distribution between the CN and MCI groups, we extend the permutation test setup from Section 6.1 and perform both a local (that is, triangle-wise) and a global two-sample test. To this end, we use the space from Example 2.2 to model the shapes of the hippocampi. Since it is a Lie group, bi-invariant tools can be applied in the tests. After encoding the shapes, the group means of the CN and

---

[3] http://adni.loni.usc.edu/

MCI sets are computed with Algorithm 2. The loss of total hippocampal volume can be observed in the qualitative comparison of both mean shapes shown in Figure 6.4.

**Local Test**

In the following, when we speak of the *j-th differential coordinate*, we mean the *j*-th element of the differential coordinates vector. For $j = 1, \ldots, 4556$ let $G_j^{(\text{CN})}$ and $G_j^{(\text{MCI})}$ denote the *j*-th differential coordinate of the mean hippocampi of the CN and MCI group, respectively. Since the choice of reference is unimportant when using our bi-invariant dissimilarity measures, we select the first mesh of the CN group. To identify subregions that contribute to differences in mean shape between the groups, we perform triangle-wise, partial tests. For every triangle *independently*, we consider the null hypothesis $H_0$ that the distribution of its differential coordinate is the same for both groups, that is, $G_j^{(\text{CN})} \overset{H_0}{\sim} G_j^{(\text{MCI})}$.

To test this hypothesis, we perform a permutation and compare the bi-invariant Hotelling $T^2$ statistic and Bhattacharyya distance as test statistics. First, we perform independent tests for all differential coordinates $j = 1, \ldots, 4556$ using the testing procedure from Section 6.1. Therefore, denoting the test statistic by $T$ again, we start by computing

$$T_0^{(j)} := T\left( \left( G_j^{(\text{CN})} \right), \left( G_j^{(\text{MCI})} \right) \right).$$

We then perform 10,000 (random) permutations of the full set

$$(Z_{j,1}, \ldots, Z_{j,52}) := \left( G_{j,1}^{(\text{CN})}, \ldots, G_{j,26}^{(\text{CN})}, G_{j,1}^{(\text{MCI})}, \ldots, G_{j,26}^{(\text{MCI})} \right);$$

here the second subscript enumerates the subjects of the groups. Then, we compute the statistics

$$T_l^{(j)} := T\left( (Z_{j,\sigma_l(1)}, \ldots, Z_{j,\sigma_l(26)}), (Z_{j,\sigma_l(27)}, \ldots, Z_{j,\sigma_l(52)}) \right), \quad l = 1, \ldots, 10000.$$

The *p*-value of the *j*-th triangle for test statistic $T$ is then given by

$$p_T^{(j)} = \frac{1}{10000} \sum_{l=1}^{10000} \mathbf{1}_{T_l^{(j)} \geq T_0^{(j)}}.$$

Because of the large number of tests, we apply Benjamini-Hochberg false discovery correction at the level $\alpha = 0.05$ to identify triangles with significant differences (see Appendix A.4 for a summary of the procedure); only for these we reject the null hypotheses. Using the bi-invariant Hotelling $T^2$ statistic and Bhattacharyya distance as test statistics, the described test is bi-invariant (in this case in $\text{GL}^+(3)$). In particular, because of right invariance, the results are independent of the reference that was chosen to compute the differential coordinates. (Left-invariance would allow to *jointly* transform the differential coordinates of the subjects by elements from $\text{GL}^+(3)$, which might, for example, be of interest for a better numerical performance of an algorithm when manipulating differential coordinates.)

In Figure 6.5, we visualize the triangles with significant differences, showing the respective *p*-values for both statistics.

Figure 6.5: Group tests for differences in the distribution of right hippocampi for cognitive normal and impaired subjects. Results for the bi-invariant Hotelling $T^2$ statistic are shown at the top and for the Bhattacharyya distance at the bottom. Each triangle of the CN mean is color-coded according to its (corrected) *p*-value using the colormap 0.0 ▬▬▬▬ 0.05.

In line with literature on MCI [171, 220], the results hint at more differentiated morphometric changes beyond the homogeneous volumetric decline of the hippocampi. The Bhattacharyya distance detects more significant differences, which is expected because it also detects differences in covariance. Interestingly, there are also some areas where the Hotelling $T^2$ statistic is more sensitive. Understanding the differing behavior seems interesting for future work.

**Global Test**

Complementing the local is a global test that is sensitive to large-scale patterns of the covariance structures. It is thus sensitive to spatial dependencies across the shape. The construction of the test follows the approach from [215].

For each triangle, we first map the $T_l^{(j)}$ to an approximate uniform distribution in $[0, 1]$ by applying the corresponding empirical cumulative distribution function (cdf) $C_j$. More precisely, we compute

$$C_j(T_l^{(j)}) := \frac{1}{10000} \sum_{r=1}^{10000} \mathbf{1}_{T_r^{(j)} \leq T_l^{(j)}}, \quad j = 1, \ldots, 4556, \quad l = 0, \ldots, 10000.$$

Then, setting $\widetilde{C}_j(T_l^{(j)}) := 0.9998\, C_j(T_l^{(j)}) - 0.00001$ and denoting the cdf of the standard normal distribution by $\phi$, the data is mapped to $U_l^{(j)} := \phi^{-1}(\widetilde{C}_j(T_l^{(j)}))$. The latter follows an approximate standard normal distribution for each triangle.

Using the sample covariance matrix $\Sigma = 1/9999\, UU^T$ (where $U = [U_1, \ldots, U_{10000}] = [U_l^{(j)}]$), a suitable test statistic is now given by the squared Mahalanobis distance; this yields

$$\mu_0 := U_0^T \Sigma^{-1} U_0, \quad \mu_l := U_l^T \Sigma^{-1} U_l, \quad l = 1, \ldots, 10000. \tag{6.1}$$
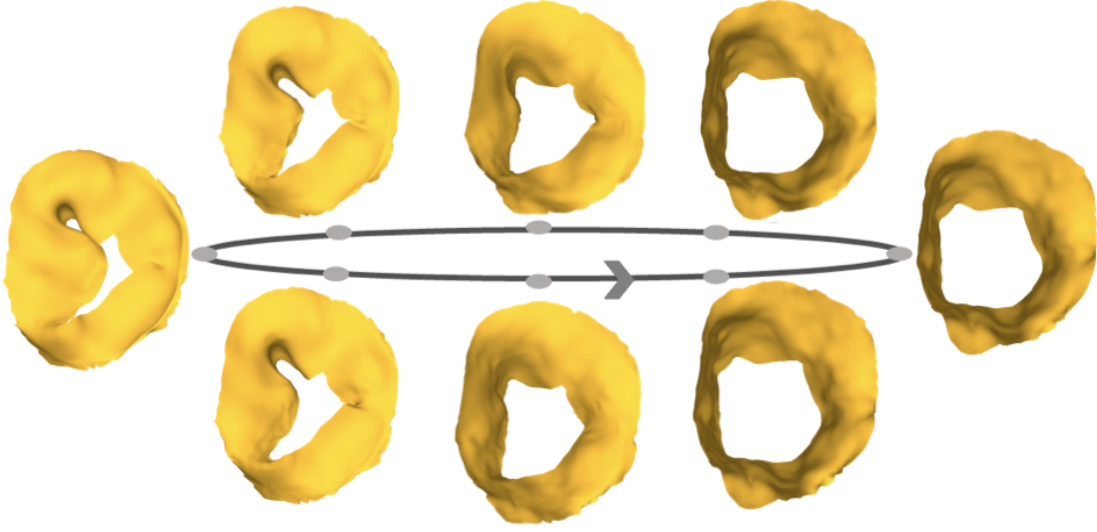
Figure 6.6: Reconstructed meshes from regression of longitudinal mitral valve data covering a full cardiac cycle. The spline consists of two cubic segments.

We then call the difference in global shape significant, when

$$p = \frac{1}{10000} \sum_{l=1}^{10000} \mathbf{1}_{\mu_l \geq \mu_0} < 0.05.$$

Note that to account for the irregularity in surface triangulations, the sample covariance operator and thus Mahalanobis distance in Equation (6.1) could be extended in terms of an adapted inner product that weights each component by the corresponding triangle's area. In this experiment, we adhere to the formulation (6.1), as the meshes of the studied hippocampi surfaces are uniform.

Applying this test to the values obtained from the local test yielded $p$-values smaller than $1/10^5$ for the bi-invariant Hotelling $T^2$ statistic and Bhattacharyya distance revealing that both are sensitive to the underlying morphological changes.

## 6.3 Reconstruction of Shape Trajectories of the Mitral Valve

We now apply regression with Bézier splines to longitudinal shape data of a mitral valve from a patient with mitral valve regurgitation. We choose to model the mitral valves in the Riemannian space of differential coordinates $\Delta$ from Example 2.3. To this end—sampling the first half of the cycle (closed to fully open) at equidistant time steps—five meshes (1,331 vertices / 2,510 faces) were extracted from a 3D+t trans-esophageal echocardiography (TEE) sequence as described in [237]. Let $q_1, \ldots, q_5 \in \Delta$ be the corresponding shapes in the space of differential coordinates. To approximate the full motion cycle we use the same five shapes in reversed order as data for the second half of the curve. Because of the periodic behavior, we choose a closed spline with two cubic segments as the model and assume an equidistant distribution of the

107

data points along the spline, that is, we employ

$$\big( (0, q_1), (1/4, q_2), (1/2, q_3), (3/4, q_4), (1, q_5), (5/4, q_4), (3/2, q_3), (7/4, q_2) \big)$$

as the full data set.

The regressed cardiac trajectory is shown in Figure 6.6. Our method successfully estimates the valve's cyclic motion capturing the prolapsing posterior leaflet. It shows the potential for improved reconstruction of mitral valve motion in presence of image artifacts like TEE shadowing and signal dropout. This, in turn, facilitates the quantification of geometric indices of valve function such as orifice area or tenting height.

## 6.4  Development of the Shape of the Distal Femur under Progressing Osteoarthritis

We now use regression with Bézier splines to investigate the development of the shape of the distal femur under OA. Therefore, we regress the 3D shape against OA severity as determined by the KL grade [141]. Our data set comprises 100 shapes (20 per grade) of randomly selected subjects from the OAI; triangle meshes of each bone were created as described in Section 6.1. We again encode the femora's shapes in the space of differential coordinates from Example 2.3.

| Order of Bézier curve | $R^2$ | $R^2_{rel}$ |
|:---:|:---:|:---:|
| 1 | 0.05 | 0.57 |
| 2 | 0.07 | 0.78 |
| 3 | 0.08 | 0.90 |

Table 6.1: The computed $R^2_{rel}$ and $R^2$ statistics of the regressed (w.r.t. KL grade) geodesic, quadratic, and cubic Bézier curve for data of distal femora

For $i = 0, \ldots, 4$, the shapes with grade $i$ are associated with the value $t_i = i/4$. We use our method to compute the best-fitting geodesic, quadratic, and cubic Bézier curves. To compare their explanatory power, we calculate for each the corresponding manifold-valued $R^2$ statistic that, for $r_1, \ldots, r_N \in M$ and *total variance* $\mathrm{var}(r_1, \ldots, r_N) := 1/N \min_{q \in M} \sum_{j=1}^{N} \mathrm{dist}(q, r_j)^2$, is defined by [90, p. 56]

$$R^2 = 1 - \frac{\text{unexplained variance}}{\text{total variance}} := 1 - \frac{2/N \, \mathcal{E}(\beta)}{\mathrm{var}(r_1, \ldots, r_N)} \in [0, 1].$$

The statistic measures how much of the data's total variance is explained by $\beta$.

For $j = 1, \ldots, 20$ and $l = 0, \ldots, 4$, let $q_j^{(l)}$ be the $j$-th femur shape with KL grade $l$. Note that, for the described setup, the unexplained variance is bounded from below by the sum of the per-grade variances, that is, $\sum_{l=0}^{4} \mathrm{var}\{q_1^{(l)}, \ldots, q_{20}^{(l)}\}$. In particular, this yields an upper bound for the $R^2$ statistic of $R^2_{opt} \approx 0.0962$ for our femur data.

Figure 6.7: Cubic regression of distal femora. Healthy regressed shape (KL = 0) together with subsequent grades overlaid wherever the distance is higher than 0.5mm, colored accordingly (0.5 ▬▬▬ 3.0).

Hence, we also provide relative values $R^2_{rel} := R^2 / R^2_{opt}$ for comparison. The results are shown in Table 6.1.

The computed cubic Bézier curve is displayed in Figure 6.7. The obtained shape changes consistently describe OA-related malformations of the femur, namely, the widening of the condyles and osteophytic growth. Furthermore, we observe only minute bone remodeling for the first half of the trajectory, while accelerated progression is visible for the second half. The substantial increase in $R^2_{rel}$ suggests that there are nontrivial higher-order phenomena involved, which are captured poorly by the geodesic model. Moreover, as time-warped geodesics are contained in the search space, we can inspect whether only a speed-up is behind the advantage of higher-order curves. This is not the case, as the control points of the cubic femoral curve do lie on a single geodesic. This confirms that there are higher-order effects beyond reparametrization.

## 6.5 Hierarchical Modeling and Analysis of Longitudinal Shape Data of the Distal Femur

In the last medical application, we investigate the shape development of the femur under OA by using data from *different* subjects. We now complement this with a group-wise analysis of femoral shape *trajectories* that we obtain by including intra-subject, longitudinal data from the OAI. We analyze this data with the hierarchical model from Section 5.2. To demonstrate that this model is not limited to the estimation of average, group-level trends, we also derive a statistical descriptor for shape trajectories in terms of the principal component scores (that is, the coefficients encoding the trajectories within the basis of principal geodesic modes [93]) and use it for *trajectory* classification.

We determined three groups of shapes trajectories: HH (healthy, that is, no OA), HD (healthy to diseased, that is, healthy onset followed by a progression to severe OA), and DD (diseased, that is, OA at baseline) according to the Kellgren-Lawrence score of grade 0 for all visits, an increase of at least three grades throughout the study,
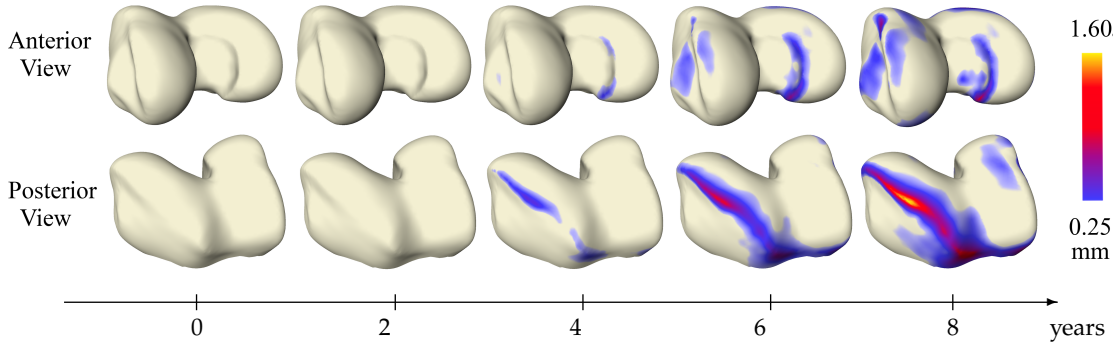
Figure 6.8: Mean of cubic femoral trends of 22 subjects evaluated at five equidistant points. The surface distance to the baseline (value of the computed mean at $t = 0$) is color-coded wherever the distance is larger than 0.25 mm. © 2022 IEEE

| actual\prediction | HH | DD | HD |
|---|---|---|---|
| HH | 19 | 1 | 2 |
| DD | 2 | 11 | 9 |
| HD | 4 | 6 | 12 |

Figure 6.9: Confusion matrix of the trained support vector machine for the classification of the trajectory type.

and grade 3 or 4 for all visits, respectively. Triangle meshes (8,988 vertices / 17,829 faces) of the femora were created as in Section 6.1. For each group, we assembled 22 trajectories (all available data for group DD except one subject that exhibited inconsistencies, and the same number for groups HD and HH, randomly selected), each of which comprises shapes of all acquired MR images, that is, at baseline, the 1-, 2-, 3-, 4-, 6-, and 8-year visits. As shape space, we again use the space of differential coordinates.

As a first application, we estimate a hierarchical model for the HD group. We choose cubic Bézier curves to model the individual trends because of the results from the last section. Time discrete computations are performed based on 2-geodesics— employing finer discretizations did not lead to further improvements for the dataset under study. The estimated group-level trend is visualized in Figure 6.8. The determined shape changes consistently expose OA-related malformations of the femur, most prominently changes along the ridge of the cartilage plate. The latter is a region known for osteophytic growth. Naturally, the changes are weaker than those of the trajectory in Figure 6.7 because the patients that are considered here only "reached" KL grade 4 at the end of the study, whereas severe malformations due to several years of grade-4-OA are included in the data that was used in the last section. Nevertheless, like in the previous experiment, only minute bone remodeling can be observed for the first half of the captured interval, whereas bone malformations develop more rapidly after four years. Therefore, also this experiment suggests that there are nontrivial higher-order phenomena involved for which geodesic models are inadequate. For the classification, we first compute the (discrete) mean trajectory of *all* 66 subjects and

use it to calculate an approximation $\widetilde{G}$ of the data's Gram matrix; see Appendix A.3 for details on the latter. For each subject trajectory, we then deduce a 65-dimensional descriptor in form of the coefficients with respect to the eigenvectors (that is, PGA modes [93]) of $\widetilde{G}$. We then train a simple support vector machine (linear kernel) on the descriptors in a leave-one-out cross-validation setup.

The percentage of correctly classified trajectories is 64%. The corresponding confusion matrix is shown in Figure 6.9. Performing the same experiment with a Euclidean model [63] results in 59% correct classifications demonstrating the advantage of our Riemannian model over shape spaces that come with the assumption that the data lies in a vector space.

# Chapter 7

# Shape Analysis in Archaeology

Artifacts, such as the physical remains of tools, weapons, clothing, adornments, and other human-made objects, represent the largest and most diverse source of evidence for archaeology. In this chapter, we show through a case study the possibilities modern shape analysis offers for information extraction from archaeological artifacts.

## 7.1 Analysis of the Shape of Ancient Sundials

### 7.1.1 Shape and Archaeology

Indeed, shape has always been fundamental to the analysis of archaeological artifacts. While this can also be said about the objects of interest of related disciplines, such as paleontology, geology, and biology, the "human dimension" of artifact creation adds complex layers of technological, economic, artistic, and other social components to the physical manifestations of shapes. As a consequence, archaeology traditionally uses the method of *typology* to impose an order, called *typological sequence* [205, Ch. 4], on a series of artifacts. It is thought to isolate the dominant morphological trends and provide a basis for further (mostly informal) analysis of the unexplained residual trends. There is a large body of published research on the process of typological ordering (often called *seriation* in archaeology and *ordination* in ecology [175]), including computational methods of varying complexity (see, for example, [152, 160, 216]). The combinatorial complexity of seriation implies that optimal results are computationally infeasible even for a relatively small collection of artifacts, whereas archaeological excavations often yield tens of thousands of them. Consequently, no single procedure is undisputed, and manual grouping and ordering based on subjective mixtures of features (for example, artistic/stylistic, technological, functional) are still common.

We use the example of ancient sundial surfaces from Italy and Greece to probe the transition from this discrete and rather subjective procedure to continuous, data-driven analysis. These are well suited to our approach because their physical shape must reflect at least one controllable trend: the (intended) geographic installation site, where the sundial functions properly.
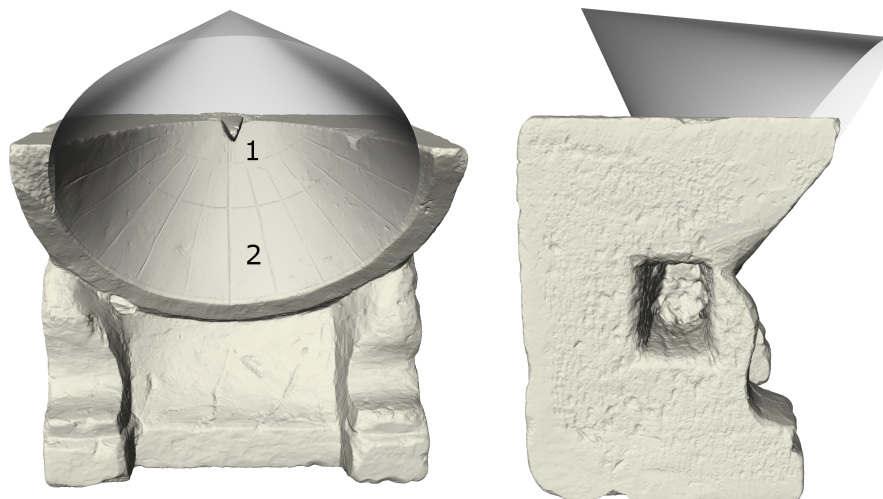
Figure 7.1: Front (left) and side (right) view of a conical sundial with gnomon (1) and shadow surface (2). To craft the shadow surface, part of a cone was cut from the stone, as indicated by the gray cone in the picture.

### 7.1.2 Problem, Analysis Task

In antiquity, sundials were mainly used to measure time [213]. By applying astronomical knowledge, sundials could be constructed to show the correct time for all seasons. While the sun is above the horizon, sundials indicate the time by the position of the shadow of a so-called *gnomon* on a shadow-receiving surface; see Figure 7.1. We adopt the term *shadow surface* for the latter. Shape-wise, at least four types of sundials can be distinguished [207]: conical, spherical, cylindrical, and planar sundials—determined solely by the shape of the shadow surface. In Figures 7.1 and 7.2, we display a conical and a spherical sundial, respectively. Independently from the type, the geographical latitude had to be considered during construction to ensure proper functioning. In particular, it is well known that the *hour lines* (that were inscribed into the shadow surface to read off time) were adapted to the latitude of the location of installation [136, 130]. An obvious question then is whether other parts were also changed: We show that, at least for the spherical type, it is very likely that the *shape* of the shadow surfaces was also adapted by the craftsman by identifying a latitude-dependent trend in the data. We demonstrate how useful this trend can be by inferring the latitude of the installation location of a sundial with an uncertain site from it. Furthermore, if one wants to compare shapes of shadow surfaces from different geographical locations, for example, for differences in construction principles, the shapes should be *normalized* to latitude; otherwise, one might misinterpret differences that are "imposed" on the craftsman by the latitude of the location. We, therefore, use the normalization method that we developed in Section 5.3 for an illustrative analysis, comparing the mean shapes of shadow surfaces from ancient Greek and Roman sundials. With more (reliable) data, such an analysis could—unbiased from latitudinal effects—reveal differences in construction principles between the groups.

To sum up: In this section, we demonstrate the use of statistical tools for shape data in archaeology by (1) identifying a latitude-driven shape trend for the shadow
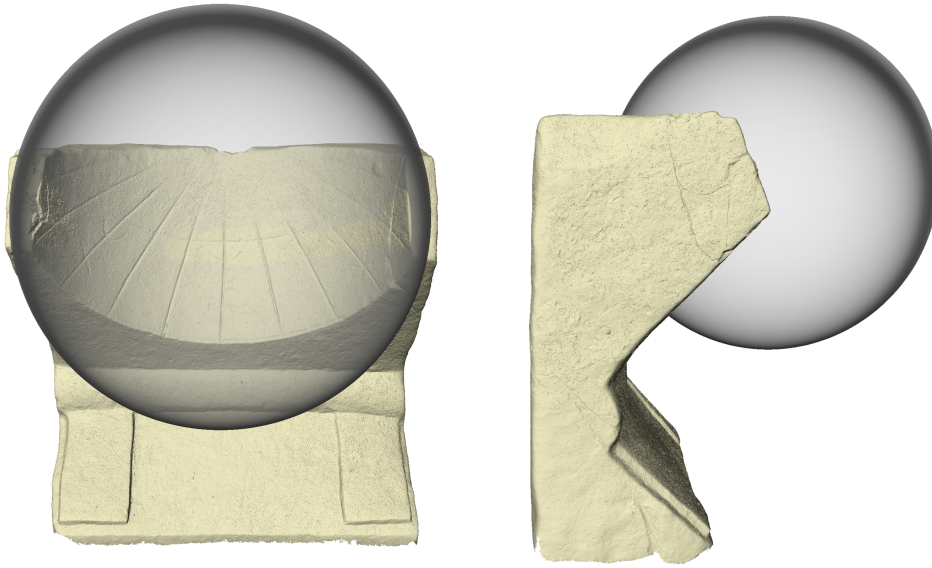
Figure 7.2: Front (left) and side (right) view of a spherical sundial with missing gnomon. When making the shadow surface, a part of a round ball was cut from the stone, as indicated by the gray sphere.

surface of ancient Roman sundials, and (2) using this trend to approximate the latitude of the place of installation of a sundial for which the latter is uncertain. Furthermore, we show a use case of our normalization method. The analysis pipeline is visualized in Figure 7.3.

### 7.1.3 Data and Data Preprocessing

Today, there are about 500 known ancient Greek or Roman sundials, and 3D models of many of them can be found in the repository of the Excellence Cluster Topoi [106]. They were reconstructed in a Structure from Motion/Multiview Stereo (SfM/MVS) procedure detailed in [96]. For this case study, we used spherical sundials from Greece and the Italian peninsula and considered all available models with well-preserved shadow surfaces (10 from the Italian peninsula and three from Greece); their IDs and sites—including longitude and latitude—are given in Appendix A.5; the sites are also depicted in Figure 7.7.

As for data preparation, we extracted the shadow surfaces with the software Amira [233]. We manually selected the shadow surfaces and corrected triangles with extreme angles near the new boundary. Furthermore, the resolution of the surface meshes was reduced to about $20\,k$ faces using the *quadric edge collapse decimation scheme* implemented in the free software MeshLab. An example of a segmented shadow surface can be seen in Figure 7.4; the same is shown after extraction and simplification in Figure 7.5.

To perform statistics on shapes, we rely on group-wise correspondence, that is, consistent point-to-point relationships of the meshes. There are many approaches to
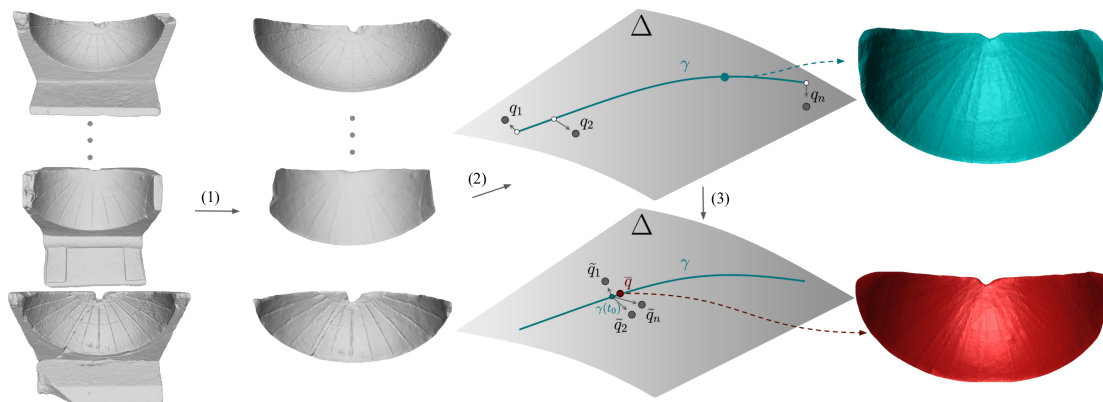
Figure 7.3: Graphic of the analysis pipeline: After digitizing $n$ sundials as triangle meshes, shadow surfaces are extracted, and a group-wise correspondence is established (1). The shapes $q_1, \ldots, q_n$ are then encoded in the shape space $\Delta$ of differential coordinates (2). The statistical analysis of the shapes of the shadow surfaces can then be performed in the manifold $\Delta$; we apply geodesic regression w.r.t. latitude, which yields the fitted geodesic $\gamma$ (as depicted). Being a function that depends on latitude, we can evaluate $\gamma$ at various latitudes. The resulting shape is a representative shadow surface at the given latitude. It can be transformed into a triangular mesh for visualization (the green shadow surface). In step (3), the novel normalization method is applied to normalize all shapes to a common latitude $t_0$, resulting in points $\widetilde{q}_1, \ldots, \widetilde{q}_n \in \Delta$. The shapes of the shadow surfaces can now be further analyzed without the additional variability introduced by the influence of latitude on the shapes; for example, we can compute the normalized mean shape $\bar{q}$. A visualization of this mean (the red shadow surface) can again be obtained when we transform $\bar{q}$ back into a triangular mesh.

establish such correspondences, for example, non-rigid registration [40, 235] and functional map–based [177] procedures. We obtained one by choosing a shadow surface as a reference and registering it with a harmonic map to all other surfaces through the following two steps: First, we established a correspondence of the boundaries using three landmarks. While the two geodesically farthest points were automatically approximated as the global minimizer and maximizer of the surface's Fiedler vector[1] (the eigenfunction that corresponds to the second smallest eigenvalue of the surface's Laplace-Beltrami operator [192]), the third landmark was manually selected as the lowest point of the gnomon hole. (The gnomon is missing at all but one of the sundials under study.) Then, we established the correspondence of the interior part using the *discrete harmonic map for disc-like topology* as described in [39]. Finally, the effects of rotation and scale were removed by a group-wise Procrustes alignment of the meshes. The processed meshes are available online [114].

Note that on many sundials, the upper left and right corners are only poorly preserved, showing a varying degree of decay. Thus, the correspondences of the worn edges (which are relatively small compared to the rest of the sundial) are only

---

[1]There are examples of graphs for which the Fiedler vector does not emphasize the vertices that are farthest apart [82], but we confirmed visually that it does so for the sundials.
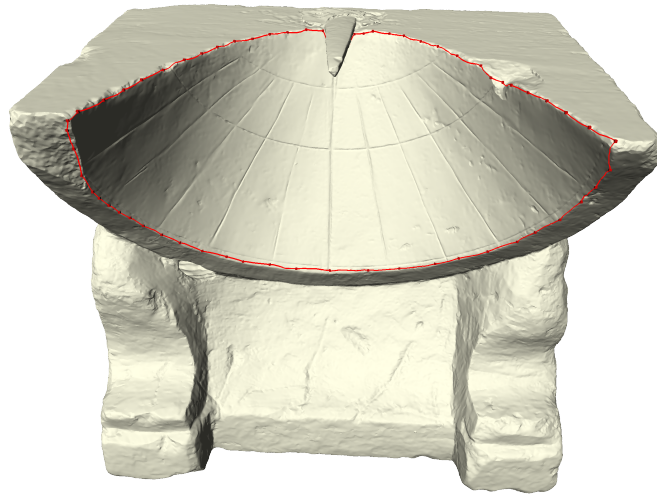
Figure 7.4: Manually segmented shadow surface



Figure 7.5: Extracted shadow surface

approximate. In the later analysis, we concentrated on the other areas (which are conserved very well).

### 7.1.4 Study

For our case study, we choose the shape space of differential coordinates from Example 2.3.

Given the shapes, we perform geodesic regression (that is, spline regression with geodesics) with the latitude of the site as the explanatory variable for both groups; the corresponding intervals are determined by the latitudes of the most southern and northern sites, that is, $I_R = [40.7030, 43.3155]$ and $I_G = [36.0917, 37.3900]$ for the Roman and Greek sundials, respectively. It seems reasonable to assume that a change in latitude causes a geodetic drift in the data: In the range of latitudes we study, shapes of shadow surfaces can be expected to change at a constant rate in a fixed direction in shape space to compensate for the latitude-dependent shift in sunlight angle.[2] Equidistant evaluations of the resulting geodesics are shown in Figure 7.6. Because of the larger number of samples, the results for the Roman group are of higher interest: They show a clear "bending" of the shadow surface that increases

---

[2]To corroborate this, we also tested regression with higher-order Bézier curves; this overfitted the data, that is, nearly interpolated the given shadow surfaces at their corresponding latitudes, *while showing unreasonable behavior in between.*
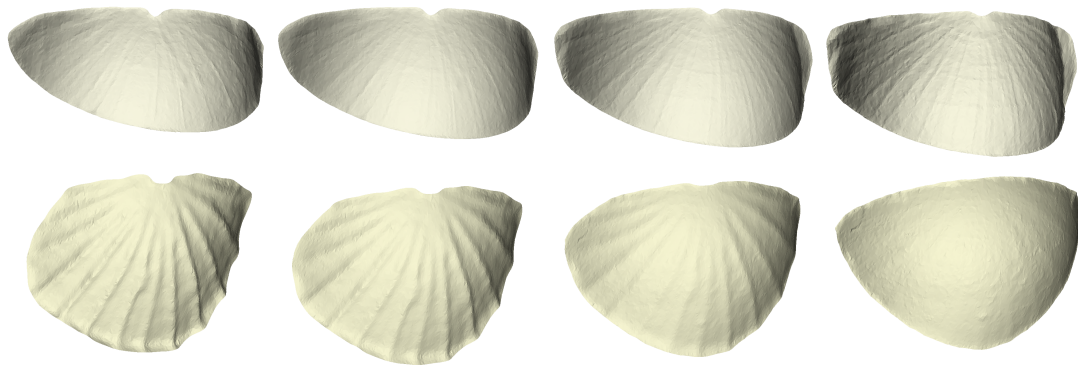
Figure 7.6: Results of geodesic regression for shadow surfaces of spherical sundials w.r.t. latitude. The results for Roman and Greek sundials are shown in the top and bottom rows, respectively. After calculating the trajectories in the space of differential coordinates, we sampled each curve at 4 points and computed the corresponding triangular meshes. The Roman geodesic models the trend for latitudes in $I_R = [40.7030, 43.3155]$ while the Greek one is defined on $I_G = [36.0917, 37.3900]$.

with latitude; it is highlighted in Figure 7.8.

The "bending" is an interesting effect: It suggests that the *form* of the whole shadow surface (and not only that of the hour lines) was adapted to the location of installation.[3] This fact can be used to place sundials of unknown or uncertain installation sites: After orthogonally projecting [12, Sec. 3.3] the shape of the sundials' shadow surface onto the regressed geodesic (which here mostly amounts to finding the most similarly bent shadow surface on the curve), we can "read off" the latitude at which the sundial probably was installed.

To demonstrate this, we selected a sundial from Italy (Dialface ID 39 in the Topoi database; see Table A.1), which at the time of writing is in a museum in Vatican City, but whose installation site is uncertain. (The inset shows a visualization of this sundial.)

Projecting the shape of its shadow surface onto the Roman geodesic gave an approximate latitude of $t^* \approx 42.1029°$, which is slightly north of Rome's latitude. Thus, our model suggests that the sundial was once located near the *solid* line shown in Figure 7.7. To investigate the accuracy of this prediction, we performed a leave-one-out cross-validation test for the Roman sundials with *known* location: Each sundial was taken out, regression was computed using all other samples, and the latitude of the left-out sundial was predicted. The *mean absolute error* (MAE) we thereby found was approximately 0.71° latitude (with a standard deviation of 0.36), which is about 80 kilometers in the north-south direction.



---
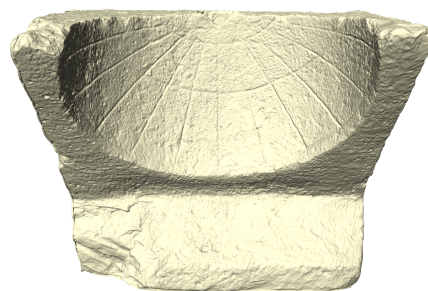
[3]It would be interesting to know how this was technically realized in production. However, we are unaware of any ancient source that provides information about this.
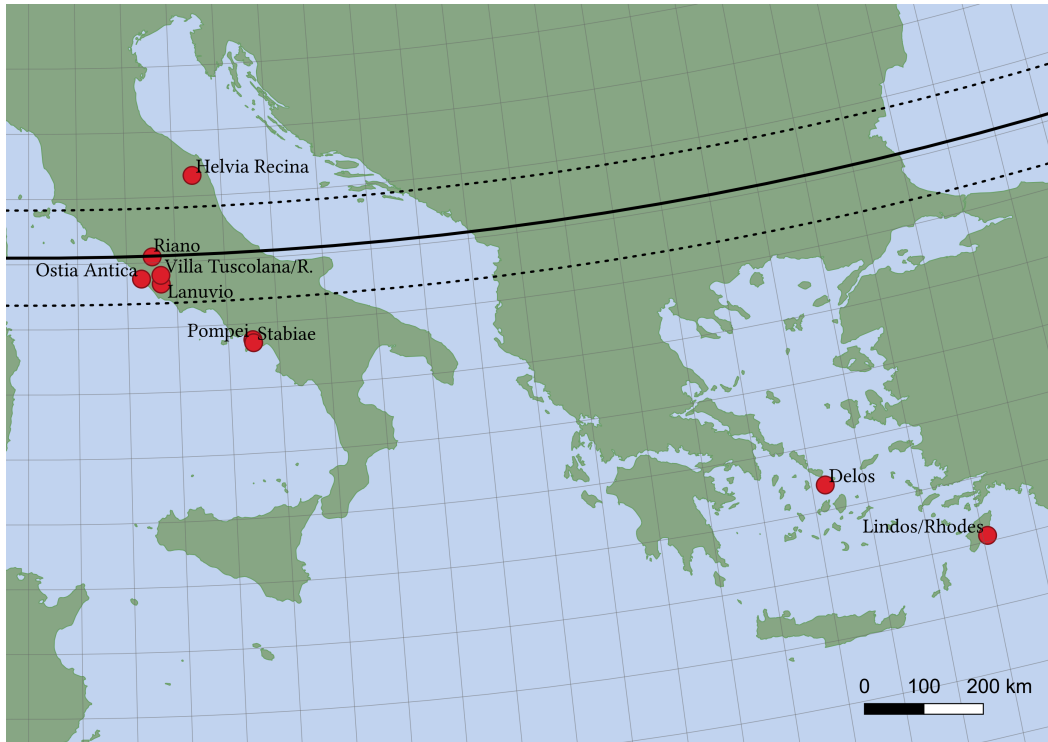
Figure 7.7: Map of the relevant area. The red circles mark the sites of the sundials. The solid line indicates the latitude where the sundial with Dialface ID 39 was probably installed. The dashed lines bound the region of uncertainty for this prediction as determined by the computed MAEs from all sundials with known installation sites.

We use the computed MAE as an uncertainty measure for the prediction and depict the latitudes $t^* \pm 0.71°$ as *dashed* lines in Figure 7.7; the location of installation was likely inside the resulting band. To compare our method against a traditional approach, we additionally applied *partial least squares regression* [253] on the Procrustes aligned coordinates of the shadow surface meshes to predict the corresponding latitude. For this, we used the "PLSRegression" module from scikit-learn 1.1.1. Setting the number of components to 1 yielded the best results. The obtained MAE was approximately 0.87° latitude (with a standard deviation of 0.73). This was about 0.16° (about 17.76 km) higher than the MAE of our proposed method. Furthermore, the standard deviation was substantially higher when using PLS regression. Sundial-wise errors and MAEs for both approaches are shown in Figure 7.9.

We now use our normalization method from Section 5.3 to compare the mean shape of shadow surfaces from the Italian peninsula with that of the samples from Greece. Controlling for latitude correctly accounts for the fact that Greece is significantly south of the Italian peninsula. Because the experiment can only be performed with three Greek samples, it should be seen as an illustrative example. Nevertheless, we mention possible research questions that could be answered with more data in the discussion section.

We have already done the first step for the normalization by performing geodesic regression with respect to latitude for both data sets. For completion, we show over-
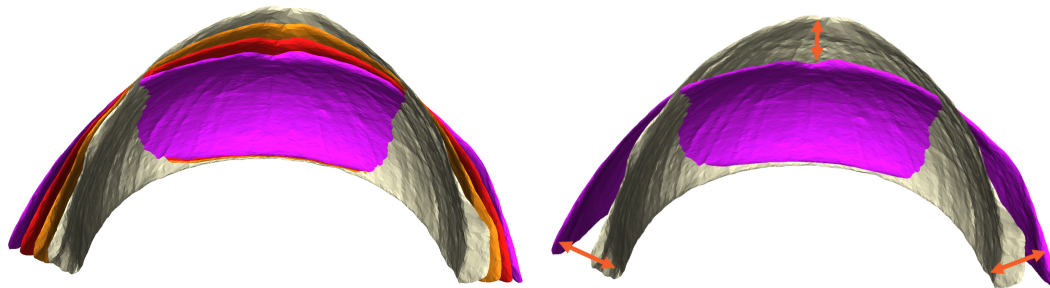
Figure 7.8: Geodesic regression w.r.t. latitude over the interval $I_R = [40.7030, 43.3155]$ for shadow surfaces of Roman sundials. We evaluated the optimal geodesic at four equidistant points and show the top view of the resulting shadow surfaces, with colors beige, orange, red, and violet ranging (in this order) from the most northern to the most southern site. On the left, the shadow surfaces are overlaid to show the bending. On the right, only the shadow surfaces corresponding to the most northern and southern latitudes are depicted. We highlight the bending with arrows there.
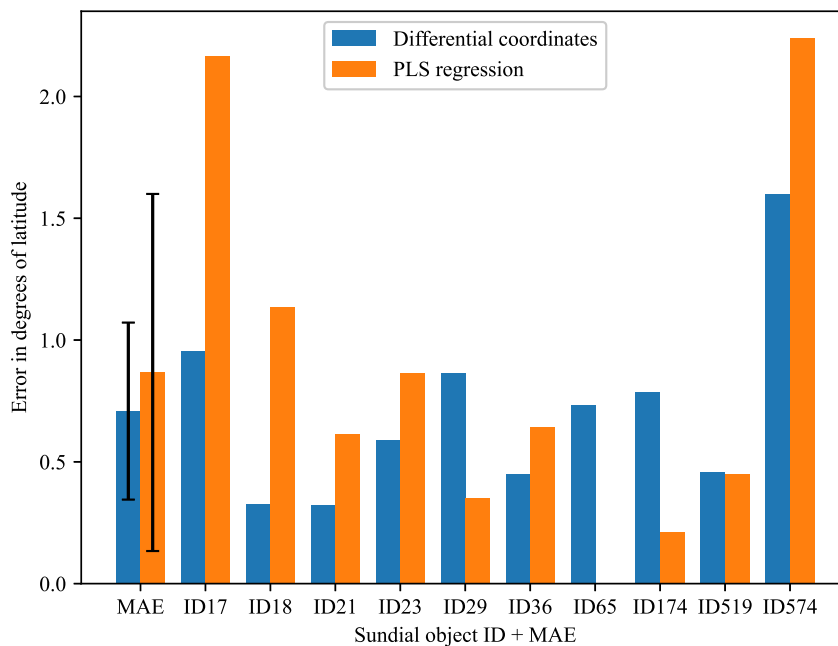


Figure 7.9: Sundial-wise and mean absolute errors obtained during latitude prediction. "Differential coordinates" refers to the proposed method; "PLS regression" stands for the partial least squares regression. MEAs are depicted with standard deviations.

laid evaluations of the Greek geodesic in Figure 7.10. Note that the shadow surface hardly bends with changing latitude contrasting our earlier result for Roman sundials.
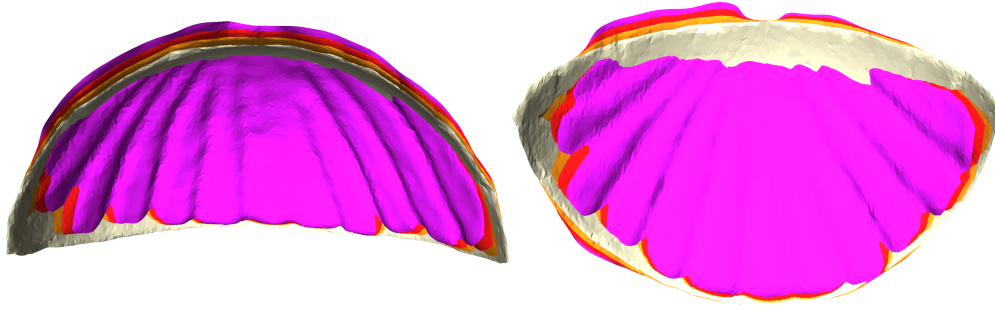
Figure 7.10: Geodesic regression w.r.t. latitude for shadow surfaces of sundials from Greece $I_G = [36.0917, 37.3900]$; the top view is on the left while the front view is on the right. We evaluated the optimal geodesic at four equidistant points, with colors beige, orange, red, and violet ranging (in this order) from the most northern to the most southern site.

Since $I_R \cap I_G = \varnothing$, we extrapolate[4] both the Roman and the Greek geodesic such that they are defined on the intervals $\widetilde{I}_R := [38.5, 43.3155]$ and $\widetilde{I}_G := [36.0917, 39]$, respectively. This allows us to normalize both data sets at different latitudes as described in Section 5.3 and, in each case, to compute the mean shape of the normalized shadow surfaces using Algorithm 1; the results are shown in Figure 7.11. Note that the means "bend" like the geodesics, which is the effect of the normalization.



Figure 7.11: Roman (left) and Greek (right) means of shadow surfaces of spherical sundials after normalizing the Roman data at 40° (red), 41° (orange) and 42° (beige) latitude and the Greek data at 37° (red), 38° (orange) and 39° (beige) latitude, respectively.

Widening the window for comparison by further extrapolation is of no avail due to increased artifact formation in the normalized data. The appearance of artifacts can be attributed to initial noise (occurring mainly at the corners, as explained in Section 7.1.3) that is amplified too much. The first signs of such artifacts can be seen in Figure 7.11, where they begin to form at the corners of the Greek mean at 39° latitude.

In Figure 7.12, we show both means after normalizing at 38.5° latitude. When overlaid they look very different, particularly in their curvatures. To quantify this, we compute the spheres that fit the shadow surfaces best[5]; the results can be seen in

---

[4]In almost all practical scenarios, geodesics can be extrapolated; see [45, Ch. 7] for details.

[5]More precisely, we computed the center $c^* \in \mathbb{R}^3$ and radius $r^* > 0$ of the sphere with the smallest
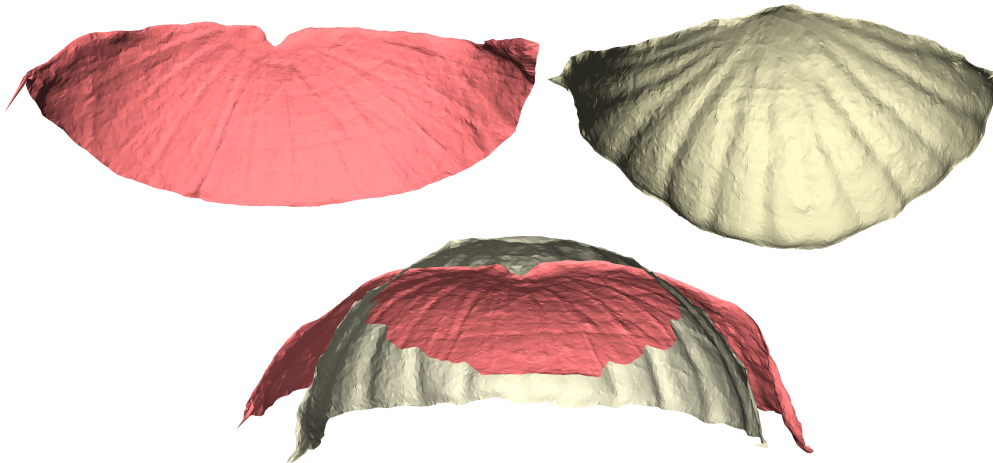
Figure 7.12: Comparison of the Roman (red) and Greek (beige) mean shapes after normalizing at 38.5°. They are shown individually (top row) and overlaid (bottom row).

Figure 7.13. Both are different in size, the Greek sphere's radius measuring 66% of the Roman. (Since the size is normalized in our study only relative notions are of interest.)
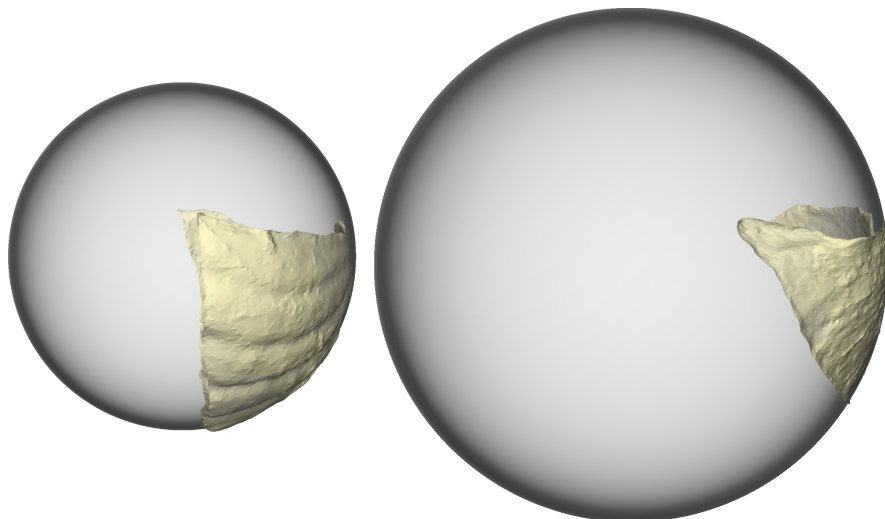


Figure 7.13: Spheres fitted to the Roman (left) and Greek (right) mean of the data that was normalized at 38.5° latitude. The center and radius of the spheres were chosen such that the average squared distances to the vertices of the triangle meshes were minimal.

The fact that the "bending" is different influences the comparison of the means.

---

sum of squared distances to the vertices of the shadow surface. Denoting the usual Euclidean norm by $\|\cdot\|$, it can be seen that the distance of a point $p \in \mathbb{R}^3$ to a sphere with center $c$ and radius $r$ is equal to $\tau(c, r; p) := |\|p - c\| - r|$. We thus computed $(c^*, r^*) := \min_{(c,r) \in \mathbb{R}^3 \times \mathbb{R}_{>0}} \sum_i \tau(c, r, v_i)^2$, where the $v_i$ are the vertices of the mesh of the shadow surface. The "L-BFGS-B" method was used for minimization. The problem is discussed in depth in [30].

As shown, at 38.5° latitude, they are differently curved; because of the stronger bending in the Roman group, the forms would converge, though, if we used this trend to compare them further north. This raises an interesting question: whether craftsmen from Greece and the Italian peninsula, when building sundials, made different adaptions to account for changes in the solar elevation angle. If backed up by more data, the results of our experiment could suggest not only that the mean shape of a shadow surface depends on the latitude of the installation site in both regions but also that this dependence *differs in strength between the groups*. For example, it could be that while craftsmen in Greece mainly adapted the hour lines, their Roman counterparts additionally adjusted the shape of the shadow surface. Examining this question further seems very interesting but requires more data, particularly from Greece.

To compare the nonlinear differential coordinates shape space with a Euclidean alternative (which is difficult, as statistically significant quantitative results cannot be expected because of the small sample size), we investigate how the Riemannian Mahalanobis distance (3.30) between the mean shapes of the Roman and Greek groups adapts to the normalization. After mapping the Roman *samples* and the Greek *mean* to the tangent space at the Roman *mean*, we compute the Mahalanobis distance[6] of the Greek mean to the distribution of the Roman samples. (We choose this direction because of the higher number of Roman samples.) This is done for both the original data and the one normalized at 38.5°. Afterward, the same computations were performed using the standard Euclidean shape space from [63]. For both shape spaces, the distance increased when the samples were normalized. This makes sense because the in-group variability (here within the Roman group) reduces through the normalization (cf. Figs. 5.4 and 5.5). Since the Mahalanobis distance weights the difference between the means (that is, the logarithm) against the inverse of the observed covariance (matrix) in the Roman group, a decrease of the latter leads to a higher value. While the distance for the nonlinear space increases by a factor of 1.73, it is multiplied by $5.1 \times 10^5$ for the Euclidean model.[7] (Both numbers are rounded.) That is, the latter loses a high amount of variability while normalizing. We think the differential coordinates conserve the finer details, whereas the Euclidean model focuses more on global features. But we believe that, for the sundials, smaller details are of high interest, particularly when additional characteristics like the hour lines are investigated as part of further analysis.

We also examine to which extent the results depend on the resolution of the meshes by simplifying them. More precisely, we reduce the number of faces of each shadow surface (further) to 10k, 2k, and 1k triangles with Meshlab's "quadric edge collapse" and compute normalized means for the Roman and Greek groups. As is visible in Figures 7.14 to 7.16, the mean shapes are effectively unchanged even with a poor resolution of only 1,000 faces. This shows that our methods are affected very little by noise from scanning devices and can also be used when only relatively poor

---

[6]As the dimension of the differential coordinate space is (much) higher than the number of samples, the covariance matrix is not invertible. Therefore, as is often done, we substituted the inverse of the covariance matrix of the Roman samples by its pseudoinverse to compute the distance.

[7]We also investigated how the geometric distance between the means behaves under the normalization. It increases by a factor of approximately 1.86 and 1.64 for the differential coordinate space and the Euclidean shape space, respectively. Therefore, the extreme increase in Mahalanobis distance for the latter cannot be explained by diverging means.

digital representations are at hand.
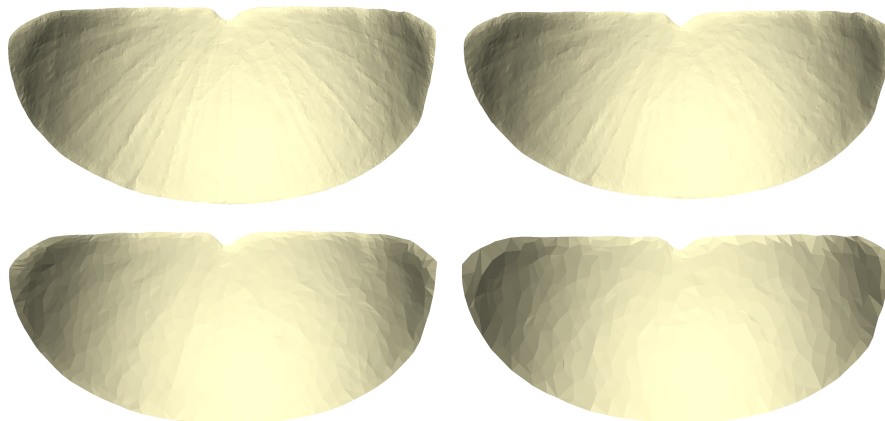


Figure 7.14: Mean shapes of the normalized (at 41°) sundials from the Italian peninsula for different mesh resolutions. Results are shown for resolutions of 20k (upper left), 10k (upper right), 2k (lower left), and 1k (lower right) faces.
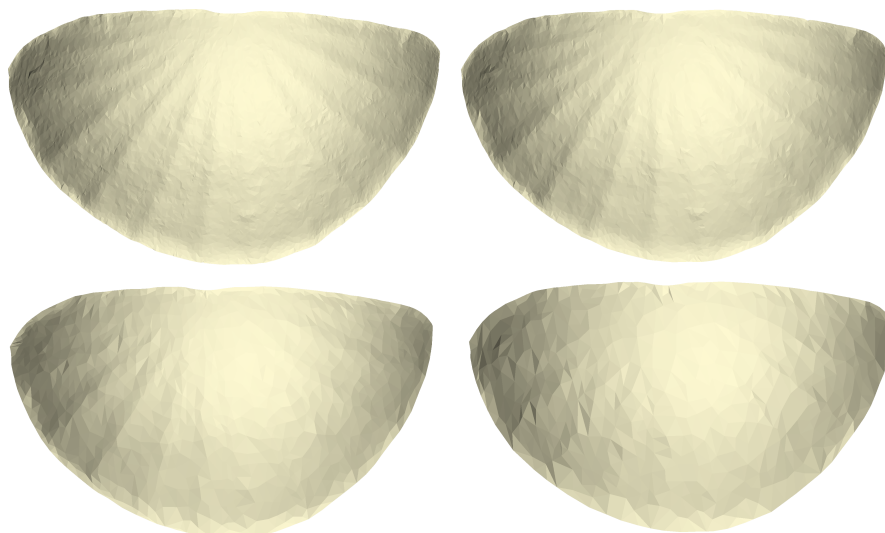


Figure 7.15: Mean shapes of the normalized (at 37°) sundials from Greece for different mesh resolutions. Results are shown for resolutions of 20k (upper left), 10k (upper right), 2k (lower left), and 1k (lower right) faces.

### 7.1.5 Discussion

The fact that the shape of the Roman shadow surfaces "bends" with latitude is a significant finding of this study. To the best of our knowledge, it has never been observed before. It seems likely that the form of the whole shadow surface was adapted to the location of installation— a conclusion that is supported by the accuracy of the derivative method for latitude prediction. Our results thus add a piece to the puzzle of how ancient Roman craftsmen created working sundials. An interesting question
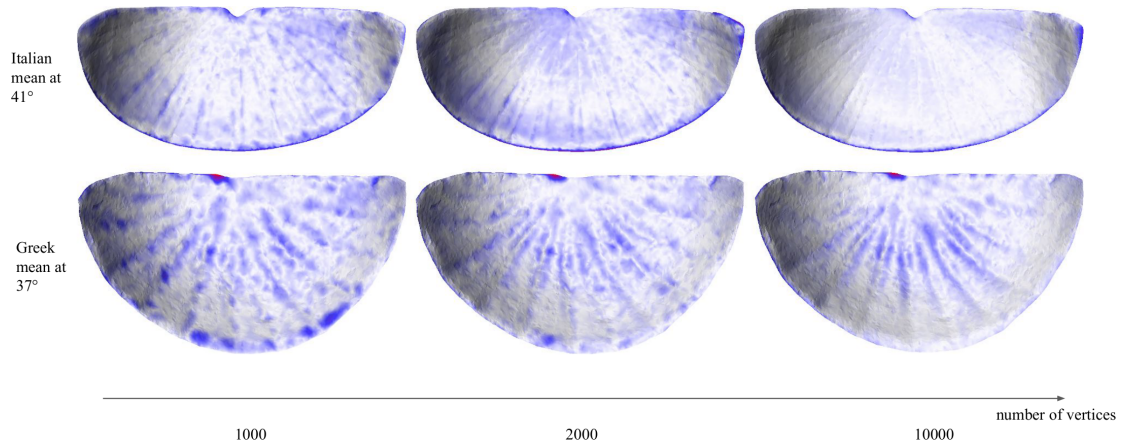
Figure 7.16: Surface distance for different mesh resolutions when computing normalized means. The respective meshes of the highest resolution are depicted with coloring (0.001 mm ▓▓▓▓ 5 mm) that encodes the distance to the corresponding mesh with the lowest resolution. Areas, where the distance is smaller than 0.001 mm, are white. As a reference, the distances between the upper left and upper right corners were about 29 and 21 cm for the Italian and Greek mean, respectively.

for future research is whether similar bending can also be observed for sundials from locations outside of Italy.

Our results also indicate that the new method to identify the latitude of the installation site is accurate; furthermore, they suggest that utilizing the non-Euclidean shape space structure for prediction is superior to the commonly employed PLS regression. The only other method for predicting a sundial's working latitude we know is the one from Pistellato, Traviglia, and Bergamasco [193]. It forecasts the working latitude by utilizing the positions of the shadow surface's inscribed hour lines. Knowledge of celestial mechanics, which they assume the craftsmen possessed, allows them to compute the latitude from these positions. While their method has the advantage that it does not need more than one sundial for prediction, it requires high mesh resolutions for accurate computations. In contrast, the above results show that the newly presented method works well even at coarse resolutions. Moreover, being purely data-driven, it requires no assumptions on the interplay between celestial mechanics and sundial shape. On the other hand, since both approaches make use of different, non-overlapping features of the shadow surface's makeup, they can complement each other and even be applied together to utilize more of the available information.

A limiting factor of this study is the small number of samples. Convincing statistical significance can only be achieved with more data. This is a common situation in archaeology. While more (digitized) ancient sundials are available, they are too damaged to be used with current methods. New approaches for statistical analysis of partial forms are needed to alleviate this problem; these could allow more existing ancient sundials to be used.

More data would also help to decide an interesting question raised by the proposed normalization method: whether craftsmen from Greece and the Italian peninsula, when building sundials, made different adaptions to account for changes in the

125

solar elevation angle. If backed up, the results of our experiment could indicate not only that the mean shape of a shadow surface depends on the latitude of the installation site in both regions, but also that the strength of dependence *varies among the groups*.

It could be, for example, that the Greek craftsmen mainly adjusted the hour lines, while their Roman counterparts also adjusted the shape of the shadow area. It seems very interesting to investigate this question further; however, it would require more data, especially from Greece.

Last but not least, we clearly showed that the proposed methods can be used even when only relatively poor digital representations are at hand.

### 7.1.6 Conclusion

Recent advances in image-based reconstruction (Structure from Motion and Multi-view Stereo) have made the acquisition of highly detailed 3D artifact models simple, fast, and economical. The sundial models explored in the present study represent a rapidly growing volume of 3D research data. This data should be explored with mathematical approaches that take full advantage of the rich information they contain and avoid information losses such as those caused by *a priori* typological grouping. We developed such a procedure by combining shape space methods and statistics on manifolds and demonstrated the potential to extract more information from groups of artifact shapes without prior assumptions. Conversely, procedures such as those discussed here can generate new, less biased, and less convoluted baselines for typological seriation, in which the ordering of data represents a better-defined (for example, geographical or temporal) trend. An obvious next step is thus to apply the proposed methods to other artifact collections and explore their full potential in archaeology.

In this study, the proposed methods revealed a latitudinal dependency on the shape of shadow surfaces in sundials of the Italian peninsula. This sheds new light on the construction principles of sundials in ancient times. We could also show how this dependence can be used to accurately determine the latitude of a sundial's installation site.

Shape spaces are a current research topic, including how best to deal with challenges particularly common in archaeological investigations, such as incompleteness and weathering of artifacts. The majority of the considered shape spaces exhibit a Riemannian structure and, hence, can be combined with the presented approach. Because of this flexibility, we expect the proposed methods will be of great importance in a wide range of application scenarios.

Finally, we demonstrated in a first practical example how to control for confounding effects with our novel method.

# Chapter 8

# Geometric Deep Learning in Neuroscience

Deep learning approaches have been successfully applied to various tasks, which has led to an explosive increase in research for more refined architectures and new applications. To this end, most of the focus has been on data from Euclidean spaces [6, 151], but more and more applications to data from non-Euclidean spaces are successfully explored [41] (see also [268] and the references therein). In this chapter, we show an application of deep learning to cognitive score prediction from SPD-valued human connectome data. With the help of a sample selection approach in SPD space, we improve over state-of-the-art methods. We thereby find further evidence that deep learning approaches benefit when they take the underlying geometry into account.

## 8.1 IQ Prediction from Brain Connectomes

### 8.1.1 Introduction

Understanding how the anatomy and dynamics of the brain influence cognitive scores such as IQ plays a vital role in understanding the working principles of the human brain. Cognitive scores are indicators of intellectual capacity. They were found to be strongly connected to social factors: While a high correlation between intelligence scores measured in childhood and educational success was observed in [61, 71], they were also linked to health and mortality [23, 103]. Motivated by this fact, many studies have investigated how far intelligence quotients can be predicted from the structure of the brain. It was found, for example, that cerebral volume positively correlates with cognitive ability [164, 204]. On a finer scale, activity and global connectivity of brain parts, especially of the lateral prefrontal cortex, are linked to IQ [59, 60, 107, 256].

Against this background, recent works have explored the possibility to predict cognitive ability scores from functional brain connectomes [68, 74, 76, 117, 133, 179]. Conventionally, connectomes are obtained from resting-state MRI and characterize the brain's network structure; they are graphs whose nodes represent regions of interest (ROIs) in the brain and whose edges correspond to correlations in activity between them [230]. To achieve better generalizability across contexts and populations, Shen et al. proposed a data-driven protocol in [222] for connectome-based predictive

modeling of brain-behavior relationships, which uses cross-validation to train a linear regression model. Building upon it, the authors of [74] improved on the results by evaluating negative and positive correlations of brain regions separately. Additionally, they performed their analysis on both neurotypical subjects and subjects with Autism Spectrum Disorder (ASD) to investigate how neural correlates of intelligence scores are altered by atypical neurodevelopmental disorders.

Although these works achieved significant success, they mainly relied on classical machine learning approaches, which do not incorporate the *graph structure* of the connectomes; therefore, the local and global topological properties of the connectomes are not leveraged. In [117], graph neural networks (GNNs) [257] were probed for the first time for IQ prediction. GNNs are deep neural networks that operate layer-wise on graphs via (graph) convolutions; their study constitutes a subfield of *geometric deep learning*, where learning is customized to non-Euclidean spaces. GNNs have already led to significant increases in performance over existing methods in many fields. For example, they have been successfully applied to classification tasks on networks [150, 199], image segmentation [198], feature matching [212], few-shot learning [98, 147], and various graph mining tasks [214, 267, 269]. A recent review [31] on GNNs examined a variety of graph-based architecture tailored for brain connectivity classification, integration, superresolution, and synthesis across time and modalities. However, no reviewed method was designed for brain graph regression for cognitive score prediction.

Here, we propose the first GNN architecture, namely *RegGNN*, that is specialized in regressing brain connectomes to a target cognitive score. Our GNN utilizes graph convolutional layers to map input connectomes onto their corresponding cognitive scores, thereby allowing the extraction of the learned weights to identify the brain connectivities between anatomical regions that fingerprint the target score.

To improve the performance of the GNN, we also propose a novel *learning-based sample selection* method. It is independent of RegGNN and can be used with any architecture or regression learner. The method identifies training samples with the highest predictive power (those most likely to predict unseen test subjects with the lowest error); only these are then used for training. Through this, we eliminate the samples that do not increase—or even decrease—the prediction success of the model and reduce the computational resources needed for training the GNN.

Within our sample selection method, we make use of the fact that the (weighted) adjacency matrix of a functional brain connectome, when modeled as a correlation matrix, is symmetric positive semi-definite; and becomes symmetric positive definite after a simple regularization step [72, 254, 264]. Using the Log-Euclidean metric, we obtain a *natural* notion of distance between two connectomes and tangent matrices that encode the paths realizing this distance.

We summarize the main contributions of our work as follows:

1. We introduce a novel, learning-based sample selection method for graph neural networks that helps to increase accuracy when predicting cognitive scores from connectomes.

2. We propose novel similarity measures between brain connectomes by combining notions from Riemannian geometry and topology of graphs. These measures

can be used in other applications whenever we deal with objects that can be interpreted as elements of Riemannian manifolds.

3. We design a pipeline, consisting of RegGNN with sample selection, which outperforms state-of-the-art models in predicting full-scale intelligence and verbal intelligence quotients from functional brain connectomes in an autism spectrum disorder cohort and achieves a competitive performance in a neurotypical cohort.

### 8.1.2 Methods

In this section, we detail the architecture of RegGNN. Furthermore, we introduce our proposed sample selection method and show how we incorporate it into the training process of the GNN.

**Preliminaries**

Apart from the Log-Euclidean structure from Example 1.7, SPD($d$) can (amongst others) also be endowed with the affine-invariant metric [167, 187]. Both structures have been applied successfully to connectomes for classification [72, 261], regression [254], fingerprint extraction [1], and statistical analysis [264]. We choose the Log-Euclidean metric because it leads to comparatively efficient algorithms.

Remember that tangent vectors of SPD($d$) are symmetric matrices. To avoid later confusion, we call them "tangent matrices" instead of "tangent vectors" in this section. We will represent the "difference" between two elements $P, Q \in$ SPD($d$) by $\text{Log}_P(Q)$. This is motivated by the fact that the latter points in the direction of $Q$, that is, it is parallel to the geodesic from $P$ to $Q$ at $P$, and its norm is equal to the distance between $P$ and $Q$. To compare the vectors, we must parallel translate them to the same tangent space. As discussed in Example 1.7, parallel transport with the Log-Euclidean structure does not depend on the path so that we can bring all vectors to $T_I$SPD($d$). We illustrate the SPD cone and parallel translation of vectors to $I$ in Figure 8.1.

We will also use three basic topological centrality measures for an undirected[1] graph $G$: degree centrality, eigenvector centrality, and closeness centrality; their definitions are given in Appendix A.6. They all measure, in different ways, how important a node is to a graph.

**RegGNN**

In the following, we denote the number of ROIs by $d$. Since adjacency matrices of connectomes are $d$-by-$d$ correlation matrices $C$, they can have zero (but no negative) eigenvalues. Therefore, we can easily regularize them to being symmetric positive definite by adding a small multiple of the identity matrix $I$, that is,

$$P := C + \varepsilon I \tag{8.1}$$

---

[1]Of course, the centrality measures can also be defined for directed graphs, but we do not need this here.
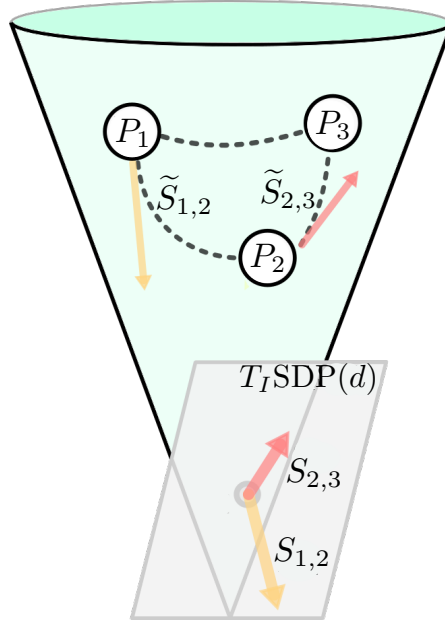
Figure 8.1: Illustration of geodesics and parallel transport of tangent matrices on the SPD cone. The dashed lines are the geodesics between the matrices $P_1, P_2, P_3 \in$ SPD$(d)$. The tangent matrices $\widetilde{S}_{1,2} := \mathrm{Log}_{P_1}(P_2)$ and $\widetilde{S}_{2,3} := \mathrm{Log}_{P_2}(P_3)$ are the yellow and red arrow, respectively; their parallel translations to the tangent space $T_I$SPD$(d)$ at the identity matrix $I$ are $S_{1,2}$ and $S_{2,3}$.

for some small $\varepsilon > 0$; see [72] or [254].

Our GNN for cognitive score prediction from connectomes, *RegGNN*, consists of two graph convolution layers, with dropout after the first layer and a downstream fully connected layer. It receives a connectome's regularized adjacency matrix $P$ and predicts the corresponding IQ score by applying graph convolutions. We use the graph convolution layers to reduce the node features from a $d$-dimensional embedding to one dimension. The obtained embedding then passes through the fully connected (linear) layer that produces the continuous scalar output. A visualization of the architecture is on the bottom left of Figure 8.2.

In the literature, various implementations of graph convolutions can be found. They mainly differ in the propagation rule. Let $H^{(i)} \in \mathbb{R}^{d,d_i}$ denote the *activation matrix at the i-th layer* for $i = 0, 1, 2$, where $d_0 := d$ and $d_2 := 1$. (The integer $d_1$ can be chosen appropriately; we use a suitable number near $d/2$.) It is propagated to the next layer according to the general rule $H^{(i+1)} = g_i(H^{(i)}, P)$ with functions $g_i : \mathbb{R}^{d,d_i} \times$ SPD$(d) \to \mathbb{R}^{d,d_{i+1}}$ for $i = 0, 1$. As initialization we choose $H^{(0)} := I$. We use the graph convolutions proposed by Kipf and Welling in [150]: Set $\widetilde{P} := P + I$, let $\widetilde{D}$ be the diagonal degree matrix of $\widetilde{P}$, and define $\mathbb{R}^{n,m} \ni A \mapsto \mathrm{ReLU}(A) := [\max(0, A_{ij})]$ element-wise. Then,

$$H^{(i+1)} = g_i(H^{(i)}, P) := \mathrm{ReLU}(\widetilde{D}^{-\frac{1}{2}} \widetilde{P} \widetilde{D}^{-\frac{1}{2}} H^{(i)} W^{(i)}), \quad i = 0, 1$$

where each $W^{(i)} \in \mathbb{R}^{d_i, d_{i+1}}$ is the to-be-learned weight matrix of the layer.
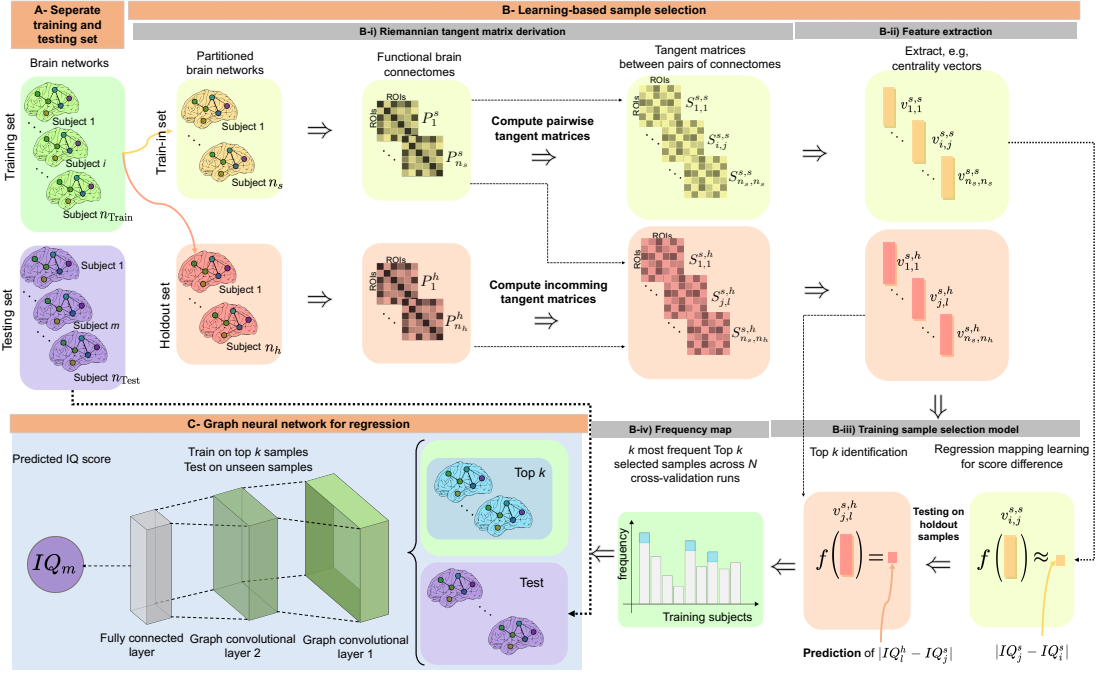
Figure 8.2: *Illustration of the proposed sample selection strategy to train our regression graph neural network RegGNN.* **A)** We split the data into training (green) and testing (violet) sets. **B-i)** After the training set is divided into a train-in (yellow) and a holdout (red) set, we extract tangent matrices for geodesics connecting elements from the train-in set (yellow) and tangent matrices encoding geodesics from elements of the train-in to elements of the holdout group (red). **B-ii)** The information from the tangent matrices is compressed into vectors through topological feature extraction. **B-iii)** With linear regression, we train a mapping *f* on the train-in-to-train-in feature vectors (yellow) to learn differences in target score and record for each element *j* of the train-in group the *k* elements from the holdout group for whom the predicted difference in target score to *j* was smallest. **B-iv)** For each sample in the holdout set we count how often it was among the top *k* predictors for a sample from the train-in group. **C)** After repeating B) in an *N*-fold cross validation manner, the *k* samples (blue) with the highest accumulated top-*k* frequency are selected. After negative correlations have been set to zero, the graph neural network is trained (only) on them. Finally, the testing set is used to evaluate the overall performance.

When training RegGNN—but not the sample selection in the next section—we set all negative eigenvalues to zero, as positive correlations have been shown to be more important in brain network analysis [94]. Indeed, in our experiments, the results improved when negative correlations were ignored.

**Learning-based Sample Selection**

We now introduce our *learning-based* sample selection strategy. The underlying idea is the following. Imagine the (rather extreme) case that in the SPD space, the adjacency

matrices of the connectomes are clustered in $k$ tight groups (possibly with outliers) according to their cognitive scores. Then, training a GNN on $k$ representatives (one from each cluster) should yield good results; it should even perform better than a GNN trained on the whole data set because it was not "distracted" by outliers during training. Ideally, as representatives, we would choose the $k$ most central samples, that is, those with the smallest average *difference in cognitive score* to the other samples in each group. But since we want to predict cognitive scores from connectomes, we do not know the differences in cognitive scores beforehand. On the other hand, existing studies validated the relationship between brain connectivity patterns and brain behavior and cognition so that we can use the former as a replacement for the latter. We elaborate on this a little more. Recent papers [68, 74, 76, 117, 133, 179, 222] have shown that the cognitive ability of a person can be predicted quite accurately from the human connectome, indicating that it is encoded in its connectivity to a measurable degree. Such predictive power was elusive if similar data inputs (here brain connectomes) could not be mapped to similar outputs (here cognitive scores). Consequently, we assume that connectomes that are close in SPD space are correlated in cognition, whereas brain connectomes that are further apart might elicit different cognitive scores. Such a hypothesis might seem somewhat simple as many other factors contribute to molding and predicting brain cognition, for example, genetics and epigenetics [70, 101, 203]. However, such factors remain out of the scope of this study. Therefore, our idea is to use the *differences* between the connectomes to learn the *differences* between the target scores. This way, we can identify the "representatives" of clusters with similar cognitive scores. Our experiments below show that this idea—to represent predicted local aggregations of data by (few) representatives and training only with them—generalizes well to real data, even when there is no discrete set of well-defined clusters. We believe that outliers tend to be filtered out, which improves the results.

Implementing the idea, we represent differences between connectomes by tangent matrices and assume that the difference in IQ between two subjects depends linearly on (notions deduced from) the tangent matrix $\text{Log}_P(Q)$ that encodes the "difference" between the corresponding connectomes $P, Q \in \text{SPD}(d)$. This model is flexible, but at the same time allows for fast computations. Our sample selection method *learns* this linear map, which we call $f$ in the following, via regression and uses it to identify the $k$ samples with the lowest predicted average difference in target score to all other samples. As motivated above, we assume that they are representative of the whole set but do not contain (most of) the outliers that hinder successful training of the GNN. The structure and terminology of our method are inspired by the work [81].

The sample selection method consists of four steps and is visualized in part B of Figure 8.2. Given a connectome data set, they are repeated in a nested $N$-fold cross-validation manner to make our selection of samples more robust. In cross-validation, we repeatedly split the data set into two groups: a training subset which we call *train-in group*, and a validation subset which we call *holdout group*. The splits are done such that each sample from the training set is exactly $N-1$ times in the train-in group. We denote the (constant) sizes of the train-in and the holdout sets by $n_s$ and $n_h$, respectively. The steps are the following.

*i) Riemannian tangent matrix derivation.* For each pair of regularized connectomes

$P_i^s$, $P_j^s \in \mathrm{SPD}(d)$ in the train-in group, we compute the tangent matrix

$$\widetilde{S}_{i,j}^{s,s} := \mathrm{Log}_{P_i^s}(P_j^s) \in T_{P_i^s}\mathrm{SPD}(d)$$

that encodes the geodesic between them and parallel translate it to $T_I\mathrm{SPD}(d)$; the resulting symmetric $d$-by-$d$ matrix is denoted by $S_{i,j}^{s,s}$. As a result, we obtain a set of $n_s(n_s-1)/2$ tangent matrices in $T_I\mathrm{SPD}(d)$ that represent the pairwise differences between the connectomes from the train-in group. (Since $S_{i,j}^{s,s} = -S_{j,i}^{s,s}$, we compute only one matrix for each pair of connectomes.) Analogously, we get a tangent matrix $S_{j,l}^{s,h} \in T_I\mathrm{SPD}(d)$ for each pair with one sample $P_i^s$ from the train-in and another sample $P_l^h$ from the holdout group; this results in another set consisting of $n_s\,n_h$ tangent matrices. The latter are the *outgoing* tangent matrices *from the train-in into the holdout set*.

*ii) Topological feature extraction from tangent matrices.* The tangent matrices are still rather high-dimensional, which leads to long computation times. Therefore, we suggest extracting topological features to encode the information in a more compact form. We select degree, closeness, eigenvector centrality, and combinations of them as our candidates for feature extraction. Remember that a tangent matrix represents the "difference" between two connectomes. The above features thus encode information on linearized *changes* in node connectivity. To the best of our knowledge, this is the first time that these notions were used in conjunction. As a result, from all in-group tangent matrices $S_{i,j}^{s,s}$ as well as outgoing tangent matrices $S_{j,l}^{s,h}$ we obtain feature vectors $v_{i,j}^{s,s}$ and $v_{i,j}^{s,h}$, respectively.

*iii) Learning a linear regression mapping for predictive sample selection.* We *learn* the linear map $f$ via regression by training to map the vectors $v_{i,j}^{s,s}$ corresponding to samples $i$ and $j$ from the train-in group to the *absolute difference in target score* $|IQ_j^s - IQ_i^s|$ between them.

We then apply the learned linear regression mapping $f$ to the vectors $v_{j,l}^{s,h}$ to predict the differences in target score between all samples $j = 1, \ldots, n_s$ from the train-in and samples $l = 1, \ldots, n_h$ from the holdout group.

*iv) Frequency map.* We record *for each* holdout sample $P_l^h$ the $k$ subjects from the train-in group with the smallest predicted difference under $f$ and increment a frequency map (a counter) that is initialized at the start of the sample selection process. The frequency value of a subject is then the number of times it was one of the top $k$ predictive samples. These frequencies give an approximated ranking, whereby the top samples' features predict the difference in target score to other samples best.

After the cross-validation, we extract the top $k$ samples[2] with the highest cumulative frequencies. We expect these samples to have the highest representative power as they consistently predicted differences in various holdout groups with low error.

**Training Process**

In the following, we explain how we integrate the sample selection method into the training process of RegGNN. The whole pipeline is shown in Figure 8.2.

---

[2]Note that we could pick a different number here. We leave exploring possible other choices for future work.

Given the data set of connectomes, our proposed training pipeline consists of the following.

**A- Training-test split:** First, we split the data set into a training and a test set. The test set is used *only* for the final evaluation of RegGNN.

**B- Learning-based sample selection:** We select the top $k$ samples with the highest representative power from the training set by applying the sample selection method from Section 8.1.2.

**C- RegGNN architecture for regression:** Finally, we train RegGNN on the top $k$ samples using cross-validation to evaluate model generalizability against perturbations of training and testing data distributions. The final testing is done on the unseen test set.

### 8.1.3  Data and Methodology

We use the pipeline from Section 8.1.2 to predict the full-scale intelligence quotient (FIQ) and the verbal intelligence quotient (VIQ) from brain connectomes for both neurotypical (NT) subjects as well as subjects with autism spectrum disorder (ASD). In the following, we summarize these experiments.

**Dataset**

We use samples from the Autism Brain Imaging Data Exchange (ABIDE) Preprocessed dataset [66] for our experiments. It contains data from 16 imaging sites, preprocessed by five different teams using four pipelines: the Connectome Computation System (CCS), the Configurable Pipeline for the Analysis of Connectomes (CPAC), the Data Processing Assistant for rs-fMRI (DPARSF), and the NeuroImaging Analysis Kit. The preprocessed data sets are available online[3]. To account for possible biases due to differences in sites, we use randomly sampled subsets of the available data for both cohorts; the same sets were also used in [74]. The NT cohort consists of 226 subjects (with mean age = (15 ± 3.6)), while the ASD cohort is made up of 202 individuals (with mean age = (15.4 ± 3.8)). FIQ and VIQ scores in the NT cohort have means 111.573 ± 12.056 and 112.787 ± 12.018, whereas FIQ and VIQ scores in the ASD cohort have means 106.102 ± 15.045 and 103.005 ± 16.874, respectively. The brain connectomes were obtained from resting-state functional magnetic resonance imaging using the parcellation from [246] with 116 ROIs. The functional connectomes are represented by 116-by-116 matrices, whose entry in row $i$ and column $j$ is the Pearson correlation between the average rs-fMRI signal measured in ROI $i$ and ROI $j$.

**Software**

All experiments are done in Python 3.7.10. We use Scikit-learn 0.24.2 [183] for machine learning models and PyTorch Geometric 1.6.3 [86] for graph neural network implementations. We rely on the Morphomatics package [10] for computations in $SPD(d)$ with the Log-Euclidean metric and extract the graph topological features from the tangent matrices with the functions provided by NetworkX [112].

---

[3]http://preprocessed-connectomes-project.org/abide/

| Method | Mean MAE ± std (min, max) | Mean RMSE ± std (min, max) |
|---|---|---|
| **NT (FIQ)** | | |
| CPM | **9.672** | **12.440** |
| CPM (a) | 11.383 ± 4.806 (9.589, 28.675) | 13.972 ± 4.986 (<u>12.075</u>, 31.911) |
| CPM (g) | 10.546 ± 1.407 (9.729, 14.989) | 13.358 ± 1.831 (12.393, 19.358) |
| CPM (tm) | 10.245 ± 2.086 (9.523, 17.745) | 12.959 ± 2.515 (12.096, 22.009) |
| CPM (dc) | 10.577 ± 2.487 (9.561, 19.394) | 13.525 ± 2.751 (12.131, 23.210) |
| CPM (ec) | 10.537 ± 2.131 (9.595, 18.159) | 13.456 ± 2.393 (12.366, 21.976) |
| CPM (cc) | <u>10.074</u> ± 1.307 (**9.370**, 14.741) | <u>12.931</u> ± 1.669 (**12.065**, 18.895) |
| CPM (cnu) | 10.895 ± 4.402 (<u>9.463</u>, 26.749) | 13.775 ± 4.633 (12.194, 30.465) |
| CPM (cns) | 11.262 ± 4.585 (9.623, 27.771) | 14.156 ± 4.822 (12.406, 31.521) |
| PNA-S | 17.217 | 21.008 |
| PNA-S (a) | <u>12.267</u> ± 0.948 (11.002, 14.480) | 15.338 ± 1.079 (13.914, 18.110) |
| PNA-S (g) | 12.305 ± 1.246 (10.814, 15.267) | 15.474 ± 1.357 (13.897, 18.777) |
| PNA-S (tm) | **11.537** ± **0.727** (**10.639**, 12.890) | **14.466** ± **0.814** (**13.322**, 15.853) |
| PNA-S (dc) | 12.211 ± 1.119 (10.751, 14.012) | 15.389 ± 1.289 (13.619, 17.191) |
| PNA-S (ec) | 12.124 ± 1.307 (10.956, 16.118) | <u>15.270</u> ± 1.466 (13.865, 19.509) |
| PNA-S (cc) | 12.166 ± 0.898 (<u>10.727</u>, 13.895) | 15.413 ± 1.098 (<u>13.615</u>, 17.309) |
| PNA-S (cnu) | 12.608 ± 1.204 (10.935, 15.944) | 15.799 ± 1.378 (13.827, 19.684) |
| PNA-S (cns) | 12.509 ± 0.963 (11.152, 14.747) | 15.737 ± 1.172 (13.960, 18.528) |
| PNA-V | 20.109 | 25.113 |
| PNA-V (a) | 15.979 ± 3.730 (11.607, 26.483) | 19.617 ± 4.111 (14.575, 30.983) |
| PNA-V (g) | <u>14.570</u> ± 4.115 (**11.014**, 26.962) | <u>18.020</u> ± 4.329 (**13.984**, 30.811) |
| PNA-V (tm) | 15.450 ± 4.313 (<u>11.259</u>, 26.906) | 19.030 ± 4.804 (<u>14.273</u>, 32.059) |
| PNA-V (dc) | **14.300** ± **2.111** (11.475, 17.750) | **17.900** ± **2.588** (14.313, 22.516) |
| PNA-V (ec) | 15.385 ± 3.141 (11.504, 21.710) | 19.294 ± 3.905 (14.779, 27.041) |
| PNA-V (cc) | 15.288 ± 3.023 (11.709, 20.592) | 18.845 ± 3.378 (14.770, 24.803) |
| PNA-V (cnu) | 16.800 ± 5.053 (12.352, 28.022) | 20.470 ± 5.736 (15.180, 32.794) |
| PNA-V (cns) | 15.910 ± 5.976 (12.507, 36.704) | 19.671 ± 6.875 (15.478, 43.397) |
| RegGNN | 9.768 | **12.270** |
| RegGNN (a) | 10.360 ± 1.090 (9.624, 14.027) | 12.997 ± 1.278 (12.158, 17.335) |
| RegGNN (g) | 10.032 ± 0.330 (9.576, 10.790) | 12.563 ± 0.310 (12.148, 13.311) |
| RegGNN (tm) | 9.820 ± 0.512 (9.525, 11.613) | <u>12.378</u> ± 0.528 (12.116, 14.259) |
| RegGNN (dc) | **9.714** ± **0.332** (9.485, 10.634) | 12.531 ± 0.458 (**12.064**, 13.716) |
| RegGNN (ec) | 9.815 ± 0.245 (9.469, 10.334) | 12.609 ± 0.334 (12.171, 13.286) |
| RegGNN (cc) | 9.777 ± 0.391 (9.461, 10.757) | 12.516 ± 0.425 (12.121, 13.523) |
| RegGNN (cnu) | 9.745 ± 0.451 (**9.438**, 11.018) | 12.482 ± 0.473 (<u>12.085</u>, 13.758) |
| RegGNN (cns) | <u>9.716</u> ± 0.292 (<u>9.452</u>, 10.474) | 12.466 ± 0.217 (12.207, 12.982) |

Table 8.1: Comparison of methods for NT cohort.

**Parameter settings**

We train our network using Adam optimizer [149] for 100 epochs with a learning rate of 0.001 and weight decay at 0.0005 based on our empirical observations. The dropout rate after the first graph convolutional layer is set to 0.1. Furthermore, we set $d_1 := 64$ in RegGNN. We regularize the adjacency matrices with $\varepsilon = 10^{-10}$ in (8.1). Finally, to explore the parameter space for the number of selected training samples, we vary $k$ between 2 and 15.

| Method | Mean MAE ± std (min, max) | Mean RMSE ± std (min, max) |
|---|---|---|
| **NT (VIQ)** | | |
| CPM | **9.517** | **12.049** |
| CPM (a) | 11.035 ± 4.970 (9.481, 28.895) | 13.728 ± 5.243 (12.031, 32.518) |
| CPM (g) | 10.363 ± 1.682 (9.549, 16.146) | 13.253 ± 2.126 (12.236, 20.633) |
| CPM (tm) | 10.194 ± 2.316 (9.503, 18.543) | 12.922 ± 2.928 (12.035, 23.477) |
| CPM (dc) | 10.065 ± 1.760 (<u>9.391</u>, 16.386) | 12.767 ± 2.386 (<u>11.877</u>, 21.332) |
| CPM (ec) | <u>9.942</u> ± 1.768 (**9.323**, 16.275) | <u>12.614</u> ± 2.209 (**11.828**, 20.532) |
| CPM (cc) | 10.233 ± 1.264 (9.514, 14.667) | 12.911 ± 1.595 (12.035, 18.505) |
| CPM (cnu) | 10.544 ± 3.585 (9.480, 23.469) | 13.256 ± 4.058 (12.022, 27.885) |
| CPM (cns) | 10.583 ± 3.560 (9.526, 23.416) | 13.326 ± 3.891 (12.108, 27.343) |
| PNA-S | 12.838 | 16.130 |
| PNA-S (a) | <u>11.846</u> ± 0.942 (10.795, 13.764) | <u>15.057</u> ± 1.099 (13.824, 17.279) |
| PNA-S (g) | 12.439 ± 1.183 (10.913, 14.774) | 15.648 ± 1.322 (13.895, 18.469) |
| PNA-S (tm) | 12.489 ± 3.099 (10.759, 23.367) | 15.884 ± 4.323 (13.983, 31.268) |
| PNA-S (dc) | **11.694** ± **0.814** (**10.302**, 12.901) | **14.707** ± **0.904** (**13.126**, 16.115) |
| PNA-S (ec) | 12.091 ± 1.006 (<u>10.543</u>, 14.032) | 15.122 ± 1.061 (<u>13.436</u>, 17.200) |
| PNA-S (cc) | 13.074 ± 1.693 (11.258, 17.486) | 16.326 ± 1.795 (14.474, 20.701) |
| PNA-S (cnu) | 12.682 ± 1.308 (10.857, 15.196) | 15.950 ± 1.536 (13.795, 18.775) |
| PNA-S (cns) | 12.014 ± 0.859 (10.545, 14.031) | 15.141 ± 0.944 (13.503, 17.302) |
| PNA-V | 14.695 | 18.903 |
| PNA-V (a) | **14.107** ± **2.081** (11.923, 19.482) | **17.696** ± **2.693** (14.832, 25.550) |
| PNA-V (g) | 14.983 ± 5.211 (11.639, 32.479) | 18.681 ± 6.224 (14.715, 39.771) |
| PNA-V (tm) | 15.489 ± 4.211 (11.717, 27.980) | 19.157 ± 4.775 (14.624, 33.183) |
| PNA-V (dc) | <u>14.332</u> ± 2.931 (**11.188**, 20.545) | <u>18.170</u> ± 3.917 (**14.284**, 26.511) |
| PNA-V (ec) | 14.924 ± 3.424 (11.360, 21.703) | 18.539 ± 4.185 (<u>14.392</u>, 27.289) |
| PNA-V (cc) | 16.049 ± 4.185 (<u>11.237</u>, 24.616) | 20.035 ± 4.966 (14.408, 29.373) |
| PNA-V (cnu) | 14.805 ± 2.587 (11.890, 21.107) | 18.344 ± 2.883 (15.171, 25.473) |
| PNA-V (cns) | 14.915 ± 2.845 (11.898, 22.045) | 18.674 ± 3.612 (14.895, 28.544) |
| RegGNN | 10.195 | 13.044 |
| RegGNN (a) | 9.587 ± 0.206 (9.477, 10.311) | 12.223 ± 0.299 (12.054, 13.261) |
| RegGNN (g) | 9.779 ± 0.126 (9.551, 9.964) | 12.530 ± 0.199 (12.190, 12.803) |
| RegGNN (tm) | **9.514** ± **0.036** (<u>9.471</u>, 9.594) | **12.041** ± **0.035** (**12.004**, 12.140) |
| RegGNN (dc) | 9.639 ± 0.161 (9.468, 10.000) | 12.213 ± 0.235 (<u>12.008</u>, 12.707) |
| RegGNN (ec) | 9.599 ± 0.243 (**9.453**, 10.189) | 12.199 ± 0.274 (12.020, 12.846) |
| RegGNN (cc) | 9.711 ± 0.193 (9.499, 10.161) | 12.313 ± 0.284 (12.023, 13.062) |
| RegGNN (cnu) | <u>9.521</u> ± 0.042 (9.473, 9.651) | <u>12.134</u> ± 0.050 (12.050, 12.266) |
| RegGNN (cns) | 9.581 ± 0.153 (9.497, 10.109) | 12.236 ± 0.199 (12.102, 12.907) |

Table 8.1: Comparison of methods for NT cohort (continued). The best-performing method for each architecture is bold while the second-best is underlined. The mean ± standard deviation as well as minima and maxima over $k = 2, \ldots, 15$ (in brackets) are given. The overall best-performing method according to mean error and the best sample selection performance are indicated in blue. Abbreviations are: (a) absolute Euclidean distance, (g) geometric Log-Euclidean distance, (tm) full tangent matrix, (dc) degree centrality, (ec) eigenvector centrality, (cc) closeness centrality, (cnu) concatenation unscaled, (cns) concatenation scaled.

**Evaluation and comparison methods**

To test the generalizability and robustness of our method, we apply 3-fold cross-validation on the NT and ASD cohorts separately for both FIQ and VIQ prediction.

| Method | Mean MAE ± std (min, max) | Mean RMSE ± std (min, max) |
|---|---|---|
| **ASD (FIQ)** | | |
| CPM | <u>12.533</u> | 15.965 |
| CPM (a) | 13.673 ± 3.748 (12.082, 26.912) | 16.783 ± 4.048 (<u>15.043</u>, 31.074) |
| CPM (g) | 13.267 ± 1.367 (12.166, 17.792) | 16.728 ± 1.672 (15.433, 22.297) |
| CPM (tm) | **12.464 ± 1.090** (12.093, 16.386) | **15.522 ± 1.336** (15.084, 20.333) |
| CPM (dc) | 12.644 ± 1.498 (<u>12.082</u>, 18.027) | <u>15.685</u> ± 1.723 (15.116, 21.887) |
| CPM (ec) | 12.825 ± 1.400 (**12.074**, 17.527) | 15.922 ± 1.654 (**15.017**, 21.252) |
| CPM (cc) | 13.310 ± 1.331 (12.430, 17.821) | 16.703 ± 1.654 (15.618, 22.335) |
| CPM (cnu) | 12.934 ± 2.460 (12.094, 21.790) | 16.039 ± 2.817 (15.086, 26.182) |
| CPM (cns) | 12.742 ± 1.527 (12.166, 18.129) | 15.796 ± 1.838 (15.077, 22.298) |
| PNA-S | 17.162 | 22.023 |
| PNA-S (a) | 15.260 ± 0.865 (13.947, 17.049) | 18.768 ± 0.940 (17.527, 20.665) |
| PNA-S (g) | 15.025 ± 1.321 (13.207, 17.762) | 18.844 ± 1.641 (16.642, 22.009) |
| PNA-S (tm) | 14.781 ± 1.182 (**13.168**, 18.488) | 18.630 ± 1.496 (16.522, 23.426) |
| PNA-S (dc) | 14.248 ± 0.533 (13.595, 15.651) | 17.715 ± 0.723 (16.654, 19.644) |
| PNA-S (ec) | 14.758 ± 1.277 (<u>13.207</u>, 18.534) | 18.263 ± 1.608 (**16.440**, 23.134) |
| PNA-S (cc) | 14.558 ± 0.666 (13.824, 16.586) | 18.126 ± 0.894 (17.296, 20.806) |
| PNA-S (cnu) | <u>14.038</u> ± 0.434 (13.353, 14.915) | **17.438 ± 0.580** (<u>16.469</u>, 18.642) |
| PNA-S (cns) | **14.023 ± 0.401** (13.465, 14.782) | <u>17.620</u> ± 0.550 (16.853, 18.605) |
| PNA-V | <u>15.671</u> | 20.085 |
| PNA-V (a) | 17.386 ± 2.428 (14.453, 23.263) | 21.234 ± 2.922 (18.006, 28.505) |
| PNA-V (g) | 17.611 ± 2.228 (14.770, 23.572) | 22.035 ± 3.023 (18.422, 30.646) |
| PNA-V (tm) | 18.458 ± 5.559 (**13.770**, 32.776) | 23.112 ± 6.498 (<u>17.326</u>, 38.914) |
| PNA-V (dc) | 16.358 ± 2.793 (14.045, 24.543) | 20.325 ± 3.344 (17.355, 29.689) |
| PNA-V (ec) | 17.085 ± 4.124 (<u>13.778</u>, 28.115) | 21.095 ± 4.885 (**17.101**, 33.584) |
| PNA-V (cc) | 16.172 ± 1.181 (14.541, 18.242) | <u>19.942</u> ± 1.346 (17.970, 22.102) |
| PNA-V (cnu) | 19.552 ± 6.530 (13.840, 33.875) | 24.470 ± 8.514 (17.399, 44.199) |
| PNA-V (cns) | **15.601 ± 2.041** (13.820, 20.748) | **19.406 ± 2.359** (17.484, 25.427) |
| RegGNN | 12.564 | 15.624 |
| RegGNN (a) | 12.588 ± 0.541 (12.170, 13.825) | 15.628 ± 0.598 (15.102, 17.006) |
| RegGNN (g) | 12.977 ± 1.073 (12.137, 16.292) | 16.194 ± 1.358 (15.146, 20.300) |
| RegGNN (tm) | **12.148 ± 0.072** (12.078, 12.379) | **15.130 ± 0.067** (15.074, 15.355) |
| RegGNN (dc) | <u>12.247</u> ± 0.167 (<u>12.073</u>, 12.667) | <u>15.214</u> ± 0.178 (<u>15.058</u>, 15.693) |
| RegGNN (ec) | 12.361 ± 0.246 (12.106, 12.905) | 15.320 ± 0.265 (**15.040**, 15.895) |
| RegGNN (cc) | 12.720 ± 0.466 (12.134, 14.008) | 15.756 ± 0.560 (15.074, 17.321) |
| RegGNN (cnu) | 12.301 ± 0.226 (**12.064**, 12.804) | 15.331 ± 0.250 (15.120, 15.881) |
| RegGNN (cns) | 12.300 ± 0.159 (12.132, 12.763) | 15.306 ± 0.201 (15.114, 15.894) |

Table 8.2: Comparison of methods for ASD cohort.

We report the mean absolute error (MAE) and the root mean squared error (RMSE) for all methods. We additionally give the mean, standard deviation, minima, and maxima over all values of *k* for the sample selection methods.

As benchmarks, we chose state-of-the-art deep learning and machine learning methods. The first baseline is *connectome-based predictive modeling* (CPM) by Shen et al. [222], which was specifically designed for behavioral score prediction on brain connectomes; the second being *principal neighborhood aggregation* (PNA) by Corso et al. [65], which outperformed common GNNs on both artificial and real-world benchmark regression tasks (but has not been applied to brain connectomes yet). PNA

| Method | Mean MAE ± std (min, max) | Mean RMSE ± std (min, max) |
|---|---|---|
| **ASD (VIQ)** | | |
| CPM | <u>14.171</u> | 18.834 |
| CPM (a) | 15.006 ± 5.390 (12.920, 34.315) | 19.034 ± 5.458 (16.887, 38.586) |
| CPM (g) | 14.336 ± 3.490 (**12.669**, 26.704) | <u>18.646</u> ± 3.660 (<u>16.845</u>, 31.560) |
| CPM (tm) | 15.496 ± 2.494 (13.694, 21.575) | 20.191 ± 3.140 (17.994, 28.150) |
| CPM (dc) | 14.449 ± 1.882 (13.289, 19.795) | 18.767 ± 2.166 (17.421, 24.648) |
| CPM (ec) | **13.697** ± **1.389** (<u>12.824</u>, 18.258) | **17.985** ± **1.737** (**16.838**, 23.693) |
| CPM (cc) | 14.954 ± 3.088 (13.432, 25.694) | 19.613 ± 3.506 (17.508, 31.570) |
| CPM (cnu) | 14.698 ± 1.565 (13.046, 19.980) | 18.917 ± 1.715 (17.412, 24.674) |
| CPM (cns) | 15.417 ± 2.761 (13.593, 22.869) | 20.041 ± 4.128 (17.648, 32.850) |
| PNA-S | 19.955 | 25.848 |
| PNA-S (a) | 15.993 ± 1.096 (**14.125**, 18.861) | 20.500 ± 1.284 (<u>18.349</u>, 23.726) |
| PNA-S (g) | **15.540** ± **1.176** (14.139, 18.017) | **20.037** ± **1.548** (**18.256**, 23.559) |
| PNA-S (tm) | 16.766 ± 1.330 (<u>15.020</u>, 20.365) | 21.548 ± 1.583 (19.513, 25.735) |
| PNA-S (dc) | 16.820 ± 2.161 (14.332, 21.360) | 21.386 ± 2.303 (18.685, 25.724) |
| PNA-S (ec) | 16.445 ± 1.598 (14.518, 21.308) | 21.150 ± 1.900 (18.760, 26.560) |
| PNA-S (cc) | 16.113 ± 1.458 (14.600, 19.985) | 21.012 ± 1.715 (19.057, 25.231) |
| PNA-S (cnu) | 16.359 ± 1.450 (14.661, 20.737) | 20.880 ± 1.714 (18.905, 26.173) |
| PNA-S (cns) | <u>15.887</u> ± 0.773 (14.621, 17.091) | <u>20.282</u> ± 0.843 (18.785, 21.847) |
| PNA-V | 22.518 | 28.804 |
| PNA-V (a) | 18.145 ± 3.017 (14.837, 26.420) | 22.937 ± 3.373 (19.233, 31.651) |
| PNA-V (g) | **16.906** ± **1.566** (14.975, 21.123) | **21.791** ± **2.210** (19.658, 28.209) |
| PNA-V (tm) | 19.040 ± 3.597 (15.631, 30.284) | 24.162 ± 4.053 (19.972, 36.701) |
| PNA-V (dc) | 19.734 ± 4.936 (15.055, 33.387) | 25.088 ± 6.380 (19.898, 43.769) |
| PNA-V (ec) | 17.837 ± 2.000 (14.973, 21.401) | 22.721 ± 2.341 (19.373, 27.477) |
| PNA-V (cc) | 18.663 ± 4.268 (14.969, 32.436) | 24.016 ± 5.063 (19.679, 40.173) |
| PNA-V (cnu) | <u>17.314</u> ± 2.329 (14.660, 22.610) | <u>22.128</u> ± 2.772 (<u>19.026</u>, 28.786) |
| PNA-V (cns) | 17.714 ± 2.833 (14.697, 24.781) | 22.630 ± 3.748 (**18.958**, 32.707) |
| RegGNN | **13.090** | **17.250** |
| RegGNN (a) | 13.356 ± 0.443 (12.834, 14.402) | <u>17.272</u> ± 0.480 (**16.782**, 18.502) |
| RegGNN (g) | <u>13.316</u> ± 0.545 (**12.687**, 14.700) | 17.474 ± 0.663 (16.877, 19.186) |
| RegGNN (tm) | 14.110 ± 1.237 (13.194, 17.014) | 18.288 ± 1.656 (17.130, 22.128) |
| RegGNN (dc) | 14.516 ± 2.043 (13.369, 19.931) | 18.381 ± 2.261 (17.140, 24.445) |
| RegGNN (ec) | 13.419 ± 0.705 (12.766, 15.061) | 17.526 ± 0.945 (<u>16.815</u>, 19.958) |
| RegGNN (cc) | 14.021 ± 1.001 (13.149, 16.812) | 18.731 ± 1.242 (17.563, 21.916) |
| RegGNN (cnu) | 14.001 ± 0.605 (<u>12.762</u>, 14.885) | 18.084 ± 0.569 (17.022, 18.924) |
| RegGNN (cns) | 14.020 ± 0.440 (12.950, 14.650) | 18.007 ± 0.434 (17.008, 18.571) |

Table 8.2: Comparison of methods for ASD cohort (continued). The best-performing method for each architecture is bold; the second-best is underlined. The mean ± standard deviation and minima and maxima over $k = 2, \ldots, 15$ (in brackets) are given. The overall best-performing method according to mean error and the best sample selection performance are indicated in blue. Abbreviations are: (a) absolute Euclidean distance, (g) geometric Log-Euclidean distance, (tm) full tangent matrix, (dc) degree centrality, (ec) eigenvector centrality, (cc) closeness centrality, (cnu) concatenation unscaled, (cns) concatenation scaled.

comes with principal neighborhood aggregation layers that are defined similarly to graph convolution operations. They are designed to increase the amount of infor-

mation used from local neighborhoods. In our experiments, we inserted PNA layers in our RegGNN architecture. We implemented both a simpler setup with sum aggregation and identity scaling only (denoted by PNA-S), as well as various aggregation (sum, mean, var, and max) and scaling (identity, amplification, and attenuation) methods (denoted by PNA-V) as detailed in [65]. The code of both CPM[4] and PNA[5] is available online.

To assess the effect of the sample selection, we also train each architecture on *all* samples as a baseline.

**Evaluation of the sample selection**

For each architecture, we compare several methods that can be used as difference measures between connectomes in the sample selection to learn the linear mapping *f* (viz., Section 8.1.2 part (*ii*)).

The first class of methods is the proposed one: tangent matrices of SPD(*d*). Thereby, we compare several methods to compress the information that is contained in the tangent matrices. As one option, we train *f* on the vectorized upper triangular part (including the diagonal) of the tangent matrix; this method is denoted by (tm). Since the matrices are symmetric, ignoring the lower part speeds up computations without reducing the amount of information. Further, we use degree centrality, eigenvector centrality, and closeness centrality (see App. A.6), and apply them to the tangent matrices; the resulting methods are denoted by (dc), (ec), and (cc), respectively. (Note that the connectivity of each connectome is not altered.) The mapping *f* is then trained on the resulting centrality vectors. Additionally, we test whether concatenating the above centrality measures into a single vector is even more informative. To this end, we use both an unscaled and a scaled version, denoted by (cnu) and (cns), respectively. The unscaled version is generated by simple concatenation of the three feature vectors. However, as the three centrality measures have different ranges, we additionally test to scale each feature vector first. For this, we use min-max scaling. Remember that min-max scaling of a vector *v* is defined element-wise by

$$\widetilde{v}_i := \frac{v_i - \max(v)}{\max(v) - \min(v)}.$$

Each centrality vector is scaled before concatenation, which then gives a vector with elements in $[0, 1]$ as data for the regression.

We complement these methods with two baselines. To check whether the additional directional information that the tangent matrices contain helps, we also test whether it suffices to train *f* on the Riemannian geometric distances $\mathrm{dist}(P_i^s, P_j^s)$ between the connectomes from the train-in group alone; this method is denoted by (g). To assess whether our results are improved by using the manifold structure of the SPD space at all, we train *f* on the Euclidean distance between the upper triangular parts $\widehat{P}_i^s, \widehat{P}_j^s$ of each pair of connectomes $P_i^s, P_j^s$, that is, on the scalars $\|\widehat{P}_i^s - \widehat{P}_j^s\|_\mathrm{F}$ (F standing for the Frobenius norm); we denote this method by (a).

We report the *p*-value between the best-performing sample selection method MAE and the baseline MAE according to a standard t-test for all architectures.

---

[4]https://github.com/esfinn/cpm_tutorial
[5]https://github.com/lukecavabarrett/pna/

### 8.1.4  Results and Discussion
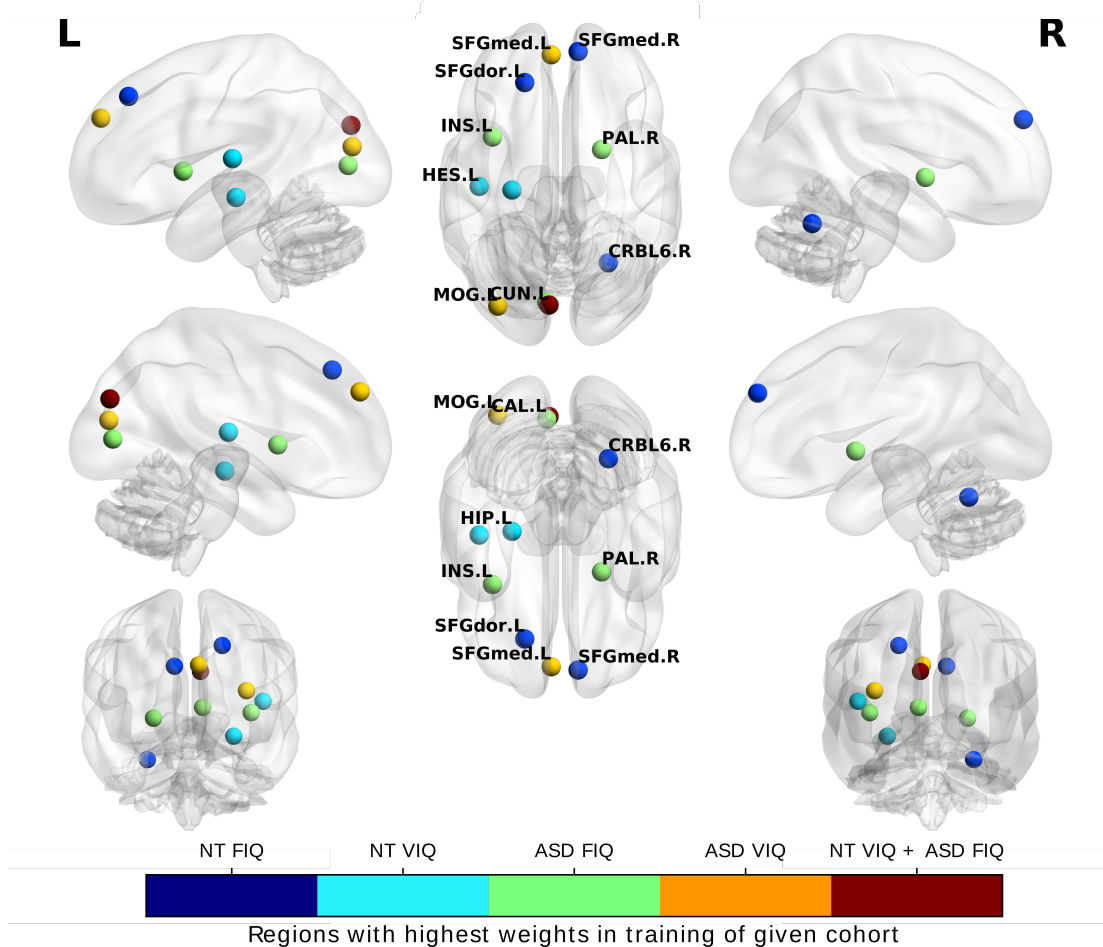


Figure 8.3: The top three most relevant brain regions for each task according to the learned weights extracted from the last layer of the trained RegGNN. The colors indicate the task. NT: neurotypical subjects. ASD: autism spectrum disorder subjects. FIQ: fluid intelligence quotient. VIQ: verbal intelligence quotient.

The results for the NT and ASD cohorts are shown in Tables 8.1 and 8.2, respectively. While the state-of-the-art machine learning model CPM surpasses naive applications of GNNs in the form of PNA, RegGNN, paired with sample selection, outperforms CPM in all tasks according to both MAE and RMSE except for the NT (FIQ) task. Improvements through our method are especially visible in the ASD cohort. Interestingly, we see that the results of all approaches are worse in the ASD cohort compared to the NT cohort. The same was also observed in [74]. We hypothesize that the difficulty of predicting IQ scores in the ASD cohort might be caused by the inter-subject heterogeneity characteristic of ASD [240]. Another factor may be that ASD samples from ABIDE are biased towards high-functioning individuals [66].

We observe further that sample selection improves the results for RegGNN in all tasks except ASD-VIQ, and for CPM in the ASD cohort. For PNA-based architectures, there are drastic improvements in NT-FIQ and ASD-VIQ and incremental improve-

ments in the other tasks. An exception to this is the PNA-V setup for ASD-FIQ, where most models with sample selection perform worse than the one trained on all samples. This might be partly explained by the more complicated structure of PNA with various aggregation models, which might demand more samples for successful training.

For all models, the minimum MAE over $k$ is lower than the MAE version trained on the whole data set, even when the mean MAE across $k$ is higher in all tasks. This indicates that improvements are highly likely with fine-tuning of parameter $k$. Our experiments did not reveal a clear trend for the value of $k$ for which the minimum was attained. Nevertheless, our observations show that the proposed RegGNN network is more stable to changes in the parameter $k$. Calculating, for each architecture, the average of the standard deviations (std) of the mean absolute error[6] (see the results table) over all feature extraction methods, we first note that the averages for RegGNN are 0.455, 0.145, 0.877, 0.369 for NT-FIQ, NT-VIQ, ASD-FIQ, and ASD-VIQ, respectively. While RegGNN, therefore, shows little sensitivity to $k$, CPM is highly sensitive to the changes of this parameter with averages of 2.901, 2.613, 1.803, and 2.754, respectively. This is approximately a 2- to 10-fold increase in variability.

Improvements to the performance of CPM are not statistically significant ($p = 0.87$ for ASD-FIQ, $p = 0.15$ for ASD-VIQ). Similarly, we observe that improvements in the performance of RegGNN are only statistically significant in the NT-VIQ task ($p < 0.05$ for NT-VIQ, $p = 0.98$ for NT-FIQ, $p = 0.64$ for ASD-FIQ). Increases in performance for PNA models are more consistent, as PNA-S improves significantly in three out of four tasks ($p < 0.01$ for NT-FIQ, $p = 0.11$ for NT-VIQ, $p < 0.05$ for ASD-FIQ, $p < 0.01$ for ASD-VIQ), and PNA-V improved significantly in two out of four tasks ($p < 0.05$ for NT-FIQ, $p = 0.21$ for NT-VIQ, $p = 0.11$ for ASD-FIQ, $p < 0.01$ for ASD-VIQ).

Looking at the sample selection pipelines, the best-performing method always utilizes the Riemannian geometric structure of the SPD space for MAE, apart from PNA-V results for the NT-VIQ task. In most cases, the approaches that rely on tangent matrices perform best, with the vectorized version of the whole tangent matrix being the best method for NT-VIQ and ASD-FIQ. We also see that the three centrality measures and concatenated versions perform well consistently.

Our results do not reveal any finer pattern among the sample selection measures. Nevertheless, we can conclude the following: *Using the Riemannian nature of connectomes for sample selection outperforms methods that do not leverage these geometric properties on average*.

Because our results do not show a clear winner among the variants of the sample selection method concerning prediction error, it is informative to look at further aspects. While the tangent matrix method performs very well, it is also the most time-consuming since no dimension reduction is exerted. On the contrary, computing centrality measures reduces the size of the matrices speeding up the process significantly. In our experiment, we observed that training linear regression models using tangent matrices took up to 16 times more time than models relying on centrality measures. A possible drawback of the centrality measures is that they have not been studied on tangent matrices; their behavior, in this case, is thus not well-understood.

---

[6]The standard deviations include variations over different $k$.

The fact that, in our application, they compress the information effectively might hint at a broader use. It is thus an interesting venue for future work.

An advantage of sample selection that crystallized is the decrease in computation time. While the exact time that is required for sample selection is heavily dependent on the hardware and varies according to the model architecture, number of epochs in training, number of training samples, and the number $k$ of samples to select, our observations during the experiments show that sample selection reduces the training time by 20% on average. Therefore, usage of our sample selection pipeline can enable the use of even deeper neural network architectures.

Finally, it should be noted that we have evaluated our method on a relatively young population only; however, RegGNN demonstrates a high degree of generalizability through cross-validation and because it yields good results for both the NC and ASD datasets. To proliferate replication studies on other cohorts, we publicly share the source code of RegGNN.[7]

**Explainability and biomarker discovery**

To identify the brain regions that influence the prediction most, we extract for each of the four tasks the learned weights of the final fully connected of *RegGNN with the best-performing sample selection method*. The learned weights of the last fully connected layer quantify the importance of its nodes in the target prediction task, thanks to the end-to-end network training, the backpropagation process, and our network design (which gives as input of the last layer an embedding of the original graph). Hence, a node with a higher weight in the fully connected layer is more influential in the prediction of the output score.

In Figure 8.3, we show the ROIs with the three highest weights averaged over $k = 2, \ldots, 15$; underlying is the AAL parcellation atlas [246].[8] For the FIQ prediction task in the NT cohort, we see that the left superior dorsal frontal gyrus (SFGdor.L), the right superior frontal medial gyrus (SFGmed.R), and the right cerebellum 6 (CRBL6.R) have the highest weights. For the VIQ prediction task in the same cohort, the left hippocampus (HIP.L), the left heschl gyrus (HES.L), and the left cuneus (CUN.L) possess the highest weights. In the ASD cohort, the left insula (INS.L), the left calcarine cortex (CAL.L), and the right pallidum (PAL.R) have the highest weights for FIQ prediction, while the highest weights for VIQ prediction are obtained by the left superior frontal medial gyrus (SFGmed.L), the left middle occipital gyrus (MOG.L), and the left cuneus (CUN.L).

According to our results, the important ROIs for IQ prediction lie in the brain's left hemisphere. Our findings are in line with other studies, which found that the insula shows higher activity in various cognitive tasks [67] and that differences in surface area in the left cuneus correlate strongly with full IQ, especially in perceptual tasks in young adults with very low birth weight [227]. Furthermore, we observe that the left cuneus was influential in predicting VIQ in both cohorts. Finally, our experiments indicate that the middle frontal gyrus is a significant region in IQ prediction, agreeing with the results from [74].

---

[7]https://github.com/basiralab/RegGNN
[8]The brain networks were visualized with the BrainNet Viewer [258].

A highly interesting venue for future work is to investigate why the sample selection method improves the prediction. It seems likely that there are aggregations within the data that can be represented by central samples. This is a challenging question that most likely requires the development of new analytical tools. Nevertheless, we think that it will be worth the effort as common structures and connections between these central samples could give us a lot more insights into the interplay between the connectivity structure of the brain and cognitive ability.

### 8.1.5 Conclusion

In this chapter, we applied RegGNN, a new graph neural network, to connectome data of neurotypical subjects and subjects with autism spectrum disorder to predict full-scale and verbal intelligence quotients. We trained it using a novel sample selection method, which tries to identify samples within the training set that are expected to better predict the cognitive scores of new subjects. This enabled us to train the network with only 15 samples or less, while the testing performance was on par or even better than state-of-the-art methods for cognitive score prediction from connectomes. Both the sample selection and RegGNN are easy to implement in open-access software and can be used in clinical practice. An open avenue for future research is the creation and adaption of methods that make the pipeline more explainable. This could help to answer why the prediction works so well with not more than 15 samples.

# Conclusion

Throughout this thesis, we investigated new methods for data analysis that can take the geometric structure of the underlying space into account. We focused on manifold-valued data and differentiated between the Riemannian perspective for general manifolds and the purely affine viewpoint with CCS connection in Lie groups. Taking the former, we derived a flexible higher-order regression scheme that can capture a much broader class of phenomena than the geodesics-only model. In contrast to other approaches, it allows for explicit computations in many relevant spaces and thus often helps to avoid expensive iterative algorithms. Building upon it, we derived a hierarchical model for the analysis of longitudinal data and a normalization technique that can be used to control for confounding variables. For bi-invariant statistics in Lie groups, we focused on dissimilarity measures of sample distributions. We derived generalizations of the Hotelling $T^2$ statistic, the Bhattacharyya, and Hellinger distance, proving the essential bi-invariance. For the Bhattacharyya distance, we also established a connection to the distributional setting, which is well-known in Euclidean space.

The analytical power of the novel notions was demonstrated for several applications in medical image analysis and archaeology. While replicating well-known findings for the shape differences of the knee and hippocampus under osteoarthritis (OA) and Alzheimer's with bi-invariant group tests, we could establish the non-geodesic nature of the shape development of the femur under progressing OA and construct shape trajectories of the mitral valve from sparse measurements with our novel regression scheme. Furthermore, when studying the shape of shadow surfaces of ancient Roman sundials, we identified a dependency of their shape on the location of installation that, to the best of our knowledge, was not known before. Apart from these successes, we hope it has become clear that the applications we investigated in this work only represent a tiny fraction of interesting real-world problems that can be tackled with our novel methods.

Thankfully (for the researcher), geometric statistics is a young field in which much is left to be discovered. This is particularly true for bi-invariant statistics in Lie groups, where we have only begun to build a theory and to find new and exciting applications. The fact that notions like the Mahalanobis distance, Hotelling $T^2$ statistic, Bhattacharyya distance, and Hellinger distance can be generalized lets us believe that similar things can be done with many other notions of interest; to us, a rich set of statistical tools for bi-invariant data analysis in Lie groups seems possible. Although, naturally, it will be hard to find generalizations of notions that explicitly involve angles (e.g., principal component analysis with its orthogonality assumptions), it should

be possible to capture the essence of a variety of tools using affine terms (similar to defining the mean as the exponential barycenter of the data). This, then, would quite possibly allow for bi-invariant extensions.

Returning to the Riemannian viewpoint, the newly defined Bézierfolds are of interest both from a theoretical and practical perspective. While understanding the structure of the spaces of higher-order Bézier curves would extend our knowledge of manifolds of curves in general, to the practitioner they offer the possibility to represent non-geodesic relationships with parametric curves that still have only a few degrees of freedom. The geometry we imposed on the (lower-dimensional) Bézierfolds can be used for further meaningful manipulations of the (splines of) Bézier curves.

Last but not least, we demonstrated that geometric information from the underlying manifold of SPD matrices, fed into a sample selection approach, helps deep learning algorithms to predict cognitive scores from brain connectomes. This result is in line with research on deep learning with non-Euclidean data and promises further success in this active research area. To further improve our understanding of the human connectome, it seems very interesting to decipher why the sample selection method improves the results.

# Appendix A

# Appendix

## A.1 Integration on Manifolds and Lie Groups

Smooth functions can be integrated on orientable manifolds with the help of differential forms. We will discuss below what orientable means. Before we can come to this, though, we give a brief introduction to the theory of differential forms and how they can be integrated. We thereby only summarize what is necessary. For more on the theory of differential forms, we refer the reader to one of the textbooks [137, 197, 252]. Apart from these, we recommend [8] as a reference on the theory of integration on manifolds.

### A.1.1 Differential Forms

Let $M$ be a $d$-dimensional manifold and $0 \leq k \leq d$. A *differential k-form* (a *k-form* in short) is a section of the *exterior k bundle*

$$\Lambda^k(M)$$

whose fiber at $p \in M$ is given by

$$\{\beta : (T_p M)^k \to \mathbb{R} \mid \beta \text{ is multilinear and alternating}\}.$$

(Here, 0-forms are simply smooth functions.) The addition of $k$-forms and multiplication with smooth functions are defined pointwise. (Note that $k$-forms are alternating $(k, 0)$-tensor fields, and 1-forms are sections of $T^*M$ as in the previous section.) Differential forms allow for the following construction. Let $N$ be another $d$-dimensional manifold, $F : M \to N$ a diffeomorphism, and $\omega$ a $k$-form on $N$. The *pullback $F^*(\omega)$ of $\omega$ along $F$* is a $k$-form on $M$ defined by

$$F^*(\omega)(X_1, \ldots, X_k) = \omega(\mathrm{d}F(X_1), \ldots, \mathrm{d}F(X_k))$$

for all $X_1, \ldots, X_k \in \Gamma(TM)$. Note that $F^*(\psi\,\omega) = (\psi \circ F)\,F^*(\omega)$ for all $\psi \in C^\infty(N)$.

We can also "multiply" a $k$- and an $\ell$-form to obtain a $(k + \ell)$-form (where $k + \ell \leq d$) as follows. Let $\omega \in \Gamma(\Lambda^k(M))$ and $\eta \in \Gamma(\Lambda^\ell(M))$. Their *wedge product* $\omega \wedge \eta \in \Gamma(\Lambda^{k+\ell}(M))$ is defined, with $S_{k+\ell}$ denoting the set of permutations of $\{1, \ldots, k + \ell\}$,

by

$$\omega \wedge \eta(X_1,\ldots,X_{k+\ell}) := \frac{1}{k!\ell!} \sum_{\sigma \in S_{k+\ell}} \mathrm{sgn}\,\sigma\, \omega(X_{\sigma(1)},\ldots,X_{\sigma(k)})\, \eta(X_{\sigma(k+1)},\ldots,X_{\sigma(k+\ell)}).$$

(A.1)

It can be shown that the pullback distributes over the wedge product, that is, for a diffeomorphism $F : M \to N$, $\omega \in \Gamma(\Lambda^k(N))$, and $\eta \in \Gamma(\Lambda^\ell(N))$ we find $F^*(\omega \wedge \eta) = F^*\omega \wedge F^*\eta$.

Let $U \subseteq M$ be a coordinate neighborhood with coordinate functions $(x_1,\ldots,x_d)$, corresponding frame $(\mathrm{d}/\mathrm{d}x_1,\ldots,\mathrm{d}/\mathrm{d}x_d)$ and coframe $(\mathrm{d}x^1,\ldots,\mathrm{d}x^d)$. From the definition of the wedge product (A.1) we get

$$\mathrm{d}x_i \wedge \mathrm{d}x_j = -\mathrm{d}x_j \wedge \mathrm{d}x_i$$

for all $i,j = 1,\ldots,d$ with $i \neq j$. Let $\omega \in \Gamma(\Lambda^k(M))$. Defining

$$\varphi_{i_1,\ldots,i_k} := \omega\left(\frac{d}{\mathrm{d}x_{i_1}},\ldots,\frac{d}{\mathrm{d}x_{i_k}}\right) \in C^\infty(U)$$

it can now be shown that every $k$-form $\omega$ can be uniquely written as the sum

$$\omega = \sum_{1 \le i_1 < \cdots < i_k \le d} \varphi_{i_1,\ldots,i_k}\, \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}.$$

In particular, if $\omega \in \Gamma(\Lambda^d(M))$ is a $d$-form, then

$$\omega = \varphi\, \mathrm{d}x^i \wedge \cdots \wedge \mathrm{d}x^d \tag{A.2}$$

in $U$.

Another operation on differential forms is the exterior derivative $\mathrm{d} : \Gamma(\Lambda^k) \to \Gamma(\Lambda^{k+1})$. It is uniquely defined by the following properties:

1. $\mathrm{d}f$ is the usual derivative for every 0-form $f \in C^\infty(M)$;

2. $(\mathrm{d} \circ \mathrm{d})f = 0$ for every 0-form $f$;

3. $\mathrm{d}(\omega \wedge \eta) = \mathrm{d}\omega \wedge \eta + (-1)^m(\omega \wedge \mathrm{d}\eta)$ for an $m$-form $\omega$ and an $\ell$-form $\eta$ with $k = m + \ell$.

## A.1.2   Integration of Differential Forms

A manifold $M$ is called *orientable* if there exists a nowhere-vanishing $d$-form. Importantly, Lie groups fulfill this requirement: When $\omega^1,\ldots,\omega^d$ is a coframe of 1-forms on a Lie group G (which we can obtain by left translating a basis of $T_e^*G$), then there is no tangent space on which the restriction of $\omega^1 \wedge \cdots \wedge \omega^d$ is the zero map.

An *orientation* on an orientable manifold is a choice of an equivalence class $[\omega]$ of nowhere-vanishing $d$-forms with the same sign (the class of $d$-forms with the opposite sign represents the opposite orientation). Any $d$-form from the chosen class determines the orientation of an *ordered* basis of a tangent space through the sign of

the value that it assigns to it. Additionally, given an orientation $[\omega]$, we say that a chart $\phi : U \to \mathbb{R}^d$, $\phi = (x_1, \dots, x_d)$, is positively oriented if

$$\omega\big|_U = \varphi \, \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^d$$

with $\varphi(p) > 0$ for all $p \in U$. Furthermore, a diffeomorphism between two oriented manifolds is said to be *orientation-preserving* if its differential takes positively oriented bases of tangent spaces to positively oriented bases. (Negative orientation is defined analogously.)

We now define integrals of differential $d$-forms on oriented manifolds. (Whenever we say that a manifold is oriented, we mean that an orientation has already been chosen.) One starts by defining integrals of differential $d$-forms in (oriented) $\mathbb{R}^d$. Let $V \subset \mathbb{R}^d$ be a "nice" subset of $\mathbb{R}^d$ and $(x_1, \dots, x_d)$ the standard coordinates. (The canonical orientation is given by the $d$-form $\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_d$.) Further, let $f \in C^\infty(V)$. The integral of the $d$-form $\varphi \, \mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_d$ is is then defined as the Lebesgue integral

$$\int_V \varphi \, \mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_d := \int_V \varphi \, \mathrm{d}x_1 \dots \mathrm{d}x_d,$$

if it exists[1].

Let now $M$ be an oriented manifold and $U \subseteq M$ a coordinate neighborhood with positively oriented chart $\phi = (x_1, \dots, x_d)$. The integral of a $d$-form $\omega$ over $U$ is defined via the pullback by

$$\int_U \omega := \int_{\phi(U)} (\phi^{-1})^* \omega$$

whenever the right-hand side exists. It can be shown that this definition does not depend on the choice of coordinates.

To go on we need *partitions of unity* [137, Lem. 1.1.1]. Remember that the support of a function $\rho \in C^\infty(M)$ is defined by $\operatorname{supp} \rho := \{p \in M \mid \rho(p) \neq 0\}$.

**Theorem A.1.1.** *Let $M$ be a manifold and $(U_i)_{i \in I}$ be an open cover of $M$. Then, there exists a (countable) partition of unity subordinate to $(U_i)_{i \in I}$. This means that there exist a locally finite refinement $(V_j)_{j \in J}$ and functions $\rho_j \in C^\infty(M)$, with $\operatorname{supp} \rho$ having compact closure, such that*
*(i) $\operatorname{supp} \rho_j \subset V_J$ for all $j \in J$;*
*(ii) $\rho_j(p) \geq 0$ for all $p \in M$ and $j \in J$;*
*(iii) for each $p \in M$, $\rho_j(p) \neq 0$ only for finitely many $j \in J$ and $\sum_{j \in J} \rho_j(p) = 1$.*

We can now define the integral of a $d$-form on the whole of $M$ using a cover $(U_j, \phi_j)_{j \in J}$ of positively oriented charts[2] and a partition of unity $(\rho_j)_{j \in J}$ subordinate to the cover $(U_j)$ to define the integral of $\omega \in \Gamma(\Lambda^d(M))$ by

$$\int_M \omega := \sum_{j \in J} \int_M \rho_j \omega$$

whenever the right-hand side exists. In other words, we "stitch integrals over coordinate neighborhoods together" to integrate over the whole domain. Note that the

---

[1] Here, we assume that the reader has some knowledge of integration theory in $\mathbb{R}^d$.
[2] Such a cover always exists on an oriented manifold.

differential forms under the integral sign are only supported in the domain of a single chart. Furthermore, it can be shown that this definition is independent of the chosen partition of unity.

Being a vast generalization of the Fundamental Theorem of Calculus, Stokes' Theorem [137, Thm. 2.1.6] relates integrals over a manifold to integrals over its boundary.

**Theorem A.1.2** (Stokes' Theorem). *Let M be an orientable d-dimensional smooth submanifold of a smooth manifold N with a smooth boundary. That is, the boundary $\partial M$ is assumed to be an $(d-1)$-dimensional smooth submanifold of N. Let further $\omega$ be a smooth $(d-1)$-form on N. Then, provided that these integrals exist, for instance, if M is compact, we have*

$$\int_M \mathrm{d}\omega = \int_{\partial M} \omega.$$

### A.1.3 Integration of Functions

Integrals of functions can be defined using the integral of differential forms. To this end, one chooses a reference measure (a non-vanishing $d$-form) and integrates "measure times function". Naturally, the value of this integral then depends on the selected measure. However, if the manifold at hand is a Lie group or is endowed with a Riemannian metric, then canonical reference measures exist, as we will see below.

For a Riemannian manifold, the canonical reference measure is given as follows. Let $M$ be an oriented Riemannian manifold. Then, there is a unique $d$-form $\mathrm{d}p \in \Gamma(\Lambda^d(M))$ such that for every $p \in M$ and for every positively oriented orthonormal basis $(v_1, \ldots, v_d)$ of $T_p M$ we have

$$\mathrm{d}p\big|_{T_p M}(v_1, \ldots, v_d) = 1.$$

The $d$-form $\mathrm{d}p$ is called the *volume form of M*. Finally, we can now define integrals of functions over Riemannian manifolds. Let $f \in C^\infty(M)$. Its integral over $M$ is now defined by

$$\int_M f(p)\mathrm{d}p := \int_M f \,\mathrm{d}p$$

(again, if the right-hand side exists).

If $M = G$ is a $d$-dimensional Lie group without Riemannian metric, then there is no canonical volume form on $G$. Hence, one must choose a reference measure in a different way. A natural choice on $G$ is a nowhere-vanishing left/right invariant $d$-form $\mathrm{d}g$, a *left/right Haar measure* consistent with the orientation of $G$. There is a 1-dimensional vector space of such $d$-forms [252, Sec. 4.11], and, hence, the resulting integrals only differ in a constant factor. Therefore, the actual choice of $\mathrm{d}g$ depends (up to considerations of machine precision when working on a computer) only on one's taste. The integral of a function $\psi \in C^\infty(G)$ is then defined by

$$\int_G \psi(g) \,\mathrm{d}g,$$

if the right-hand side exists. If $\mathrm{d}g$ is left-invariant,

$$\int_G \psi(g) \,\mathrm{d}g = \int_G (\psi \circ L_f)(g) \,\mathrm{d}g$$

holds for all $f \in G$, that is, the integral is also left invariant. However, in general, it is not right invariant. Lie groups with bi-invariant Haar measures (which lead to bi-invariant integrals) are called *unimodular*.

## A.2 Existence and Uniqueness of the Group Mean

Let $G$ be a Lie Group with an affine connection. If $U \subseteq G$ is a normal convex neighborhood and $g_1, \ldots, g_m \in U$, then there always exists a group mean $\overline{g} \in U$ [188, Thm. 5.3]. For uniqueness to be ensured, we need the following notion of convexity that was proposed by Arnaudon and Li in [17]. In the following, we summarize their ideas; for more details, including examples, we refer to their article or [188] and the references therein.

First, we need the following definitions. (Remember that a real-valued function on a smooth manifold is called convex if its restriction to any geodesic is convex.)

**Definition A.2.1.** *A function $\rho : U \times U \to \mathbb{R}_{\geq 0}$ that is convex with respect to the product structure is called* separating function *if it vanishes on the diagonal of $U \times U$ and only there.*

**Definition A.2.2** (*p*-convexity). *Let $\rho$ and* dist *be a smooth separating function and an auxiliary Riemannian distance function on $U$, respectively. We say that $U$ has a $p$-convex geometry if there are constants $c, C \in \mathbb{R}$ with $0 < c < C$ and an even integer $p \geq 2$ such that*

$$c \operatorname{dist}(f, g)^p \leq \rho(f, g) \leq C \operatorname{dist}(f, g)^p$$

*for all $f, g \in U$.*

For general smooth manifolds, it is known that not all normal convex neighborhoods have a $p$-convex geometry. On the other hand, Whitehead's theorem ensures that each point in $G$ has a 2-convex neighborhood. Importantly, if the normal convex neighborhood $U$ has a $p$-convex geometry for any $p \in 2\mathbb{N}$, then this is enough to ensure the uniqueness of the group mean. Nevertheless, there is a weaker condition that still yields uniqueness.

**Definition A.2.3.** *(CSLCG neighborhood [17]) We say that $U$ is* convex with semilocal convex geometry (CSLCG) *if every compact subset $K \subset U$ has a relatively compact neighborhood $U_K$ with $p_K$-convex geometry for some $p_K \in 2\mathbb{N}$ depending on $K$.*

Observe that if $U$ has a $p$-convex geometry for some $p \in 2\mathbb{N}$, then it is CSLCG. Further examples of CSLCG neighborhoods are open hemispheres of $S^n$ endowed with the Levi-Civita connection of the embedding metric. Remarkably, these do not possess a $p$-convex geometry for any $p \in 2\mathbb{N}$; see [144]. We have the following result, which is a special case of [17, Prop. 2.4].

**Proposition A.2.4.** *Let $G$ be a Lie group with affine connection. Further, let $U \subseteq G$ be a CSLCG neighborhood and $(g_1, \ldots, g_m) \overset{i.i.d.}{\sim} \mathcal{G}$ data in $U$. Then, there exists a unique group mean $\overline{g} \in U$ of $g_1, \ldots, g_m$.*

The following lemma provides a property of CSLCG neighborhoods that is needed in this work.

**Lemma A.2.5.** *Let $G$ be a $d$-dimensional Lie group with affine connection and $U \subseteq G$ be a CSLCG neighborhood. Then, $U$ is diffeomorphic to $\mathbb{R}^d$.*

*Proof.* Let $g \in U$. Since $U$ is a normal convex neighborhood, $\mathrm{Log}_g(U) \subseteq T_g G$ is well-defined and star-shaped about $0 \in T_g G$. The claim now follows, since any star-shaped domain in a $d$-dimensional vector space is diffeomorphic to $\mathbb{R}^d$ [38, Thm. 5.1]. $\square$

## A.3 Discrete geodesic calculus

In the following, we summarize important aspects of discrete geodesic calculus [210] as given in [119]. Let $M$ be a Riemannian manifold. Then, the *energy* of a curve $\alpha : [0, 1] \to M$ is defined by

$$\mathrm{energy}(\alpha) := \int_0^1 \langle \alpha'(t), \alpha'(t) \rangle_{\alpha(t)} \mathrm{d}t.$$

Given $p, q \in M$, a curve $\gamma$ that minimizes $E$ over all curves $\alpha : [0, 1] \to M$ with $\alpha(0) = p$ and $\alpha(1) = q$ is a geodesic and we have

$$\mathrm{dist}(p, q)^2 = \mathrm{energy}(\gamma).$$

Furthermore, for $\alpha : [0, 1] \to M$ and $\ell \in \mathbb{N}_{>0}$, we find for the interpolated points $a_k := \alpha(k/\ell), k = 0, \dots, \ell$,

$$\mathrm{energy}(\alpha) \geq \ell \sum_{k=1}^{\ell} \mathrm{dist}(a_{k-1}, a_k)^2$$

with equality if and only if $\alpha = \gamma$ is a geodesic. This motivates the definition of the *discrete path energy*

$$E(\mathfrak{a}) := \ell \sum_{k=1}^{\ell} \mathrm{dist}(a_{k-1}, a_k)^2, \tag{A.3}$$

where $\mathfrak{a}$ is a polygonal curve with vertices $a_k := \mathfrak{a}(k/\ell)$. Minimizers of Equation (A.3) with fixed end points $a_0$ and $a_n$ are called *discrete $\ell$-geodesics*. In [210], it is shown that discrete $\ell$-geodesics converge to smooth geodesics for $\ell \to \infty$ under suitable assumptions.

Let now $q_1, \dots, q_n \in M$ be given. The authors of [119] define the *discrete $\ell$-mean* of $q_1, \dots, q_n$ by

$$\overline{q} := \arg\min_{q \in M} \sum_{i=1}^{n} \min_{\substack{\mathfrak{a}^{(i)}(0) = q \\ \mathfrak{a}^{(i)}(1) = q_i}} E(\mathfrak{a}_i).$$

Here, for each $i = 1, \dots, n$ the inner minimization is done over all polygonal curves $\mathfrak{a}_i$ with $\ell$ vertices, $\mathfrak{a}_i(0) = q$, and $\mathfrak{a}_i(1) = q_i$. The minimizing curves are the discrete $\ell$-geodesics between $\overline{q}$ and $q_i$.

We also summarize how Gram's matrix $G$ corresponding to data $q_1, \dots, q_n \in M$ can be approximated using discrete geodesic calculus [120]. In the smooth setting,

assuming a unique Fréchet mean $\overline{q}$, we have

$$
G := \begin{bmatrix}
\left\langle \mathrm{Log}_{\overline{q}}(q_1), \mathrm{Log}_{\overline{q}}(q_1) \right\rangle_{\overline{q}} & \cdots & \left\langle \mathrm{Log}_{\overline{q}}(q_1), \mathrm{Log}_{\overline{q}}(q_n) \right\rangle_{\overline{q}} \\
\vdots & \ddots & \vdots \\
\left\langle \mathrm{Log}_{\overline{q}}(q_n), \mathrm{Log}_{\overline{q}}(q_1) \right\rangle_{\overline{q}} & \cdots & \left\langle \mathrm{Log}_{\overline{q}}(q_n), \mathrm{Log}_{\overline{q}}(q_n) \right\rangle_{\overline{q}}
\end{bmatrix} \in \mathbb{R}^{n,n}.
$$

Let $\overline{\mathfrak{q}}$ be the discrete geodesic average, $\mathfrak{a}^i$ the discrete $\ell$-geodesics between $\overline{\mathfrak{q}}$ and $q_i$. Define $a^i := \mathfrak{a}^i(1/\ell)$ and $d^i := \mathrm{dist}(\overline{\mathfrak{q}}, a^i)^2$ for all $i = 1, \ldots, n$. Then, we can substitute $G$ by the approximation

$$
\widetilde{G} := \frac{n}{2} \begin{bmatrix}
2d^1 & \cdots & d^1 + d^n - \mathrm{dist}(a^1, a^n)^2 \\
\vdots & \ddots & \vdots \\
d^n + d^1 - \mathrm{dist}(a^n, a^1)^2 & \cdots & 2d^n
\end{bmatrix}.
$$

Note that $\widetilde{G}$ is symmetric just like $G$.

## A.4 Benjamini-Hochberg Procedure

When testing multiple independent or positively correlated hypotheses, the method of Benjamini and Hochberg [26, 27] allows controlling the false discovery rate, that is, the expected proportion of falsely rejected null hypotheses. Let $V$ and $S$ be the (random) numbers of false rejections of the null and true discoveries, respectively. The ratio of false discoveries among all discoveries is then given by $Q := V/(S + V)$ (defined to be zero if $V + S = 0$). The *false discovery rate* (FDR) is the expected value $\mathbb{E}[Q]$.

Suppose $n$ hypotheses were tested. The following procedure controls the FDR among the tests at level $\alpha$ (that is, with $n_0$ being the (unknown) number of true null hypotheses, we have $\mathbb{E}[Q] < \frac{n_0}{n}\alpha < \alpha$):

1. Order the $p$-values that were obtained from the hypothesis tests in ascending order; this gives a sequence $p_1, \ldots, p_n$ of $p$-values with corresponding null hypotheses $H_1, \ldots, H_n$.

2. Let $k \in \mathbb{N}$ be the largest integer such that

$$
p_i < \frac{i}{n}\alpha
$$

for all $i = 1, \ldots, k$. Reject $H_1, \ldots, H_k$ and do not reject $H_{k+1}, \ldots, H_n$.

## A.5 List of Sundials

The sundials used in the study are listed in Table A.1 along with the latitude and longitude of their sites. They are part of the Topoi database. The triangle meshes of the sundials can be downloaded from http://repository.edition-topoi.org/collection/BSDP. The metadata can also be found there.

Table A.1: **Spherical sundials used in the study**

| Object ID | Dialface ID | Latitude | Longitude | Site | Dating | Type (Edition Topoi) |
|---|---|---|---|---|---|---|
| | | | | **Italy** | | |
| 574 | 623 | 43.3155 | 13.4082 | Helvia Recina | N/A | sphere, spherical |
| 17 | 17 | 42.0913 | 12.5231 | Riano | N/A | sphere, spherical-cut dials—central gnomon point |
| 36 | 35 | 41.8034 | 12.6890 | Villa Tuscolana/Villa Rufinella[5] (near Rome) | N/A | sphere, spherical-cut dials—central gnomon point |
| 65 | 62 | 41.7561 | 12.2927 | Ostia Antica (near Rome) | N/A | sphere, spherical-cut/ quarter dis-shaped |
| 18 | 18 | 41.6700 | 12.6900 | Lanuvio | N/A | sphere, spherical-cut dials—central gnomon point |
| 21 | 21 | 40.7503 | 14.4871 | Pompei | before 79 CE | sphere, spherical-cut dials—central gnomon point |
| 23 | 23 | 40.7503 | 14.4871 | Pompei | before 79 CE | sphere, spherical-cut dials—central gnomon point |
| 174 | 173 | 40.7503 | 14.4871 | Pompei | N/A | sphere, spherical-hemispherical—central gnomon point |
| 29 | 29 | 40.7503 | 14.4871 | Pompei | before 79 CE | sphere, spherical-cut dials—central gnomon point |

| Object ID | Dialface ID | Latitude | Longitude | Site | Dating | Type (Edition Topoi) |
|-----------|-------------|----------|-----------|------|--------|----------------------|
| 519 | 559 | 40.7030 | 14.4988 | Stabiae | N/A | sphere, spherical |
| 40 | 39 | N/A | N/A | N/A[6] | N/A | sphere, spherical-cut dials—central gnomon point |
| | | | | **Greece** | | |
| 76 | 73 | 37.3900 | 25.2600 | Delos | N/A | sphere, spherical-hemispherical—central gnomon point |
| 77 | 74 | 37.3900 | 25.2600 | Delos | N/A | sphere, spherical-transposed hemispherical |
| 546 | 583 | 36.0917 | 28.0881 | Lindos on Rhodes[7] | Hellenistic | sphere, spherical |

---

[5]This site is not 100% certain as the only reference is a dissertation from 1764.

[6]Although in the database the site is stated to be Vatican City, we did not find evidence for this. Therefore, we consider the site to be uncertain.

[7]The longitude and latitude of this sundial are wrong in the database. The references stated there confirm that the sundial is from Lindos. We used the correct values that are given in this table.

## A.6 Topological Centrality Measures for Graphs

We now recall three important centrality measures for undirected graphs that are widely used. A good reference on the topic is [94].

Let $A$ be the (weighted) adjacency matrix of a graph $G$, $V$ the set of vertices of $G$, and $v \in V$.[3] The *degree centrality $D(v)$ of $v$* is defined by

$$D(v) := \sum_{\substack{w \in V \\ w \neq v}} A_{vw},$$

that is, it assigns to each node its weighted sum of neighbors.

Let $x$ be the unit norm eigenvector of $A$ that corresponds to the largest eigenvalue $\lambda_1$ and has only non-negative entries. The *eigenvector centrality $E(v)$ of $v$* is the $v$-th entry of $x$; that is,

$$E(v) := \frac{1}{\lambda_1} \sum_{w \in V} A_{vw} x_w,$$

$$\text{s.t. } \|x\|_2 = 1 \text{ and } x_w \geq 0 \text{ for all } w \in V.$$

It measures, in a relative sense, how influential a node is in the network. Intuitively, a high score means that a node has many neighbors that themselves have high eigenvector centrality scores.

Let $l_{vw}$ be the length of the shortest path between two nodes $v$ and $w$, and $n = |V|$. The *closeness centrality $C(v)$ of $v$* is defined by

$$C(v) := \frac{n-1}{\sum_{\substack{w \in V \\ w \neq v}} l_{vw}},$$

that is, as the inverse of the average distance of $v$ to all other nodes.

## A.7 Technical Acknowledgements

---

[3]With a slight abuse of notation, we identify nodes $v \in V$ and the integers we assign to them to construct $A$, e.g., we write $A_{vw}$ for the entry that corresponds to the edge between nodes $v$ and $w$.

# A.8 Curiculum Vitae

# Martin Hanik

## Email

Email:                   hanik@zib.de

## Work Experience

**RESEARCH STAFF, FREIE UNIVERSITÄT BERLIN, BERLIN, 2022 - TODAY**

Research into generalizations of statistical methods for non-Euclidean data, in particular, for geometric deep learning.

**RESEARCH STAFF, ZUSE INSTITUTE BERLIN, BERLIN, 2019 - 2021**

Research into generalizations of statistical methods for non-Euclidean data; in particular, generalizations of higher order regression in the setting of Riemannian manifolds and bi-invariant statistical methods for Lie groups.

**MENTOR, BRAIN AND SIGNAL RESEARCH & ANALYSIS LABORATORY, ISTANBUL,  2021 - TODAY**

Mentor undergraduate students in research.

## Publications

Martin Hanik, Hans-Christian Hege, and Christoph von Tycowicz. "Bi-invariant Dissimilarity Measures for Sample Distributions in Lie Groups." *SIAM Journal on Mathematics of Data Science*. 2022.

Martin Hanik, Hans-Christian Hege, and Christoph von Tycowicz. "A Hierarchical Model for Longitudinal Data on Manifolds." In *International Symposium on Biomedical Imaging*. IEEE, 2022.

Martin Hanik, Mehmet Arif Demirtaş, Mohammed Amine Gharsallaoui, Islem Rekik. "Predicting cognitive scores with graph neural networks through sample selection learning." *Brain Imaging and Behavior*. 2022

Adele Meyers, Martin Hanik et al. "ICLR 2022 Challenge for Computational Geometry & Topology: Design and Results." In *Proceedings of Topological, Algebraic, and Geometric Learning Workshops 2022*. PMLR, 2022

Marco Emporio, Martin Hanik et al. "SHREC 2022 track on online detection of heterogeneous gestures." *Computers & Graphics, 2022*

Felix Ambellan, Martin Hanik, Christoph von Tycowicz. "Morphomatics: Geometric morphometrics in non-Euclidean shape spaces." https://morphomatics.github.io, 2021

Ivan Sipiran, Martin Hanik et al. "SHREC 2021: Retrieval of cultural heritage objects." *Computers & Graphics, 2021*

Martin Hanik, Hans-Christian Hege, and Christoph von Tycowicz. "Bi-invariant Two-Sample Tests in Lie Groups for Shape Analysis." In *International Workshop on Shape in Medical Imaging*. Springer, 2020.

Martin Hanik, et al. "Nonlinear Regression on Manifolds for Shape Analysis using Intrinsic Bézier Splines." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020.

## Awards

Best-Paper Award for "Bi-invariant Two-Sample Tests in Lie Groups for Shape Analysis" at the *International Workshop on Shape in Medical Imaging (ShapeMI)*

Student Travel Award for the paper "Nonlinear Regression on Manifolds for Shape Analysis using Intrinsic Bézier Splines" at the *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*

Special Mention for contribution "Sasaki Metric and Applications in Geodesic Analysis" to *ICLR Computational Geometry & Topology Challenge 2022*

## A.9 Zusammenfassung

Datenanalyse ist aus unserer Zeit nicht mehr wegzudenken. Die dabei eingesetzten Methoden sind vielfältig, haben jedoch fast immer eines gemein: die Annahme, dass die Daten aus einem Vektorraum mit Euklidischer Struktur stammen. Ausgehend von Forschungsfeldern wie Bildverarbeitung, (medizinische) Formenanalyse und Netzwerkanalyse sind in letzter Zeit jedoch Daten aus nichteuklidischen Räumen - insbesondere gekrümmten Mannigfaltigkeiten - immer mehr in den Fokus geraten. Die Methoden des so entstandenen Forschundsfelds der *geometrischen Datenanalyse* zeichnen sich dadurch aus, dass sie die Geometrie des zugrundeliegenden Raumes berücksichtigen. Sie führen einerseits zu besseren Ergebnissen in vielen Anwendungen, in denen Daten näherungsweise Euklidisch sind und altbekannte Methoden der Datenanalyse angewandt werden können; andererseits eröffnen sie auch komplett neue Anwendungsgebiete. Diese Arbeit trägt zur weiteren Entwicklung der geometrischen Datenanalyse bei, indem sie sowohl theoretische Beiträge liefert als auch neue Anwendungen ergründet. Die Methodologie wird erweitert, indem mehrere wohlbekannte Methoden der multivariaten Datenanalyse auf Mannigfaltigkeiten verallgemeinert werden. Dabei konzentrieren wir uns auf zwei verschiedene Blickwinkel. Zuerst nutzen wir den der Riemannschen Mannigfaltigkeit, um basierend auf Splines aus generalisierten Bézierkurven Regressionsanalyse höherer Ordnung auf allgemeine Mannigfaltigkeiten zu heben. Dadurch ist es möglich nichtgeodätische Zusammenhänge (z.B. zeitliche Zusammenhänge mit Sättigungseffekten oder zyklische Zusammenhänge) zu modellieren. Außerdem erlaubt uns die neue Regressionsmethodik ein hierarchisches statistisches Modell zur Analyse von longitudinalen Daten und eine Methode zur Normalisierung von Daten bezüglich unerwünschter Einflussgrößen zu entwickeln. Neben dem Riemannschen Blickwinkel liegt der Fokus in dieser Arbeit auf Daten aus Lie Gruppen. Dies ist in Anwendungen regelmäßig der Fall und es ist wünschenswert über Analysemethoden zu verfügen, die auch die Gruppenstruktur nicht außer Acht lassen. Letzteres bedeutet, dass die Methoden invariant/äquivariant unter Translationen der Daten sein müssen. Dies kann nur erreicht werden, wenn eine bestimmte affine Struktur verwendet wird, die im Allgemeinen nicht Riemannsch ist. Indem wir diese Struktur verwenden, leiten wir Verallgemeinerung von bekannten Ähnlichkeitsmaßen von Datenverteilungen her, die für eine Vielzahl von Aufgaben (z.B. Hypothesentests) verwendet werden können. Wir beweisen die Invarianz dieser neuen Größen und leiten Verbindungen zu kontinuierlichen Verteilungen her. Wir zeigen Anwendungen von all unseren methodologischen Beiträgen an Hand von realen Daten aus der medizinischen Formenanalyse und der Archäologie. Wir können dabei sowohl wohlbekannte Effekte in einigen medizinischen Anwendungen replizieren, als auch neue Einsichten in einer archäologischen Anwendung gewinnen. Zuletzt wird auch der Nutzen unserer Methoden im Rahmen des tiefen Lernens an einem Beispiel verdeutlicht.

# Bibliography

[1]   K. Abbas et al. "Geodesic distance on optimally regularized functional connectomes uncovers individual fingerprints". In: *Brain Connect.* 0.0 (2021), null. DOI: 10.1089/brain.2020.0881.

[2]   P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton University Press, 2007. DOI: 10.1515/9781400830244.

[3]   R. L. Adler et al. "Newton's method on Riemannian manifolds and a geometric model for the human spine". In: *IMA J. Numer. Anal.* 22.3 (2002), pp. 359–390. DOI: 10.1093/imanum/22.3.359.

[4]   B. Afsari. "Riemannian $L^p$ center of mass: existence, uniqueness, and convexity". In: *Proc. Am. Math. Soc.* 139.2 (2011), pp. 655–673. DOI: 10.1090/S0002-9939-2010-10541-5.

[5]   B. Afsari, R. Tron, and R. Vidal. "On the convergence of gradient descent for finding the Riemannian center of mass". In: *SIAM J. Control. Optim.* 51.3 (2013), pp. 2230–2260. DOI: 10.1137/12086282X.

[6]   C. C. Aggarwal. *Neural Networks and Deep Learning*. Berlin: Springer, 2018. DOI: 10.1007/978-3-319-94463-0.

[7]   C. C. Aggarwal, H. Wang, et al. *Managing and Mining Graph Data*. Vol. 40. Berlin: Springer, 2010. DOI: 10.1007/978-1-4419-6045-0.

[8]   H. Amann and J. Escher. *Analysis III*. Berlin: Springer, 2009. DOI: 10.1007/978-3-7643-7480-8.

[9]   K. Amasyali and N. M. El-Gohary. "A review of data-driven building energy consumption prediction studies". In: *Renew. Sust. Energ. Rev.* 81 (2018), pp. 1192–1205. DOI: 10.1016/j.rser.2017.04.095.

[10]  F. Ambellan, M. Hanik, and C. von Tycowicz. *Morphomatics: Geometric morphometrics in non-Euclidean shape spaces*. https://morphomatics.github.io/. 2021. DOI: 10.12752/8544.

[11]  F. Ambellan, S. Zachow, and C. von Tycowicz. "An as-invariant-as-possible $GL^+(3)$-based statistical shape model". In: *Proc. 7th MICCAI workshop on Mathematical Foundations of Computational Anatomy*. Springer, 2019, pp. 219–228. DOI: 10.1007/978-3-030-33226-6_23.

[12]  F. Ambellan, S. Zachow, and C. von Tycowicz. "Geodesic B-score for improved assessment of knee osteoarthritis". In: *Information Processing in Medical Imaging – 27th International Conference, IPMI 2021*. Springer, 2021, pp. 177–188. DOI: 10.1007/978-3-030-78191-0_14.

[13]  F. Ambellan, S. Zachow, and C. von Tycowicz. "Rigid motion invariant statistical shape modeling based on discrete fundamental forms". In: *Med. Image Anal.* 73 (2021), p. 102178. DOI: 10.1016/j.media.2021.102178.

[14]  F. Ambellan et al. "Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: data from the Osteoarthritis Initiative". In: *Med. Image Anal.* 52 (2019), pp. 109–118. DOI: 10.1016/j.media.2018.11.009.

[15]  N. Arden and M. C. Nevitt. "Osteoarthritis: epidemiology". In: *Best Pract. Res. Clin. Rheumatol.* 20.1 (2006), pp. 3–25. DOI: 10.1016/j.berh.2005.09.007.

[16]  M. A. Armstrong. *Basic Topology*. Berlin: Springer, 2013. DOI: 10.1007/978-1-4757-1793-8.

[17]  M. Arnaudon and X.-M. Li. "Barycenters of measures transported by stochastic flows". In: *Ann. Probab.* 33.4 (2005), pp. 1509–1543. DOI: 10.1214/009117905000000071.

[18]  V. Arsigny et al. "Log-Euclidean metrics for fast and simple calculus on diffusion tensors". In: *Magn. Reson. Med.* 56.2 (2006), pp. 411–421. DOI: 10.1002/mrm.20965.

[19]  V. Arsigny et al. "Geometric means in a novel vector space structure on symmetric positive-definite matrices". In: *SIAM J. Matrix Anal. Appl.* 29.1 (2007), pp. 328–347. DOI: 10.1137/050637996.

[20]  M. Bacák et al. "A second order nonsmooth variational model for restoring manifold-valued images". In: *SIAM J. Sci. Comput.* 38.1 (2016), A567–A597. DOI: 10.1137/15M101988X.

[21]  F. Barbaresco. "Innovative tools for radar signal processing based on Cartan's geometry of SPD matrices & information geometry". In: *2008 IEEE Radar Conference*. IEEE, 2008, pp. 1–6. DOI: 10.1109/RADAR.2008.4720937.

[22]  F. Barbaresco. "Lie Group Statistics and Lie group machine learning based on Souriau Lie groups thermodynamics & Koszul-Souriau-Fisher metric: new entropy definition as generalized Casimir invariant function in coadjoint representation". In: *Entropy* 22.6 (2020), p. 642. DOI: 10.3390/e22060642.

[23]  G. D. Batty, I. J. Deary, and L. S. Gottfredson. "Premorbid (early life) IQ and later mortality risk: systematic review". In: *Ann. Epidemiol.* 17.4 (2007), pp. 278–288. DOI: 10.1016/j.annepidem.2006.07.010.

[24]  M. Bauer, M. Bruveris, and P. W. Michor. "Overview of the geometries of shape spaces and diffeomorphism groups". In: *J. Math. Imaging Vis.* 50.1–2 (2014), pp. 60–97. DOI: 10.1007/s10851-013-0490-z.

[25] E. J. Bekkers et al. "Roto-translation covariant convolutional networks for medical image analysis". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 440–448. DOI: 10.1007/978-3-030-00928-1_50.

[26] Y. Benjamini and Y. Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *J. R. Stat. Soc. Series B Stat. Methodol.* 57.1 (1995), pp. 289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

[27] Y. Benjamini and D. Yekutieli. "The control of the false discovery rate in multiple testing under dependency". In: *Ann. Stat.* 29.4 (2001), pp. 1165–1188. DOI: 10.1214/aos/1013699998.

[28] S. Benoit, D. D. Holm, and V. Putkaradze. "Helical states of nonlocally interacting molecules and their linear stability: a geometric approach". In: *J. Phys. A Math* 44.5 (2011), p. 055201. DOI: 10.1088/1751-8113/44/5/055201.

[29] R. Bergmann and P.-Y. Gousenbourger. "A variational model for data fitting on manifolds by minimizing the acceleration of a Bézier curve". In: *Front. Appl. Math. Stat.* 4 (2018), pp. 1–16. DOI: 10.3389/fams.2018.00059.

[30] Mark Berman and David Culpin. "The statistical behaviour of some least squares estimators of the centre and radius of a circle". In: *J. R. Stat. Soc. Series B Stat. Methodol.* 48.2 (1986), pp. 183–196. DOI: 10.1111/j.2517-6161.1986.tb01401.x.

[31] A. Bessadok, M. Ali Mahjoub, and I. Rekik. *Graph neural networks in network neuroscience*. 2021. arXiv: 2106.03535.

[32] R. Bhattacharya and V. Patrangenaru. "Large sample theory of intrinsic and extrinsic sample means on manifolds". In: *Ann. Stat.* 31.1 (2003), pp. 1–29. DOI: 10.1214/aos/1046294456.

[33] R. Bhattacharya and V. Patrangenaru. "Large sample theory of intrinsic and extrinsic sample means on manifolds: II". In: *Ann. Stat.* (2005), pp. 1225–1259. DOI: 10.1214/009053605000000093.

[34] S. Bi, M. Broggi, and M. Beer. "The role of the Bhattacharyya distance in stochastic model updating". In: *Mech. Syst. Signal Process.* 117 (2019), pp. 437–452. DOI: 10.1016/j.ymssp.2018.08.017.

[35] S. Bi et al. "The Bhattacharyya distance: Enriching the P-box in stochastic sensitivity analysis". In: *Mech. Syst. Signal Process.* 129 (2019), pp. 265–281. DOI: 10.1016/j.ymssp.2019.04.035.

[36] J. Boisvert et al. "Geometric variability of the scoliotic spine using statistics on articulated shape models". In: *IEEE Trans. Med. Imaging* 27.4 (2008), pp. 557–568. DOI: 10.1109/TMI.2007.911474.

[37] A. Bône, O. Colliot, and S. Durrleman. "Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2018, pp. 9271–9280. DOI: 10.1109/CVPR.2018.00966.

[38] R. Bott and L. W. Tu. *Differential Forms in Algebraic Topology*. Vol. 82. Berlin: Springer, 1982. DOI: 10.1007/978-1-4757-3951-0.

[39] A. D. Brett and C. J. Taylor. "Construction of 3D shape models of femoral articular cartilage using harmonic maps". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2000, pp. 1205–1214. DOI: 10.1007/978-3-540-40899-4_129.

[40] Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. *Numerical geometry of non-rigid shapes*. New York: Springer, 2008. DOI: 10.1007/978-0-387-73301-2.

[41] M. M. Bronstein et al. "Geometric deep learning: going beyond Euclidean data". In: *IEEE Signal Process. Mag.* 34.4 (2017), pp. 18–42. DOI: 10.1109/MSP.2017.2693418.

[42] P. Bühlmann and S. Van De Geer. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Berlin: Springer, 2011. DOI: 10.1007/978-3-642-20192-9.

[43] E. Calabi et al. "Differential and numerically invariant signature curves applied to object recognition". In: *Int. J. Comput. Vis.* 26.2 (1998), pp. 107–135. DOI: 10.1023/A:1007992709392.

[44] K. M. Campbell and P. T. Fletcher. "Nonparametric aggregation of geodesic trends for longitudinal data analysis". In: *Shape in Medical Imaging: Proceedings. International Workshop on Shape in Medical Imaging*. Springer, 2018, pp. 232–243. DOI: 10.1007/978-3-030-04747-4_22.

[45] M. P. do Carmo. *Riemannian Geometry*. 2nd ed. Mathematics: Theory and Applications. Cambridge, MA, USA: Birkhäuser Boston, 1992.

[46] E. Cartan and J. Shouten. "On the geometry of the group-manifold of simple and semi-groups". In: *Proc. Akad. Wetensch., Amsterdam* 29 (1926), pp. 803–815.

[47] J. Ćesić et al. "Radar and stereo vision fusion for multitarget tracking on the special Euclidean group". In: *Rob. Auton. Syst.* 83 (2016), pp. 338–348. DOI: 10.1016/j.robot.2016.05.001.

[48] R. Chakraborty, M. Banerjee, and B. C. Vemuri. "Statistics on the space of trajectories for longitudinal data analysis". In: *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*. IEEE, 2017, pp. 999–1002. DOI: 10.1109/ISBI.2017.7950684.

[49] Y. Chandrashekhar, S. Westaby, and J. Narula. "Mitral stenosis". In: *The Lancet* 374.9697 (2009), pp. 1271–1283. DOI: 10.1016/S0140-6736(09)60994-6.

[50] G. Chen et al. "Automated change-point detection of EEG signals based on structural time-series analysis". In: *IEEE Access* 7 (2019), pp. 180168–180180. DOI: 10.1109/ACCESS.2019.2956768.

[51] V. Cherkassky and F. M. Mulier. *Learning From Data: Concepts, Theory, and Methods*. Chichester, UK: John Wiley and Sons, 2007. DOI: 10.1002/9780470140529.

[52]   E. Chevallier and N. Guigui. "A bi-Invariant statistical model parametrized by mean and covariance on rigid Motions". In: *Entropy* 22.4 (2020). DOI: 10.3390/e22040432.

[53]   E. Chevallier et al. "Exponential-Wrapped Distributions on Symmetric Spaces". In: *SIAM J. Math. Data Sci.* 4.4 (2022), pp. 1347–1368. DOI: 10.1137/21M1461551.

[54]   G. S. Chirikjian. *Stochastic Models Information Theory and Lie Groups Volume 2: Analytic Methods and Modern Applications*. Cambridge, MA, USA: Birkhäuser Boston, 2012. DOI: 10.1007/978-0-8176-4944-9.

[55]   E. Choi and C. Lee. "Feature extraction based on the Bhattacharyya distance". In: *Pattern Recognit.* 36.8 (2003), pp. 1703–1709. DOI: 10.1016/S0031-3203(03)00035-9.

[56]   M. Christ. "The strong maximal function on a nilpotent group". In: *Trans. Am. Math. Soc.* 331.1 (1992), pp. 1–13. DOI: 10.2307/2153994.

[57]   D. Cieslak et al. "Hellinger distance decision trees are robust and skew-insensitive". In: *Data. Min. Knowl. Discov.* 24 (2011), pp. 136–158. DOI: 10.1007/s10618-011-0222-1.

[58]   T. S. Cohen and M. Welling. "Group equivariant convolutional networks". In: ICML'16. 2016, pp. 2990–2999.

[59]   M. W. Cole, T. Ito, and T. S. Braver. "Lateral prefrontal cortex contributes to fluid intelligence through multinetwork connectivity". In: *Brain Connect.* 5.8 (2015), pp. 497–504. DOI: 10.1089/brain.2015.0357.

[60]   M. W. Cole et al. "Global connectivity of prefrontal cortex predicts cognitive control and intelligence". In: *J. Neurosci.* 32.26 (2012), pp. 8988–8999. DOI: 10.1523/JNEUROSCI.0536-12.2012.

[61]   R. Colom et al. "Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents". In: *Pers. Individ. Differ.* 42.8 (2007), pp. 1503–1514. DOI: 10.1016/j.paid.2006.10.023.

[62]   D. J. Cook and L. B. Holder. *Mining Graph Data*. Chichester, UK: John Wiley and Sons, 2006.

[63]   T. F. Cootes et al. "Active shape models-their training and application". In: *Comput. Vis. Image Underst.* 61.1 (1995), pp. 38–59. DOI: 10.1006/cviu.1995.1004.

[64]   E. Cornea et al. "Regression models on Riemannian symmetric spaces". In: *J. R. Stat. Soc. Series B Stat. Methodol.* 79.2 (2017), pp. 463–482. DOI: 10.1111/rssb.12169.

[65]   G. Corso et al. "Principal neighbourhood aggregation for graph nets". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 13260–13271.

[66]   C. Craddock et al. "The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives". In: *Front. Neuroinform.* 7 (2013). DOI: 10.3389/conf.fninf.2013.09.00041.

[67] H. D. Critchley et al. "The functional neuroanatomy of social behaviour: changes in cerebral blood flow when people with autistic disorder process facial expressions". In: *Brain* 123.11 (2000), pp. 2203–2212. DOI: 10.1093/brain/123.11.2203.

[68] K. Dadi et al. "Benchmarking functional connectome-based predictive models for resting-state fMRI". In: *NeuroImage* 192 (2019), pp. 115–134. DOI: 10.1016/j.neuroimage.2019.02.062.

[69] B. C. Davis et al. "Population shape regression from random design data". In: *2007 IEEE 11th International Conference on Computer Vision*. IEEE, 2007, pp. 1–7. DOI: 10.1109/ICCV.2007.4408977.

[70] I. J. Deary, F. M. Spinath, and T. C. Bates. "Genetics of intelligence". In: *Eur. J. Hum. Genet.* 14.6 (2006), pp. 690–700. DOI: 10.1038/sj.ejhg.5201588.

[71] I. J. Deary et al. "Intelligence and educational achievement". In: *Intelligence* 35.1 (2007), pp. 13–21. DOI: 10.1016/j.intell.2006.02.001.

[72] L. Dodero et al. "Kernel-based classification for brain connectivity graphs on the Riemannian manifold of positive definite matrices". In: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015, pp. 42–45. DOI: 10.1109/ISBI.2015.7163812.

[73] D. L. Donoho. "High-dimensional data analysis: The curses and blessings of dimensionality". In: *AMS math challenges lecture* 1.2000 (2000), p. 32.

[74] E. Dryburgh, S. McKenna, and I. Rekik. "Predicting full-scale and verbal intelligence scores from functional connectomic data in individuals with autism spectrum disorder". In: *Brain. Imaging. Behav.* 14 (2020), pp. 1769–1778. DOI: 10.1007/s11682-019-00111-w.

[75] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis, with Applications in R.* 2nd ed. Chichester, UK: John Wiley and Sons, 2016. DOI: 10.1002/9781119072492.

[76] J. Dubois et al. "A distributed brain network predicts general intelligence from resting-state human neuroimaging data". In: *Philos. Trans. R. Soc. B* 373.1756 (2018), p. 20170284. DOI: 10.1098/rstb.2017.0284.

[77] R. Duits and B. Burgeth. "Scale spaces on Lie groups". In: *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 2007, pp. 300–312. DOI: 10.1007/978-3-540-72823-8_26.

[78] A. Edelman, T. A. Arias, and S. T. Smith. "The geometry of algorithms with orthogonality constraints". In: *SIAM J. Matrix Anal. Appl* 20.2 (1998), pp. 303–353. DOI: 10.1137/S0895479895290954.

[79] B. Eltzner and S. F. Huckemann. "A smeary central limit theorem for manifolds with application to high-dimensional spheres". In: *Ann. Stat.* 47.6 (2019), pp. 3360–3381. DOI: 10.1214/18-AOS1781.

[80] M. Enriquez-Sarano, C. W. Akins, and A. Vahanian. "Mitral regurgitation". In: *The Lancet* 373.9672 (2009), pp. 1382–1394. DOI: 10.1016/S0140-6736(09)60692-9.

[81]  F. Errica et al. *A fair comparison of graph neural networks for graph classification*. 2019. arXiv: 1912.09893.

[82]  L. C. Evans. *The Fiedler rose: on the extreme points of the Fiedler vector*. 2011. arXiv: 1112.6323.

[83]  D. Ezuz and M. Ben-Chen. "Deblurring and denoising of maps between shapes". In: *Comput. Graph. Forum* 36.5 (2017), pp. 165–174. DOI: 10.1111/cgf.13254.

[84]  G. Farin. *Curves and Surfaces for Computer-Aided Geometric Design*. 3rd ed. London: Academic Press, 1993. DOI: 10.1016/C2009-0-22351-8.

[85]  A. Feragen and T. Nye. "Statistics on stratified spaces". In: *Riemannian Geometric Statistics in Medical Image Analysis*. London: Academic Press, 2020, pp. 299–342. DOI: 10.1016/B978-0-12-814725-2.00016-9.

[86]  M. Fey and J. E. Lenssen. *Fast graph representation learning with PyTorch Geometric*. 2019. arXiv: 1903.02428.

[87]  M. Finzi et al. "Generalizing convolutional neural networks for equivariance to Lie groups on arbitrary continuous data". In: *Proceedings of the 37th International Conference on Machine Learning*. PMLR, 2020, pp. 3165–3176.

[88]  S. Fiori and T. Tanaka. "An algorithm to compute averages on matrix Lie groups". In: *IEEE Trans. Signal Process.* 57.12 (2009), pp. 4734–4743. DOI: 10.1109/TSP.2009.2027754.

[89]  P. T. Fletcher. "Geodesic regression and the theory of least squares on Riemannian manifolds". In: *Int. J. Comput. Vis.* 105.2 (2013), pp. 171–185. DOI: 10.1007/s11263-012-0591-y.

[90]  P. T. Fletcher. "2 - Statistics on manifolds". In: *Riemannian Geometric Statistics in Medical Image Analysis*. Academic Press, 2020, pp. 39–74. DOI: 10.1016/B978-0-12-814725-2.00009-1.

[91]  P. T. Fletcher, C. Lu, and S. Joshi. "Statistics of shape via principal geodesic analysis on Lie groups". In: *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.* Vol. 1. IEEE, 2003, pp. I–I. DOI: 10.1109/CVPR.2003.1211342.

[92]  P. T. Fletcher, S. Venkatasubramanian, and S. Joshi. "The geometric median on Riemannian manifolds with application to robust atlas estimation". In: *NeuroImage* 45.1 (2009), S143–S152. DOI: 10.1016/j.neuroimage.2008.10.052.

[93]  P. T. Fletcher et al. "Principal geodesic analysis for the study of nonlinear statistics of shape". In: *IEEE Trans. Med. Imaging* 23.8 (2004), pp. 995–1005. DOI: 10.1109/TMI.2004.831793.

[94]  A. Fornito, A. Zalesky, and E. Bullmore. *Fundamentals of Brain Network Analysis*. London: Academic Press, 2016. DOI: 10.1016/C2012-0-06036-X.

[95]  M. Fréchet. "Les éléments aléatoires de nature quelconque dans un espace distancié". In: *Annales de l'institut Henri Poincaré*. Vol. 10. 4. 1948, pp. 215–310.

[96]    B. Fritsch, E. Rinner, and G. Graßhoff. "3D Models of ancient sundials: a comparison". In: *Int. J. Herit. Digit. Era* 2.3 (2013), pp. 361–373. DOI: 10.1260/2047-4970.2.3.361.

[97]    Y. Fujikoshi, V. V. Ulyanov, and R. Shimizu. *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Vol. 760. Chichester, UK: John Wiley and Sons, 2011. DOI: 10.1002/9780470539873.

[98]    V. Garcia and J. Bruna. *Few-shot learning with graph neural networks*. 2017. arXiv: 1711.04043.

[99]    G. Gerig, J. Fishbaugh, and N. Sadeghi. "Longitudinal modeling of appearance and shape and its potential for clinical use". In: *Med. Image Anal.* 33 (2016), pp. 114–121. DOI: 10.1016/j.media.2016.06.014.

[100]   S. Glyn-Jones et al. "Osteoarthritis". In: *The Lancet* 386.9991 (2015), pp. 376–387. DOI: 10.1016/S0140-6736(14)60802-3.

[101]   T. E. Goldberg and D. R. Weinberger. "Genes and the parsing of cognitive processes". In: *Trends Cogn. Sci.* 8.7 (2004), pp. 325–335. DOI: 10.1016/j.tics.2004.05.011.

[102]   C. Goodall. "Procrustes methods in the statistical analysis of shape". In: *J. R. Stat. Soc. Series B Stat. Methodol.* 53.2 (1991), pp. 285–321. DOI: 10.1111/j.2517-6161.1991.tb01825.x.

[103]   L. S. Gottfredson and I. J. Deary. "Intelligence predicts health and longevity, but why?" In: *Curr. Dir. Psychol. Sci.* 13.1 (2004), pp. 1–4. DOI: 10.1111/j.0963-7214.2004.01301001.x.

[104]   F. Goudail, P. Réfrégier, and G. Delyon. "Bhattacharyya distance as a contrast parameter for statistical processing of noisy optical images". In: *J. Opt. Soc. Am. A* 21.7 (2004), pp. 1231–1240. DOI: 10.1364/JOSAA.21.001231.

[105]   P.-Y. Gousenbourger, E. Massart, and P.-A. Absil. "Data fitting on manifolds with composite Bézier-like curves and blended cubic splines". In: *J. Math. Imaging Vision* 61.5 (2019), pp. 645–671. DOI: 10.1007/s10851-018-0865-2.

[106]   G. Graßhoff et al. *Ancient Sundials*. Edition Topoi. 2016. DOI: 10.17171/1-1.

[107]   J. R. Gray, C. F. Chabris, and T. S. Braver. "Neural mechanisms of general fluid intelligence". In: *Nat. Neurosci.* 6.3 (2003), pp. 316–322. DOI: 10.1038/nn1014.

[108]   P. Grohs, M. Holler, and A. Weinmann, eds. *Handbook of Variational Methods for Nonlinear Geometric Data*. Berlin: Springer, 2020. DOI: 10.1007/978-3-030-31351-7.

[109]   N. Guigui and X. Pennec. "Numerical accuracy of ladder schemes for parallel transport on manifolds". In: *Found. Comput. Math.* (2021). DOI: 10.1007/s10208-021-09515-x.

[110]   X. Guorong, C. Peiqi, and W. Minhui. "Bhattacharyya distance feature selection". In: *Proceedings of 13th International Conference on Pattern Recognition*. Vol. 2. IEEE, 1996, pp. 195–199. DOI: 10.1109/ICPR.1996.546751.

[111] C. Hafemeister and R. Satija. "Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression". In: *Genome Biol.* 20.1 (2019), p. 269. DOI: 10.1186/s13059-019-1874-1.

[112] A. A. Hagberg, D. A. Schult, and P. J. Swart. "Exploring network structure, dynamics, and function using NetworkX". In: *Proceedings of the 7th Python in Science Conference*. Pasadena, CA USA, 2008, pp. 11–15.

[113] M. Hallin and J. Jurečková. "Equivariant estimation: theory". In: *Wiley StatsRef: Statistics Reference Online*. Atlanta: American Cancer Society, 2014. DOI: 10.1002/9781118445112.stat07378.

[114] M. Hanik and C. von Tycowicz. *Triangle meshes of shadow-recieving surfaces of ancient sundials*.
https://www.zib.de/ext-data/sundial_shadow_surfaces/meshes.zip. 2022. DOI: 10.12752/8425.

[115] M. Hanik et al. "Nonlinear regression on manifolds for shape analysis using intrinsic Bézier splines". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Springer, 2020, pp. 617–626. DOI: 10.1007/978-3-030-59719-1_60.

[116] W. K. Härdle and L. Simar. *Applied Multivariate Statistical Analysis*. Berlin: Springer, 2019. DOI: 10.1007/978-3-030-26006-4.

[117] T. He et al. "Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics". In: *NeuroImage* 206 (2020), p. 116276. ISSN: 1053–8119. DOI: 10.1016/j.neuroimage.2019.116276.

[118] B. Heeren et al. "Exploring the geometry of the space of shells". In: *Comput. Graph. Forum* 33.5 (2014), pp. 247–256. DOI: 10.1111/cgf.12450.

[119] B. Heeren et al. "Splines in the space of shells". In: *Proceedings of the Symposium on Geometry Processing*. SGP '16. Eurographics Association, 2016, pp. 111–120.

[120] B. Heeren et al. "Principal geodesic analysis in the space of discrete shells". In: *Comput. Graph. Forum* 37.5 (2018), pp. 173–184. DOI: 10.1111/cgf.13500.

[121] S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*. Vol. 34. Graduate Studies in Mathematics. Providence, RI, USA: American Mathematical Society, 2001. DOI: 10.1090/gsm/034.

[122] E. Hellinger. "Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen." In: *J. für die Reine und Angew. Math.* 1909.136 (1909), pp. 210–271. DOI: 10.1515/crll.1909.136.210.

[123] H. Hendriks and Z. Landsman. "Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions". In: *J. Multivar. Anal.* 67.2 (1998), pp. 227–243. DOI: 10.1006/jmva.1998.1776.

[124] N. J. Higham. *Functions of Matrices: Theory and Computation*. Philadelphia: SIAM, 2008. DOI: 10.1137/1.9780898717778.

[125]   J. Hinkle, P. T. Fletcher, and S. Joshi. "Intrinsic polynomials for regression on Riemannian manifolds". In: *J. Math. Imaging Vision* 50.1 (2014), pp. 32–52. DOI: 10.1007/s10851-013-0489-5.

[126]   Y. Hong et al. "Group testing for longitudinal data". In: *International Conference on Information Processing in Medical Imaging*. Springer, 2015, pp. 139–151. DOI: 10.1007/978-3-319-19992-4_11.

[127]   Z. Huang et al. "Deep learning on Lie groups for skeleton-based action recognition". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 1243–1252. DOI: 10.1109/CVPR.2017.137.

[128]   S. F. Huckemann and B. Eltzner. "Data analysis on nonstandard spaces". In: *Wiley Interdiscip. Rev. Comput. Stat.* 13.3 (2021), e1526. DOI: 10.1002/wics.1526.

[129]   S. F. Huckemann, T. Hotz, and A. Munk. "Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions". In: *Stat. Sin.* (2010), pp. 1–58.

[130]   M. Hüttig. "The conical sundial from Thyrrheion—reconstruction and error analysis of a displaced antique sundial". In: *Arch. Hist. Exact Sci.* 55.2 (2000), pp. 163–176.

[131]   A. Iserles et al. "Lie-group methods". In: *Acta Numerica* 9 (2000), pp. 215–365. DOI: 10.1017/S0962492900002154.

[132]   K. A. Janes and M. B. Yaffe. "Data-driven modelling of signal-transduction networks". In: *Nat. Rev. Mol. Cell Biol.* 7.11 (2006), pp. 820–828. DOI: 10.1038/nrm2041.

[133]   R. Jiang et al. "Gender differences in connectome-based predictions of individualized intelligence quotient and sub-domain scores". In: *Cereb. Cortex* 30.3 (2020), pp. 888–900. DOI: 10.1093/cercor/bhz134.

[134]   J. Jin and J. Shi. "Automatic feature extraction of waveform signals for in-process diagnostic performance improvement". In: *J. Intell. Manuf.* 12.3 (2001), pp. 257–268. DOI: 10.1023/A:1011248925750.

[135]   I. M. Johnstone and D. M. Titterington. "Statistical challenges of high-dimensional data". In: *Philos. Trans. Royal Soc. A* 367.1906 (2009), pp. 4237–4253. DOI: 10.1098/rsta.2009.0159.

[136]   A. Jones. "Down to the Hour: Short Time in the Ancient Mediterranean and Near East". In: Leiden, Netherlands: Brill, 2019. Chap. Greco-Roman sundials: precision and displacement, pp. 125–157. DOI: 10.1163/9789004416291_006.

[137]   J. Jost. *Riemannian Geometry and Geometric Analysis*. 7th ed. Berlin: Springer, 2017. DOI: 10.1007/978-3-319-61860-9.

[138]   T. Kailath. "The divergence and Bhattacharyya distance measures in signal selection". In: *IEEE Trans. Commun.* 15.1 (1967), pp. 52–60. DOI: 10.1109/TCOM.1967.1089532.

[139]   H. Karcher. "Riemannian center of mass and mollifier smoothing". In: *Commun. Pure Appl. Math.* 30.5 (1977), pp. 509–541. DOI: 10.1002/cpa.3160300502.

[140] H. Karcher. *Riemannian center of mass and so called Karcher mean*. 2014. arXiv: 1407.2087.

[141] J. H. Kellgren and J. S. Lawrence. "Radiological assessment of osteo-arthrosis." In: *Ann. Rheum. Dis.* 16 4 (1957), pp. 494–502. DOI: 10.1136/ard.16.4.494.

[142] D.G. Kendall et al. *Shape and Shape Theory*. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: Wiley, 2009. DOI: 10.1002/9780470317006.

[143] W. S. Kendall. "Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence". In: *Proc. London Math. Soc.* s3-61.2 (1990), pp. 371–406. DOI: 10.1112/plms/s3-61.2.371.

[144] W. S. Kendall. "Convexity and the hemisphere". In: *J. London Math. Soc.* s2-43.3 (1991), pp. 567–576. DOI: 10.1112/jlms/s2-43.3.567.

[145] A. Kheyfets, W. A. Miller, and G. Newton. "Schild's ladder parallel transport procedure for an arbitrary connection". In: *Int. J. Theor. Phys.* 39 (Dec. 2000), pp. 2891–2898. DOI: 10.1023/A:1026473418439.

[146] H. J. Kim et al. "Multivariate general linear models (MGLM) on Riemannian manifolds with applications to statistical analysis of diffusion weighted images". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 2705–2712. DOI: 10.1109/CVPR.2014.352.

[147] J. Kim et al. "Edge-labeling graph neural network for few-shot learning". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2019, pp. 11–20. DOI: 10.1109/CVPR.2019.00010.

[148] K.-R. Kim et al. "Smoothing splines on Riemannian manifolds, with applications to 3D shape space". In: *J. R. Stat. Soc. Series B Stat. Methodol.* 83.1 (2021), pp. 108–132. DOI: 10.1111/rssb.12402.

[149] D. P. Kingma and J. Ba. *Adam: a method for stochastic optimization*. 2017. arXiv: 1412.6980.

[150] T. N. Kipf and M. Welling. *Semi-supervised classification with graph convolutional networks*. 2017. arXiv: 1609.02907.

[151] M. Kubat. *An Introduction to Machine Learning*. Berlin: Springer, 2017. DOI: 10.1007/978-3-319-63913-0.

[152] R. R. Laxton and J. Restorick. "Seriation by similarity and consistency". In: *CAA 89. Computer Applications and Quantitative Methods in Archaeology*. BAR International Series 48. Oxford: Oxbow Books, 1989, pp. 215–225.

[153] H. Le. "Locating Fréchet means with application to shape spaces". In: *Adv. Appl. Probab.* 33.2 (2001), pp. 324–338. DOI: 10.1017/s0001867800010818.

[154] H. Le. "Estimation of Riemannian barycentres". In: *LMS J. Comput. Math.* 7 (2004), pp. 193–200. DOI: 10.1112/S1461157000001091.

[155] John M Lee. *Introduction to smooth manifolds*. Berlin: Springer, 2013. DOI: 10.1007/978-1-4419-9982-5.

[156] J. J. Locascio and A. Atri. "An overview of longitudinal data analysis methods for neurological research". In: *Dementia and geriatric cognitive disorders extra* 1.1 (2011), pp. 330–357. DOI: 10.1159/000330228.

[157] M. Loeve. *Probability theory*. Mineaola, NY, USA: Dover Publications, 2017.

[158] M. Louis et al. "A fanning scheme for the parallel transport along geodesics on Riemannian manifolds". In: *SIAM J. Numer. Anal.* 56.4 (2018), pp. 2563–2584. DOI: 10.1137/17M1130617.

[159] A. Madesis et al. "Review of mitral valve insufficiency: repair or replacement". In: *J. Thorac. Dis.* 6.Suppl 1 (2014), S39. DOI: 10.3978/j.issn.2072-1439.2013.10.20.

[160] T. Madsen. "Seriation and multivariate statistics". In: *CAA 89. Computer Applications and Quantitative Methods in Archaeology*. BAR International Series 48. Oxford: Oxbow Books, 1989, pp. 205–214.

[161] A. Mallasto and A. Feragen. "Wrapped Gaussian process regression on Riemannian manifolds". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA: IEEE, 2018, pp. 5580–5588. DOI: 10.1109/CVPR.2018.00585.

[162] L. C. Marsh and D. R. Cormier. *Spline Regression Models*. 137. Thousand Oaks, CA, USA: Sage, 2001. DOI: 10.4135/9781412985901.

[163] R. L. Mason and J. C. Young. *Multivariate Statistical Process Control with Industrial Applications*. Philadelphia: SIAM, 2002. DOI: 10.1137/1.9780898718461.

[164] M. Mcdaniel. "Big-brained people are smarter: a meta-analysis of the relationship between in vivo brain volume and intelligence". In: *Intelligence* 33 (2005), pp. 337–346. DOI: 10.1016/j.intell.2004.11.005.

[165] N. Miolane and X. Pennec. "Computing bi-invariant pseudo-metrics on Lie groups for consistent statistics". In: *Entropy* 17 (2015), pp. 1850–1881. DOI: 10.3390/e17041850.

[166] M. Moakher. "Means and averaging in the group of rotations". In: *SIAM J. Matrix Anal. Appl* 24.1 (2002), pp. 1–16. DOI: 10.1137/S0895479801383877.

[167] M. Moakher. "A differential geometric approach to the geometric mean of symmetric positive-definite matrices". In: *SIAM J. Matrix Anal. Appl* 26.3 (2005), pp. 735–747. DOI: 10.1137/S0895479803436937.

[168] A. H. Al-Mohy and N. J. Higham. "Computing the Fréchet derivative of the matrix exponential, with an application to condition number estimation". In: *SIAM J. Matrix Anal. Appl.* 30.4 (2009), pp. 1639–1657. DOI: 10.1137/080716426.

[169] A. H. Al-Mohy, N. J. Higham, and S. D. Relton. "Computing the Fréchet derivative of the matrix logarithm and estimating the condition number". In: *SIAM J. Sci. Comput.* 35.4 (2013), pp. C394–C410. DOI: 10.1137/120885991.

[170] L. Mucke. "Alzheimer's disease". In: *Nature* 461.7266 (2009), pp. 895–897. DOI: 10.1038/461895a.

[171] S. G. Mueller et al. "Hippocampal atrophy patterns in mild cognitive impairment and Alzheimer's disease". In: *Hum. Brain Mapp.* 31.9 (2010), pp. 1339–1347. DOI: 10.1002/hbm.20934.

[172] P. Muralidharan and P. T. Fletcher. "Sasaki metrics for analysis of longitudinal data on manifolds". In: *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 1027–1034. DOI: 10.1109/CVPR.2012.6247780.

[173] E. Nava-Yazdani, H.-C. Hege, and C. von Tycowicz. "A Hierarchical Geodesic Model for Longitudinal Analysis on Manifolds". In: *J. Math. Imaging Vision* (2022). DOI: 10.1007/s10851-022-01079-x.

[174] M. Niethammer, Y. Huang, and F.-X. Vialard. "Geodesic regression for image time-series". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*. Springer, 2011, pp. 655–662. DOI: 10.1007/978-3-642-23629-7_80.

[175] J. O'Brien and R. L. Lyman. *Seriation, Stratigraphy, and Index Fossils. The Backbone of Archaeological Dating*. Berlin: Springer, 1998.

[176] P. Olver, G. Sapiro, and A. Tannenbaum. "Differential invariant signatures and flows in computer vision: a symmetry group approach". In: *Geometry-driven diffusion in computer vision*. Berlin: Springer, 1994, pp. 255–306. DOI: 10.1007/978-94-017-1699-4_11.

[177] Maks Ovsjanikov et al. "Computing and processing correspondences with functional maps". In: *SIGGRAPH ASIA 2016 Courses*. New York, NY, USA: ACM, 2016, pp. 1–60. DOI: 10.1145/3084873.3084877.

[178] G. Pai et al. "On geometric invariants, learning, and recognition of shapes and forms". In: *Handbook of Variational Methods for Nonlinear Geometric Data*. Berlin: Springer, 2020, pp. 443–461. DOI: 10.1007/978-3-030-31351-7_16.

[179] G. S. P. Pamplona et al. "Analyzing the association between functional connectivity of the brain and intellectual performance". In: *Front. Hum. Neusci.* 9 (2015), p. 61. DOI: 10.3389/fnhum.2015.00061.

[180] L. Pardo. *Statistical Inference Based on Divergence Measures*. London: Chapman and Hall, 2005. DOI: 10.1201/9781420034813.

[181] F. C. Park, J .E. Bobrow, and S. R. Ploen. "A Lie group formulation of robot dynamics". In: *Int. J. Robot. Res.* 14.6 (1995), pp. 609–618. DOI: 10.1177/027836499501400606.

[182] V. Patrangenaru and L. Ellingson. *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. Boca Raton, FL, USA: CRC Press, 2015. DOI: 10.1201/b18969.

[183] F. Pedregosa et al. "Scikit-learn: machine learning in Python". In: *J. Mach. Learn. Res.* 12 (2011), pp. 2825–2830.

[184] X. Pennec. "Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements". In: *J. Math. Imaging Vision* 25 (2006), pp. 127–154. DOI: 10.1007/s10851-006-6228-4.

[185] X. Pennec. "Barycentric subspace analysis on manifolds". In: *The Annals of Statistics* 46.6A (2018), pp. 2711–2746. DOI: 10.1214/17-AOS1636.

[186] X. Pennec and V. Arsigny. "Exponential barycenters of the canonical Cartan connection and invariant means on Lie groups". In: *Matrix Information Geometry*. Berlin: Springer, 2013, pp. 123–166. DOI: 10.1007/978-3-642-30232-9_7.

[187] X. Pennec, P. Fillard, and N. Ayache. "A Riemannian framework for tensor computing". In: *Int. J. Comput. Vis.* 66.1 (2006), pp. 41–66. DOI: 10.1007/s11263-005-3222-z.

[188] X. Pennec and M. Lorenzi. "5 - Beyond Riemannian geometry: the affine connection setting for transformation groups". In: *Riemannian Geometric Statistics in Medical Image Analysis*. London: Academic Press, 2020, pp. 169–229. DOI: 10.1016/B978-0-12-814725-2.00012-1.

[189] X. Pennec, S. Sommer, and P. T. Fletcher, eds. *Riemannian Geometric Statistics in Medical Image Analysis*. London: Academic Press, 2020. DOI: 10.1016/C2017-0-01561-6.

[190] P. Petersen. *Riemannian Geometry*. Berlin: Springer, 2006. DOI: 10.1007/978-3-319-26654-1.

[191] G. Peyré. "Manifold models for signals and images". In: *Comput. Vis. Image Underst.* 113.2 (2009), pp. 249–260. DOI: 10.1016/j.cviu.2008.09.003.

[192] U. Pinkall and K. Polthier. "Computing discrete minimal surfaces and their conjugates". In: *Exp. Math.* 2.1 (1993), pp. 15–36. DOI: 10.1080/10586458.1993.10504266.

[193] M. Pistellato, A. Traviglia, and F. Bergamasco. "Geolocating time: digitisation and reverse engineering of a Roman sundial". In: *European Conference on Computer Vision*. Springer. 2020, pp. 143–158. DOI: 10.1007/978-3-030-66096-3_11.

[194] S. M. Pizer et al. "6 - Object shape representation via skeletal models (s-reps) and statistical analysis". In: *Riemannian Geometric Statistics in Medical Image Analysis*. London: Academic Press, 2020, pp. 233–271. DOI: 10.1016/B978-0-12-814725-2.00014-5.

[195] T. Popiel and L. Noakes. "Bézier curves and C2 interpolation in Riemannian manifolds". In: *J. Approx. Theory* 148.2 (2007), pp. 111–127. DOI: 10.1016/j.jat.2007.03.002.

[196] F. Porikli, O. Tuzel, and P. Meer. "Covariance tracking using model update based on Lie algebra". In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Vol. 1. IEEE, 2006, pp. 728–735. DOI: 10.1109/CVPR.2006.94.

[197] M. M. Postnikov. *Geometry VI: Riemannian Geometry*. Encyclopaedia of Mathematical Sciences. Berlin: Springer, 2013. DOI: 10.1007/978-3-662-04433-9.

[198] X. Qi et al. "3D graph neural networks for RGBD semantic segmentation". In: *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, 2017, pp. 5199–5208. DOI: 10.1109/ICCV.2017.556.

[199] M. Qu, Y. Bengio, and J. Tang. "GMNN: graph Markov neural networks". In: *International conference on machine learning*. PMLR, 2019, pp. 5241–5250.

[200] I. U. Rahman et al. "Multiscale representations for manifold-valued data". In: *Multiscale Model. Simul.* 4.4 (2005), pp. 1201–1232. DOI: 10.1137/050622729.

[201] C. A. Raji et al. "Age, Alzheimer disease, and brain structure". In: *Neurology* 73.22 (2009), pp. 1899–1905. DOI: 10.1212/WNL.0b013e3181c3f293.

[202] C. R. Rao. "A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance". In: (1995).

[203] A. Reichenberg, J. Mill, and J. H. MacCabe. "Epigenetics, genomic mutations and cognitive function". In: *Cogn. Neuropsychiatry* 14.4–5 (2009), pp. 377–390. DOI: 10.1080/13546800902978417.

[204] A. L. Reiss et al. "Brain development, gender and IQ in children: a volumetric imaging study". In: *Brain* 119.5 (1996), pp. 1763–1774. DOI: 10.1093/brain/119.5.1763.

[205] C. Renfrew and P. Bahn. *Archaeology: Theories, Methods and Practice*. 4th ed. London: Thames and Hudson, 2004.

[206] Q. Rentmeesters. "Algorithms for data fitting on some common homogeneous spaces". PhD thesis. Université Catholique de Louvain, Louvain, Belgium, 2013.

[207] E. Rinner, B. Fritsch, and G. Graßhoff. "Die unvollendete Sonnenuhr von der Agora der Italiker auf Delos". In: *eTopoi. J. Anc. Studies*. Berlin Studies of the Ancient World 2 (2013), pp. 111–130. DOI: 10.17171/4-2-4.

[208] R.-A. Rohan. "Some remarks on the exponential map on the groups SO(n) and SE(n)". In: *Proceedings of the Fourteenth International Conference on Geometry, Integrability and Quantization*. 2013, pp. 160–175. DOI: 10.7546/giq-14-2013-160-175.

[209] V. K. Rohatgi and A. K. Md. Ehsanes Saleh. *An introduction to probability and statistics*. Chichester, UK: John Wiley and Sons, 2015. DOI: 10.1002/9781118799635.

[210] M. Rumpf and B. Wirth. "Variational time discretization of geodesic calculus". In: *IMA J. Numer. Anal.* 35.3 (2014), pp. 1011–1046. DOI: 10.1093/imanum/dru027.

[211] C. Samir et al. "A gradient-descent method for curve fitting on Riemannian manifolds". In: *Found. Comput. Math.* 12.1 (2012), pp. 49–73. DOI: 10.1007/s10208-011-9091-7.

[212] P.-E. Sarlin et al. "Superglue: learning feature matching with graph neural networks". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. IEEE, 2020, pp. 4938–4947. DOI: 10.1109/CVPR42600.2020.00499.

[213] K. Schaldach. "Time and Cosmos in Greco-Roman Antiquity". In: Princeton University Press, 2016. Chap. Measuring the hours: sundials, water clocks, and portable sundials, pp. 63–93. DOI: 10.1163/9789004416291_002.

[214] M. Schlichtkrull et al. "Modeling relational data with graph convolutional networks". In: *European semantic web conference*. Springer, 2018, pp. 593–607. DOI: 10.1007/978-3-319-93417-4_38.

[215] J. Schulz et al. "Non-linear hypothesis testing of geometric object properties of shapes applied to hippocampi". In: *J. Math. Imaging Vision* 54 (2016), pp. 15–34. DOI: 10.1007/s10851-015-0587-7.

[216] A. Scott. "A parametric approach to seriation". In: *Computing the Past, CAA 92. Computer Applications and Quantitiative Methods in Archaeology*. Aarhus: Aarhus University Press, 1992, pp. 317–324.

[217] G. A. F. Seber. *Multivariate Observations*. Vol. 252. Chichester, UK: John Wiley and Sons, 2009. DOI: 10.1002/9780470316641.

[218] H. Sengar et al. "Detecting VoIP floods using the Hellinger distance". In: *IEEE Trans. Parallel Distrib. Syst.* 19.6 (2008), pp. 794–805. DOI: 10.1109/TPDS.2007.70786.

[219] H. Shappell et al. "A paradigm for longitudinal complex network analysis over patient cohorts in neuroscience". In: *Netw. Sci.* 7.2 (2019), pp. 196–214.

[220] K.-k. Shen et al. "Detecting global and local hippocampal shape changes in Alzheimer's disease using statistical shape models". In: *Neuroimage* 59.3 (2012), pp. 2155–2166. DOI: 10.1016/j.neuroimage.2011.10.014.

[221] X. Shen et al. "Normalization and integration of large-scale metabolomics data using support vector regression". In: *Metabolomics* 12.5 (2016), p. 89. DOI: 10.1007/s11306-016-1026-5.

[222] X. Shen et al. "Using connectome-based predictive modeling to predict individual behavior from brain connectivity". In: *Nat. Protoc.* 12.3 (2017), pp. 506–518. DOI: 10.1038/nprot.2016.178.

[223] X. Shi et al. "Intrinsic regression models for manifold-valued data". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2009, pp. 192–199. DOI: 10.1007/978-3-642-04271-3_24.

[224] K. Siddiqi and S. M. Pizer, eds. *Medial representations: mathematics, algorithms and applications*. Berlin: Springer, 2008. DOI: 10.1007/978-1-4020-8658-8.

[225] N. Singh et al. "A Hierarchical Geodesic model for diffeomorphic longitudinal shape analysis". In: *Information Processing in Medical Imaging*. Springer, 2013, pp. 560–571. DOI: 10.1007/978-3-642-38868-2_47.

[226] N. Singh et al. "Hierarchical geodesic models in diffeomorphisms". In: *Int. J. Comput. Vis.* 117.1 (2016), pp. 70–92. DOI: 10.1007/s11263-015-0849-2.

[227] J. Skranes et al. "Cortical surface area and IQ in very-low-birth-weight (VLBW) young adults". In: *Cortex* 49.8 (2013), pp. 2264–2271. DOI: 10.1016/j.cortex.2013.06.001.

[228] S.T. Smith. "Covariance, subspace, and intrinsic Cramér-Rao bounds". In: *IEEE Trans. Signal Process.* 53.5 (2005), pp. 1610–1630. DOI: 10.1109/TSP.2005.845428.

[229] S. Sommer, F. Lauze, and M. Nielsen. "Optimization over geodesics for exact principal geodesic analysis". In: *Adv. Comput. Math.* 40.2 (2014), pp. 283–313. DOI: 10.1007/s10444-013-9308-1.

[230] O. Sporns, G. Tononi, and R. Kötter. "The human connectome: a structural description of the human brain". In: *PLoS Comput. Biol.* 1.4 (2005). DOI: 10.1371/journal.pcbi.0010042.

[231] A. Srivastava and E. Klassen. *Functional and Shape Data Analysis*. Berlin: Springer, 2016. DOI: 10.1007/978-1-4939-4020-2.

[232] Blausen.com staff. "Medical gallery of Blausen Medical 2014". WikiJournal of Medicine 1 (2), ISSN 2002-4436 - Own work. License: CC BY 3.0. 2014. DOI: 10.15347/wjm/2014.010. URL: https://commons.wikimedia.org/w/index.php?curid=33041242.

[233] D. Stalling, M. Westerhoff, and H.-C. Hege. "Amira: a highly interactive system for visual data analysis." In: *The Visualization Handbook*. London: Academic Press, 2005. Chap. 38, pp. 749–767. DOI: 10.1016/B978-012387582-2/50040-X.

[234] H. Tabia et al. "Covariance descriptors for 3D shape matching and retrieval". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 4185–4192. DOI: 10.1109/CVPR.2014.533.

[235] Gary KL Tam et al. "Registration of 3D point clouds and meshes: A survey from rigid to nonrigid". In: *IEEE Trans. Vis. Comput. Graph.* 19.7 (2012), pp. 1199–1217. DOI: 10.1109/TVCG.2012.310.

[236] A. Tarighati, J. Gross, and J. Jaldén. "Decentralized hypothesis testing in energy harvesting wireless sensor networks". In: *IEEE Trans. Signal Process.* 65.18 (2017), pp. 4862–4873. DOI: 10.1109/TSP.2017.2716909.

[237] L. Tautz et al. "Combining position-based dynamics and gradient vector flow for 4D mitral valve segmentation in TEE sequences". In: *Int. J. Comput. Assist. Radiol. Surg.* 15.1 (2020), pp. 119–128. DOI: 10.1007/s11548-019-02071-4.

[238] C. J. Taylor and D. J. Kriegman. *Minimization on the Lie group SO(3) and related manifolds*. Tech. rep. 9405. Yale University, 1994.

[239] D. W. Thompson. *On Growth and Form*. Cambridge, UK: Cambridge University Press, 1992. DOI: 10.1017/cbo9781107325852.

[240] S. Tordjman et al. "Repint of "Reframing autism as a behavioral syndrome and not a specific mental disorder: implications of genetic and phenotypic heterogeneity"". In: *Neurosci. Biobehav. Rev.* 89 (2018), pp. 132–150. DOI: 10.1016/j.neubiorev.2018.01.014.

[241] D. Tosato et al. "Multi-class classification on Riemannian manifolds for video surveillance". In: *European Conference on Computer Vision*. Springer, 2010, pp. 378–391. DOI: 10.1007/978-3-642-15552-9_28.

[242]  A. Trouvé. "Diffeomorphisms groups and pattern matching in image analysis". In: *Int. J. Comput. Vis.* 28.3 (1998), pp. 213–221. DOI: 10.1023/A:1008001603737.

[243]  A. Trouvé and L. Younes. "Metamorphoses through Lie group action". In: *Found. Comput. Math.* 5.2 (2005), pp. 173–198. DOI: 10.1007/s10208-004-0128-z.

[244]  O. Tuzel, F. Porikli, and P. Meer. "Region covariance: a fast descriptor for detection and classification". In: *European Conference on Computer Vision*. Springer, 2006, pp. 589–600. DOI: 10.1007/11744047_45.

[245]  C. von Tycowicz et al. "An efficient Riemannian statistical shape model using differential coordinates". In: *Med. Image Anal.* 43 (2018), pp. 1–9. DOI: 10.1016/j.media.2017.09.004.

[246]  N. Tzourio-Mazoyer et al. "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain". In: *NeuroImage* 15.1 (2002), pp. 273–289. DOI: 10.1006/nimg.2001.0978.

[247]  V. Veeriah, N. Zhuang, and G.-J. Qi. "Differential recurrent neural networks for action recognition". In: *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, 2015, pp. 4041–4049. DOI: 10.1109/iccv.2015.460.

[248]  R. Vemulapalli, F. Arrate, and R. Chellappa. "Human action recognition by representing 3D skeletons as points in a Lie group". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 588–595. DOI: 10.1109/CVPR.2014.82.

[249]  R. Vemulapalli and R. Chellapa. "Rolling rotations for recognizing human actions from 3D skeletal data". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016, pp. 4471–4479. DOI: 10.1109/CVPR.2016.484.

[250]  F. Wahid et al. "A multiple regression approach to normalization of spatiotemporal gait features". In: *J. Appl. Biomech.* 32.2 (2016), pp. 128–139. DOI: 10.1123/jab.2015-0035.

[251]  J. Wallner, E. Nava-Yazdani, and P. Grohs. "Smoothness properties of Lie group subdivision schemes". In: *Multiscale Model. Simul.* 6.2 (2007), pp. 493–505. DOI: 10.1137/060668353.

[252]  F. W. Warner. *Foundations of Differentiable Manifolds and Lie Groups*. Vol. 94. Berlin: Springer, 2013. DOI: 10.1007/978-1-4757-1799-0.

[253]  J. A. Wegelin. *A Survey of Partial Least Squares (PLS) Methods, with Emphasis on the Two-Block Case*. Tech. rep. 371. University of Washington, Department of Statistics, 2000. URL: https://stat.uw.edu/sites/default/files/files/reports/2000/tr371.pdf.

[254]  E. Wong et al. "Riemannian regression and classification models of brain networks applied to autism". In: *Connectomics in NeuroImaging*. Springer, 2018, pp. 78–87. DOI: 10.1007/978-3-030-00755-3_9.

[255]    R. P. Woods. "Characterizing volume and surface deformations in an atlas framework: theory, applications, and implementation". In: *NeuroImage* 18.3 (2003), pp. 769–788. DOI: 10.1016/s1053-8119(03)00019-3.

[256]    A. Woolgar et al. "Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex". In: *Proc. Natl. Acad. Sci. U.S.A.* 107.33 (2010), pp. 14899–14902. DOI: 10.1073/pnas.1007928107.

[257]    Z. Wu et al. "A comprehensive survey on graph neural networks". In: *IEEE Trans. Neural Netw. Learn. Syst.* 32.1 (2020), pp. 4–24. DOI: 10.1109/TNNLS.2020.2978386.

[258]    M. Xia, J. Wang, and Y. He. "BrainNet Viewer: a network visualization tool for human brain connectomics". In: *PloS one* 8.7 (2013), e68910. DOI: 10.1371/journal.pone.0068910.

[259]    G. Xuan et al. "Feature selection based on the Bhattacharyya distance". In: *18th International Conference on Pattern Recognition (ICPR'06)*. Vol. 4. IEEE, 2006, pp. 957–957. DOI: 10.1109/ICPR.2006.557.

[260]    X. Xue et al. "Model-based condition monitoring of PEM fuel cell using Hotelling T2 control limit". In: *J. Power Sources* 162.1 (2006), pp. 388–399. DOI: 10.1016/j.jpowsour.2006.07.004.

[261]    M. A. Yamin et al. "Geodesic clustering of positive definite matrices for classification of mental disorder using brain functional connectivity". In: *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–5. DOI: 10.1109/IJCNN48605.2020.9207687.

[262]    N. Ye et al. "Multivariate statistical analysis of audit trails for host-based intrusion detection". In: *IEEE Trans. Comput.* 51.7 (2002), pp. 810–820. DOI: 10.1109/TC.2002.1017701.

[263]    S. Yin et al. "A review on basic data-driven approaches for industrial process monitoring". In: *IEEE Trans. Ind. Electron.* 61.11 (2014), pp. 6418–6428. DOI: 10.1109/TIE.2014.2301773.

[264]    K. You and H.-J. Park. "Re-visiting Riemannian geometry of symmetric positive definite matrices for the analysis of functional connectivity". In: *NeuroImage* 225 (2021), p. 117464. DOI: 10.1016/j.neuroimage.2020.117464.

[265]    L. Younes. *Shapes and Diffeomorphisms*. Berlin: Springer, 2019. DOI: 10.1007/978-3-662-58496-5.

[266]    Y. Yuan et al. "Local polynomial regression for symmetric positive definite matrices". In: *J. R. Stat. Soc. Series B Stat. Methodol.* 74.4 (2012), pp. 697–719. DOI: 10.1111/j.1467-9868.2011.01022.x.

[267]    S. Yun et al. "Graph transformer networks". In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc., 2019.

[268]    S. Zafeiriou et al. "Guest editorial: non-Euclidean machine learning". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 44.02 (2022), pp. 723–726. DOI: 10.1109/TPAMI.2021.3129857.

*Bibliography*

[269]  C. Zhang et al. "Heterogeneous graph neural network". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2019, pp. 793–803. DOI: 10.1145/3292500.3330961.

[270]  H. Zhu et al. "Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging". In: *J. Am. Stat. Assoc.* 104.487 (2009), pp. 1203–1212. DOI: 10.1198/jasa.2009.tm08096.