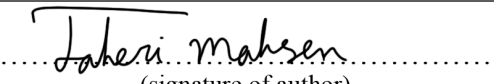




Universitetet
i Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study programme/specialisation: Masters of Mathematics and Physics Statistics	Spring / Autumn semester, 2020 Open/ Confidential
Author: Mohsen Taheri Shalmani	 (signature of author)
Supervisor(s): Jörn Schulz	
Title of master's thesis: Statistical Shape Analysis of Brain Structures	
Credits: 60	
Keywords: Skeletal representation, Hypothesis test, Hippocampus, s-rep, SPHARM-PDM, Composite Principal Nested Spheres, Weighted alignment, Distance matrix, Nonlinear PGA, Permutation	Number of pages:92..... + supplemental material/other: Stavanger, 14/06/2020 date/year



Faculty of Science and Technology
Department of Mathematics and Physics

Statistical Shape Analysis of Brain Structures

Master's Thesis in Mathematics and Physics

by

Mohsen Taheri Shalmani

Supervisor

Jörn Schulz

“If you can't explain it simply, you don't understand it well enough.”

Albert Einstein

Abstract

The purpose of this work is to study structural differences of the left hippocampus between patients with Parkinson's disease (PD) and healthy control group (CG) based on shape models like skeletal representation (s-rep) and spherical harmonics point distribution model (SPHARM-PDM). We apply a permutation test on the s-reps of CG and PD to detect significant differences between the means of their geometric object properties (GOPs). We also introduce a parametric test for s-rep, constructed on multivariate Hotelling's T^2 test. We discuss different methods of alignment, their impact on the result, and propose the elimination algorithm to have an adequate alignment. To make the test independent from the alignment, we propose a method according to distance matrices. We explain possible approaches to define mean and variation of directional data, including principal nested spheres (PNS), and principal geodesic analysis (PGA). Besides, we propose a non-linear PGA (NLPGA) on rotating tangent space of the unit sphere. Finally, we discuss the results of the hypothesis tests and show there are statistically significant differences between PD and CG.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Jörn Schulz, for his invaluable feedback and continued guidance throughout the year. I would like to thank Prof. Stephen Pizer for providing comprehensive discussions and imparting his knowledge in this study. I would like to thank Prof. Guido Alves for allowing me to access and analyze data from the ParkWest study. I would also like to acknowledge Prof. Jan Terje Kvaløy and Prof. Ian Dryden for providing constructive advice and guidance. I would like to thank Zhiyuan Liu for providing help and detailed instructions about the s-rep model fitting.

I wish to express my appreciation to my parents and my family: Mohammad Taheri, Zahra Khojasteh, Dr. Masoumeh Taheri, and Dr. Mahasti Taheri for their care and support.

Finally, I like to thank my beautiful wife, Paria, for her patience, support, and love.

Contents

Abstract	iii
Acknowledgements	iv
Abbreviations	viii
Symbols	ix
1 Introduction	1
1.1 Motivation	1
1.2 Scope	2
1.3 Contributions	3
2 Shape and Shape Space	5
2.1 Shape Representations	5
2.1.1 Landmark modeling and PDM	5
2.1.2 Medial representation m-rep	6
2.1.3 Skeletal representation s-rep	7
2.1.4 SPHARM-PDM	9
2.1.5 SPHARM-PDM plus skeletal PDM	11
2.1.6 Flat slabbed-shape middle part parameterization	11
2.2 Shape space	13
2.3 Manifolds	16
2.4 Shape distances	19
2.5 Alignment	20
2.5.1 Alignment of configurations	20
2.5.2 s-rep space	21
2.5.3 s-rep alignment	21
2.5.4 Weighted Procrustes alignment	23
2.5.5 Elimination algorithm	24
2.5.6 Weighted alignment by elimination	26
3 Dimensionality Reduction	29
3.1 Principal Component Analysis (PCA)	29

3.2	Principal Geodesic Analysis (PGA)	31
3.3	Principal Nested Sphere (PNS)	32
3.3.1	PNG	33
3.4	Non-Linear PGA (NLPGA)	33
3.5	Composite Principal Nested Spheres (CPNS)	37
3.5.1	Composite Principal Nested Great Sphere (CPNG)	39
3.5.2	s-rep classification based on CPNG	40
4	Hypothesis Testing	41
4.1	Overview of hypothesis testing	41
4.1.1	Parametric vs. non-parametric approach	42
4.1.2	p -value	42
4.2	Multiple comparisons problem	43
4.2.1	FDR and FWER	43
4.3	Permutation test	44
4.3.1	Permutation distribution	44
4.4	Mean s-rep and s-rep test statistic	45
4.4.1	Mean s-rep by CPNS	45
4.4.2	Mean s-rep by GPA and PGA	47
4.4.3	s-rep test statistic	48
4.5	Non-parametric approach for s-reps	49
4.5.1	Partial test with FWER	49
4.5.2	Partial test with FDR	50
4.5.3	Global test for non-parametric approach	51
4.6	Parametric approach	53
4.6.1	Hotelling's T^2 test	53
4.6.2	s-rep partial tests with parametric approach	54
4.6.3	Global test for the parametric approach	55
4.6.4	SPHARM-PDM partial test with parametric approach	55
4.6.5	Partial test for triangle mesh normal vectors	55
4.7	Hypothesis test independent from the alignment	56
4.7.1	Hypothesis test with distance	56
4.7.2	Hypothesis test with angle for planer shapes	60
4.7.3	Example	61
5	Data Analysis	63
5.1	Data	63
5.2	Software and model fitting	64
5.3	Analysis	65
5.3.1	Global tests	65
5.3.2	s-rep analysis	66
5.3.3	SPHARM-PDM analysis	69
5.3.4	Further analysis	71
5.3.5	s-rep classification	71
5.3.6	PD with dementia	72
6	Discussion and Future Work	76

6.1	Discussion	76
6.2	Future work	77
6.2.1	Analysis of other brain parts and their relationship	77
6.2.2	Distance and Force	78
6.2.3	Classification improvement	79
6.2.4	Alignment	79
List of Figures		79
List of Tables		81
A Appendix		82
A.1	Analysis	82
A.1.1	Non-parametric result for SPHARM-PDM	82
A.1.2	Analysis of boundary normal vectors	82
A.1.3	Middle Surface PDM analysis	83
A.1.4	Analysis of SPHARM-PDM plus skeletal PDM	84
A.2	Example	84
A.2.1	Helmert submatrix	84
Bibliography		85

Abbreviations

ASL	A chieved S ignificance L evel
CDF	C umulative D istribution F unction
CG	C ontrol G roup
CPNG	C omposite P rincipal N ested G reat sphere
CPNS	C omposite P rincipal N ested S phere
FDR	F alse D iscovery R ate
FWER	F amily W ise E rror R ate
GLM	G eneralized L inear M odel
GOP	G eometric O bject P roperitiy
GPA	G eneralized P rocrustes A nalysis
i.i.d	independent and identically d istributed
m-rep	m edial r epresentation
MCF	M ean C urvature F low
MRI	M agnetic R esonance I maging
NLPCA	N on L inear P rincipal C omponent A nalysis
NLPGA	N on L inear P rincipal G eodesic A nalysis
PCA	P rincipal C omponent A nalysis
PD	P arkinson's D isease
PDD	P arkinson's D isease with D ementia
PGA	P rincipal G eodesic A nalysis
PNG	P rincipal N ested G reat sphere
PNS	P rincipal N ested S phere
SPHARM-PDM	S pherical H armonic P oint D istribution M odel
s-rep	s keletal r epresentation

Symbols

Symbol	Description
\mathbf{v}	Vectors and tuples are denoted by boldface letters, such as \mathbf{v} or \mathbf{V} .
$Cov(x, y)$	Covariance between x and y .
$\frac{d}{dx}$	Derivative with respect to x .
$\frac{\partial}{\partial t}$	Partial derivative with respect to t .
$E(x)$	Expectation of x .
$det(X)$	Determinant of X .
$I(\alpha)$	Indicator function, i.e. $I(\alpha) = 1$ if α is true, and 0 otherwise.
\mathbb{R}^m	The m -dimensional real coordinate system.
\forall	Logical sign "for all".
$\ \mathbf{x}\ $	Norm of \mathbf{x} .
s^\dagger	Discrete s-rep.
$\{x_i\}_{i=1}^n$	Set $\{x_1, \dots, x_n\}$.
$ X _c$	Cardinality of set X .
$ x $	Absolute value of x .
X^T	Transpose of X .
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2 .
$\mathcal{N}_m(\boldsymbol{\mu}, \Sigma)$	m -dimensional multivariate Normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ .
$P(X)$	Probability of event X .
$Pr\{X\}$	Probability statement relative to X .

Chapter 1

Introduction

1.1 Motivation

With the improvement of non-invasive medical imaging techniques like Magnetic Resonance Imaging (MRI) to picture the internal human organs, medical scientists became increasingly interested in studying shape and structure of internal organs in order to improve treatment and diagnosis. Often the goal is to scrutinize the morphological changes of a target object during a period or to detect the structural differences between organs among healthy controls and patients. In this manner, statistical shape analysis as a field of research to study shapes and objects within a population became the center of attention. However, data collection, shape visualization, generating models, and data analysis require collaboration between medical specialists, computer scientists, and statisticians. In this regard, there are evolving interdisciplinary research areas, including mathematical and medical sciences, to assist physicians by shedding light on the hidden aspects of diseases such as dementia.

Various studies show there is a direct relationship between the aetiology of mental diseases and shrinkage or deformation of brain parts. For example, a comprehensive study about chronic neurodegenerative disorders such as Alzheimer's disease (AD) and their psychological affection on patients' behavior and memory loss can be found in (Budson and Solomon, 2011, 2015). Although there are many reports about the brain lesion and abnormality in such diseases (Perl, 2010; Barber et al., 2000; Snowden et al., 1997), in many cases including early Parkinson's disease (PD) global brain atrophy and white matter hyperintensities (WMH) do not indicate significant differences between the normal control group (CG) and patients e.g. (Dalaker et al., 2009). To improve diagnosis, it is reasonable to focus on the most vulnerable brain structures. *"Hippocampus can be damaged by a variety of stimuli, e.g., stress, hypoxia, hypoperfusion, hypoglycemia, and seizures"* (Dhikav and Anand, 2011). Consequently, in neuropathological studies, hippocampus has been one of the first candidates for the analysis. Besides, there is strong evidence that dementia can

cause the hippocampus as part of the medial temporal lobe structure to lose volume substantially (Csernansky et al., 2005; Fox et al., 1996; Ikeda et al., 1994; Acton, 2012). Despite the fact that neurologists categorize Parkinson’s disease as a progressive neurodegenerative disorder that can be diagnosed after months of progression, it is still challenging to recognize and confirm it at early stages. Also, autopsy studies show that clinical diagnosis of PD, even with correctly identified symptoms, might be incorrect up to 25% (Pfeiffer et al., 2012, ch.51). Thereby, developing methods for early PD diagnosis, to start the treatment as soon as possible and increase the diagnosis accuracy is vital. Conventionally, medical scientists use scale measuring and volumetric analysis to study the structure of the hippocampus in PD (de la Monte et al., 1989; Junqué et al., 2005), and early PD (Brück et al., 2004). Although these types of analyses highlight the changes of hippocampal degeneration and atrophy, they are not suitable for detecting precise location of the differences specifically for the early PD, where the degeneration process has started just recently. One possible accurate approach is based on the radial distance (i.e., distance from the object’s medial core to the surface points) (Thompson et al., 2004). (Apostolova et al., 2012) applied radial distance analysis and showed a significant level of hippocampal atrophy in PD. There are some disadvantages in radial distance analysis, which we discuss briefly in Section 5.1. On the other hand, (Nobis et al., 2019) studied the hippocampal volume of over 19,700 people in the UK. The study discussed various parameters like sex, age, smoking, etc. affect the volume of the hippocampus, so the size of the hippocampus or temporal lobe varies even among healthy people. For this reason, we need more accurate object analysis methods to capture small disparities between corresponding elements of the object even when we remove the scale of the shapes. The question is whether there are any differences between shapes of study after removing size and scale or not. Therefore, the first motivation of this thesis is to implement a sensitive method of hypothesis testing (with and without removing scale) to detect locational differences of left hippocampi among a healthy control group and a group of patients with early PD.

To have an accurate testing method, we need suitable models that reflect accurate dissimilarity between the corresponding geometric object properties (GOP). One such model is the skeletal representation (s-rep) (Siddiqi and Pizer, 2008) introduced in Section 2.1.3 that characterizes the interior as well as the exterior properties of an object by a set of GOP including positions, radii (local widths), and directions. For statistical analysis of s-rep, (Schulz et al., 2016) designed a permutation test, which might be computationally expensive in practice. As a result, the second motivation of this thesis is to improve the current hypothesis testing method.

1.2 Scope

The next chapter intends to acquaint the reader with some advanced types of object representations for shape analysis, e.g., landmark modeling, point distribution model (PDM), s-rep, and

spherical harmonics point distribution model (SPHARM-PDM). These shape representations help us to compare objects with a high level of precision. Then it discusses the general definition of shape, shape space, shape distance, and mean shape on manifolds as prerequisites of shape analysis. Follow, different alignment methods are elaborated that is a necessary pre-processing step for the shape analysis.

Chapter 3 outlines several dimension reduction methods for data on Euclidean and non-Euclidean spaces, including Principal Component Analysis (PCA), Principal Geodesic Analysis (PGA), and Principal Nested Spheres Analysis (PNS). It also introduces a Non-Linear PGA (NLPGA) method on rotating tangent space of the unit sphere S^2 . Given that, analyzing s-rep data needs a method that covers both Euclidean and non-Euclidean data, Composite PNS (CPNS) will also be discussed.

Chapter 4 introduces the general concepts of hypothesis testing, p -value, multiple comparisons problem, and controlling false positives. It discusses different hypothesis testing methods suitable for each shape representation and tries to improve the available hypothesis testing method for s-rep. Furthermore, as a fast approach, it introduces an alternative testing method for s-rep with the normality assumption. Finally, in this chapter, we propose a novel hypothesis testing method independent from the alignment based on distance matrices.

In chapter 5, the proposed methods and concepts are applied to find the difference between the hippocampi of healthy controls and PD patients. The data set consists of MRI images provided by ParkWest ([ParkWest, 2020](#)) in collaboration with Stavanger University Hospital ([Helse Stavanger, 2020](#)). Data analysis workflow consists of three main parts: 1. Model fitting. 2. Implementation of different methods of hypothesis testing. 3. Analyzing tests' results and classification.

Chapter 6 discusses the results, summarises the study, and suggests possible directions for future work.

1.3 Contributions

In Section [2.1.6](#), we introduce a shape representation called Middle Surface PDM, for flat slabbed-objects. This representation parameterizes the middle part of the object, fits a nonlinear surface inside the shape, and approximates the s-rep skeletal locus's location. The idea to design this representation is to compare and support the analysis result of the s-rep skeletal PDM.

We introduce the elimination algorithm for PDMs and landmark models in Section [2.5.5](#). The algorithm detects suspicious points that cause the mean shape differences. Since the result of the hypothesis test depends on the alignment, we use the unsuspecting points to design a covariance

matrix for the weighted Procrustes alignment. This approach, by using weighted alignment, reduces the variance of the points with more similarity among the groups (see Section 2.5.6).

In this work, we propose a dimension reduction method in Section 3.4, analogous to PGA called NLPGA. The method fits a principal curve to the log-mapped data on the rotating tangent space. It tries to find a geodesic submanifold on the sphere, which maximizes the variance of the projected data. Further, we discuss the advantage of this method to find the sample mean on the sphere in comparison with Fréchet mean and PNS mean.

The computation cost to calculate mean s-rep by CPNS is exceptionally high. In this regard, to reduce the time of computation, we introduce a mean s-rep by substituting the PNS mean of the skeletal PDM with the Procrustes mean (see Section 4.4.2). This approach is helpful when we want to apply the permutation test to detect partial differences. In Section 4.6.2, we propose a parametric hypothesis testing method for s-rep as a fast approach by the normality assumption.

In Section 4.7.1 to make the hypothesis tests independent from the alignment, we propose a testing method based on GOP distances. The method tries to rank the GOPs (e.g., points and directions) by using distance matrices. Further, we use Kernel Density Function (KDE) on p -values to standardize the ranking procedure. In addition, we discuss a similar approach for planner shapes, which uses angles (instead of distances) to rank the points.

Chapter 2

Shape and Shape Space

In this chapter, we introduce shape representations relative to our study. We also become familiar with shape and shape space. Then we explain different methods of alignment and propose a weighted alignment method based on the elimination algorithm. Most of the definitions in this chapter are taken from (Dryden and Mardia, 2016; Pizer et al., 2013; Fletcher et al., 2004).

2.1 Shape Representations

One of the obstacles in statistical shape analysis is how to model the objects to be correspondent. In other words, we need to parameterize object representations by using geometric models that relate GOPs together. Among a wide variety of modeling methods, we mention some of them relevant to this work, including Landmark modeling, PDM, m-rep, s-rep, and SPHARM-PDM. Later in Chapter 5, we use s-rep and SPAHRM-PDM for data analysis.

2.1.1 Landmark modeling and PDM

To understand a **landmark model**, we begin with the most fundamental definition.

Definition 2.1. *"A **landmark** is a point of correspondence on each object that matches between and within populations"* (Dryden and Mardia, 2016, ch.1).

The landmark model introduced in (Kendall, 1984; Bookstein et al., 1986), tries to model an object with k landmarks by a $k \times m$ **configuration matrix** $X = (\mathbf{x}_1, \dots, \mathbf{x}_k)^T$, where $\mathbf{x}_i \in \mathbb{R}^m$ is the cartesian coordinate of the i th landmark, and m is equal to 2 or 3. (see Figure 2.1(a)). In some sections of this work, we use the notation l_i instead of \mathbf{x}_i as the i th landmark.

Although landmark modeling is one of the oldest shape representations, it is still important. In this respect, the book (Dryden and Mardia, 2016) presents landmark modeling with applications in R programming that shows the power of this representation.

As a drawback, one of the main difficulties in landmark modeling is to find suitable, corresponding landmarks in a population. Conventionally, anatomists locate landmarks based on the organs' biological properties, which could be an overwhelming task. Also, in some organs like the prostate, it is challenging or hardly possible to identify landmarks.

PDM

Researchers commonly use PDM (point distribution model) instead of landmark modeling. A PDM is defined as a discrete set of points sampled from the object (Tu et al., 2017). Although both of the models consist of a set of corresponding points, the methods of generating these points are different. Also, a landmark usually represents a characteristic of the object, but in PDM, points are sampled data. However, in this work, based on the definition of landmark, we consider PDM as a set of landmarks and vice versa.

2.1.2 Medial representation m-rep

The medial representation introduced in (Siddiqi and Pizer, 2008, ch.1), describes a model according to the **medial locus** of the objects. The medial locus is formed by the center of all interior spheres bitangent or multitangent to the object boundary. Vectors connecting the center of an inscribed sphere to the boundary and are tangent to the boundary are called **spokes**. By sampling the inscribed spheres, we can define **medial atoms**. A medial atom in \mathbb{R}^m ($m = 2$ or $m = 3$), is a tuple $\mathbf{m} = (\mathbf{p}, \mathbf{u}_0, \mathbf{u}_1, r)$, where $\mathbf{p} \in \mathbb{R}^m$ is located on the medial locus and represents the center of an inscribed sphere, $\mathbf{u}_0, \mathbf{u}_1 \in S^{(m-1)}$ are spoke directions where $S^{(m-1)} = \{\mathbf{x} \in \mathbb{R}^m \mid \|\mathbf{x}\| = 1\}$ is the unit sphere, and $r \in \mathbb{R}^+$ is the radius of the inscribed sphere represents the common spoke length. In this sense, an m-rep model is a collection of finite number of medial atoms (Fletcher et al., 2004; Siddiqi and Pizer, 2008) (see Figures 2.1(b,c)).

Have a good insight into the medial locus would be constructive in understanding the medial representation. A formulated definition of the medial locus of the object Ω is given by

$$M_\Omega = \{\mathbf{p} \in \mathbb{R}^m \mid |\{\mathbf{q} \in \Omega_B \mid \|\mathbf{p} - \mathbf{q}\| = d(\mathbf{p}, \Omega_B)\}|_c > 1\},$$

where $d(\mathbf{p}, \Omega_B)$ is the minimum Euclidean distance between the point \mathbf{p} to the object boundary Ω_B , and $|\cdot|_c$ is the cardinality sign (Schulz, 2013, part.1).

The advantage of m-rep is that it captures the interior part of object. Therefore it is more stable in comparison with models that only capture the surface.

As a negative point, m-rep suffers from branching instability. In fact, boundary protrusions result in additional medial branches. Therefore usually, an m-rep without a number of skeletal branches cannot capture the inner part of a non-branching object (Breuß et al., 2013, ch.5). To fix the extra branching problem, (Pizer et al., 2013) introduced **skeletal representation** (s-rep).

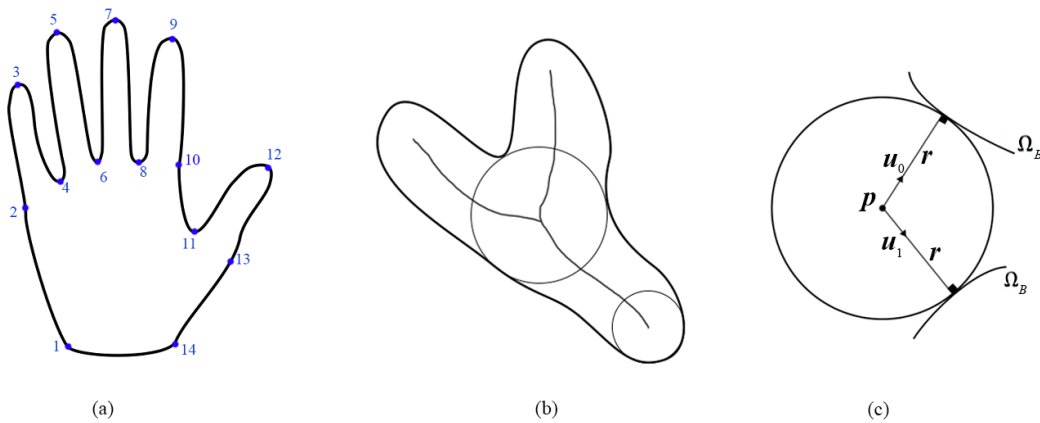


Figure 2.1: Landmark model and m-rep.

(a) Landmark model of a hand with 14 landmarks. (b) Illustration of medial locus formed by the center of 2D spheres bitangent or multitanent to the boundary. (c) Medial atom with the position p and two equal-length spokes of size r . The spokes are tangent to the boundary Ω_B , and u_0, u_1 represent their directions.

2.1.3 Skeletal representation s-rep

s-rep is an evolved version of m-rep introduced by (Siddiqi and Pizer, 2008; Pizer et al., 2013). Here we define the continuous form of s-rep, then to obey the correspondence, we define and use **discrete s-rep**. Notice that s-rep can be defined for all shapes with the topology of the sphere, but in this work, we focus on slabular s-rep, which is suitable for slab-shaped objects such as the hippocampus.

Continuous s-rep for slab-shaped objects is a locus of vectors we call spokes (p, v) with tail at p and tip at $p + v$ such that

1. Spokes do not cross each other.
2. The union of the spokes forms the interior part of the object and the union of the spoke tips forms the object boundary.
3. The union of the spoke tails, forms the skeletal locus which is a double-sided sheet with a cyclic fold curve.

(Pizer et al., 2013).

For the parameterization assume ζ with the topology of a sphere. Let the equator of the parametrizing sphere map onto the fold of the skeletal locus. Thus, two sides of the skeletal locus associate with the north and south of the sphere. For each ζ , we have a spoke $\mathbf{v}(\zeta)$ with its tail on $\mathbf{p}(\zeta)$ and its tip on an implied boundary point $\mathbf{b}(\zeta)$. ζ parametrizes the skeletal locus as $\mathbf{p}(\zeta)$. It also parametrizes the object boundary as $\mathbf{b}(\zeta)$, and the spokes as $\mathbf{v}(\zeta)$. The lengths of the spokes, is $r(\zeta) = \|\mathbf{v}(\zeta)\|$, and the directions of the spokes, is $\mathbf{u}(\zeta) = \mathbf{v}(\zeta)/r(\zeta)$. As a result, ζ parametrizes the whole s-rep $\mathbf{s}(\zeta) = (\mathbf{p}(\zeta), \mathbf{u}(\zeta), r(\zeta))$ (Pizer et al., 2013).

Now we categorize the spokes in three groups: 1. Crest spokes with tail position on the fold of the skeletal sheet. 2. Up spokes with tail position on the skeletal sheet and tips on the upper boundary of the object relative to the north side of the sheet. 3. Down spokes similar to up spokes but point to the south part of the sheet. Since the crest spoke tail is on the fold of the skeletal sheet, the inverse of its length represents the curvature of the object boundary at its tip. In other words, $\kappa = 1/r$, where κ is the curvature of the object surface Ω_B at the tip of the crest spoke, and r is the length of the crest spoke (Pizer et al., 2013; Siddiqi and Pizer, 2008).

For correspondence, by sampling, we restrict continuous s-rep to have a finite number of spokes. A discrete s-rep s^\dagger with n_s spokes can be described as a feature vector

$$s^\dagger = (\mathbf{p}_i, \mathbf{u}_i, r_i), \quad i = 1, 2, \dots, n_s, \quad (2.1)$$

where for the i th spoke, $\mathbf{p}_i \in \mathbb{R}^3$ is the tail position or **skeletal position** on the skeletal locus, $\mathbf{u}_i \in S^2$ is the spoke direction (S^2 is the unit sphere), and $r_i \in \mathbb{R}^+$ is the spoke length. (Breuß et al., 2013, ch.5). Note that depending on the method of model fitting, some spokes may have identical skeletal positions thus, the number of skeletal positions would be less than the number of spokes (i.e., $s^\dagger = (\mathbf{p}_1, \dots, \mathbf{p}_{n_p}, \mathbf{u}_1, \dots, \mathbf{u}_{n_s}, r_1, \dots, r_{n_s}), n_p \leq n_s$). Figure 2.2(c) illustrates a discrete s-rep fitted in a left hippocampus.

In this work we use skeletal PDM or skeletal positions to denote all the spokes' tail positions of a discrete s-rep.

s-rep model fitting

(Hong, 2018) proposed a model fitting procedure based on the Mean Curvature Flow (MCF). MCF is a geometric flow method where the object boundary deformation according to time t is given by

$$\begin{aligned} \frac{\partial}{\partial t} \Omega_B(\mathbf{x}, t) &= \mathcal{C}(\mathbf{x}, t) \mathcal{O}(\mathbf{x}, t), \\ \Omega_B(\mathbf{x}, 0) &= \Omega_{B0}(\mathbf{x}), \end{aligned}$$

where Ω_{B_0} is the initial object boundary, Ω_B is the deforming boundary, \mathbf{x} is a point on Ω_B , $\mathcal{C}(\mathbf{x}, t)$ is the mean curvature of Ω_B at \mathbf{x} at time t , $\mathcal{O}(\mathbf{x}, t)$ is the boundary normal of Ω_B at \mathbf{x} at time t , and Ω_B will converge to a sphere as $t \rightarrow \infty$ (Hong, 2018).

Briefly, the model fitting algorithm consists of two steps 1. Initialization, and 2. Refinement. In the initialization, we use an algorithm that iteratively deforms the object boundary by MCF. At each iteration, we check whether the deformed surface is approximately ellipsoidal, and if it is, we fit an s-rep to the ellipsoid. Then we apply the reverse MCF so that the s-rep fits the original boundary. Finally, in the refinement step, we refine the s-rep to fit tighter to the initial object boundary (Pizer et al., 2020; Hong, 2018). Figure 2.2 shows the workflow of s-rep model fitting.

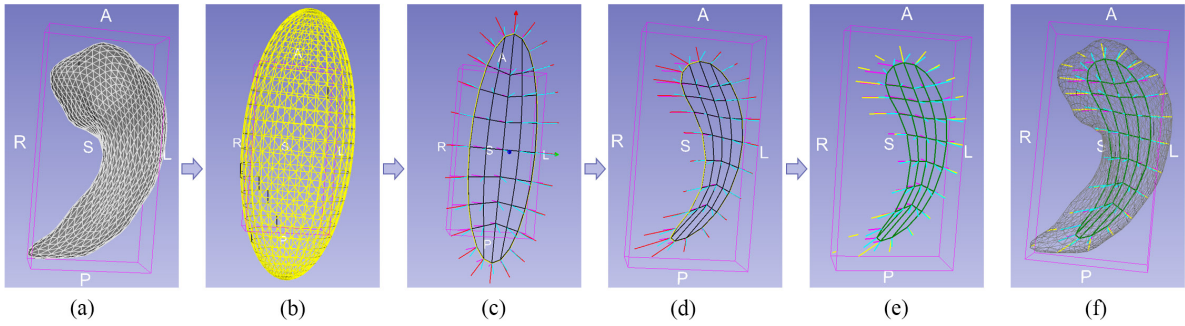


Figure 2.2: Workflow of s-rep model fitting.

(a) Target object. (b) Approximated ellipsoid by MCF. (c) The ellipsoid s-rep with the grid of 5×9 as the skeletal positions. (d) Fitted s-rep by the reverse MCF. (e) Refined s-rep. Cyan, magenta, and yellow indicate up, down, and crest spokes, respectively. Grid vertices in green are the positions of the up and down spokes. (f) Overlaid images of the target object and the refined s-rep.

One important outcome of this model fitting is a point located approximately at the center of the object. Usually, in c-shape objects (e.g., hippocampus), the **centroid** (center of gravity, see Equation 2.5) of the boundary is located outside or close to the boundary. By deforming the s-rep of the ellipsoid to fit the target object, the center of the s-rep, which represents the center of the ellipsoid (see the blue point in Figure 2.2(c)) moves to locate relatively at the center of the object. Later in Section 2.5, we will see how the Procrustes alignment uses the centroid to remove the shape location (i.e., after the alignment, the centroid of all shapes coincide with each other). Since the centroid is not necessarily in the middle of the object, it seems reasonable to align c-shape objects based on the middle point of the s-rep rather than their centroid. This type of alignment could be the subject of further study.

2.1.4 SPHARM-PDM

Spherical harmonics point distribution model (SPHARM-PDM) introduced in (Styner et al., 2006), uses spherical harmonics basis functions as presented by (Brechtbühler et al., 1995) to estimate the object boundary and generate a PDM for the objects with the spherical topology.

In summary, according to (Styner et al., 2006) and (Gerig et al., 2001), we assume spherical harmonics basis functions \mathcal{Y}_ℓ^n , $-\ell \leq n \leq \ell$ of degree ℓ and order n defined on $\theta \in [0, \pi] \times \phi \in [0, 2\pi]$ such that

$$\mathcal{Y}_\ell^n(\theta, \phi) = \sqrt{\frac{2\ell + 1}{4\pi} \frac{(\ell - n)!}{(\ell + n)!}} \mathcal{P}_\ell^n(\cos\theta) e^{in\phi},$$

where i is the imaginary unit number, and \mathcal{P}_ℓ^n denotes the associated Legendre polynomials as

$$\mathcal{P}_\ell^n(w) = \frac{(-1)^n}{2^\ell \ell!} (1 - w^2)^{\frac{n}{2}} \frac{d^{n+1}}{dw^{n+1}} (w^2 - 1)^\ell,$$

where d/dw indicates the derivative with respect to w .

Therefore, we can estimate the surface as

$$\hat{\Omega}_B(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{n=-\ell}^{\ell} \mathbf{c}_\ell^n \mathcal{Y}_\ell^n(\theta, \phi),$$

where coefficients \mathbf{c}_ℓ^n are obtainable by solving a least-squares problem. The sampling points \mathbf{x}_i at the locations (θ_i, ϕ_i) are given by

$$\mathbf{x}_i = \sum_{\ell=0}^{\infty} \sum_{n=-\ell}^{\ell} \mathbf{c}_\ell^n \mathcal{Y}_\ell^n(\theta_i, \phi_i),$$

(Styner et al., 2006).

Triangle mesh

SPHARM-PDM provides triangulated surfaces that form the object triangle mesh. These triangulated surfaces are correspondent. We can calculate the perpendicular unit normal vector of them and analyze them as directional data. The unit normal vector of a triangulated surface with vertices $\mathbf{a}, \mathbf{b}, \mathbf{c}$ is $\mathbf{n} = \mathbf{v}/\|\mathbf{v}\|$, where $\mathbf{v} = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1)^T$ is the normal vector perpendicular to the surface. The size of \mathbf{v} reflects the area of the triangle (i.e., $\|\mathbf{v}\|/2$). Thus, we can also analyze the size of the corresponding triangulated surfaces as Euclidean data. Figure 2.3 illustrates a sample of SPHARM-PDM and a triangle mesh with the unit normal vectors.

The detailed explanation of the correspondence and model fitting procedure of s-rep and SPHARM-PDM is beyond the scope of this thesis. For more detail, see (Pennec et al., 2019; Styner et al., 2007; Hong, 2018).

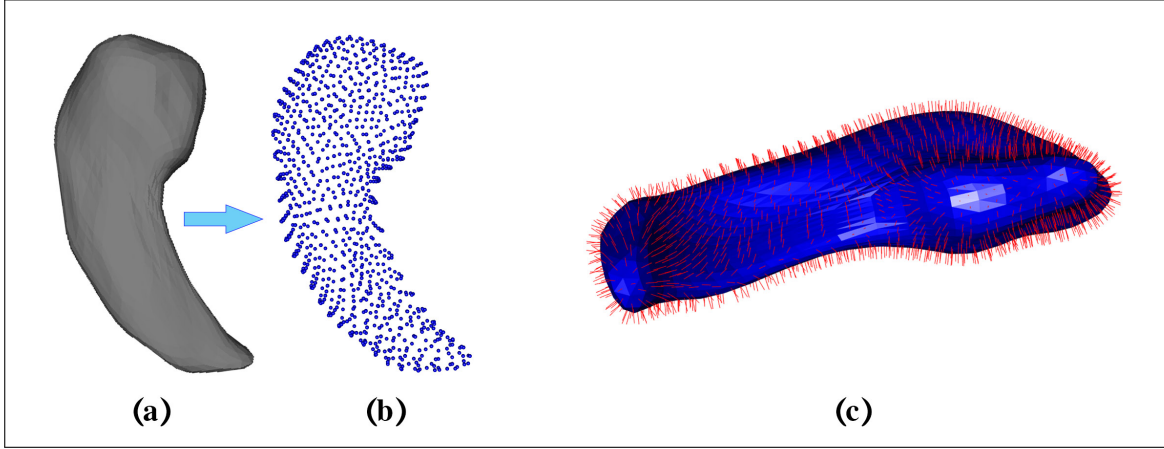


Figure 2.3: SPHARM-PDM.

(a) Target object. (b) SPHARM-PDM with 1002 points. (c) Triangle mesh in blue and unit normal vectors in red.

2.1.5 SPHARM-PDM plus skeletal PDM

A possible shape representation could be a PDM as a combination of SPHARM-PDM and s-rep skeletal PDM. Analyzing this enriched PDM that represents the internal part and the object's boundary is easier and computationally less expensive than s-rep, where directional data is included. Also, this PDM contains skeletal middle point (see Section 2.1.3), which may help us to align boundary PDMs more appropriately. The black and blue points in Figure 2.4(c) illustrates this model.

2.1.6 Flat slabbed-shape middle part parameterization

To parameterize the middle part of the objects, we use non-linear surfaces. Non-linear surfaces can be generated by the polynomial regressions. Here we briefly explain the polynomial regressions.

Polynomial regression

With n observation $(x_1, y_1)^T, \dots, (x_n, y_n)^T$, the standard normal linear regression model $\mathbf{y} = \beta X + \varepsilon$ describes a relationship between the covariate matrix X and the response vector $\mathbf{y} = (y_1, \dots, y_n)^T$, where β is a vector of unknown regression coefficients, and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ is the vector of random errors (e.g., $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, $i = 1, \dots, n$). Univariate polynomial regression is a specific case of linear regression in a quadratic form such that,

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^m \\ 1 & x_2 & x_2^2 & \dots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

As a matter of fact, we model the expected value of the variable y as an m -degree polynomial $y = \sum_{i=0}^m \beta_i x^i + \varepsilon$ (i.e., y is a function of x , $y = f(x)$) where $y, x, \varepsilon \in \mathbb{R}$, and $\forall i \beta_i \in \mathbb{R}$. Regression coefficients can be estimated by the method of least squares estimation $\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}$. To cover the 3D data $(x_1, y_1, z_1)^T, \dots, (x_n, y_n, z_n)^T$ we need to work with bivariate polynomials as

$$z = \sum_{i,j} \beta_{i,j} x^i y^j + \varepsilon, \quad (2.2)$$

where $x, y, z \in \mathbb{R}$. Note that z is a function of x and y (i.e., $z = f(x, y)$). Polynomial regression is a popular method, details can be found in (Montgomery et al., 2015, ch.7); (Jorgensen, 1993, ch.4).

Polynomial regression of SPHARM-PDM

By using PCA (introduced in detail in Section 3.1), we can align the object according to the principal axes. Assume SPHARM-PDM of a flat slabbed-shape object (e.g., left hippocampus). Since the object is slabular, the surface points are distributed along first and second principal axes rather than the third one. Besides, we know SPHARM-PDM points are relatively distributed uniformly on the boundary. Thus, we can fit the bivariate polynomial regression model, according to the third principal axis. The fitted non-linear surface crosses the middle part of the shape and parameterize it by coefficients $\beta_{i,j}$ from Equation (2.2).

In practice, we observe the fitted polynomial surface with a suitable degree (in this work, we used 4-degree polynomial) is very close to the skeletal positions of the fitted s-rep. Now, if we project the SPHARM-PDM on the fitted polynomial surface, we have a PDM called Middle Surface PDM at the middle part of the shape (see Figure 2.4). The obtained PDM inherit the correspondence from the SPHARM-PDM. In this thesis, the motivation for generating the Middle Surface PDM is to compare and support the outcome of the skeletal PDM analysis.

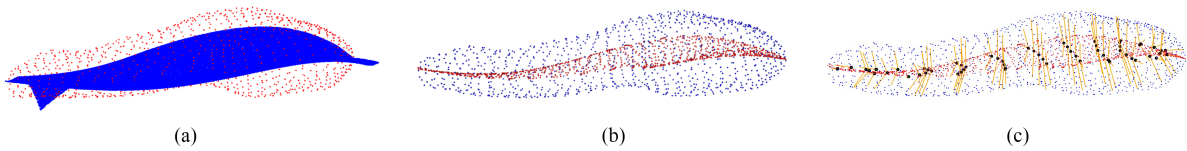


Figure 2.4: Middle Surface PDM.

(a) Fitted polynomial surface to a left hippocampus in blue, and the SPHARM-PDM in red. (b) Middle Surface PDM (i.e., Projection of SPHARM-PDM points on the polynomial surface) in red. (c) Overlaid SPHARM-PDM in blue, Skeletal PDM in black, s-rep spokes in yellow, and Middle Surface PDM in red.

2.2 Shape space

In statistical shape analysis, the study of mean and variation of GOPs within a shape distribution is desired. Therefore, having a clear explanation for shape is necessary. For this purpose (Kendall, 1977) defined shape.

Definition 2.2. "*Shape is all the geometrical information that remains when location, scale and rotational effects are removed from an object*" (Dryden and Mardia, 2016, ch.1).

In other words, shape of an object is invariant under the act of translation, rotation and scaling.

In medical research scale or the size of the objects could be important in some aspects (e.g., study the size of brain tumors), as a result (Dryden and Mardia, 1998) introduced size-and-shape.

Definition 2.3. "*size-and-shape is all the geometrical information that remains when location and rotational effects are removed from an object*" (Dryden and Mardia, 2016, ch.1).

Shape definition is a prerequisite for shape analysis. Before taking any action based on the shape definition, we need to remove translation, rotation, and scale from the shape representations. Here for a better explanation of the shape, size-and-shape, and pre-shape, we stick to the landmark analysis.

Definition 2.4. An $m \times m$ matrix Γ is a **rotation matrix** if $\Gamma^T \Gamma = \Gamma \Gamma^T = I_m$ and $\det(\Gamma) = 1$ (Dryden and Mardia, 2016, ch.3).

If $m = 2$ then $\theta \in [-\pi, \pi]$ parameterizes the rotation matrix Γ as

$$\Gamma = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}, \quad (2.3)$$

where the rotation takes place clockwise around the origin. If $m = 3$ we can parameterize the rotation matrix by three angles $\theta_1, \theta_2, \theta_3$ around z-axis, y-axis, and x-axis by

$$\Gamma = \begin{bmatrix} \cos \theta_3 & \sin \theta_3 & 0 \\ -\sin \theta_3 & \cos \theta_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ \cos \theta_2 & \sin \theta_2 & 0 \\ -\sin \theta_2 & \cos \theta_2 & 0 \end{bmatrix} \begin{bmatrix} \cos \theta_1 & \sin \theta_1 & 0 \\ -\sin \theta_1 & \cos \theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Special orthogonal group $SO(m)$ is the set of all $m \times m$ rotation matrices (Dryden and Mardia, 2016, ch.3).

Definition 2.5. Given a set of points $\mathbf{x}_i \in \mathbb{R}^m$, $i = 1, \dots, k$. The point $\bar{\mathbf{x}}$ that minimizes the sum of squared Euclidean distances to the given points is the **arithmetic mean** i.e.,

$$\bar{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^m} \sum_{i=1}^k \|\mathbf{x} - \mathbf{x}_i\|^2, \quad (2.4)$$

(Fletcher et al., 2004).

Euclidean distance between two points is a length of a straight line segment connecting them. Thus, Equation (2.4) is equivalent to

$$\bar{\mathbf{x}} = \frac{1}{k} \sum_{i=1}^k \mathbf{x}_i, \quad (2.5)$$

(Fletcher et al., 2004).

The centroid (i.e., the center of gravity) of a set of points is the arithmetic mean of the points.

Definition 2.6. "The **Euclidean similarity transformations** of a configuration matrix X are the set of translated, rotated and isotropically rescaled X , that is

$$\{\beta X\Gamma + \mathbf{1}_k \boldsymbol{\gamma}^T : \beta \in \mathbb{R}^+, \Gamma \in SO(m), \boldsymbol{\gamma} \in \mathbb{R}^m\}, \quad (2.6)$$

where β is the scale, Γ is a rotation matrix, $\boldsymbol{\gamma}$ is a translation m -vector", and $\mathbf{1}_k$ is $k \times 1$ vector of ones. By omitting the scaling factor from Euclidean similarity transformations (i.e., $\beta = 1$) we have **rigid-body transformations** (Dryden and Mardia, 2016, ch.3).

With the definition of Euclidean similarity transformations and Rigid-body transformations the shape, and size-and-shape of any configuration matrix is obtainable. We remove translation of the configuration matrix X based on the centroid. This can be done by using the **centring matrix**

$$C = I_k - \frac{1}{k} \mathbf{1}_k \mathbf{1}_k^T, \quad (2.7)$$

where I_k is the $k \times k$ identity matrix. Thus, the **centered landmark coordinates** of X is

$$X_C = CX. \quad (2.8)$$

The result of CX is the same as subtracting the elements of the X by its centroid (Dryden and Mardia, 2016, ch.2).

Alternatively, we can remove translation by using **Helmert submatrix**. Helmert submatrix H is a Helmert matrix without the first row. A Helmert matrix is an orthogonal square $k \times k$ matrix where the elements of its first row are equal to $1/\sqrt{k}$, and the remaining rows are orthogonal to the first row. The i th row of H is given by

$$(h_i, \dots, h_i, -ih_i, 0, \dots, 0)^T, h_i = -(i(i+1))^{-\frac{1}{2}}, i = 1, \dots, k-1.$$

Then similar to Equation (2.8) the **Helmertized landmark coordinates** of X is given by

$$X_H = HX \in \mathbb{R}^{(k-1)m} \setminus \{\mathbf{0}\},$$

(Dryden and Mardia, 2016, ch.2). An example of Helmert submatrix is presented in the Appendix A.2.1.

Note that $H^T H = C$, and we can transfer Helmertized landmark coordinates to centered landmark coordinates by

$$H^T HX = CX,$$

(Dryden and Mardia, 2016, ch.3).

For removing the scale factor, we standardize the configuration matrix X by its size. We define the size of the configuration matrix $X = (\mathbf{x}_1, \dots, \mathbf{x}_k)^T$ as its **centroid size**. The centroid size is the sum of squared Euclidean distances from the centroid to each landmark and is given by

$$S(X) = \|CX\| = \sqrt{\sum_{i=1}^k \|\mathbf{x}_i - \bar{\mathbf{x}}\|^2}, \quad (2.9)$$

where $\bar{\mathbf{x}}$ is the centroid (i.e., arithmetic mean) of $\mathbf{x}_1, \dots, \mathbf{x}_k$, and $\|CX\| = \sqrt{\text{trace}((CX)^T(CX))}$ (Dryden and Mardia, 2016, ch.2).

By removing scale from the configuration matrix X we have the **pre-shape** of X as

$$Z_C = \frac{CX}{\|CX\|}. \quad (2.10)$$

An alternative representation of the pre-shape is

$$Z_H = \frac{X_H}{\|X_H\|}.$$

(Dryden and Mardia, 2016, ch.3).

So far, we managed to remove the scale and location of the configuration matrices. Now we define pre-shape space, which is an essential concept in shape analysis.

Definition 2.7. "The **pre-shape space** is the space of all pre-shapes. Formally, the pre-shape space S_m^k is the orbit space of the non-coincident k point set configurations in \mathbb{R}^m under the action of translation and isotropic scaling" (Dryden and Mardia, 2016, ch.3).

Since $\|Z_C\| = 1$, the pre-shape space is a unit hypersphere as

$$S^{km-1} = \{\mathbf{x} \in \mathbb{R}^{km} \mid \|\mathbf{x}\| = 1\}. \quad (2.11)$$

By removing rotation from a pre-shape of a configuration we have the shape of a configuration. Thus "*the **shape of a configuration** matrix X is all the geometrical information of X that is invariant under Euclidean similarity transformations. The shape of X can be represented by the set $[X]$ as follows*

$$[X] = \{Z_C \Gamma : \Gamma \in SO(m)\},$$

where Z_C is the pre-shape of X " (Dryden and Mardia, 2016, ch.3).

Note that a shape of a configuration matrix is a set of configuration matrices not a single one.

The **Shape space** is the space of all configurations with the same number of landmarks. Assume k -landmark model in \mathbb{R}^m . We have overall km coordinates. By reducing m dimensions for location, one dimension for scale, and $\frac{m(m-1)}{2}$ for rotation, the dimension of the shape space is

$$m_s = km - m - 1 - \frac{m(m-1)}{2}, \quad (2.12)$$

(Dryden and Mardia, 2016, ch.3).

2.3 Manifolds

In Section 2.2 we saw the shapes and pre-shapes live on hyperspheres. Therefore it is necessary to know how to calculate the mean and variation of data on manifolds to analyze the shape distributions. In this regard, we describe manifolds and mean data on manifolds.

"A **Manifold** M is a space which can be viewed locally as Euclidean space" (Dryden and Mardia, 2016, ch.3). More precisely, "an m -dimensional manifold M is a topological Hausdorff space with a countable basis such that each point on M has a neighborhood that is locally homeomorphic to an open subset of \mathbb{R}^m ." A one-dimensional manifold is a curve, and a two-dimensional manifold is a surface, e.g., sphere S^2 is a two-dimensional manifold (Adhikari, 2016, ch.1).

There are different ways to define distance on manifolds, but a standard definition comes from the Riemannian metric. In this regard, we define tangent space.

If \mathbf{q} is a point in the manifold M and $\gamma(t) \in M$ is a differentiable curve where $t \in \mathbb{R}$, and $\gamma(0) = \mathbf{q}$. Then the tangent vector at \mathbf{q} is $\gamma'(0) = \lim_{t \rightarrow 0} \frac{d\gamma}{dt}$ and the unit tangent vector is $\boldsymbol{\xi} = \frac{\gamma'(0)}{\|\gamma'(0)\|}$. **Tangent space** $T_{\mathbf{q}}(M)$ of M at point \mathbf{q} is the set of all tangent vectors $\gamma'(0)$ of all curves passing through \mathbf{q} (Dryden and Mardia, 2016, ch.3).

If $\mathbf{x}, \mathbf{y} \in M$, the **Riemannian distance** between \mathbf{x} and \mathbf{y} is the minimum length over all possible smooth curves on M connecting \mathbf{x} and \mathbf{y} . A curve on M that locally minimizes the length between two points is a **geodesic** curve (Fletcher et al., 2004).

Assume a tangent vector $\mathbf{v} \in T_{\mathbf{q}}(M)$ at point \mathbf{q} . There is a unique geodesic $\gamma(t)$ on the manifold M passing through \mathbf{q} , with the initial tangent vector $\gamma'(0) = \mathbf{v}$. Then, the **exponential map** from the tangent space to the manifold is $Exp_{\mathbf{q}}(\mathbf{v}) = \gamma(1)$ and the inverse exponential map or the **logarithmic map** from the manifold to the tangent space is $Log_{\mathbf{q}}(\gamma(1)) = \mathbf{v}$. In a neighborhood of zero the exponential map is a diffeomorphism (i.e. map between manifolds which is differentiable and has a differentiable inverse). Therefore, the geodesic distance (i.e., Riemannian distance) between \mathbf{x} and \mathbf{y} by assuming \mathbf{x} as the base point is given by

$$d_g(\mathbf{x}, \mathbf{y}) = \|Log_{\mathbf{x}}(\mathbf{y})\|, \quad (2.13)$$

(Dryden and Mardia, 2016, ch.3); (Fletcher et al., 2004).

In a case that the manifold is the unit sphere S^m . The geodesic distance is the the arc length of the shortest great circle segment connecting \mathbf{x} and \mathbf{y} , given by

$$d_g(\mathbf{x}, \mathbf{y}) = \cos^{-1}(\mathbf{x}^T \mathbf{y}), \quad (2.14)$$

(Jung et al., 2012).

Assume $\mathbf{q} = (0, 0, 1)^T$ i.e., north pole of the unit sphere. The geodesics at the base point \mathbf{q} are the meridians (i.e., great circles) passing through \mathbf{q} . Now, if $\mathbf{v} = (v_1, v_2, 0)^T$ is a tangent vector in $T_{\mathbf{q}}S^2$, then the exponential map to the sphere is given by

$$Exp_{\mathbf{q}}(\mathbf{v}) = \left(v_1 \cdot \frac{\sin\|\mathbf{v}\|}{\|\mathbf{v}\|}, v_2 \cdot \frac{\sin\|\mathbf{v}\|}{\|\mathbf{v}\|}, \cos\|\mathbf{v}\| \right)^T, \quad (2.15)$$

where $\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2}$. The corresponding inverse log map for a point $\mathbf{x} = (x_1, x_2, x_3)^T \in S^2$ to the tangent plane is given by

$$Log_{\mathbf{q}}(\mathbf{x}) = \left(x_1 \cdot \frac{\theta}{\sin(\theta)}, x_2 \cdot \frac{\theta}{\sin(\theta)} \right)^T, \quad (2.16)$$

where $\theta = \arccos(x_3)$ is the angle between \mathbf{q} and \mathbf{x} (Fletcher et al., 2004).

Definition 2.8. The **intrinsic mean** of a set of points $\mathbf{x}_1, \dots, \mathbf{x}_k \in M$ is a point that minimizes the sum of squared Riemannian distances to the given points

$$\boldsymbol{\mu} = \arg \min_{\mathbf{x} \in M} \sum_{i=1}^k d_g(\mathbf{x}, \mathbf{x}_i)^2, \quad (2.17)$$

where $d_g(\cdot, \cdot)$ is the Riemannian distance on M (Fletcher et al., 2004).

The intrinsic mean is known as the **Fréchet mean** because the general idea was first introduced by (Fréchet, 1948). To calculate the Fréchet mean, (Fletcher et al., 2004) presented a gradient

descent algorithm (see Algorithm 2.1). The algorithm starts by a point on the manifold μ_0 as an initial mean. Next, based on μ_0 , log maps the data to the tangent space and calculate the arithmetic mean of the data on the tangent space. Then, maps back the data to the manifold and considers the exponential map of the arithmetic mean as the updated mean. The algorithm iterates until the distance between two sequential means become very small.

Algorithm 2.1 Fréchet Mean.

Input: $\mathbf{x}_1, \dots, \mathbf{x}_k \in M$

Output: $\mu \in M$, is the Fréchet mean

- 1: $\mu_0 \leftarrow \mathbf{x}_1, j \leftarrow 0$
- 2: While $\Delta\mu > \varepsilon$ Do
- 3: $\mu_{j+1} \leftarrow \text{Exp}_{\mu_j} \left(\frac{\tau}{k} \sum_{i=1}^k \text{Log}_{\mu_j} \mathbf{x}_i \right)$
- 4: $\Delta\mu \leftarrow d_g(\mu_{j+1}, \mu_j)$
- 5: $j \leftarrow j + 1$

(d_g is the geodesic distance, τ is the step size usually equal to 1, and ε is a small number.)

Definition 2.9. A differentiable manifold G is a **Lie group** that forms an algebraic group, with the group operations of multiplication and inverse such that

$$\tau : (x, y) \rightarrow xy : G \times G \rightarrow G,$$

$$\iota : x \rightarrow x^{-1} : G \rightarrow G,$$

(Fletcher et al., 2004).

Generally, we use Lie groups to define the transformation of smooth manifolds. For example, affine transformations, rotations, and scaling of \mathbb{R}^m , all form Lie groups. (Fletcher et al., 2004).

Another way to define a mean on the manifold M is to embed M in a Euclidean space, find the mean, and then project the mean back to the manifold. The obtained mean is called extrinsic mean.

Definition 2.10. Assume the embedding $\Phi : M \rightarrow \mathbb{R}^m$, the **extrinsic mean** of a set of points $\mathbf{x}_1, \dots, \mathbf{x}_k \in M$ is given by

$$\mu_\Phi = \arg \min_{\mathbf{x} \in M} \sum_{i=1}^k \|\Phi(\mathbf{x}) - \Phi(\mathbf{x}_i)\|^2,$$

(Fletcher et al., 2004). Or equivalently, we can calculate the arithmetic mean of $\Phi(\mathbf{x}_i)$ in \mathbb{R}^m by (2.5), and project it on the embedded manifold. Let $\varphi(\mathbf{x}) : \mathbb{R}^m \rightarrow G$ be a projection mapping to the lie group G (embedded manifold) as $\varphi(\mathbf{x}) = \arg \min_{\mathbf{y} \in M} \|\Phi(\mathbf{y}) - \mathbf{x}\|^2$, then the extrinsic mean is

$$\mu_\Phi = \varphi \left(\frac{1}{k} \sum_{i=1}^k \Phi(\mathbf{x}_i) \right),$$

(Fletcher et al., 2004). For more detail see (Srivastava and Klassen, 2002).

2.4 Shape distances

First we introduce shape distances only for landmark models with configuration matrices, but for the m-rep and s-rep, we need to define shape distances in a different manner. (Fletcher et al., 2004) proposed an algorithm to find the optimal distance between m-reps. In Section 2.5.3 we follow the same procedure to find the minimum distance for s-reps.

The idea to measure the shape distance is to find the minimum distance between two pre-shapes, by rotating or scaling one relative to another. Notice that shape distance is different from the distance between data on manifolds. Actually, by transferring two configurations to the pre-shape space, it is possible to measure the distance between them, but the obtained distance is not necessarily infimum. Therefore, we define shape distance as the infimum distance between pre-shapes.

Definition 2.11. The **partial Procrustes distance** can be estimated by matching the pre-shapes Z_{C_1} and Z_{C_2} of X_1 and X_2 over rotations as closely as possible. Thus,

$$d_p(X_1, X_2) = \inf_{\Gamma \in SO(m)} \|Z_{C_2} - Z_{C_1}\Gamma\|, \quad (2.18)$$

(Dryden and Mardia, 2016, ch.4).

By adding the scaling factor to Equation (2.18) we have the full Procrustes distance.

Definition 2.12. The **full Procrustes distance** between two configuration matrices X_1 and X_2 is

$$d_F(X_1, X_2) = \inf_{\Gamma \in SO(m), \beta \in \mathbb{R}^+} \|Z_{C_2} - \beta Z_{C_1}\Gamma\|, \quad (2.19)$$

where Z_{C_1} and Z_{C_2} are the pre-shapes of X_1 and X_2 respectively (Dryden and Mardia, 2016, ch.4).

Note that scaling factor β in (2.19) is different from the centroid size.

Definition 2.13. Let Z_{C_1} and Z_{C_2} be the pre-shapes of X_1 and X_2 respectively. **Riemannian distance of configurations** $d_\rho(X_1, X_2)$ is the closest geodesic distance between Z_{C_1} and Z_{C_2} on the pre-shape sphere where the minimization is carried out by rotation (Dryden and Mardia, 2016, ch.4).

Note that Definition 2.13 consider the minimum Riemannian distance, and is different from Definition 2.8.

We consider Riemannian distance of configurations as an intrinsic distance, and partial and full Procrustes distances as extrinsic distances. The range of partial Procrustes, full Procrustes,

and Riemannian distance of configurations are $0 \leq d_F \leq 1$, $0 \leq d_p \leq \sqrt{2}$, and $0 \leq d_\rho \leq \pi/2$ respectively (Dryden and Mardia, 2016, ch.4).

2.5 Alignment

2.5.1 Alignment of configurations

As we mentioned in Section 2.4, to find the infimum distance between shapes (e.g., Procrustes distance), we need to rotate and scale them relative to each other. Therefore, shapes are aligned if they have minimized distances. In shape analysis, before taking any actions, we need to align the shapes. The main reason is after the alignment, the distances between shapes are minimized concerning translation, rotation, and occasionally scaling. We will see in Section 2.5.6 that the alignment is even more crucial when we attempt to detect the partial differences.

A fast approach to align a set of configuration matrices and estimate the extrinsic mean shape is the full **Generalized Procrustes Analysis** (GPA). The purpose of the GPA is to minimize a total sum of squares distances of a set of configurations (of k landmarks) relative to each other. Thus, we minimize

$$G(X_1, \dots, X_n) = \sum_{i=1}^n \|(\beta_i X_i \Gamma_i + \mathbf{1}_k \gamma_i^T) - \mu\|^2, \quad (2.20)$$

with respect to β_i , Γ_i , γ_i , and μ , where $i = 1, \dots, n$. Then, the Procrustes fits (i.e., aligned configurations) are given by

$$X_i^P = \hat{\beta}_i X_i \hat{\Gamma}_i + \mathbf{1}_k \hat{\gamma}_i, \quad (2.21)$$

(Dryden and Mardia, 2016, ch.7).

Result 1. The shape of the full Procrustes mean

$$\hat{\mu}_F = \arg \inf_{\mu} \frac{1}{n} \sum_{i=1}^n d_F^2(X_i, \mu),$$

has the same shape as the arithmetic mean of the Procrustes fits

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i^P, \quad (2.22)$$

(Dryden and Mardia, 2016, page.135).

Result 1 is intuitively important as it shows the full Procrustes mean is obtainable by calculating the arithmetic mean of the corresponding landmarks after the alignment.

We became familiar with the shape space of landmark models and PDMs such as SPHARM-PDM. Now we explain s-rep shape space.

2.5.2 s-rep space

In general, a discrete s-rep s^\dagger (2.1) with n_s spokes is living in a manifold $M(n_s)$ as a product of Euclidean and non-Euclidean spaces, $\mathbb{R}^{3n_s} \times (S^2)^{n_s} \times \mathbb{R}_+^{n_s}$ where \mathbb{R}^{3n_s} is the space of n_s skeletal positions, $(S^2)^{n_s}$ is the space of n_s spokes' directions, and $\mathbb{R}_+^{n_s}$ is the space of n_s spokes' lengths. We know from Equation (2.11) that the pre-shape space of n_s points in \mathbb{R}^3 is S^{3n_s-1} (N.B. we consider pre-shape as Z_C not Z_H). Thus, the pre-shape space of centered s-reps can be expressed by $S^{3n_s-1} \times (S^2)^{n_s} \times \mathbb{R}_+^{n_s}$. Further, we may need to remove scale so the s-reps become commensurate. By considering the scale factor as a property of the s-reps, the feature space of the scaled s-reps is $S^{3n_s-1} \times \mathbb{R}_+^{n_s+1} \times (S^2)^{n_s}$. Based on Equation (2.12), after removing the rotation from the s-reps, shape space dimension of the skeletal positions would be $3n_s - 7$. Thus, the s-rep shape space is a product of Riemannian symmetric spaces as $\mathbb{R}^{3n_s-7} \times (S^2)^{n_s} \times \mathbb{R}_+^{n_s+1}$ or more precisely $S^{3n_s-8} \times (S^2)^{n_s} \times \mathbb{R}_+^{n_s+1}$ (Schulz et al., 2016).

2.5.3 s-rep alignment

For the m-rep alignment (Fletcher et al., 2004) defined Riemannian distance between medial atoms and minimized overall distances between a collection of m-reps with action of Euclidean similarity transformation on medial atoms. Here we follow the same procedure to align the s-reps.

Consider two spokes $\mathbf{s}_1 = (\mathbf{p}_1, \mathbf{u}_1, r_1)$ and $\mathbf{s}_2 = (\mathbf{p}_2, \mathbf{u}_2, r_2)$ from two discrete s-reps s_1^\dagger and s_2^\dagger . The spoke space is the manifold $M(1) = \mathbb{R}^3 \times S^2 \times \mathbb{R}^+$. Let $\mathbf{q} = (\mathbf{0}, \mathbf{u}_0, 1) \in M(1)$ be the base point where $\mathbf{0} = (0, 0, 0)^T$ is the base point of positional element, and $\mathbf{u}_0 = (0, 0, 1)^T$ is the base point of the spherical element. Assume a tangent vector $\mathbf{w} \in T_{\mathbf{q}}M(1)$ as $\mathbf{w} = (\mathbf{p}, \mathbf{v}, \rho)$, where $\mathbf{p} \in \mathbb{R}^3$ is the positional element, $\mathbf{v} \in \mathbb{R}^2$ is the spherical element, and $\rho \in \mathbb{R}$ is the radius element of \mathbf{w} on the tangent space. The direct product of the exponential map for each element defines the exponential map for $M(1)$. Thus, we define the exponential map for \mathbb{R}^3 as the identity map, for S^2 as the spherical exponential map given by Equation (2.15), and for \mathbb{R} as the exponential function $f(\rho) = e^\rho$. Thus, we have

$$Exp_{\mathbf{q}}(\mathbf{w}) = (\mathbf{p}, Exp_{\mathbf{u}_0}(\mathbf{v}), e^\rho),$$

where the $Exp_{\mathbf{u}_0}(\mathbf{v})$ is the spherical exponential map of \mathbf{v} from (2.15). Then, the *log* map of a point $\mathbf{s} = (\mathbf{p}, \mathbf{u}, r) \in M(1)$ is the direct product of the *log* map for each element

$$Log_{\mathbf{q}}(\mathbf{s}) = (\mathbf{p}, Log_{\mathbf{u}_0}(\mathbf{u}), \log(r)),$$

where the $Log_{\mathbf{u}_0}(\mathbf{u})$ is the spherical log map of \mathbf{u} from (2.16). As a result, direct product of n_s exponential or logarithmic map for the spoke space $M(1)$ defines the exponential or logarithmic map for the s-rep space $M(n_s)$.

In the tangent space the positional element \mathbf{p} , spherical tangent element \mathbf{v} , and radius ρ are in different units. To commensurate \mathbf{v} and ρ with the positional element \mathbf{p} , we scale the \mathbf{v} and ρ with the average spoke length over all corresponding spokes in the population. Thus, the norm of the vector $\mathbf{w} \in T_q M(1)$ becomes

$$\|\mathbf{w}\|_T = (\|\mathbf{p}\|^2 + \bar{r}^2(\|\mathbf{v}\|^2 + \rho^2))^{\frac{1}{2}}, \quad (2.23)$$

where \bar{r} is the average spokes' length over all corresponding spokes.

By using the norm (2.23), the Riemannian distance of (2.13) between two spokes $\mathbf{s}_1, \mathbf{s}_2 \in M(1)$ is given by

$$d(\mathbf{s}_1, \mathbf{s}_2) = \|Log_{\mathbf{s}_1}(\mathbf{s}_2)\|_T. \quad (2.24)$$

Let $\mathcal{T} = (\beta, \Gamma, \gamma)$ be an Euclidean similarity transformation (2.6). We define the action of \mathcal{T} on a spoke $\mathbf{s}_i = (\mathbf{p}_i, \mathbf{u}_i, r_i)$, $i \in \{1, \dots, n_s\}$ of an s-rep by,

$$\mathcal{T} \cdot \mathbf{s}_i = \mathcal{T} \cdot (\mathbf{p}_i, \mathbf{u}_i, r_i) = (\beta \mathbf{p}_i^T \Gamma + \gamma, \mathbf{u}_i^T \Gamma, \beta r_i),$$

The action of \mathcal{T} on the s-rep $\mathbf{s}^\dagger = \{\mathbf{s}_i : i = 1, \dots, n_s\}$ is,

$$\mathcal{T} \cdot \mathbf{s}^\dagger = \{\mathcal{T} \cdot \mathbf{s}_i : i = 1, \dots, n_s\}.$$

Assume a set of s-reps $\mathbf{s}_1^\dagger, \dots, \mathbf{s}_N^\dagger \in M(n_s)$ that we want to align. Let $\mathbf{s}_{\alpha i}$ be the α th spoke in the i th s-rep. The squared distance metric between s-reps \mathbf{s}_i^\dagger and \mathbf{s}_j^\dagger is

$$d(\mathbf{s}_i^\dagger, \mathbf{s}_j^\dagger)^2 = \sum_{\alpha=1}^{n_s} d(\mathbf{s}_{\alpha i}, \mathbf{s}_{\alpha j})^2,$$

where $d(.,.)$ for spokes on the right-hand side is given by (2.24).

For the s-reps alignment, we calculate the set of similarity transforms $\mathcal{T}_1, \dots, \mathcal{T}_N$ that minimize the following total sum of squares distances between the s-reps

$$d(\mathcal{T}_1, \dots, \mathcal{T}_N; \mathbf{s}_1^\dagger, \dots, \mathbf{s}_N^\dagger) = \sum_{i=1}^N \sum_{j \neq i} d(\mathcal{T}_i \cdot \mathbf{s}_i^\dagger, \mathcal{T}_j \cdot \mathbf{s}_j^\dagger)^2. \quad (2.25)$$

For removing translation, we multiply centering matrix to the skeletal PDMs (see Equation (2.8)). Removing rotation and scale is an optimization problem that can be solved by the gradient

descent algorithm to minimize the metric of (2.25). Notice that there is no guaranty for the convergence of the gradient descent algorithm, and it is important to find a suitable starting point. For that reason, the alignment can be initialized by aligning the s-rep skeletal PDMs with GPA. For more detail see (Fletcher et al., 2004).

Alternatively, by treating spokes as the joined pair of tips and tails, then an s-rep with n_s spokes consists of $2n_s$ positions in \mathbb{R}^3 (i.e., the collection of skeletal and implied boundary points). Thus, we can express the distance of two s-reps as the Procrustes distance between two configuration matrices where each one represents a PDM with $2n_s$ points. As a result, we can use GPA (2.20) to align s-reps. Note that after the alignment, it is crucial to update the spokes' directions because the rotation of the skeletal sheet affects the spokes' directions. Also, we need to update spokes' lengths if we remove scaling by GPA (this can be done by scaling skeletal PDMs, preserving the scaling factor for each s-rep and then using this scaling factor to resize the spokes' lengths). In a case that we define the distance between two s-reps only based on the skeletal positions, we can align s-reps by applying GPA on skeletal PDM and then update the directions. Note that after aligning the s-reps with this method (i.e., alignment only by the skeletal positions), as we ignore the spokes' directions and lengths, we may observe less variation in skeletal positions and more variation in spokes' directions in comparison with other methods of alignment.

Although it is common practice to align objects by GPA, we will discuss other possible approaches, such as weighted alignments in the next section.

2.5.4 Weighted Procrustes alignment

In Procrustes alignment we use the Euclidean norm $\|X\|^2 = \text{trace}(X^T X)$ and treat all the landmarks equally as uncorrelated. In weighted Procrustes, the aim is to estimate the covariance relation between landmarks then, use this covariance structure for the alignment. For this purpose we substitute the Euclidean norm with the Mahalanobis norm

$$\|X\|_{\Sigma}^2 = \text{vec}(X)^T \Sigma^{-1} \text{vec}(X), \quad (2.26)$$

where Σ is the covariance matrix, and $\text{vec}(X)$ is the vectorize operator of the configuration X with columns $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ as $\text{vec}(X) = (\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_m^T)^T$ (Dryden and Mardia, 2016, ch.7, ch.4).

Generally, estimating the covariance relation is a matter of debate. (Goodall, 1991) estimated the covariance matrix as

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \text{vec}(X_i^P - \bar{X}) \text{vec}(X_i^P - \bar{X})^T,$$

where \bar{X} is the mean shape, and X_i^P , $i = 1, \dots, n$ are the Procrustes fits from Equation (2.21). After estimating the covariance matrix, we can use $\text{GPA}(\hat{\Sigma})$ to weight the landmarks and align the shapes based on the Mahalanobis norm (2.26), where $\text{GPA}(\hat{\Sigma})$ denotes the weighted Procrustes with the covariance matrix $\hat{\Sigma}$. Also, without estimating Σ , we can specify a covariance matrix and use GPA to align the shapes based on it.

Notice that ordinary Procrustes alignment is weighted Procrustes alignment with the identity matrix as the covariance matrix $\Sigma = I$.

To have a weighted alignment, (Brignell et al., 2016) introduced an algorithm starts by aligning the shapes with $\text{GPA}(I)$. Then, iteratively calculates $\hat{\Sigma}$ by a suitable estimator and aligns shapes by $\text{GPA}(\hat{\Sigma})$ until convergence. For further detail about the suitable estimators and the algorithm see (Dryden and Mardia, 2016, page.148).

Most of the methods in this area, consider only one group of shapes and try to align shapes based on the group covariance matrix. However, there might be two groups, and we want to find the differences between them. (Goodall, 1991) suggested to estimate the covariance matrix of each group, find the mean of the groups based on the weighted alignment, and superimpose obtained mean to see the differences. The drawback of this method is that we estimate the covariance matrix of each group without considering the structure of the other group. In this respect, in Section 2.5.6 we specify a new covariance matrix for the weighted alignment that decreases the variance of landmarks, which are similar among two groups of shapes. For this purpose first, we need to introduce a new algorithm.

2.5.5 Elimination algorithm

Assume two groups of k -landmark models A_1 and A_2 . In the proposed method, we remove landmarks one by one from both groups and measure the Procrustes distance of the groups' means by Equation (2.19). In each step, we eliminate a landmark from both groups, which is associated with the greatest Procrustes distance while it was involved in the models. We repeat the procedure until two landmarks remain. The idea is to eliminate a landmark in each iteration which causes more difference between the means. In this sense, the distance between means decreases and they become more similar after each round. Now, assume a small number $\delta \in [0, 1]$ as the threshold ($d_F \in [0, 1]$ so $\delta \in [0, 1]$, see Section 2.4). The threshold divides the measured distances set in two subsets with the value greater or smaller than δ . Each measured distance is associated with an eliminated landmark. Thus we classify landmarks in two groups of "suspicious" and "unsuspicious" according to their corresponding distances. In other words, by removing suspicious landmarks, the Procrustes distance between the means is less than δ . The value of δ can be the subject of further study. In our analysis, $\delta = 0.03$ has shown good

results. As an effect in Section 2.5.6, we will see the alignment is more robust against outliers. Algorithm 2.2 explains the method.

Algorithm 2.2 Elimination Algorithm.

Input: (i) Two groups of k -landmark models A_1 and A_2 (ii) Threshold δ (iii) $x = \{1, \dots, k\}$

Output: (i) Suspicious-landmarks (ii) Unsuspicious-landmarks

- 1: Sorted-Landmarks $\leftarrow []$, distances $\leftarrow []$, $s \leftarrow 0$, $j \leftarrow 1$
- 2: While $|x|_c > 2$ Do:
- 3: $d_{min} \leftarrow \infty$, distances[j] $\leftarrow procdis(\bar{A}_1, \bar{A}_2)$
- 4: Iterate through x by i
- 5: $d \leftarrow procdis(A_1 \setminus l_i, A_2 \setminus l_i)$
- 6: If $d < d_{min}$
- 7: $d_{min} \leftarrow d$, $s \leftarrow i$
- 8: Sorted-Landmarks[j] $\leftarrow l_s$, $A_1 \leftarrow A_1 \setminus l_s$, $A_2 \leftarrow A_2 \setminus l_s$, $x \leftarrow x \setminus s$, $j \leftarrow j + 1$
- 9: Sorted-Landmarks $\leftarrow l_x$, distances $\leftarrow (0, 0)$
- 10: $y \leftarrow$ indices of distances' elements which are greater than δ
- 11: Suspicious-landmarks \leftarrow Sorted-Landmarks[y],
Unsuspicious-landmarks \leftarrow Sorted-Landmarks[$\{1, \dots, k\} \setminus y$]

(l_i is the i th landmark. $A \setminus l_i$ is A without l_i . \bar{A} is the mean shape of A . $procdis$ is the Procrustes distance. \leftarrow and \leftarrow indicate assign and push operators respectively. $[]$ indicates empty array. $A[x]$ is the elements of A with the indices of set x . $|x|_c$ is the cardinality of x .)

Example (Dryden and Mardia, 2016, p.190-196) tested the mean shape similarity of 8-landmark models of male versus female chimpanzee skulls (panm vs. panf) and male versus female gorilla skulls (gorf vs. gorm). The overall result shows no significant difference between females and males chimpanzees but a significant difference in the gorilla group. Also (Dryden and Mardia, 2016, ch.9) discussed the mean difference between 13-landmark models of Schizophrenia vs. The Control group and concluded that different methods of testing provide different results. Some tests show there is a significant difference between Schizophrenia and Control, and some show no significant difference. However it was stated (Dryden and Mardia, 2016, page.257) the shape change is related to the splenium and the two landmarks just below it (i.e., landmarks l_1, l_{13}, l_6). We use these three pared groups of shapes to test the elimination method. Table 2.1 shows the sorted landmarks and Procrustes distances of the elimination algorithm for the mentioned groups. By using the threshold $\delta = 0.03$, there is no suspicious point in panm vs. panf. In other words, even without removing any points, the distance between the mean shapes is less than δ . In gorf vs. gorm, we classify landmarks (l_8, l_4, l_5, l_7) as suspicious landmarks. In Schizophrenia vs. Control the algorithm classifies landmarks (l_1, l_{13}) as suspicious, which is compatible with the observation of the book. Figure 2.5 shows the pooled mean shape of pared groups, and red dots indicate suspicious landmarks.

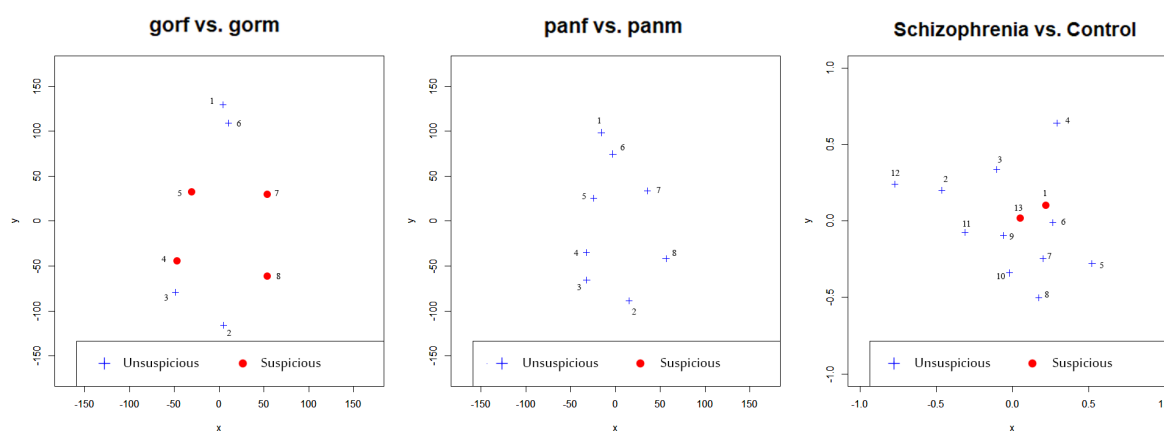
Female chimpanzee vs. male chimpanzee								
Sorted Landmarks	4	7	3	6	2	1	5	8
Proc Distance	0.022	0.020	0.015	0.009	0.007	0.004	0	0

Female gorilla vs. Male gorilla								
Sorted Landmarks	8	4	5	7	3	1	2	6
Proc Distance	0.058	0.051	0.045	0.037	0.025	0.015	0	0

Schizophrenia vs. Control													
Sorted Landmarks	1	13	3	4	11	10	12	7	5	9	2	6	8
Proc Distance	0.038	0.033	0.028	0.022	0.016	0.011	0.009	0.007	0.004	0.003	0.001	0	0

Table 2.1: Elimination examples.

Sorted landmarks and associated Procrustes distances from the elimination algorithm.

**Figure 2.5:** Suspicious landmarks.

Suspicious landmarks depicted in red. Threshold $\delta = 0.03$. Left: female gorilla skull vs. male gorilla skull. Middle: female chimpanzee skull vs. male chimpanzee skull. Right: Schizophrenia vs. Control. Data are taken from (Dryden, 2018).

2.5.6 Weighted alignment by elimination

Often researchers use Procrustes alignment to compare shapes and to detect differences. The critical issue is that Procrustes alignment introduces undesirable differences as it tries to minimize the dissimilarities between shapes. As a result, we may fail to expose actual differences. For instance, we know the main difference between cats' heads and dogs' heads is in the muzzle. Assume landmark models as depicted in Figure 2.6. In this case, the muzzle landmark acts as leverage to rotate and translates other landmarks. Thereby, the hypothesis test introduces non-existing differences (i.e., increase in the false positive), or in some cases fail to introduce real ones (i.e., increase in the false negative). In this regard, the idea is to ignore parts of the shapes that cause differences and align the remaining shapes. Thereby, we can have a better alignment concerning the similarity of the mean shapes. Figure 2.6 illustrates the idea which we call weighted alignment by elimination.

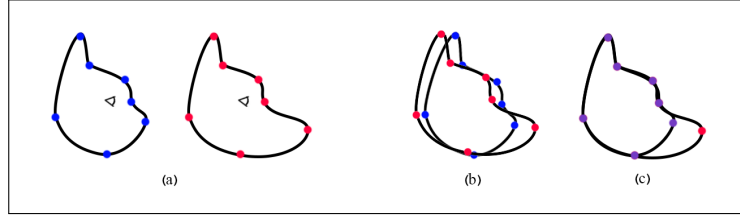


Figure 2.6: Comparison of alignments.

(a) Profile sketch of a cat and a dog with landmarks. (b) Visualization of the Procrustes alignment. (c) Desired alignment to detect muzzle as the main difference obtainable by alignment without the muzzle landmark.

To reach the desired alignment, we detect unsuspecting points by the elimination algorithm from Section 2.5.5. Then, we define a covariance matrix and take advantage of weighted Procrustes alignment as described in Section 2.5.4. Assume shapes with k landmarks $l_i, i = 1, \dots, k$. In ordinary Procrustes alignment the covariance matrix is $\Sigma = I$. We can stabilize unsuspecting landmarks by adjusting Σ . This can be done by decreasing the associated diagonal element of Σ with unsuspecting landmarks. Thus, we have

$$\Sigma = \begin{bmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_k \end{bmatrix} \text{ s.t. } \begin{cases} a_i = 1, & \text{if } l_i \text{ is suspicious} \\ a_i = \varepsilon, & \text{if } l_i \text{ is unsuspecting} \end{cases}, \quad i = 1, \dots, k,$$

where ε is a small number depending on how intensively we want to stabilize unsuspecting points relative to the suspicious points. Based on our observation $\varepsilon = 10^{-4}$ seems to be suitable.

Example As an intuitive example, we create two groups of landmarks by modifying female gorilla skulls (taken from (Dryden, 2018)) in a way that two groups are identical in five bottom landmarks (l_2, l_3, l_4, l_6, l_7) but different in two top landmarks (l_1, l_5). Also we remove one landmark (landmark no. 3 in the original data) to have a better illustration. By applying elimination algorithm we obtain landmarks l_1 and l_5 as suspicious. Thus, we define covariance matrix as $\Sigma = \text{diag}(1, \varepsilon, \varepsilon, \varepsilon, 1, \varepsilon, \varepsilon)$. Obviously, by Procrustes alignment shapes are different in all the landmarks. But the weighted alignment by elimination aligns the identical landmarks and exposes the real differences (see Figure 2.7).

A major drawback of this method is the cost of computation for shapes with a large number of landmarks (e.g., SPHARM-PDM) unless we eliminate more than one landmark in each iteration of the elimination algorithm.

Note that by applying this method, we decrease and increase the variation of unsuspecting and suspicious landmarks, respectively. This change in variation may bias the result of the hypothesis testing. We discuss this issue in Section 5.3.2.

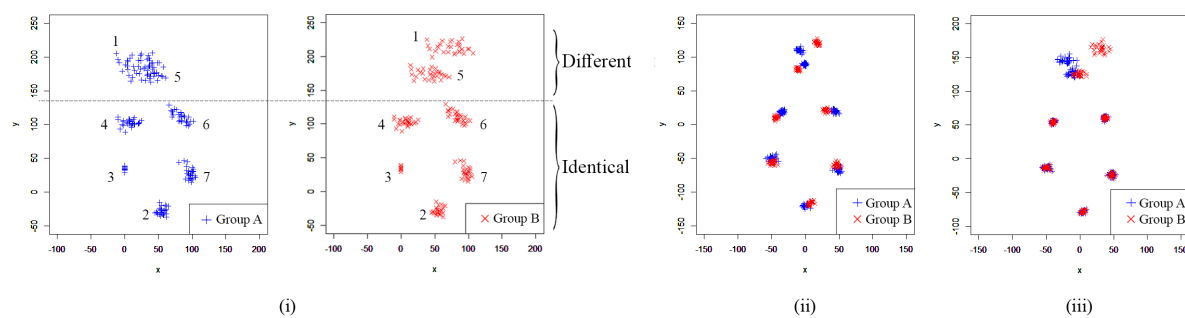


Figure 2.7: Procrustes vs. weighted alignment by elimination.

- (i) Two groups of landmark models which are identical at l_2, l_3, l_4, l_6, l_7 and different at l_1, l_5 . (ii) Ordinary Procrustes alignment. (iii) Suggested weighted alignment by elimination with $\delta = 0.03$ and $\varepsilon = 10^{-4}$.

Data are modified version of female gorilla skulls taken from (Dryden, 2018).

An alternative method to reveal local differences is deformation (Dryden and Mardia, 2016, ch.12). The idea is to deform mean shapes of two groups to the pooled mean shape, and then to use a thin-plate spline to observe the differences. However, depending on the alignment method, the mean shapes vary, and again it is important to have an appropriate alignment.

Chapter 3

Dimensionality Reduction

In this chapter, we discuss different methods of dimensionality reduction, including Principal Component Analysis (PCA), Principal Geodesic Analysis (PGA), and Principle Nested Sphere Analysis (PNS). Then we introduce Non-Linear PGA (NLPGA) to find the mean of the spherical data based on the rotation of the tangent space. Finally, we explain the Euclideanization of the s-rep data by Composite PNS (CPNS).

3.1 Principal Component Analysis (PCA)

PCA is one of the oldest multivariate techniques and the most widely used dimensionality reduction method in data analysis. It has a long history, but the modern formalization is introduced by (Hotelling, 1933).

Given a set of observations $\mathbf{y}_i \in \mathbb{R}^m$, $i = 1, \dots, k$. We assume the arithmetic mean $\bar{\mathbf{y}} = \frac{1}{k} \sum_{i=1}^k \mathbf{y}_i$ as the PCA mean. Similar to (2.8) we center the data to the origin based on $\bar{\mathbf{y}}$. Let $\mathbf{x}_i = \mathbf{y}_i - \bar{\mathbf{y}}$, $i = 1, \dots, k$ be the centered data. PCA determines an $n \leq m$ dimensional subspace such that after projecting data on this subspace, the variation of the data points is optimally retained. This subspace is defined based on n orthogonal axes, known as principal axes or principal components (i.e., eigenmodes) (Theodoridis, 2015, page.940). To derive the principal axes, first assume $n = 1$, the aim is to find a direction in \mathbb{R}^m so that the variance of the corresponding projections of the data points on this direction is maximized. Let \mathbf{v}_1 denote the first principal axes. The variance of the projected data on \mathbf{v}_1 is given by

$$\mathcal{J}(\mathbf{v}_1) = \frac{1}{k} \sum_{i=1}^k (\mathbf{v}_1^T \mathbf{x}_i)^2 = \frac{1}{k} \sum_{i=1}^k (\mathbf{v}_1^T \mathbf{x}_i)(\mathbf{x}_i^T \mathbf{v}_1) = \mathbf{v}_1^T \hat{\Sigma} \mathbf{v}_1,$$

where $\hat{\Sigma}$ is the sample covariance matrix of the of observed data

$$\hat{\Sigma} = \frac{1}{k} \sum_{i=1}^k \mathbf{x}_i \mathbf{x}_i^T.$$

Since we maximize the variance, the optimization problem will be

$$\mathbf{v}_1 = \arg \max_{\mathbf{v}} \mathbf{v}^T \hat{\Sigma} \mathbf{v}, \quad \text{s.t. } \mathbf{v}^T \mathbf{v} = 1.$$

The corresponding Lagrangian is given by

$$\mathcal{L}(\mathbf{v}, \lambda) = \mathbf{v}^T \hat{\Sigma} \mathbf{v} - \lambda(\mathbf{v}^T \mathbf{v} - 1). \quad (3.1)$$

Taking the gradient of (3.1) with respect to \mathbf{v} and setting it to zero we have

$$\mathbf{v}^T \hat{\Sigma} \mathbf{v} = \lambda.$$

In other words, the first principal direction is an eigenvector corresponding to the maximum eigenvalue of $\hat{\Sigma}$. The second principle component \mathbf{v}_2 can be selected in a way that \mathbf{v}_2 is orthogonal to \mathbf{v}_1 and \mathbf{v}_2 maximizes the variance after projecting the data on this direction. With the above procedure and the orthogonality condition $\mathbf{v}_2^T \mathbf{v}_1 = 0$, we can show that \mathbf{v}_2 is the eigenvector corresponding to the second largest eigenvalue of $\hat{\Sigma}$. We continue the process until we obtain m principal components. In this regard, the obtained principal axes are corresponding to the sorted eigenvalues $\lambda_1 > \lambda_2 > \dots > \lambda_m$ of $\hat{\Sigma}$ (Theodoridis, 2015, Chapter 19); (Jolliffe, 2013).

PCA of m -dimensional data can be calculated in forward and backward directions. We already explained the forward approach in a Euclidean space where we start by calculating the sample mean (arithmetic mean) as the best fitted 0-dimensional subspace. Mean with the first eigenvector describe a line as the best fitted one-dimensional subspace. Then mean with the first and second eigenvectors describe the best-fitted plane as a two-dimensional subspace. Consequently, we can increase the dimension of the best fitting sub-spaces. However, PCA could also be calculated in a backward direction. We start with the best fitted $(m - 1)$ -dimensional subspace, and then within that, we find the best fitted $(m - 2)$ -dimensional subspace, etc. which results in a sequence of best fitted subspaces (Damon and Marron, 2014).

Although, PCA is simple, fast and easily applicable data analysis for data lying in the Euclidean vector space \mathbb{R}^m , it is not directly applicable for data on manifolds. From Chapter 2 we know that shapes are living on manifolds. As an extrinsic approach if we embed the shape space manifold in an Euclidean space (e.g., \mathbb{R}^m) then, it is possible to use PCA on the embedded space.

Various methods are accessible to analyse the data on manifolds. In Section 2.5.2 we discussed the s-rep shape space which is a combination of Euclidean space with spheres and a hypersphere. Thus, we prefer to use rather new methods for spherical data like PNS (Jung et al., 2012).

Before discussing PNS we briefly explain PGA to have a better intuition about PNS.

3.2 Principal Geodesic Analysis (PGA)

PGA is a generalization of PCA to manifolds and is a forward approach. PGA introduced by (Fletcher et al., 2004) is a general method applicable on manifolds, and it is not restricted to the spheres.

In Section 2.3 we discussed the Fréchet mean of a data set $\mathbf{x}_i \in M$, $i = 1, \dots, k$ in a manifold M . Similar to PCA, the aim is to find a sequence of nested geodesics as the manifold subspaces that maximize the variance of the projected data. (Fletcher et al., 2004) showed that by applying PCA on the tangent space we can approximate PGA of the data. In this sense, first we use the Fréchet mean $\boldsymbol{\mu} \in M$ as the base point of the tangent space $T_{\boldsymbol{\mu}}M$. Let $U \cap T_{\boldsymbol{\mu}}M$ be a neighborhood of the base point, and assume the data is localized enough to lie within U . With a well-defined logarithmic mapping we can map the data from the manifold to the tangent space as $Log_{\boldsymbol{\mu}}(\mathbf{x}_i)$. Then we use PCA to find principal directions $\mathbf{v}_1, \dots, \mathbf{v}_m \in T_{\boldsymbol{\mu}}M$ of the tangent space (m is the dimension of the tangent space). The obtained vectors form a basis of the tangent space, and generate a sequence of nested subspaces $V_n = span(\{\mathbf{v}_1, \dots, \mathbf{v}_n\}) \cap U$ where $n = 1, \dots, m$. Finally, the principal geodesics submanifolds are the projection of V_n with the exponential map $Exp_{\boldsymbol{\mu}}(V_n)$ on the manifold. If the manifold is the unit sphere S^2 , as a straightforward approach we can translate data to the north pole by Fréchet mean (see Equation 3.4). Then log map the data to the tangent space by (2.16) and use PCA to analyse the data. The Algorithm 3.1 explains the procedure (Fletcher et al., 2004).

Algorithm 3.1 PGA approximation.

Input: $\mathbf{x}_1, \dots, \mathbf{x}_k \in M$

Output: Principal components $\mathbf{v}_n \in T_{\boldsymbol{\mu}}M$, Variances $\lambda_n \in \mathbb{R}$

- 1: $\boldsymbol{\mu} \leftarrow$ Fréchet mean of $\{\mathbf{x}_i\}_{i=1}^k$ (Algorithm 2.1)
 - 2: $\mathbf{u}_i \leftarrow Log_{\boldsymbol{\mu}}(\mathbf{x}_i)$, $i = 1, \dots, k$
 - 3: $\hat{\Sigma} \leftarrow \frac{1}{k} \sum_{i=1}^k \mathbf{u}_i \mathbf{u}_i^T$
 - 4: $\{\mathbf{v}_n, \lambda_n\} \leftarrow$ eigenvectors and eigenvalues of $\hat{\Sigma}$
-

3.3 Principal Nested Sphere (PNS)

Basically, the idea of PNS is based on fitting a geodesic to a data set on a manifold, then by projecting the data on the geodesic, the mean can be estimated by minimizing the sum of squares distances of the projected data within the geodesic path. (Jung et al., 2012) generalized the idea as a backward approach for high-dimensional spherical data.

Assume an m -dimensional unit sphere S^m . By using a decomposition of S^m , PNS tries to capture the maximal non-geodesic variation in an $(m-1)$ -dimensional subsphere. Sequentially for $i = m - 1, \dots, 1, 0$ the decomposition provides the best i -dimensional approximation Ψ_i of the spherical data. The subsphere Ψ_i is called the i -dimensional principal nested sphere. Ψ_i is a submanifold of Ψ_j , $j = i + 1, \dots, m$. Thus we have

$$\Psi_0 \subset \Psi_1 \subset \dots \subset \Psi_{m-1} \subset S^m.$$

The fitting procedure reduces the dimension of the data iteratively. First we fit an $(m - 1)$ -dimensional subsphere Ψ_{m-1} of S^m which best approximate the data. Therefore, we can associate each data point to a residual based on the its distance to Ψ_{m-1} . In this sense, the residual is a signed geodesic distance to its projection on Ψ_{m-1} . Then we find the best fitting subsphere Ψ_{m-2} of Ψ_{m-1} , and continue until the best fitting subsphere has zero dimension. Note that 0-dimensional subsphere Ψ_0 is a point, and we consider it as the Fréchet mean of the projected data on one-dimensional subsphere (circle) Ψ_1 (Jung et al., 2012).

Definition 3.1. If $\mathbf{v} \in S^m$ and $\rho \in (0, \pi/2]$ then a **subsphere** \mathcal{A}_{m-1} of S^m is defined as

$$\mathcal{A}_{m-1}(\mathbf{v}, \rho) = \{\mathbf{x} \in S^m \mid d_g(\mathbf{v}, \mathbf{x}) = \rho\},$$

where d_g is the geodesic distance function from Equation (2.14). Thus, \mathcal{A}_{m-1} can be considered as an intersection of S^m and an m -dimensional hyperplane $\{\mathbf{x} \in \mathbb{R}^{m+1} \mid \mathbf{v}^T \mathbf{x} - \cos(\rho) = 0\}$, where \mathbf{v} is the axis of \mathcal{A}_{m-1} and also the unit normal vector of the hyperplane (Jung et al., 2012).

\mathcal{A}_{m-1} is a unit sphere, but its axis does not coincide with the axis of S^{m-1} (i.e., north pole). By a transformation function $f_1 : \mathcal{A}_{m-1} \rightarrow S^{m-1}$ we rotate \mathcal{A}_{m-1} such that its axis \mathbf{v} reach the north pole (for the definition of f see (Jung et al., 2012)).

Definition 3.2. An $(m - i)$ -dimensional **nested sphere** Ψ_{m-i} of S^m is

$$\Psi_{m-i} = \begin{cases} f_1^{-1} \circ \dots \circ f_{i-1}^{-1}(\mathcal{A}_{m-i}) & (i = 2, \dots, m - 1), \\ \mathcal{A}_{m-1} & (i = 1), \end{cases} \quad (3.2)$$

where f_i^{-1} is the inverse transformation of $f_i : \mathcal{A}_{m-i} \rightarrow S^{m-i}$ (Jung et al., 2012).

Let $\mathbf{x}_1, \dots, \mathbf{x}_k$ be a set of data on S^m . An estimation of the best fitting subsphere $\hat{\mathcal{A}}_{m-1} = \mathcal{A}_{m-1}(\hat{\mathbf{v}}_1, \hat{\rho}_1)$ can be found by minimizing the sum of squares of the residuals ξ_j , $j = 1, \dots, k$ (i.e., sum of squares of the geodesic distances between data and $\hat{\mathcal{A}}_{m-1}$). Thus, to find $\hat{\mathbf{v}}_1$ and $\hat{\rho}_1$ we minimize

$$\sum_{j=1}^k \xi_j(\mathbf{v}_1, \rho_1)^2 = \sum_{j=1}^k (d_g(\mathbf{x}_j, \mathbf{v}_1) - \rho_1)^2, \quad (3.3)$$

with respect to $\mathbf{v}_1 \in S^m$ and $\rho_1 \in (0, \pi/2]$ (Jung et al., 2012).

After finding $\hat{\mathcal{A}}_{m-1}$, we project $\{\mathbf{x}_j\}_{j=1}^k$ on $\hat{\mathcal{A}}_{m-1}$ along the minimal geodesic joining \mathbf{x}_j to $\hat{\mathcal{A}}_{m-1}$. The projection function is given by

$$proj\{\mathbf{x}; \hat{\mathcal{A}}_{m-1}(\mathbf{v}, \rho)\} = \frac{\sin(\rho)\mathbf{x} + \sin\{d_g(\mathbf{x}, \mathbf{v}) - \rho\}\mathbf{v}}{\sin\{d_g(\mathbf{x}, \mathbf{v})\}}.$$

By repeating the above procedure and by using the transformation functions (3.2), we obtain a sequence of sample principal nested spheres $\hat{\Psi}_{m-1}, \dots, \hat{\Psi}_1$ of S^m . At the lowest level, $\hat{\Psi}_1$ is a circle, and we consider $\hat{\Psi}_0$ as the Fréchet mean of the projected data (Jung et al., 2012).

3.3.1 PNG

In PNS, we use a sequence of non-geodesic nested great or small spheres. It means radii of the nested subspheres could be one or less than one. If we use only subspheres with radius one, then we have Principal Nested Great-spheres (PNG). Note that in dealing with S^2 , nested spheres are great or small circles (i.e., 2D subspheres).

3.4 Non-Linear PGA (NLPGA)

Although PNS has an outstanding performance for girdle and von Mises distributions, it can be shown that in some cases (e.g., spherical semi-ellipse or s-like distributions) the obtained mean is far from our expectation. Inspired from principal flows (Panaretos et al., 2014), which determines curves on manifolds as principal components, we introduce a nonlinear dimensionality reduction method on the tangent space of S^2 . This method is analogous to PGA from Section 3.2 but instead of using PCA on the tangent space we define and use a specific form of Non-Linear PCA (NLPCA) based on the principal curve. "Principal curves are smooth one-dimensional curves that pass through the middle of a m -dimensional data set, providing a nonlinear summary of the data" (Hastie and Stuetzle, 1989). A comprehensive explanation about NLPCA can be found in (Kruger et al., 2008; Scholz, 2002).

Imagine we have a set of spherical data samples $\mathbf{z}_i \in S^2$, $i = 1, \dots, k$, with the Fréchet mean $\boldsymbol{\mu}_F$. Given a spherical rotation matrix $R(c, \theta)$, the rotation of $\boldsymbol{\mu}_F$ along its meridian to the north pole is $R\boldsymbol{\mu}_F = (0, 0, 1)^T$. The spherical rotation matrix is given by

$$R(c, \theta) = I_3 + \sin\theta[c]_{\times} + (1 - \cos\theta)(cc^T - I_3), \quad (3.4)$$

where I_3 is the 3×3 identity matrix, $[c]_{\times}$ is the cross product matrix satisfying $[c]_{\times}\mathbf{v} = c \times \mathbf{v} \forall \mathbf{v} \in \mathbb{R}^3$, and $\theta \in [0, \pi/2]$ is the rotation angle (i.e., the angle between $\boldsymbol{\mu}_F$ and the north pole) (Schulz et al., 2016).

Consequently $R\mathbf{z}_i$ is the transformation of \mathbf{z}_i , $i = 1, \dots, k$. Let $\forall i$, $\mathbf{y}_i = \text{Log}_{\mathbf{q}}(R\mathbf{z}_i)$ be the log-map of $R\mathbf{z}_i$ to the tangent space by Equation (2.16), where $\mathbf{q} = (0, 0, 1)^T$. By applying PCA on the log-mapped data \mathbf{y}_i , $i = 1, \dots, k$, we find $\bar{\mathbf{y}}$ as the PCA mean (i.e., arithmetic mean on the tangent space). Also, we can express \mathbf{y}_i in the PCA coordinate based on the principal axes PC1 and PC2. Let \mathbf{x}_i indicate the PCA coordinate of \mathbf{y}_i . Similar to PGA we transformed data from spherical coordinate to the PCA coordinate on the tangent space. Also we know due to the *Log* and *Exp* map, the tangent space of S^2 is a circular disc with radius π inscribed in a square of the width 2π . Thus, we can consider the centered data (i.e., centered by PCA) inscribed in a square with the same size. The two line segments connecting the middle points of the square's parallel sides indicate PC1 and PC2. In this coordinate system we implement curve fitting techniques to find the principal curve. In this work, to make sure we can define isomorphic transformation between sphere and the tangent space, we only use the polynomial regression method (see Section 2.1.6), however it is possible to apply other techniques. As a result, an m -degree polynomial $f(x) = \sum_{i=1}^m \beta_i x^i + \varepsilon$ with the domain $[-\pi, \pi]$ represent the fitted curve. The principal curve (i.e., fitted polynomial) defines the first principal axis of an NLPCA.

Let $\tilde{\mathbf{x}}_i$ be the projection of \mathbf{x}_i along PC2 to the polynomial. The principal curve f is a one-dimensional manifold thus, the Fréchet mean of projected data on f is given by

$$\boldsymbol{\mu} = \arg \min_{\mathbf{x} \in f} \sum_{i=1}^k d_c(\mathbf{x}, \tilde{\mathbf{x}}_i)^2,$$

where $d_c(\mathbf{u}, \mathbf{v}) = \int_{u_1}^{v_1} \sqrt{1 + \left(\frac{df(x)}{dx}\right)^2} dx$ is the curve distance between points $\mathbf{u} = (u_1, f(u_1))^T$ and $\mathbf{v} = (v_1, f(v_1))^T$.

Note that the domain of the polynomial is limited and relatively small so we can estimate the Fréchet mean by the Monte Carlo random sampling method. We can produce $N + 1$ large enough uniformly distributed real numbers $v_j \in \mathbb{R}$, $j = 0, \dots, N$ such that $v_j \in [-\pi, \pi]$, or simply choose $v_j = -\pi + 2\pi j/N$, $j = 0, \dots, N$. The projection of v_j on f along PC2 is $f(v_j)$, and $\mathbf{v}_j = (v_j, f(v_j))^T$ is a point of tangent space located on f . Therefore, the estimation of Fréchet

mean on the principle curve is given by

$$\hat{\boldsymbol{\mu}} = \arg \min_{\boldsymbol{v}} \sum_{j=0}^N \sum_{i=1}^k d_c(\boldsymbol{v}_j, \tilde{\boldsymbol{x}}_i)^2.$$

Recall $\boldsymbol{r}_i = \boldsymbol{x}_{i2} - \tilde{\boldsymbol{x}}_{i2}$ is the regression residual of sample \boldsymbol{x}_i . We consider, $d_i = d_c(\hat{\boldsymbol{\mu}}, \tilde{\boldsymbol{x}}_i)$ and r_i as the first and second coordinates of \boldsymbol{x}_i in NLPCA respectively. In this regard, $\hat{\boldsymbol{\mu}}$ is the origin of NLPCA axes, and the reduced dimensional data are located on the principle curve. We consider PC2 as the second principle axis of NLPCA. Furthermore, let Γ_{pca} be the matrix of ordered PCA eigenvectors known as PCA rotation matrix. To map back $\hat{\boldsymbol{\mu}}$ to the sphere, we rotate and translate $\hat{\boldsymbol{\mu}}$ to the north pole with $\hat{\boldsymbol{\mu}}^T \Gamma_{pca}^{-1} + \bar{\boldsymbol{y}}$ (i.e., $\bar{\boldsymbol{y}}$ coincide with $(0, 0, 1)^T$). Applying Exp map from (2.15) and multiplying the result with R^{-1} , map the $\hat{\boldsymbol{\mu}}$ on the sphere. Therefore we have

$$\boldsymbol{\mu}^* = R^{-1}(Exp_q(\hat{\boldsymbol{\mu}}^T \Gamma_{pca}^{-1} + \bar{\boldsymbol{y}})), \quad (3.5)$$

where $\boldsymbol{\mu}^*$ is the obtained NLPGA mean, and $\boldsymbol{q} = (0, 0, 1)^T$.

In Figure 3.1 \boldsymbol{x} (the solid points) indicates a data point on the tangent space, $\tilde{\boldsymbol{x}}$ is the projection of \boldsymbol{x} along PC2, $\hat{\boldsymbol{x}}$ is the projection of \boldsymbol{x} based on the Mean Squared Error (MSE) (i.e., $e = \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2$), $\boldsymbol{\mu}$ is the Fréchet mean on the principal curve, and d^\pm and r^\pm indicate the NLPCA coordinates with positive and negative signs.

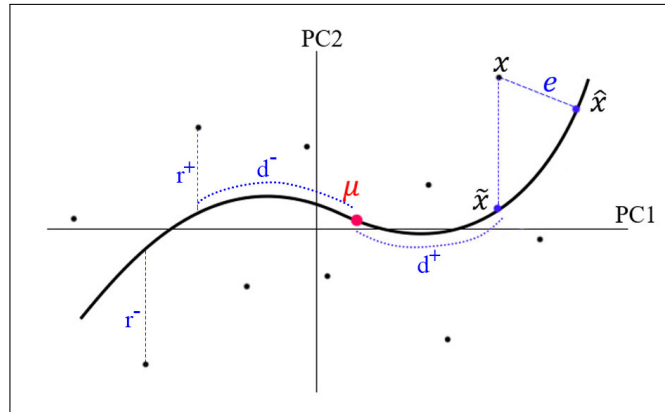


Figure 3.1: Illustration of NLPCA.

Solid points indicate log-mapped data on the tangent space, $\tilde{\boldsymbol{x}}$ is the projection of \boldsymbol{x} along PC2, $\hat{\boldsymbol{x}}$ is the projection of \boldsymbol{x} based on MSE, $\boldsymbol{\mu}$ is the Fréchet mean on the fitted curve, and d and r indicate the NLPCA coordinates.

Rotation to improve fitting

In most cases, the performance of NLPGA is not satisfactory. To improve the method, by a small degree of rotation (e.g. $\theta = 2\pi n/360, n = 1, \dots, 360$), we rotate the PCA coordinate on the tangent space from 0 to 2π . The rotation take place around $\bar{\boldsymbol{y}}$ by the rotation matrix (2.3). During the rotation we find the principle curve and form NLPCA. Further, we estimate

the performance of the NLPCA in order to find the best NLPCA coordinate system. The performance of the NLPCA like PNS and PGA can be measured by the percentage contribution of the eigenmodes. Note that this approach is different from performance estimation of the principal curve by MSE, because we consider the both residuals r^\pm and d^\pm . After a complete rotation we choose an NLPCA with the best performance. Figure 3.2 illustrate the procedure.

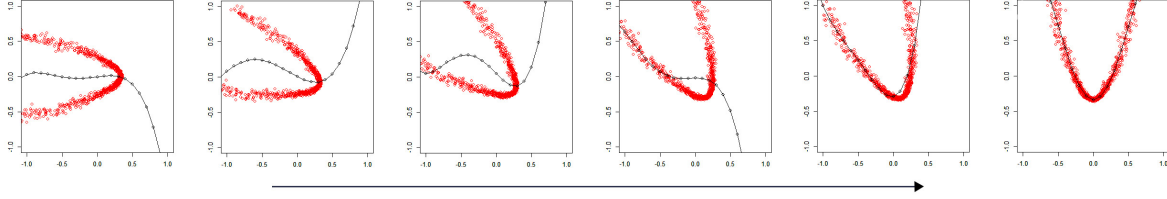


Figure 3.2: Rotation of the tangent space.

Figures from left to right include data points and fitted 4-degree polynomial during the rotation of the tangent space. The right figure visualize the best fit for this sample.

Since we rotated the tangent space, to calculate NLPGA mean, we need to include this rotation in Equation (3.5). Assume $\hat{\Gamma}_{nl}$ represent the estimated rotation matrix corresponding to an NLPCA with the best performance, and $\hat{\boldsymbol{\mu}}_{nl}$ be the corresponding estimated mean. Then, map of the $\hat{\boldsymbol{\mu}}_{nl}$ on the sphere, provides the NLPGA mean as

$$\boldsymbol{\mu}_{nl}^* = R^{-1}(\text{Exp}_q(\hat{\boldsymbol{\mu}}_{nl}^T \hat{\Gamma}_{nl}^{-1} \Gamma_{pca}^{-1} + \bar{\mathbf{y}})),$$

where $\boldsymbol{\mu}_{nl}^*$ indicates NLPGA mean after rotation.

The same way $\forall \mathbf{x} \in T_{\boldsymbol{\mu}_F} S^2$ we define mapping from the tangent space to the sphere as

$$\mathbf{z} = R^{-1}(\text{Exp}_q(\mathbf{x}^T \hat{\Gamma}_{nl}^{-1} \Gamma_{pca}^{-1} + \bar{\mathbf{y}})),$$

where $\mathbf{z} \in S^2$. As a result, the NLPCA axes on the tangent space correspond to two geodesic submanifolds on the unit sphere where one of them is a great circle and the other one is the image of the principal curve.

Note that if we choose to project the data on the fitted curve by MSE which is always positive, we may fail to define an isomorphic transformation between spherical coordinate and NLPCA coordinate. Also, if we try to fit a closed curve (e.g., ellipse) we may encounter a similar issue. Because, it would require to specify which points belong to the inside or to the outside of the curve that could be a difficult task specifically for the complicated curves. In Figure 3.3 we compare the NLPGA with PNS for some artificial s-shape, v-shape and w-shape distributions. In the left column of Figure 3.3, blue lines on the spheres indicate the mapping of the principal curve. Choosing the degree of the polynomial not only here but in the most regression problems is a matter of controversy. For our examples we chose polynomials of degrees three and four. Also to increase the performance of NLPGA we can map tangent space as a circular disk to a

square as discussed in (Fong, 2015), and after finding the best fitted curve inverse map data from square to the disk but there are cautions about the distance preservation which merit the further study.

The spokes' directions of the fitted s-reps in this thesis have concentrated von Mises distributions. As a result the application of NLPGA has no superiority over PNG or PGA and we ignore to use it.

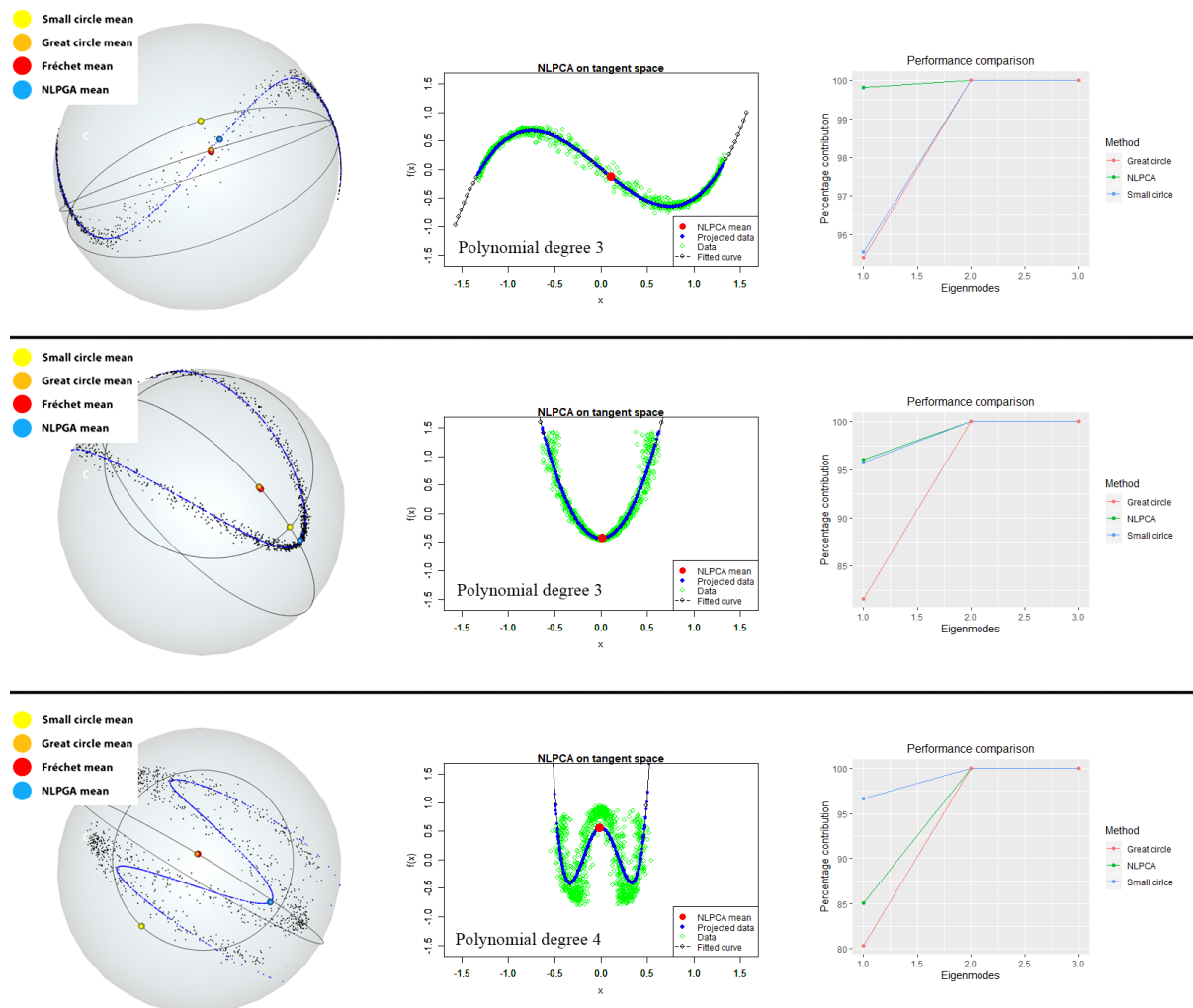


Figure 3.3: PNS vs. NLPGA.

Left column indicates the PNS means of the fitted small and great circles, Fréchet mean, and NLPGA mean of the fitted curve on S^2 . Middle column shows the NLPGA mean and fitted polynomial on the tangent space. Right column compares the performance of PNS, PNG and NLPGA.

3.5 Composite Principal Nested Spheres (CPNS)

As we discussed in Section 2.5.2, an s-rep consists of a set of points, directions and lengths living on a product of Euclidean and non-Euclidean spaces. For analysing the s-reps variability one

possible option is to consider each s-rep component separately. First, using PNS to produce Euclidean scores (i.e., residuals) from spherical data, namely the pre-shape of skeletal positions and the spokes' directions. Second, compose the scores with the Euclidean data (i.e., logarithms of the scale factors and spoke lengths). Finally this composition can be used to present the mean and variation of the s-reps. This method is called Composite Principal Nested Spheres (CPNS) in (Jung, 2011; Pizer et al., 2013) and is discussed in more details in the following.

Note that in this part we follow the similar mathematical notation used in (Jung, 2011; Pizer et al., 2013). Assume a set of N centered and aligned discrete s-reps $\{s_1^\dagger, \dots, s_N^\dagger\}$, each one with n_s spokes as defined in Section 2.1.3 such that

$$s_i^\dagger = (P_i, \mathbf{u}_1(i), \dots, \mathbf{u}_{n_s}(i), r_1(i), \dots, r_{n_s}(i)), \quad (3.6)$$

where $P_i = (\mathbf{p}_1(i), \dots, \mathbf{p}_{n_s}(i))^T$ represent the skeletal positions of the i th s-rep, and $i = 1, \dots, N$.

Since we assumed aligned and centered s-reps, the centroid size of P_i is $\gamma_i = \sqrt{\sum_{j=1}^{n_s} \|\mathbf{p}_j(i)\|^2}$. Let $\forall i, P_{C_i} = P_i/\gamma_i$ be the scaled skeletal positions of P_i . The centroid size of P_{C_i} is 1 thus, P_{C_i} is the pre-shape of P_i (see Equation 2.10).

We know from Section 2.5.2 that the space of the skeletal positions is a hypersphere $P_{C_i} \in S^{3n_s-1}$. Therefore we can use PNS to analyse the skeletal position of N s-reps. By using PNS we obtain principal scores as an $m \times N$ matrix $Z = (z_{ki})$ where $m = \min(N-1, 3n_s-1)$ is the number of nontrivial components, and z_{ki} is the i th sample's deviation from the PNS mean along the k th principal subsphere. The principal nested spheres can be represented by a collection of axes and radii as

$$\{(\mathbf{v}_1, \rho_1), \dots, (\mathbf{v}_{3n_s-2}, \rho_{3n_s-2}), \mathbf{v}_{3n_s-1}\}.$$

Since $\gamma_i \in \mathbb{R}^+$, we use the geometric mean $\bar{\gamma} = (\prod_{i=1}^N \gamma_i)^{\frac{1}{N}}$ to define $\gamma_i^* = \log(\gamma_i/\bar{\gamma})$. Hence, the scale factors are normalized by their geometric mean. Let $\boldsymbol{\gamma} = (\gamma_1^*, \dots, \gamma_N^*)^T$ denotes a vector of all scale factors. To incorporate the correlation between pre-shapes and scale factors we scale Z by

$$Z_s = \bar{\gamma} \begin{pmatrix} Z \\ \boldsymbol{\gamma} \end{pmatrix}_{(m+1) \times N}.$$

In this way, the normalized scale factor of i th sample is located below the corresponding column of Z .

Spoke directions are spherical data, so for the j th spoke direction the PNS input is $\mathbf{u}_j(i) \in S^2$, $i = 1, \dots, N$. We obtain principal scores as a $2 \times N$ matrix $Z^{Sj} = (z_{ki}^{Sj})_{k=1,2}^{i=1,\dots,N}$. The principal nested spheres can be represented as a collection of two axes and a radius $\{(\mathbf{v}_1^j, \rho_1^j), \mathbf{v}_2^j\}$.

Spoke lengths can be treated just like the scale factors. Let $\bar{r}_j = (\prod_{i=1}^N r_j(i))^{\frac{1}{N}}$ be the geometric mean of the spoke lengths and define $r_j^*(i) = \log(r_j(i)/\bar{r}_j)$ and $R_j^* = (r_j^*(1), \dots, r_j^*(N))^T$. Now we can scale the score variables Z^{Sj} and R_j^* by multiplying with \bar{r}_j .

Data are Euclideanized and commensurate, and we can compose them in a Euclidean matrix.

Composite space

We compose the s-rep Euclideanized data from the previous Section 3.5 in a matrix as follows

$$Z_{comp} = \begin{bmatrix} Z_s \\ \bar{r}_1 Z^{S1} \\ \vdots \\ \bar{r}_{n_s} Z^{Sn_s} \\ \bar{r}_1 R_1^* \\ \vdots \\ \bar{r}_{n_s} R_{n_s}^* \end{bmatrix} \begin{array}{l} \longrightarrow (m+1) \times N \quad (\text{skeletal positions and scale factors}) \\ \longrightarrow 2 \times N \quad (1st \text{ spoke direction}) \\ \vdots \\ \longrightarrow 2 \times N \quad (n_s \text{th spoke direction}) \\ \longrightarrow 1 \times N \quad (1st \text{ spoke length}) \\ \vdots \\ \longrightarrow 1 \times N \quad (n_s \text{th spoke length}) \end{array}, \quad (3.7)$$

where columns of Z_{comp} belong to the vector space $E_{comp} \subset \mathbb{R}^{m+1+3n_s}$.

Similar to PCA we can form the covariance matrix $\frac{1}{N-1} Z_{comp} Z_{comp}^T = U \Lambda U^T$ where columns of U are the principal components and the diagonal elements of Λ are the eigenvalues of Z_{comp} . Again the eigenvectors and corresponding eigenvalues represent the direction and magnitude of data variation in the space of E_{comp} . Note that the mean of the Z_{comp} is at the origin of E_{comp} because the Z_{comp} variables are centered (Jung, 2011; Pizer et al., 2013; Jung, 2014).

3.5.1 Composite Principal Nested Great Sphere (CPNG)

In Section 3.3.1 we introduced PNG. By substituting PNS with PNG in CPNS, we have CPNG. In practice, CPNG is a faster approach in comparison with CPNS. Because it avoids a sequential testing procedure to check whether a small or great sphere is a good choice. Besides, CPNG assures better permutation test correspondence because it guarantees correspondences between PNS decompositions for each iteration (permutation is discussed in Section 4.3). However, we show in Section 4.4, because of the cost of computation, applying permutation with CPNG for s-reps with a large number of skeletal positions is not reasonable (e.g., for 10000 permutation we need 8772 hours, see Table 4.2). Moreover, in this study, CPNS and CPNG have no superiority over PGA to capture the s-rep data (see Figure 5.3). Thus, we use CPNG only for the classification but not the permutation.

3.5.2 s-rep classification based on CPNG

The objective of this study is the hypothesis testing. However, we briefly discuss s-rep classification.

As we explained in Section 3.5, columns of the Z_{comp} matrix (3.7) belongs to the Euclidean space E_{comp} . In this regard, we consider E_{comp} as a feature space and the columns of Z_{comp} as the feature vectors. Each column of Z_{comp} is associated with a unique s-rep. If we have two or more groups of s-reps, we can form a pooled Z_{comp} and label the columns of the pooled matrix to indicate the group that each s-rep belongs to it. Now we can implement different methods of classification (e.g., Bayesian classifier).

Note that in the classification problem, we deal with a High Dimensional Low Sample Size (HDLSS) problem because the dimension of the feature space is higher than the number of available samples. Therefore, applying common classification techniques like Support Vector Machin (SVM) might be inappropriate as we may encounter data piling issue. (Marron et al., 2007) introduced Distance Weighted Discrimination (DWD) to overcome the data piling obstacle in SVM. (Hong, 2018) applied DWD and SVM to classify the s-rep data. Similarly in this work, we use both methods to classify the s-reps.

Detail of the SVM and DWD algorithms is beyond the scope of this work. For more detail see (Marron et al., 2007), (Pennec et al., 2019, ch.6), and (Theodoridis, 2015, ch.11).

Chapter 4

Hypothesis Testing

In this chapter, we discuss parametric and non-parametric hypothesis testing approaches to compare the mean of two data sets. Next, we extend the non-parametric approach to find locational differences between the GOPs of two groups of s-reps. We then discuss the global test based on empirical p-values. Also, we introduce a parametric approach for s-rep hypothesis testing under the normality assumption. Finally, we propose a method of hypothesis testing independent from the alignment.

4.1 Overview of hypothesis testing

For the hypothesis testing, assume we have two independent random samples $A_1 = \{\mathbf{x}_1, \dots, \mathbf{x}_{n_1}\}$ and $A_2 = \{\mathbf{y}_1, \dots, \mathbf{y}_{n_2}\}$ from distributions F_{A_1} and F_{A_2} respectively. Then the objective is to test the null hypothesis $H_0 : F_{A_1} = F_{A_2}$ against alternative $H_1 : F_{A_1} \neq F_{A_2}$. In other words, under the null hypothesis, A_1 and A_2 and the pooled sample $A = A_1 \cup A_2$ are all from the same distribution F_{A_1} . Moreover, under the null hypothesis, any subset $A_1^* \subset A$ of size n_1 from the pooled sample A , and its complement $A_2^* = A \setminus A_1^*$, represent independent random samples from distribution F_{A_1} . Let T be a two-sample statistic that measure the distance between F_{A_1} and F_{A_2} (e.g., $T = \|\mu_{F_{A_1}} - \mu_{F_{A_2}}\|$ distance between sample means). Without loss of generality, we can consider large values of test statistic T support the alternative $F_{A_1} \neq F_{A_2}$ (Rizzo, 2007, page.219).

In this study, we focus on testing the mean of the populations. The general form of the hypothesis test can be expressed as follows

$$H_0 : \mu_{A_1} = \mu_{A_2} \text{ vs. } H_1 : \mu_{A_1} \neq \mu_{A_2}.$$

4.1.1 Parametric vs. non-parametric approach

The parametric approach is based on making assumptions about the distributions and relative parameters. In contrast, the non-parametric approach does not make such assumptions. The Student's t-test (known as t-test) is an example of the parametric approach. Assume random samples from normal distributions with equal variances $x_1, \dots, x_{n_1} \sim \mathcal{N}(\mu_x, \sigma^2)$ and $y_1, \dots, y_{n_2} \sim \mathcal{N}(\mu_y, \sigma^2)$. Then, a classical parametric approach to test the hypothesis $H_0 : \mu_x = \mu_y$ vs. $H_1 : \mu_x \neq \mu_y$ is a two-sample t-test. The test statistic is

$$T = \frac{\bar{x} - \bar{y}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where $\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i$ and $\bar{y} = \frac{1}{n_2} \sum_{i=1}^{n_2} y_i$ are the arithmetic means of the samples. The pooled standard deviation is given by

$$S_p = \sqrt{\frac{(n_1 - 1)S_x^2 + (n_2 - 1)S_y^2}{n_1 + n_2 - 2}},$$

where $S_x^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2$ and $S_y^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2$ are unbiased estimators of the variances.

Under the null hypothesis H_0 , T follows the Student's t-distribution $T \sim t(n_1 + n_2 - 2)$. If the obtained T is unlikely we reject null hypothesis. For more detail see (de Winter and Cahusac, 2014, ch.6).

To reject the null hypothesis, we need a scale to measure how much T is unlikely or extreme. For this purpose, we define p -value and level of significance.

4.1.2 p -value

As an informal definition, "a p -value is the probability under a specified statistical model that a statistical summary of the data would be equal to or more extreme than its observed value" (Wasserstein et al., 2016). Thus, by assuming α as a significance level of the test, we reject the null hypothesis when $p\text{-value} < \alpha$. As an example, the p -value for two-sample t-test can be defined as $p\text{-value} = P(|T| > T_{obs} | H_0)$, where T_{obs} is the observed test statistic.

The significance level α is the probability of making a type I error. In other words, the probability that we falsely reject a true null hypothesis (Kim and Bang, 2016). In most cases, it is acceptable to consider $\alpha \in [0.01, 0.05]$. Still, in multiple comparisons problem (see Section 4.2), to control false positives, we need to define the significance level.

4.2 Multiple comparisons problem

Assume testing n simultaneously null hypotheses that n_0 of them are true. Let u be the number of truly declared non-significant, v be the number of type I errors (i.e., false positive), r be the total number of rejected null hypotheses, and t be the number of type II errors (i.e., false negative). The situation can be summarized the in Table 4.1 (Benjamini and Hochberg, 1995).

	Declared non-significant	Declared significant	Total
True null hypotheses	u	v	n_0
Non-true null hypotheses	t	s	$n - n_0$
Total	$n - r$	r	n

Table 4.1: Multiple testing.

Number of errors committed by testing n null hypotheses.

Table is taken from (Benjamini and Hochberg, 1995).

4.2.1 FDR and FWER

Basically, family wise error rate (FWER) can be considered as "the probability of at least one type I error", i.e., $\text{FWER} = P(v \geq 1)$, while false discovery rate (FDR) is "the expected proportion of type I errors among all the rejected hypotheses", i.e.,

$$\text{FDR} = E(v/r; r > 0) = E(v/r | r > 0)P(r > 0),$$

(Candes, 2018).

A well-known method to control the probability of type I error is based on controlling FWER proposed by (Bonferroni, 1936). Bonferroni's method, tests individually each hypothesis at level α/n and guaranties that $P(v \geq 1) \leq \alpha$. This method is highly conservative because by having a large number of hypotheses, significance level is $\alpha/n \ll 1$, which increases the chance of committing type II error (i.e., fail to reject the false null hypotheses).

(Benjamini and Hochberg, 1995) proposed a more moderate method than Bonferroni by controlling FDR. Consider n tested null hypotheses H_{01}, \dots, H_{0n} and their corresponding p -values p_1, \dots, p_n . Let $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$ be the ordered p -values. Assume the null hypothesis $H_{0(i)}$ corresponds to $p_{(i)}$. For controlling FDR by δ (e.g., $\delta \leq 1.5$) let k be the largest i for which $p_{(i)} \leq \frac{i}{n} \delta$, then reject all $H_{0(i)}$, $i = 1, \dots, k$.

The following theorem states the assumption of independence for implementing Benjamini-Hochberg method.

Theorem 4.1. "For independent test statistics and for any configuration of false null hypotheses, the above procedure (i.e., Benjamini-Hochberg) controls the FDR at δ " (Benjamini and Hochberg, 1995).

In practice, it is common to use **adjusted p -values**. There are variety of techniques for the adjustment. In a simple way, assume we have a set of p -values p_1, \dots, p_n , and the corresponding ordered p -values $p_{(1)}, \dots, p_{(n)}$. By the Bonferroni's method $\{np_1, \dots, np_n\}$, and by the Benjamini-Hochberg method $\{\frac{np_{(1)}}{1}, \frac{np_{(2)}}{2}, \dots, \frac{np_{(n)}}{n}\}$, are the adjusted p -values respectively.

A drawback of controlling false positives (e.g., by FWER or FDR) is the increase of false negatives. The consequences of false negatives can be higher than false positives (Kim and Bang, 2016). Therefore in medical research judging the results based on the adjusted p -values is not always reasonable.

4.3 Permutation test

A permutation test is a non-parametric approach with minimum assumptions, which is applicable when we can randomly rearrange the data under the null hypothesis in any order. In general, permutation tests and inferences for a sample space χ (i.e., set of all possible outcomes) are architected based on the following principle.

***Permutation Testing Principle** If two experiments, taking values on the same sample space χ and respectively with underlying distribution F_1 and F_2 , both members of F , give the same data set x , then the two inferences conditional on x and obtained using the same test statistic must be the same, provided that the exchangeability of data with respect to groups is satisfied in the null hypothesis. Consequently, if two experiments, with underlying distributions F_1 and F_2 give respectively x_1 and x_2 , and $x_1 \neq x_2$, then the two conditional inferences may be different.* (Pesarin, 2001, page.6)

4.3.1 Permutation distribution

Suppose we observed two independent random samples $A_1 = \{\mathbf{x}_1, \dots, \mathbf{x}_{n_1}\}$ and $A_2 = \{\mathbf{y}_1, \dots, \mathbf{y}_{n_2}\}$ from the distributions F_{A_1} and F_{A_2} , respectively. Assume an ordered set $A = \{\mathbf{z}_1, \dots, \mathbf{z}_{n_1+n_2}\}$, indexed by $v = \{1, \dots, n_1, n_1 + 1, \dots, n_1 + n_2\}$. Let $n = n_1 + n_2$ then, $\mathbf{z}_i = \mathbf{x}_i$ if $1 \leq i \leq n_1$, and $\mathbf{z}_i = \mathbf{y}_{i-n_1}$ if $n_1 + 1 \leq i \leq n$. Thus, A is the pooled sample of A_1 and A_2 . Let $A^* = (A_1^*, A_2^*)$ be a partition of A , where A_1^* and A_2^* are of sizes n_1 and n_2 respectively. Then, A^* corresponds to a permutation π of the integer set v . Let $A_{\pi(v)} = A^*$, i.e, $A_{\pi(v)}$ is a partition of A corresponds to $\pi(v)$. The number of possible partitions is equal to $\binom{n}{n_1}$ (i.e., there are $\binom{n}{n_1}$ possible ways to partition the pooled sample A into two subsets of size n_1 and n_2). If $T(A_1, A_2) = T(A_v)$ be the observed test statistic, then the **permutation distribution** of T^* is the distribution of replicates $\{T^*\} = \{T(A_{\pi_j(v)}), j = 1, \dots, \binom{n}{n_1}\}$ where $\pi_j(v)$ is the j th permutation of v . The

cumulative distribution function of T^* is given by

$$F_{T^*}(t) = P(T^* \leq t) = \frac{n_1!n_2!}{n!} \sum_{j=1}^n I(T_j \leq t), \quad (4.1)$$

where $I(X)$ is the indicator function on the set X (i.e., $I(x) = 1$ if $x \in X$ and $I(x) = 0$ if $x \notin X$) (Rizzo, 2007, ch.8).

The achieved significance level (ASL) is given by

$$P(T^* \geq T) = \frac{n_1!n_2!}{n!} \sum_{j=1}^n I(T_j \geq T),$$

where $T = T(A_v)$ is the observed test statistic (Rizzo, 2007, ch.8).

In practice for large numbers of permutations, we use the empirical p -value described in (Davison et al., 1997, Ch.4) as the ASL given by

$$\eta = \frac{1 + \sum_{j=1}^B I(T_j \geq T)}{B + 1}, \quad (4.2)$$

where B (usually $B \geq 10000$) is the number of permutations, and $\forall j$, $T_j = T(A_{\pi_j(v)})$ is the permutation test statistic. Thus, we reject the null hypothesis at significance level α if $\eta \leq \alpha$ (Rizzo, 2007, ch.8).

4.4 Mean s-rep and s-rep test statistic

Before testing the mean differences of s-rep GOPs, we need to calculate the mean s-rep. This can be done by different methods. In this section we explain two possible approaches.

4.4.1 Mean s-rep by CPNS

Assume a set of scaled and aligned s-reps $s_1^\dagger, \dots, s_N^\dagger$. As discussed in Section 3.5, to build the Z_{comp} matrix we apply PNS on skeletal positions and directions to calculate the residuals. Then we find the geometric mean of the spokes' lengths and scale factors. Finally we scaled the data to become commensurate. In the same manner we can define mean s-rep $(\bar{s}^\dagger, \bar{\gamma})$ as follows

$$(\bar{s}^\dagger, \bar{\gamma}) = (\bar{\gamma}, \bar{P}, \bar{\mathbf{u}}_1, \dots, \bar{\mathbf{u}}_{n_s}, \bar{r}_1, \dots, \bar{r}_{n_s}), \quad (4.3)$$

where $\bar{\gamma}$ is the geometric mean of scale factors (i.e., mean of centroid sizes), \bar{P} is the PNS mean of skeletal positions, $\bar{\mathbf{u}}_i, i = 1, \dots, n_s$ is the PNS mean of i th spoke's direction and $\bar{r}_i, i = 1, \dots, n_s$

is the geometric mean of i th spoke's length. In this sense, we defined the mean s-rep element wise.

(Jung, 2011, 2014; Pizer et al., 2013) by applying PNS, defined transformations between the s-rep space \mathcal{S} and the Euclidean space E_{comp} such that, $\mathcal{H} : \mathcal{S} \rightarrow E_{comp}$ and $\tilde{\mathcal{H}} : E_{comp} \rightarrow \mathcal{S}$, where $\tilde{\mathcal{H}}$ is the inverse transformation of \mathcal{H} . The definitions of \mathcal{H} and $\tilde{\mathcal{H}}$ are given by

$$\mathcal{H}((s^\dagger, \gamma)) = \begin{bmatrix} \bar{\gamma} h_P(PC) \\ \bar{\gamma} \log(\gamma/\bar{\gamma}) \\ \bar{r}_1 h_1(\mathbf{u}_1) \\ \vdots \\ \bar{r}_{n_s} h_{n_s}(\mathbf{u}_{n_s}) \\ \bar{r}_1 \log(r_1/\bar{r}_1) \\ \vdots \\ \bar{r}_{n_s} \log(r_{n_s}/\bar{r}_{n_s}) \end{bmatrix} \in E_{comp}, \quad \tilde{\mathcal{H}}(\mathbf{z}) = \begin{bmatrix} \tilde{h}_P((z_1, \dots, z_m)^T/\bar{\gamma}) \\ \bar{\gamma} \exp(z/\bar{\gamma}) \\ \tilde{h}_1(z^{s_1}/\bar{r}_1) \\ \vdots \\ \tilde{h}_{n_s}(z^{s_{n_s}}/\bar{r}_{n_s}) \\ \bar{r}_1 \exp(z/\bar{r}_1) \\ \vdots \\ \bar{r}_{n_s} \exp(z/\bar{r}_{n_s}) \end{bmatrix} \in \mathcal{S},$$

where h_P, h_1, \dots, h_{n_s} are transformation functions to calculate PNS residuals, $\tilde{h}_P, \tilde{h}_1, \dots, \tilde{h}_{n_s}$ are transformation functions to return spherical data based on the Euclidean data (e.g, PNS residuals), and the rest of the variables are defined in Section 3.5.

Since the mean of the Z_{comp} is at the origin (i.e., Z_{comp} mean is a zero vector). The mean s-rep can be defined as the inverse transformation of the zero vector $\tilde{\mathcal{H}}((0, \dots, 0)^T) \in \mathcal{S}$. We can show, this approach could be identical to computing mean s-rep element wise. Assume \mathcal{H} and $\tilde{\mathcal{H}}$ define an isomorphism. We search for an s-rep \hat{s}^\dagger such that $\mathcal{H}((\hat{s}^\dagger, \hat{\gamma})) = (0, \dots, 0)^T$. By inserting the s-rep $(\bar{s}^\dagger, \bar{\gamma})$ from (4.3) in \mathcal{H} we have

$$\mathcal{H}((\bar{s}^\dagger, \bar{\gamma})) = \begin{bmatrix} \bar{\gamma} h_P(\bar{P}) \\ \bar{\gamma} \log(\bar{\gamma}/\bar{\gamma}) \\ \bar{r}_1 h_1(\bar{\mathbf{u}}_1) \\ \vdots \\ \bar{r}_{n_s} h_{n_s}(\bar{\mathbf{u}}_{n_s}) \\ \bar{r}_1 \log(\bar{r}_1/\bar{r}_1) \\ \vdots \\ \bar{r}_{n_s} \log(\bar{r}_{n_s}/\bar{r}_{n_s}) \end{bmatrix} = \begin{bmatrix} \bar{\gamma} h_P(\bar{P}) \\ 0 \\ \bar{r}_1 h_1(\bar{\mathbf{u}}_1) \\ \vdots \\ \bar{r}_{n_s} h_{n_s}(\bar{\mathbf{u}}_{n_s}) \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Furthermore, the outcomes of h_P, h_1, \dots, h_{n_s} , are equal to zero because the principle scores (residuals) of the PNS means are zero. Thus,

$$\begin{aligned} \mathcal{H}((\bar{s}^\dagger, \bar{\gamma})) = (0, \dots, 0)^T &\Rightarrow \tilde{\mathcal{H}}(\mathcal{H}((\bar{s}^\dagger, \bar{\gamma}))) = \tilde{\mathcal{H}}((0, \dots, 0)^T) \Rightarrow \\ (\bar{s}^\dagger, \bar{\gamma}) = \tilde{\mathcal{H}}((0, \dots, 0)^T) &\Rightarrow (\bar{s}^\dagger, \bar{\gamma}) = (\hat{s}^\dagger, \hat{\gamma}). \end{aligned}$$

Therefore, the inverse transformation of the zero vector is $(\bar{s}^\dagger, \bar{\gamma})$.

The definition of transformation functions (i.e., h, \tilde{h}) is beyond the scope of this thesis. R implementation of them can be found in `shapes::PNSs2e` and `shapes::PNSe2s` from the package "shapes" (Dryden, 2018). For more detail see (Jung, 2011, 2014; Pizer et al., 2013; Jung et al., 2012).

Note that as we discussed in Section 3.5.1 we prefer to find mean s-rep by CPNG rather than CPNS and the procedure is the same.

4.4.2 Mean s-rep by GPA and PGA

(Dryden and Mardia, 2016, ch.6) calculated the mean of the landmark models by different methods and compared the results. Based on both intrinsic and extrinsic distances the book concluded the means are very similar unless we have outliers. Also, (Jung, 2011, ch.7) confirms that GPA mean and PNG mean in most cases are virtually the same. On the other hand, by increasing the number of spokes, the dimension of the pre-shape space S^{3n_s-1} increases, and the PNG has to fit higher dimensional sub-spheres. Therefore, by adding new spokes to the model, the PNG performance becomes noticeably slower. Since as a part of our analysis, we want to apply a permutation test on mean s-reps, it is reasonable to reduce cost of computation. In this regard, to find mean skeletal PDM we substitute PNG mean with GPA mean (2.22). Also, for spokes' directions if they have concentrated distributions, we can use PGA mean instead of PNG mean. Note that if we have concentrated spherical data then, Fréchet mean, PGA mean and the projection of the center of gravity on the sphere (extrinsic mean (Mardia et al., 1982, ch.15)) are close. Again it is logical to choose the fastest approach (e.g., Fréchet mean for the spokes' directions). Table 4.2 indicates the average computation time to calculate mean s-rep in R.

Method to calculate mean s-rep	Computation time
CPNG mean (Max iteration 300 default)	~3158 sec.
CPNG mean (Max iteration 50)	~793 sec.
GPA mean of skeletal positions + PNG mean of directions + Geometric mean of spokes' lengths and scale factors	~60 sec.
GPA mean of skeletal positions + Fréchet mean of directions + Geometric mean of spokes' lengths and scale factors	~10 sec.

Table 4.2: Mean s-rep computation time.

Computation time to find mean of 108 s-reps (control group) with 114 spokes in R (CPU: Core i5). Max iteration is associated with nls.lm function of "shapes" package in shapes::pns::getSubSphere::sphereFit.

4.4.3 s-rep test statistic

To define test statistics, first we need to define s-rep distance element-wise. (Schulz et al., 2016) introduced distance for scaled s-reps with respect to the s-rep GOPs as follows:

$$\begin{aligned}
d_{s^\dagger}((s_1^\dagger, \gamma_1), (s_2^\dagger, \gamma_2)) &= (d(\log(\gamma_1), \log(\gamma_2)), \\
& d(\mathbf{p}_{11}, \mathbf{p}_{21}), \dots, d(\mathbf{p}_{1n_s}, \mathbf{p}_{2n_s}), \\
& d_g(\mathbf{u}_{11}, \mathbf{u}_{21}), \dots, d_g(\mathbf{u}_{1n_s}, \mathbf{u}_{2n_s}), \\
& d(\log(r_{11}), \log(r_{21})), \dots, d(\log(r_{1n_s}), \log(r_{2n_s})))^T,
\end{aligned} \tag{4.4}$$

where d is the Euclidean distance, and d_g is the geodesic distance.

In Section 4.4, we discussed how to calculate mean s-rep of a population, and d_{s^\dagger} (4.4) defines the distance between two s-reps according to their GOPs' differences. Based on d_{s^\dagger} , the test statistic T consists of K partial tests, where K is the number of GOPs (i.e., skeletal positions, spokes' directions, and spokes' lengths) plus the scale factor.

Assume two sample groups of s-reps $A_1 = \{s_{A_1 1}^\dagger, \dots, s_{A_1 n_1}^\dagger\}$ and $A_2 = \{s_{A_2 1}^\dagger, \dots, s_{A_2 n_2}^\dagger\}$ of sizes n_1 and n_2 . The test statistic can be defined as a vector by

$$\mathbf{T}(A_1, A_2) = d_{s^\dagger} \left((\bar{s}_{A_1}^\dagger, \bar{\gamma}_{A_1}), (\bar{s}_{A_2}^\dagger, \bar{\gamma}_{A_2}) \right),$$

where $(\bar{s}_{A_1}^\dagger, \bar{\gamma}_{A_1})$ and $(\bar{s}_{A_2}^\dagger, \bar{\gamma}_{A_2})$ are the observed sample mean of A_1 and A_2 . Let T_k be the k th element of $\mathbf{T}(A_1, A_2)$, $k = 1, \dots, K$. Obviously, $\mathbf{T}(A_1, A_2)$ consists of K partial test statistics T_k , associated with partial distances of d_{s^\dagger} . Consequently we can design k partial tests for each GOP to test the mean difference. Let $s^\dagger(k)$ be the k th GOP of the s-rep s^\dagger . The partial test is given by

$$H_{0k} : \bar{s}_{A_1}^\dagger(k) = \bar{s}_{A_2}^\dagger(k) \quad \text{vs.} \quad H_{1k} : \bar{s}_{A_1}^\dagger(k) \neq \bar{s}_{A_2}^\dagger(k)$$

4.5 Non-parametric approach for s-reps

4.5.1 Partial test with FWER

Consider s-reps with K number of GOPs. For the partial test, we have the test statistics $\mathbf{T}_0 = \mathbf{T}(A_1, A_2)$ of the observed samples and $\mathbf{T}_i = \mathbf{T}(A_{1i}, A_{2i})$, $i = 1, \dots, B$ from B permuted samples. $\mathbf{T}_i = (T_{i1}, \dots, T_{iK})^T$, $\forall i \in \{0, 1, \dots, B\}$. (Schulz et al., 2016) proposed a non-parametric method of testing based on GOPs differences and controlling by FWER. Since test statistic \mathbf{T}_i measures the GOP differences in different units, its elements are not commensurate. We map the normalized GOP differences to a common coordinate system in a way that preserves the multivariate dependence structure between GOPs. In this regard, at the first step, we map all GOP differences to an approximately uniform distribution on the interval $[0, 1]$ by using a cumulative distribution function (CDF) of T_{ik} as

$$C_k(T_{ik}) = \frac{1}{B} \sum_{j=1}^B I(T_{jk} \leq T_{ik}), \quad k = 1, \dots, K,$$

where I is the indicator function (see Equation (4.1)). Now we can map $C_k(T_{ik})$ to $\mathcal{N}(0, 1)$ and represent GOP differences as standard normally distributed variables by

$$U_{ik} = \Phi^{-1}(\tilde{C}_k(T_{ik})), \quad (4.5)$$

where Φ^{-1} is the inverse CDF of standard normal distribution, $k = 1, \dots, K$, $i = 1, \dots, B$, and

$$\tilde{C}_k(T_{ik}) = \frac{sc - 2}{sc} C_k(T_{ik}) + \frac{1}{sc}, \quad sc = 10,000.$$

Φ^{-1} is a monotonic function on $(-\infty, \infty)$ and $\Phi^{-1}(x) \rightarrow \pm\infty$ when $x \rightarrow 0, 1$. we use $\tilde{C}_k(T_{ik})$ as scaled form of $C_k(T_{ik})$ to make sure the input value of Φ^{-1} is between zero and one. The marginal distribution of U_{ik} is standard normal $U_{ik} \sim \mathcal{N}(0, 1)$, $k = 1, \dots, K$. With the same procedure we map the observed GOP differences T_{0k} to U_{0k} .

We design an FWER method to control type I error for all GOPs simultaneously. Assume a threshold δ for all $U_{ik} \sim \mathcal{N}(0, 1)$. The aim is to find GOPs such that

$$P(U_{01} \leq \delta, \dots, U_{0K} \leq \delta) = 1 - \frac{\alpha}{2}, \quad (4.6)$$

where α is the level of significance (e.g., $\alpha = 0.05$).

Theorem 4.2. (Multivariate Central Limit Theorem) "Let $\mathbf{U}_1, \dots, \mathbf{U}_B$ be i.i.d random vectors $\mathbf{U}_i = (U_{i1}, \dots, U_{iK})^T \in \mathbb{R}^K$, $i = 1, \dots, B$ with mean vector $\mathbf{0}$ and covariance matrix Σ .

Then

$$\frac{1}{\sqrt{B}} \sum_{i=1}^B \mathbf{U}_i \xrightarrow{w} \mathcal{N}_K(\mathbf{0}, \Sigma),$$

where \mathcal{N}_K is a K -dimensional multivariate normal distribution with mean $\mathbf{0}$, covariance Σ , and \xrightarrow{w} denote the weak convergence i.e., convergence in distribution" (Schulz et al., 2016).

From (4.5) we have B independent and identically distributed vectors $\mathbf{U}_i = (U_{i1}, \dots, U_{iK})^T$, $i = 1, \dots, B$. Therefore by theorem 4.2, \mathbf{U}_i is approximately distributed as $\mathcal{N}_K(\mathbf{0}, \hat{\Sigma}_U)$ where

$$\hat{\Sigma}_U = \frac{1}{B-1} U^T U, \quad U = (\mathbf{U}_1, \dots, \mathbf{U}_B)^T. \quad (4.7)$$

Each marginal distribution is $U_{ik} \sim \mathcal{N}(0, 1)$. With the level of significance α , similar to (4.6) we have

$$P(U_{i1} < \delta, \dots, U_{iK} < \delta) = 1 - \frac{\alpha}{2}, \quad i = 1, \dots, B.$$

For the variable $\mathbf{U}_i \sim \mathcal{N}_K(\mathbf{0}, \hat{\Sigma}_U)$ the probability P is a multiple integral from $-\infty$ to δ . Thus we can consider it as a function of δ like

$$\mathcal{G}(\delta) = P(U_{i1} < \delta, \dots, U_{iK} < \delta).$$

\mathcal{G} is a monotonically increasing function with a horizontal asymptote at 1. By assuming $\alpha \leq 0.05$ the threshold δ will be near 1, because $\mathcal{G}(\delta) = 1 - \alpha/2$ and $\forall i, U_{ik} \sim \mathcal{N}(0, 1)$. Now let $\delta_1 = \Phi^{-1}(1 - \frac{\alpha}{2})$ be the threshold for perfectly correlated U_{ik} and $\delta_2 = \Phi^{-1}\left(\left(1 - \frac{\alpha}{2}\right)^{1/K}\right)$ be the threshold for independent U_{ik} . Since $\delta \gg 0$, we can assume $0 < \delta_1 \leq \delta \leq \delta_2$. (Schulz et al., 2016) showed that $\mathcal{G}(\delta)$ is concave downwards in the interval $[\delta_1, \delta_2]$. By a given δ , the value of $\mathcal{G}(\delta)$ can be estimated by N large enough number (e.g., 200,000) of random samples $\mathbf{y}_j \sim \mathcal{N}_K(\mathbf{0}, \hat{\Sigma}_U)$, $j = 1, \dots, N$ as follows

$$\hat{\mathcal{G}}(\delta) = \frac{\sum_{j=1}^N I_\delta(\mathbf{y}_j)}{\sum_{j=1}^N \vartheta(\mathbf{y}_j)}, \quad I_\delta(\mathbf{y}_j) = \begin{cases} \vartheta(\mathbf{y}_j) & \vartheta(\mathbf{y}_j) < \delta, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbf{y}_j = (y_{j1}, \dots, y_{jK})^T$, and ϑ is the density function of $\mathcal{N}_K(\mathbf{0}, \hat{\Sigma}_U)$. With the estimated values $\hat{\mathcal{G}}(\delta_1)$ and $\hat{\mathcal{G}}(\delta_2)$, and by applying the iterative Regula Falsi method on the interval $[\delta_1, \delta_2]$, we can solve the equation $\mathcal{G}(\delta) = 1 - \alpha/2$. At the end, we consider GOPs with $|U_{0k}| \geq \delta$, $k = 1, \dots, K$ as significant (Schulz et al., 2016).

4.5.2 Partial test with FDR

By the strong assumption of independence as we discussed in Section 4.2, we can implement FWER or FDR methods on the empirical p -values to control type I error.

Assume we have K independent partial tests. A classical approach to test partial differences is based on calculating empirical p -values from Equation (4.2). Assume B is the number of permutations. The empirical p -values for the partial tests are given by

$$\eta_k = \frac{1 + \sum_{i=1}^B I(T_{ik} \geq T_{0k})}{B + 1}, \quad k = 1, \dots, K.$$

A set of K obtained empirical p -values are significant if $\{\eta_j\}_{j \in J} \leq \alpha$ where $J \subseteq \{1, \dots, K\}$, and α is the significance level (e.g., $\alpha = 0.05$). Since we have multiple testing, we use FWER or FDR (e.g., Benjamini-Hochberg method) to control false positives. Also, it would be easier to work with adjusted p -values. As a result, with the Benjamini-Hochberg method if δ be the FDR rate, we mark a GOP as statistically significant if the corresponding adjusted empirical p -value is less than δ .

4.5.3 Global test for non-parametric approach

Generally, if we have n individual null hypotheses H_{01}, \dots, H_{0n} , the global hypothesis test is to test the global null $H_0 = \cap_{i=1}^n H_{0i}$. Thus, we reject global null if at least one of the individual nulls is rejected (Candes, 2018). In s-rep we designed K partial tests (see Section 4.4.3) so we have K null hypothesis and the global null is $\cap_{k=1}^K H_{0k}$. With the assumption that partial tests are marginally consistent, significant, and unbiased, a suitable combining function would produce an unbiased test for the global test (Schulz et al., 2016).

For the permutation testing, (Pesarin, 2001, ch.6) discussed different combining functions to define global test for the permutation.

Assume two groups of s-reps A_1, A_2 . Let $A = A_1 \cup A_2$ be the ordered pooled group, and A^* be a partition of A where the random samples are s-reps with K number of GOPs. As discussed in Section 4.3.1 the vector of observed test statistic is $\mathbf{T}_0 = \mathbf{T}_A$ where $\mathbf{T}_A = \mathbf{T}(A_1, A_2)$. Let $\mathbf{T}^* = \mathbf{T}_{A^*}$ be the vector of test statistic correspond to a partition A^* , and $r = 1, \dots, B$ be the r th permutation. Then $\forall x \in \mathbb{R}$ and $i = 1, \dots, K$,

$$\hat{L}_i(x|A) = \frac{1/2 + \sum_{r=1}^B I(T_{ir}^* \geq x)}{(B + 1)},$$

gives an estimate of the marginal permutation significant level function $L_i(x|A) = Pr\{T_i^* \geq x|A\}$. Therefore, the k observed p -values are estimated on A by $\hat{\eta}_i = \hat{L}_i(T_{i0}|A)$ where T_{i0} is the observed test statistic of the i th GOP, $i = 1, \dots, K$. Also, $\eta_{ir}^* = \hat{L}_i(T_{ir}^*|A)$ represent the empirical p -value associated with the i th GOP in the r th permutation (Pesarin, 2001, ch.6); (SALMASO and Brombin, 2013, ch.2).

Now, assume ψ as a combining function, then $T_0'' = \psi(\hat{\eta}_1, \dots, \hat{\eta}_K)$ and $T_r''^* = \psi(\eta_{1r}^*, \dots, \eta_{Kr}^*)$ represent the combined p -values of observed values and vector statistics respectively. Hence, the p -value of combined test T'' can be estimated by

$$\hat{\eta}_\psi'' = \sum_{r=1}^B I(T_r''^* \geq T_0'')/B.$$

We reject the global null hypothesis H_0 at α level if $\hat{\eta}_\psi'' \leq \alpha$ (Pesarin, 2001, ch.6).

Combining functions

Assume $i = 1, \dots, K$, where K is the number of s-rep GOPs plus scale factor. The combining functions we use in this thesis (discussed in more detail in (Pesarin, 2001, ch.6)) are as follows:

- Fisher omnibus combining function: $T_F'' = -2\sum_{i=1}^K \log(\eta_i)$.
- Liptak combining function: $T_L'' = \sum_{i=1}^K \Phi^{-1}(1 - \eta_i)$ where Φ is the standard normal CDF.
- Logistic Liptak combining function: $T_L'' = \sum_{i=1}^K \log((1 - \eta_i)/\eta_i)$.
- Tippett combining function: $T_T'' = \max_{1 \leq i \leq K} (1 - \eta_i)$.
- Mahalanobis quadratic combining function: $T_Q'' = \mathbf{U}^T (R_U^*)^{-1} \mathbf{U}$ where

$$\mathbf{U}^T = (\Phi^{-1}(1 - \eta_1), \dots, \Phi^{-1}(1 - \eta_K))^T, \quad (4.8)$$

Φ^{-1} is the inverse of standard normal CDF, and R_U^* is the correlation matrix of the transformed permutation p -values. The transformation of the permutation p -values $\boldsymbol{\eta}_r^* = (\eta_{1r}^*, \dots, \eta_{Kr}^*)^T$, $r = 1, \dots, B$ is similar to (4.8). R_U^* assumed to be positive definite and can be estimated by $R_{U_{ij}}^* = \sum_{r=1}^B U_{jr}^* \cdot U_{ir}^* / B$ where $U_{ij}^* = \Phi^{-1}(1 - \eta_{ij}^*)$, $i, j = 1, \dots, K$ (Pesarin, 2001, ch.6).

Note that in practice we substitute η_i with $\hat{\eta}_i$.

Tables 4.3 and 4.4 from (Pesarin, 2001, page.146-150) summarize the non-parametric combination method for the global test.

In addition, (Schulz, 2013) defined global test for s-rep based on Mahalanobis distance. Assume $M_0 = \mathbf{U}_0^T \hat{\Sigma}_U^{-1} \mathbf{U}_0$ and $M_i = \mathbf{U}_i^T \hat{\Sigma}_U^{-1} \mathbf{U}_i$ where $\hat{\Sigma}_U^{-1}$, \mathbf{U}_0 and \mathbf{U}_i come from (4.6) and (4.7). Then the empirical global p -value is

$$\hat{\eta} = \frac{1}{B} \sum_{i=1}^B I(M_i \geq M_0)$$

we reject the global null if $\hat{\eta} \leq \alpha$. According to our observation, the result of this approach is close to the result of the global test by the Mahalanobis quadratic combining function.

Z	Z_1^*	\dots	Z_r^*	\dots	Z_B^*
T_{10}	T_{11}^*	\dots	T_{1r}^*	\dots	T_{1B}^*
\vdots	\vdots		\vdots		\vdots
T_{K0}	T_{K1}^*	\dots	T_{Kr}^*	\dots	T_{KB}^*

↓

$\hat{\eta}_1$	η_{11}^*	\dots	η_{1r}^*	\dots	η_{1B}^*
\vdots	\vdots		\vdots		\vdots
$\hat{\eta}_K$	η_{K1}^*	\dots	η_{Kr}^*	\dots	η_{KB}^*

↓

T_0''	$T_1''^*$	\dots	$T_r''^*$	\dots	$T_B''^*$
---------	-----------	---------	-----------	---------	-----------

Table 4.3: Non-parametric combination.

Z	Z_1^*	\dots	Z_r^*	\dots	Z_B^*
T_{10}	T_{11}^*	\dots	T_{1r}^*	\dots	T_{1B}^*
\vdots	\vdots		\vdots		\vdots
T_{K0}	T_{K1}^*	\dots	T_{Kr}^*	\dots	T_{KB}^*

↓

$\hat{\eta}_1$	η_{11}^*	\dots	η_{1r}^*	\dots	η_{1B}^*
\vdots	\vdots		\vdots		\vdots
$\hat{\eta}_K$	η_{K1}^*	\dots	η_{Kr}^*	\dots	η_{KB}^*

↓

U_{10}	U_{11}^*	\dots	U_{1r}^*	\dots	U_{1B}^*
\vdots	\vdots		\vdots		\vdots
U_{K0}	U_{K1}^*	\dots	U_{Kr}^*	\dots	U_{KB}^*

↓

T_{Q0}''	$T_{Q1}''^*$	\dots	$T_{Qr}''^*$	\dots	$T_{QB}''^*$
------------	--------------	---------	--------------	---------	--------------

Table 4.4: Quadratic combination.

4.6 Parametric approach

We can see from Table 4.2 the non-parametric hypothesis test for s-rep is hugely time-consuming. In this regard, we attempt to introduce a parametric approach to test both partial and global differences as a fast approach. Before going any further, let us discuss the Hotelling's T^2 test for multivariate data.

4.6.1 Hotelling's T^2 test

Hotelling's T^2 test is the generalization of the t-test (discussed in Section 4.1.1) for multivariate normal distributions. Suppose $\mathbf{x}_1, \dots, \mathbf{x}_{n_1} \sim \mathcal{N}_m(\boldsymbol{\mu}_1, \Sigma)$ and $\mathbf{y}_1, \dots, \mathbf{y}_{n_2} \sim \mathcal{N}_m(\boldsymbol{\mu}_2, \Sigma)$ are two independent random samples, where $\mathcal{N}_m(\boldsymbol{\mu}, \Sigma)$ is an m -dimensional multivariate distribution with mean $\boldsymbol{\mu}$, and covariance matrix Σ . The aim is to test the mean difference as

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ vs. } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

Let $\bar{\mathbf{x}} = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{x}_i$, and $\bar{\mathbf{y}} = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{y}_i$ be sample means, and

$$\hat{\Sigma}_x = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T, \quad \hat{\Sigma}_y = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T,$$

be sample covariance matrices. Assume $n = n_1 + n_2$ then, sample Mahalanobis distance is given by

$$d_{\mathcal{M}}^2 = (\bar{\mathbf{x}} - \bar{\mathbf{y}})^T \hat{\Sigma}_p^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}}), \quad (4.9)$$

where $\hat{\Sigma}_p = (n_1 \hat{\Sigma}_x + n_2 \hat{\Sigma}_y)/(n - 2)$ is an unbiased estimate of common covariance matrix. By using test statistic

$$T = \frac{n_1 n_2 (n - m - 1)}{n(n - 2)m} d_{\mathcal{M}}^2 \sim F_{(m, n-m-1)}, \quad (4.10)$$

we reject H_0 for large value of T , where $F_{(m, n-m-1)}$ is the F distribution with m and $(n - m - 1)$ degrees of freedom (Mardia et al., 1982, ch.3), (Dryden and Mardia, 2016, page.188).

4.6.2 s-rep partial tests with parametric approach

A closer look at Equation (2.22) helps us to understand the skeletal PDM mean shape after the alignment and inspires us to design a parametric hypothesis test for s-rep.

Assume after the alignment we have two sample groups of s-reps $A_1 = \{s_{A_1 1}^\dagger, \dots, s_{A_1 n_1}^\dagger\}$, and $A_2 = \{s_{A_2 1}^\dagger, \dots, s_{A_2 n_2}^\dagger\}$. Let $\mathbf{p}_{ji}^{A_1}$, and $\mathbf{p}_{jk}^{A_2}$, be the j th spoke positions of all s-reps in A_1 and A_2 respectively, where $i = 1, \dots, n_1$, $k = 1, \dots, n_2$, and $j = 1, \dots, n_s$. Assume $\forall j$ $\mathbf{p}_{j1}^{A_1}, \dots, \mathbf{p}_{jn_1}^{A_1} \sim \mathcal{N}_3(\boldsymbol{\mu}_j^{A_1}, \Sigma_j)$, and $\mathbf{p}_{j1}^{A_2}, \dots, \mathbf{p}_{jn_2}^{A_2} \sim \mathcal{N}_3(\boldsymbol{\mu}_j^{A_2}, \Sigma_j)$ where \mathcal{N}_3 is the 3-dimensional multivariate normal distribution. Then n_s partial hypothesis tests can be defined as

$$H_{0j} : \boldsymbol{\mu}_j^{A_1} = \boldsymbol{\mu}_j^{A_2} \quad \text{vs.} \quad H_{1j} : \boldsymbol{\mu}_j^{A_1} \neq \boldsymbol{\mu}_j^{A_2}, \\ j = 1, \dots, n_s.$$

We assumed points have multivariate normal distributions. Thus we can implement Hotelling's T^2 test and calculate the p -values. Also, we have n_s corresponding spoke directions. Similar to the spoke positions let $\mathbf{u}_{ji}^{A_1}$, and $\mathbf{u}_{jk}^{A_2}$, be the j th spoke directions of all s-reps in A_1 and A_2 respectively. As we discussed PGA in Section 3.2, $\forall j$, we can project directions to their tangent spaces $T_{\boldsymbol{\mu}_{uj}} S^2$ with their pooled means $\boldsymbol{\mu}_{uj}$, $j = 1, \dots, n_s$. The tangent spaces are two-dimensional plane. Assume $\forall j$, $\log_{\boldsymbol{\mu}_{uj}}(\mathbf{u}_{j1}^{A_1}), \dots, \log_{\boldsymbol{\mu}_{uj}}(\mathbf{u}_{jn_1}^{A_1}) \sim \mathcal{N}_2(\boldsymbol{\mu}_{uj}^{A_1}, \Sigma_j)$, and $\log_{\boldsymbol{\mu}_{uj}}(\mathbf{u}_{j1}^{A_2}), \dots, \log_{\boldsymbol{\mu}_{uj}}(\mathbf{u}_{jn_2}^{A_2}) \sim \mathcal{N}_2(\boldsymbol{\mu}_{uj}^{A_2}, \Sigma_j)$. The partial tests become $H_{0j} : \boldsymbol{\mu}_{uj}^{A_1} = \boldsymbol{\mu}_{uj}^{A_2}$ vs. $H_{1j} : \boldsymbol{\mu}_{uj}^{A_1} \neq \boldsymbol{\mu}_{uj}^{A_2}$, $j = 1, \dots, n_s$. Therefore, we can apply Hotelling's T^2 test on the log-mapped data and calculate the p -values. ((Dryden and Mardia, 2016, ch.9) used similar method to hypothesize mean shapes on the pre-shape tangent space). Next, we use a t-test for log scale factors and log spoke lengths and compute the associated p -values. As a result we have $(3n_s + 1)$ p -values, (i.e., n_s for skeletal positions, n_s for directions, n_s for spokes' lengths and 1 for scale factor). Finally, we have a multiple comparisons problem, and by the assumption of independence, we can use the Benjamini-Hochberg method to control the false positives. Analogous to this method, (McClure et al., 2013) applied FDR in the study of localized differences in caudate and hippocampal of schizophrenia based on m-rep.

Although this method considers strong assumptions of normality and independence, it could be helpful as it is swift (i.e., computation time is less than a minute) and approximates the result of the permutation test.

4.6.3 Global test for the parametric approach

We can use the most popular methods for the global test: Fisher's combination test and Bonferroni's method.

Bonferroni's method

Bonferroni's method is simple and straightforward but conservative. Assume n null hypotheses H_{01}, \dots, H_{0n} . Let p -value p_i corresponds to H_{0i} then, we reject global null H_0 at α level if $\min_i p_i \leq \alpha/n$ (Candes, 2018).

Fisher combination test

Suppose we have n independent p -values p_1, \dots, p_n then we reject global null hypothesis for the large value of the test statistics $T = -2 \sum_{i=1}^n \log(p_i) \sim \chi_{2n}^2$, where χ_{2n}^2 is the Chi-square distribution with $2n$ degrees of freedom. (Mosteller and Fisher, 1948; Candes, 2018).

4.6.4 SPHARM-PDM partial test with parametric approach

In SPHARM-PDM with k points, we have k partial tests. (Styner et al., 2006) used the Hotelling T^2 metric (i.e., Mahalanobis distance (4.9)) in the permutation. Then applied the Benjamini-Hochberg method to control false positive. In practice, the results of the permutation and the parametric approach (with normality assumption and Hotelling's T^2 test) are close. In this regard, the parametric method might be preferred as a noticeably faster approach (compare Figures 5.9 and A.1). Also, by having k p -values, we can apply the Bonferroni's method and Fisher combination test from Section 4.6.3 to test the global differences.

4.6.5 Partial test for triangle mesh normal vectors

From Section 2.1.4, we saw that SPHARM-PDM provides us with corresponding triangulated surfaces and, consequently, corresponding unit normal vectors. The parametric hypothesis approach to test the mean direction of the unit normal vectors could be the same as in the hypothesis test for s-rep spokes' directions in Section 4.6.2. Note that it is crucial to align meshes before extracting the directional information of normal vectors.

4.7 Hypothesis test independent from the alignment

Most of the hypothesis testing methods are dependent on alignment. We will see later (see Figure 5.7) different methods of alignment cause different results. In this regard, we propose a novel hypothesis test independent from the alignment by considering the intrinsic properties of shapes.

4.7.1 Hypothesis test with distance

One intrinsic property of the shapes is the distance between GOPs. Distance is a general concept and may vary depending on the space. For example, we can use Euclidean distance between pair of points, Manhattan distance between pair of pixels, or geodesic distance between directions. For more detail see (Christensen, 2010) which discussed different norms in vector spaces, and (Pennec et al., 2019) which studied the shape statistics with geodesic distance.

The distance between shape GOPs (e.g., the distance between two landmarks of a shape) is an intrinsic property independent from the shape location and orientation. We build a hypothesis test to look at locational mean differences among shape populations. For further elaboration, let us first define the distance matrix for a landmark model.

Definition 4.3. Assume a configuration matrix $X = (\mathbf{x}_1, \dots, \mathbf{x}_k)^T$, $\mathbf{x}_i \in \mathbb{R}^m$. The **distance matrix** D_X for X is given by

$$D_X = \begin{bmatrix} d(\mathbf{x}_1, \mathbf{x}_1) & \dots & d(\mathbf{x}_1, \mathbf{x}_k) \\ \vdots & \ddots & \vdots \\ d(\mathbf{x}_k, \mathbf{x}_1) & \dots & d(\mathbf{x}_k, \mathbf{x}_k) \end{bmatrix}, \quad (4.11)$$

where d is a selected distance measure (e.g., Euclidean distance). Similar definition for the Euclidean distance matrix can be found in (Dryden and Mardia, 2016, ch.15).

Obviously, by considering the geodesic distance, we can define a geodesic distance matrix. The geodesic distance matrix can be used for the directions of shape representations containing directional data (e.g., s-rep). On the other hand, one may use the effect of distance on the GOPs. For instance, in Newton's universal gravitation law, $F = G \frac{m_1 m_2}{r^2}$, objects impose more gravitational force on each other when they are closer (Goldstein et al., 2013). The inverse squared distance defines the force between identical masses where the force is a function of distance. We will explain some possible ways to determine distances and forces briefly in Section 6.2.2. Study of the distance on shape requires extensive discussion and could be the subject of future study.

Whether we choose a distance matrix or a force matrix, the hypothesis test that we provide has the same procedure. Therefore, for simplicity, we focus on the distance matrix with the Euclidean distance. Note that for s-rep analysis, we substitute the Euclidean distance matrix with the geodesic distance matrix to analyze spokes' directions.

Any infinitesimal change in the position of a landmark is traceable from the distance matrix. In other words, if we transform shape to another one, during the transformation, the elements of the distance matrix start to change. The change is traceable, so we can find which points have more effect on the transformation based on the differences between the initial distance matrix and the last one. For example, in Figure 4.1 we can see, by transforming the shape 1 to the shape 2 which are different in one landmark, the corresponding distance matrices become different in the row and column associated with that landmark (i.e., here landmark number 1).

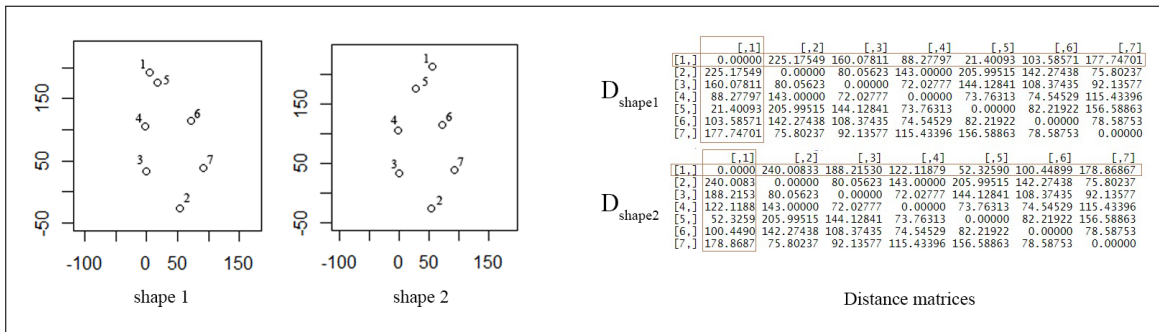


Figure 4.1: Distance matrices.

Left: Similar shapes with the difference in the first landmark. Right: Distance matrices highlight the difference of distances associate to the first landmark. (Illustrated shapes are modified version of female gorilla skull sample no.1, taken from (Dryden, 2018))

Assume two sets of k -landmark models $A_1 = \{X_1, \dots, X_{n_1}\}$, and $A_2 = \{Y_1, \dots, Y_{n_2}\}$. Let $\{D_{X_1}, \dots, D_{X_{n_1}}\}_{A_1}$, $\{D_{Y_1}, \dots, D_{Y_{n_2}}\}_{A_2}$ be the corresponding sets of distance matrices. Then, $(D_{X_i})_{rs}$ and $(D_{Y_j})_{rs}$ represent the elements of two 3D-arrays of distances where $i = 1, \dots, n_1$, $j = 1, \dots, n_2$, and $r, s = 1, \dots, k$. The 3D-arrays contain all distances of A_1 and A_2 in a form of arranged distance matrices. We design k^2 partial tests based on the distance between points r and s as follows

$$H_0(rs) : \mu(\{(D_{X_i})_{rs}\}_{i=1}^{n_1}) = \mu(\{(D_{Y_j})_{rs}\}_{j=1}^{n_2}) \text{ vs. } H_1(rs) : \mu(\{(D_{X_i})_{rs}\}_{i=1}^{n_1}) \neq \mu(\{(D_{Y_j})_{rs}\}_{j=1}^{n_2}).$$

The calculation of the mean depend on the distance measure, and if we consider Euclidean distance the geometric mean might be preferred.

If we assume the distances are normally distributed then we can apply t-test and assign p_{rs} as the p -value associated with the distance between points r and s . (N.B. in the absence of normality assumption, we can apply a permutation test). By applying the same hypothesis test for all pairs of points, we form a square p -value matrix $(\Pi)_{rs}$, $r, s = 1, \dots, k$. Each row of Π corresponds to the distances of a landmark to all the other landmarks. By having the significant distances at

α level, we count the number of significant elements of Π in each row, and associate the number as a rank to the corresponding point. By this method, we rank all the points of the model in a way that the points with higher rank have more significant connections with other points among the population. Note that by considering Euclidean distance, we have $d(\mathbf{x}_i, \mathbf{x}_j) = d(\mathbf{x}_j, \mathbf{x}_i)$, but this reciprocal relation between \mathbf{x}_i and \mathbf{x}_j is not necessarily true in a case of using force or other types of distances. For example, a point may impose greater forces to the other one depending on the force definition so $F(\mathbf{x}_i, \mathbf{x}_j) \neq F(\mathbf{x}_j, \mathbf{x}_i)$. Also, in the directed graph, the distance definition between two points is based on the graph directions and $d(\mathbf{x}_i, \mathbf{x}_j) \neq d(\mathbf{x}_j, \mathbf{x}_i)$ (e.g., the distance that one travels from the home to a shopping center is not the same as shopping center to home depending on the road he/she chooses).

In general Π has diagonal elements equal to 1 and contains $k^2 - k$ p -values. By considering Euclidean distance, Π becomes symmetric with $\frac{k^2 - k}{2}$ p -values. Also, it is possible to adjust the obtained p -values as described in Section 4.2.1 and based on the FDR rate count the significant adjusted p -values. To elaborate more, the Euclidean distance Π is symmetric and can be shown as $\Pi = U^T + I + U$, where U is an upper triangular matrix with diagonal elements equal to zero, and I is the identity matrix. By ignoring the zero elements we can adjust the remaining part of U (i.e., unique p -values). The adjusted p -value matrix can be shown as $\Pi_{adj} = U_{adj}^T + I + U_{adj}$, where U_{adj} is U after the adjustment. Then we count the number of adjusted p -values smaller than the FDR rate in each row. Figure 4.2 illustrates the workflow.

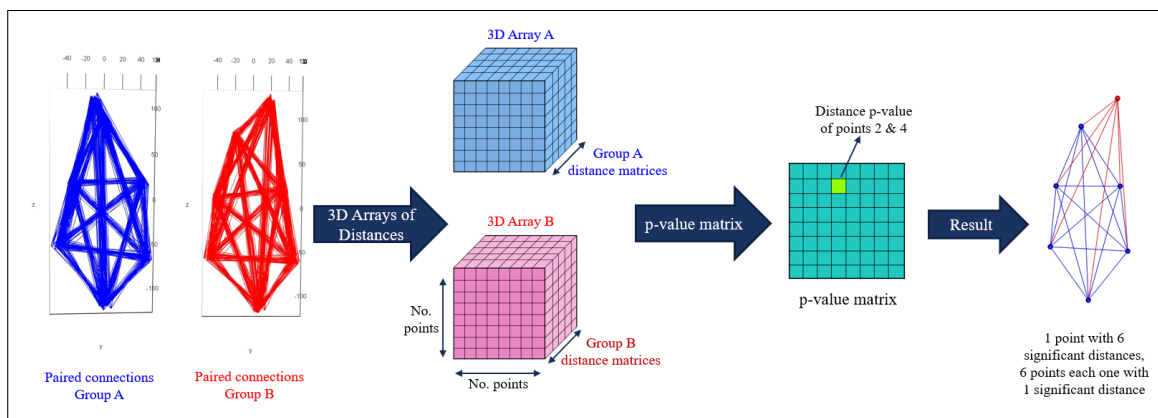


Figure 4.2: Workflow of the hypothesis test independent from alignment.

Two groups of shapes Group A and Group B with 7 landmarks are identical except in one landmark. Test shows one point with 6 significant distances and 6 point each one with one significant distance.

(The illustrated groups of shapes are modified version of female gorilla skulls taken from (Dryden, 2018)).

KDE on p -values

Counting the number of significant distances is not suitable, specifically in shape analysis. The reason is after removing scale, we may need to reduce the level of significance extremely to capture any differences. Another possible option is to rank the landmarks based on the smallest

p -value obtained from the combined test (e.g., Fisher combination test from Section 4.6.3). Again this method is not appropriate since a point with only one extremely significant p -value could be ranked before the points with multiple small p -values but without an extreme one. In this regard, we introduce a method based on kernel density estimation (KDE).

In feature space \mathbb{R}^m we express the KDE of a sample $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^m$ as

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^m} \sum_{i=1}^n \mathcal{K}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right),$$

where $\mathcal{K} : \mathbb{R}^m \rightarrow \mathbb{R}$ is the kernel function and h is the smoothing parameter known as the bandwidth (Terrell and Scott, 1992). By choosing \mathcal{K} as a Gaussian probability distribution function (pdf) then we have the KDE method known as Parzen approximation as

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^m} \sum_{i=1}^n \frac{1}{(2\pi)^{m/2}} \exp\left(-\frac{(\mathbf{x} - \mathbf{x}_i)^T(\mathbf{x} - \mathbf{x}_i)}{2h^2}\right). \quad (4.12)$$

(Theodoridis, 2015, ch.3).

Note that in (4.12), $0 < \exp\left(-\frac{(\mathbf{x} - \mathbf{x}_i)^T(\mathbf{x} - \mathbf{x}_i)}{2h^2}\right) \leq 1$ thus, the maximum possible value of $\hat{f}(\mathbf{x})$ is $1/(h^m(2\pi)^{m/2})$.

Let $\mathbf{p}_{(i)} = \{p_{i1}, \dots, p_{ik}\}$, where p_{i1}, \dots, p_{ik} are the i th row elements of the p -value matrix Π , $i = 1, \dots, k$. Assume $\mathbf{p}_{(i)} \cup -\mathbf{p}_{(i)}$ (mirrored p -values) as a set of one-dimensional data points in $[-1, 1]$. We calculate $\hat{f}_i(0)$ by (4.12), where $\hat{f}_i(0)$ is the KDE of $\mathbf{p}_{(i)} \cup -\mathbf{p}_{(i)}$ at zero, and the maximum possible value of $\hat{f}_i(0)$ is $1/(h\sqrt{2\pi})$. Therefore, we have a set of estimated KDEs $\hat{f}_1(0), \dots, \hat{f}_k(0)$, which based on their magnitudes we can rank the landmarks. For this purpose, let $\kappa = 1/(h\sqrt{2\pi}) + \varepsilon$ where $\varepsilon \ll 1$ and divide the interval $[0, \kappa]$ into $b \in \mathbb{N}$ sub intervals $[0, \frac{\kappa}{b}), \dots, [\frac{(b-1)\kappa}{b}, \kappa]$. Then, the rank of the landmark l_i is j if $\hat{f}_i(0) \in [\frac{(j-1)\kappa}{b}, \frac{j\kappa}{b})$, $j \in \{1, \dots, b\}$.

This test could be applied with or without removing scale. To remove scale, we can use GPA to minimize the scale among the population. Alternatively, to be independent of the Procrustes analysis we can scale shapes by their centroid size (see Equation (2.10)).

Note that it is possible to apply this method for the adjusted p -values but, since the nature of this method is based on ranking the GOPs rather than classifying them as significant and non-significant, we prefer to use raw p -values.

For the visualization, we can define a color map based on the rank of the landmarks (see Figure 4.4).

Pros and cons Hypothesis test with distance is a powerful approach for shapes as it includes all possible distances between points. But, it is a weak test when the two groups are symmetric

relative to each other. For example, beaks of pied avocet, and white-faced ibis (two types of birds) are relatively symmetric as one turned upward and the other downward (see Figure 4.3). Possibly, by removing scale, the hypothesis test with distance is incapable of detecting differences even though they are apparent. To overcome this issue, we can add more landmarks to the object to violate the symmetry. In bird example, it can be done by adding a landmark to the bird's neck.

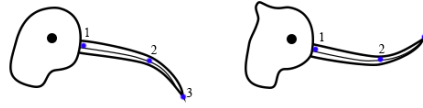


Figure 4.3: Symmetric objects.

Illustration of 3-landmark models of a white-faced ibis at left and a pied avocet at right.

Another intrinsic shape property is the angles between the landmarks which is invariant under the act of rotation, translation, and scale. In this regard, we propose a hypothesis testing method to find the significant landmarks based on the analysis of the angles.

4.7.2 Hypothesis test with angle for planer shapes

For 2D-shapes by using angle instead of distance, a similar procedure as described in Section 4.7.1 can be applied. Obviously, a planer shape with k landmarks consists of $\binom{k}{3}$ triangles and consequently $s = 3\binom{k}{3}$ counterclockwise angles $\theta_i \in [0, 2\pi]$, $i = 1, \dots, s$ (N.B., $\angle abc \neq \angle cba$). Thus, we can assign a vector of angles $\boldsymbol{\theta} = (\theta_1, \dots, \theta_s)^T$ to a k -landmark model. Assume two groups of k -landmark models $A_1 = \{X_1, \dots, X_{n_1}\}$, $A_2 = \{Y_1, \dots, Y_{n_2}\}$. We can arrange the angles of the groups in two matrices as follows:

$$\Theta_{A_1} = \begin{bmatrix} (\theta_{X_1})_1 & \dots & (\theta_{X_1})_s \\ \vdots & \ddots & \vdots \\ (\theta_{X_{n_1}})_1 & \dots & (\theta_{X_{n_1}})_s \end{bmatrix}, \quad \Theta_{A_2} = \begin{bmatrix} (\theta_{Y_1})_1 & \dots & (\theta_{Y_1})_s \\ \vdots & \ddots & \vdots \\ (\theta_{Y_{n_2}})_1 & \dots & (\theta_{Y_{n_2}})_s \end{bmatrix},$$

where $(\theta_{X_i})_j$ and $(\theta_{Y_{i'}})_j$ represent the j th angle of the configuration matrices X_i and $Y_{i'}$ respectively, $i = 1, \dots, n_1$, $i' = 1, \dots, n_2$, and $j = 1, \dots, s$.

We design s partial hypothesis tests based on the corresponding columns of Θ_{A_1} and Θ_{A_2} such that,

$$H_0(r) : \mu\left(\{(\theta_{X_i})_r\}_{i=1}^{n_1}\right) = \mu\left(\{(\theta_{Y_{i'}})_r\}_{i'=1}^{n_2}\right) \quad \text{vs.} \quad H_1(r) : \mu\left(\{(\theta_{X_i})_r\}_{i=1}^{n_1}\right) \neq \mu\left(\{(\theta_{Y_{i'}})_r\}_{i'=1}^{n_2}\right),$$

where $r = 1, \dots, s$.

Note that the angles could be acute or obtuse, so to test the mean difference, we need to treat angles as directional data. Similar to the parametric test for s-rep directions in Section 4.6.2, we project data to the tangent line (one-dimensional space) of the unit circle with the pooled Fréchet mean. Then we apply t-test with normality assumption on the log-mapped data to obtain p -values. Thus, we have $\mathbf{p} = (p_1, \dots, p_s)^T$ as a vector of all p -values. Each landmark associates with $\binom{k-1}{2}$ angles so we can arrange the elements of \mathbf{p} in a p -value matrix Π with k rows and $\binom{k-1}{2}$ columns where the rows are associated with the landmarks. Again we rank the landmarks as we described in Section 4.7.1.

Pros and cons Similarity transformations (i.e., scale, rotation, and translation) do not affect the angles, so the method is suitable for shape analysis rather than size-and-shape. A drawback is by increasing the number of landmarks, the number of angles, and consequently, the cost of computation increase dramatically.

Applying this method for SPHARM-PDM and s-rep skeletal needs generalization for 3D shapes, and we avoid using it for the thesis data analysis.

4.7.3 Example

To detect locational differences, we apply the introduced tests with distance and angle on the chimpanzee skull data (i.e., panf and panm) from (Dryden, 2018) (discussed before in Section 2.5.5). Then, we compare the result with the conventional method of alignment, as discussed in Section 4.6. For the hypothesis test, we consider shapes with and without removing scale. In Section 2.5.5, we mentioned that by removing scale, there are no significant differences between the groups' means. Furthermore, we can check the means are significantly different in size-and-shape analysis. Although we expect to see no local differences in shapes after removing the scale, the Hotelling's T^2 test (after controlling false positive by FDR) shows four significant landmarks that this might be an unfavorable effect of alignment (see Figure 4.4(v)). KDE plots in Figure 4.5 for the distance test clearly explain the situation. The plot shows that before removing the scale, for most of the landmarks, we have a concentration of p -values near zero but, by removing the scale, the distances become similar, and consequently, the height of the KDEs decrease. This fact is reflected in Figure 4.4, where we use a color map to rank landmarks based on the KDE magnitude at zero such that brighter colors indicate lower ranks. In the test with distance, after removing the scale, the color of landmarks becomes brighter, which means they become more similar. Besides, landmarks' colors of test with angles are not so much dark, which indicates similarity in angles.

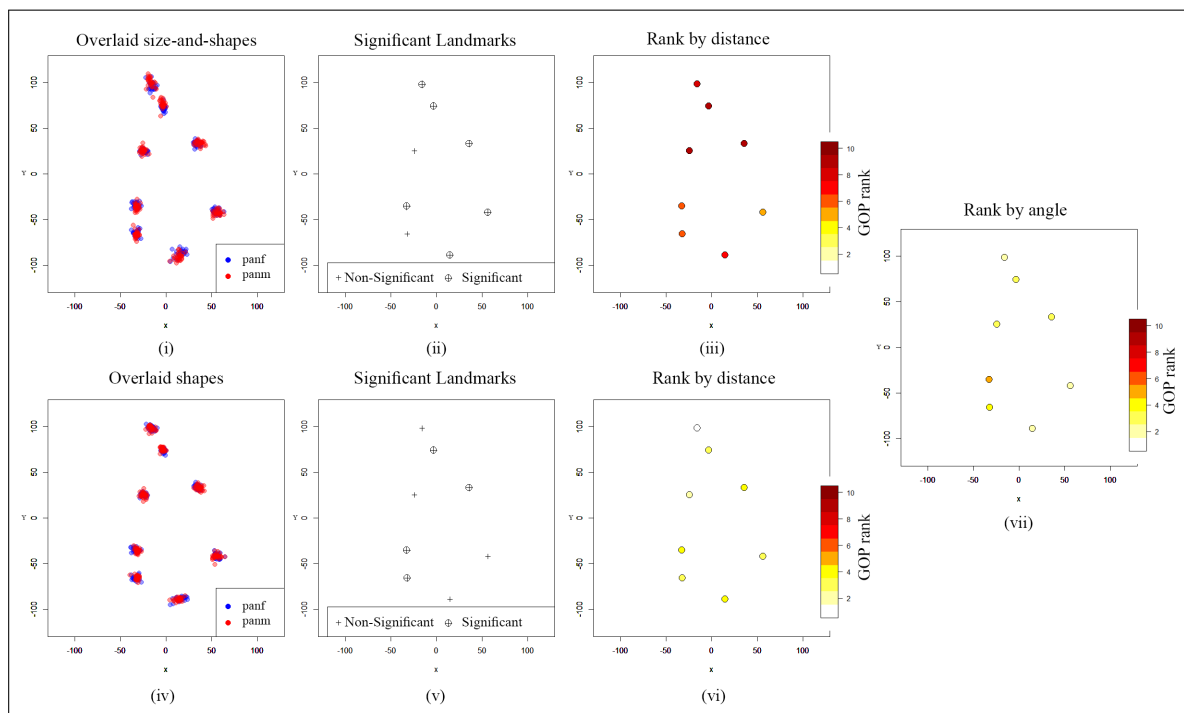


Figure 4.4: Test with alignment vs. Test with distance and angle.

(i) Overlaid size-and-shapes. (ii) Circles indicate significant landmarks of size-and-shape analysis. (iii) Result based on ranking distances of size-and-shapes. (iv) Overlaid shapes. (v) Circles indicate significant landmarks of shape analysis. (vi) Result based on ranking distances of shapes. (vii) Landmarks are ranked by color based on test with angle. In (iii), (vi), and (vii) brighter and darker colors respectively indicate lower and higher KDE value at zero. (Data are taken from (Dryden, 2018))

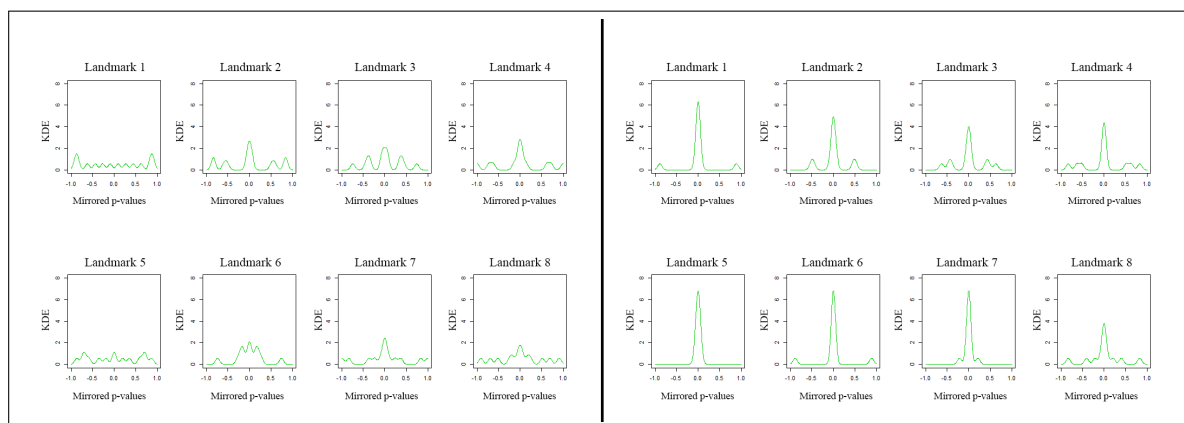


Figure 4.5: KDE on mirrored p -values.

KDE on mirrored p -values of test with distance for chimpanzee skulls with eight landmarks. On the left after removing the scale and on the right without removing the scale.

Chapter 5

Data Analysis

In this chapter, we analyze data from the ParkWest study (<http://www.parkvest.no>) provided by The Norwegian Centre for Movement Disorders at the Stavanger University Hospital (<https://helse-stavanger.no>). For the analysis, we apply different methods of hypothesis testing, including permutation test, parametric test with normality assumption, and test by distance matrices from Chapter 4. We compare the results for different alignment methods, as discussed in Chapter 2. Also, we analyze the performance of CPNG, CPNS, and PNG, from Chapter 3. In addition, we implement DWD and SVM to classify s-reps based on the CPNG data as we explained in Section 3.5.2.

5.1 Data

ParkWest provided overall 667 MR images with the segmentation of the healthy controls and patients with early PD. MR images were taken at baseline and after three and five years. The baseline contains 108 images for CG and 182 for PD. Figure 5.1 shows one sample of the baseline data. In previous studies on the same set of data, (Apostolova et al., 2012) applied radial distance analysis and showed the hippocampal atrophy in PD, and (Beyer et al., 2013) investigated the association between memory performance and hippocampal radial distance reduction. Although, radial distance approach is more accurate than volumetric analysis, there are some disadvantages, including the method of alignment and the centroid movement. In radial distance analysis, we first align and scale shapes based on the brain size and orientation, which is not suitable when our concentration is on a specific part of the brain. Then, radial distances are measured slice-by-slice, i.e., for each slice, the distance between the boundary points to the centroid of the slice is measured. The problem is, the centroid is not a good reference as its position is highly dependent on the shape formation. Therefore we may report inflation (increase in radial distances) as atrophy (decrease in radial distances) and vice versa. Also, parallel slicing of curvy

shapes such as the hippocampus may mislead us to associate irrelevant object parts. In this work, we study the structural contrast of the left hippocampus between early PD and normal control with s-rep, SPHARM-PDM, and middle surface PDM. The same method of hypothesis testing applies to other brain parts, which can be the subject of further study, see Section 6.2.1.

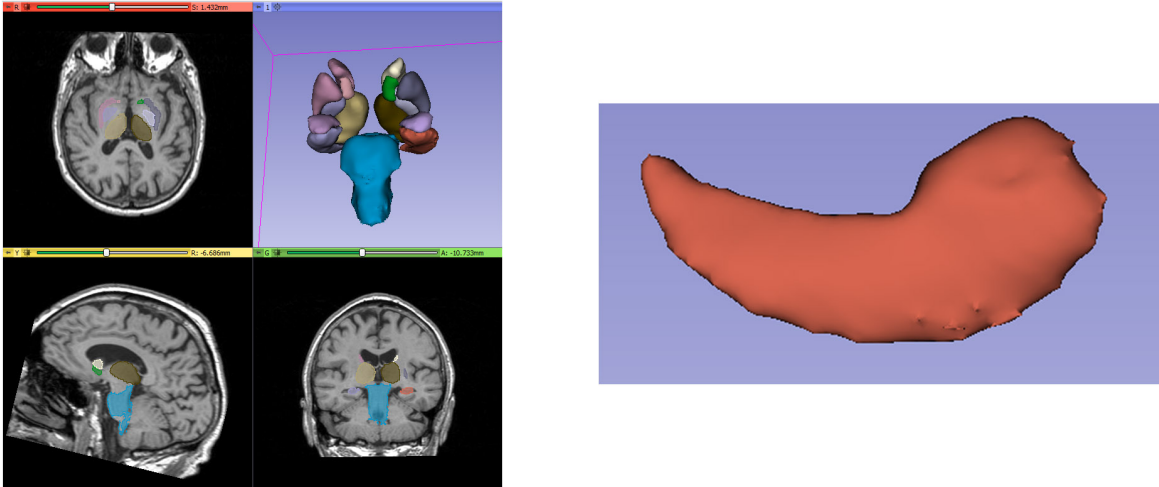


Figure 5.1: ParkWest data.

Left: Views of an MR image with the segmentations including Left and Right (LR) thalamus-proper, LR caudate, LR putamen, LR pallidum, LR hippocampus, LR amygdala, LR accumbens-area and brain-Stem. Right: Left hippocampus

5.2 Software and model fitting

The used software in this work for the model fitting is a part of SlicerSalt (<http://salt.slicer.org>) and 3DSlicer (<https://www.slicer.org>). As the object meshes were not smooth enough for the model fitting, we used Meshmixer (<http://www.meshmixer.com>) to repair holes and sharp protrusions (see Figure 5.2). We repaired meshes one by one manually, which needs time and patience. Then, we used 3DSlicer shape analysis extension to generate SPHARM-PDMs and suitable triangle meshes for the s-rep model fitting.

The s-rep model fitting consists of two steps: 1. Initialization, 2. Refinement. In this work, we kept the default parameters for the initialization and chose the skeletal positions as a grid of 5×9 . As a result, we obtained discrete s-reps with 114 spokes (i.e., 45 up spokes, 45 down spokes, 24 crest spokes). The model fitting procedure is based on MCF (see Section 2.1.3), so the positions of the up and down spokes coincide. Therefore, we have 69 points as the skeletal positions (i.e., 45 up and down spokes, 24 crest spokes). For the refinement, we used default parameters except for Image-Match-Weight=5, which cause spokes' tips to become closer to the boundary (see Figure 2.2).

In this study, we used R and Python programming languages. For s-reps, SlicerSalt provided *.vtp files associated with Up, Down, and Crest spokes. We used Python to extract the information

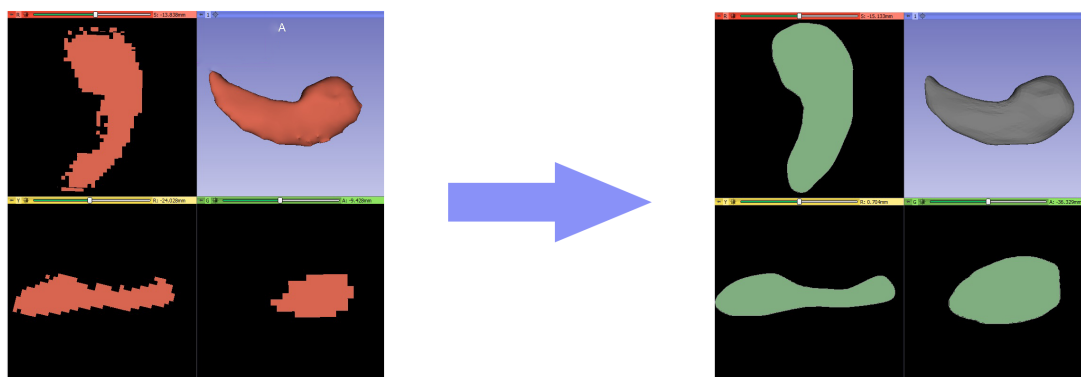


Figure 5.2: SPHARM-PDM smooth triangle mesh.

Left: Jagged mesh of a left hippocampus with holes and sharp protrusions. Right: Smooth SPHARM-PDM triangle mesh.

from these files. The output is a *.csv file containing all the information of the fitted s-reps. By importing the *.csv file as a data frame into the R studio environment, we could analyze the s-rep data. We did a similar procedure for the SPHARM-PDMs. Two R packages we frequently used in this work are "shapes" (Dryden, 2018) (for the analysis), and "rgl" (for the visualization).

Medical data of this work are confidential. However, some codes we used for the analysis accompanied by some examples are available at

(<https://www.dropbox.com/sh/bs18vmg44beb6bb/AADSnPq1mfEIBIOf8jrcuQJAa?dl=0>).

5.3 Analysis

In some sense, the centroid size of SPHARM-PDM and s-rep boundary points (spokes' tips), reflect the volume of the object. Thus, let us test the mean difference of the centroid sizes between CG and PD before we dive deep into the data analysis. t-test indicates no significant differences between the means of the two groups (see Table 5.1). Furthermore, s-rep skeletal alignment by GPA provides us with scale factors (i.e., GPA scale factors relevant to the mean shape of the pooled group). The p -value of the t-test on s-rep skeletal scale factor is 0.15, which supports the test on the centroid size.

Note that by increasing the number of points, the centroid size increases. Therefore, the centroid sizes of SPHARM-PDMs are greater than the centroid sizes of the s-reps spokes' tips.

5.3.1 Global tests

The global test for the parametric approach on SPHARM-PDM, Middle Surface PDM, and s-rep even with Bonferroni as a conservative method indicates a significant difference between PD

T-test on centroid size					
Model	Mean CG	Mean PD	SD CG	SD PD	p-value
SPHARM-PDM	437.52	440.16	24.30	23.83	0.37
s-rep boundary	146.46	147.36	8.09	7.94	0.35

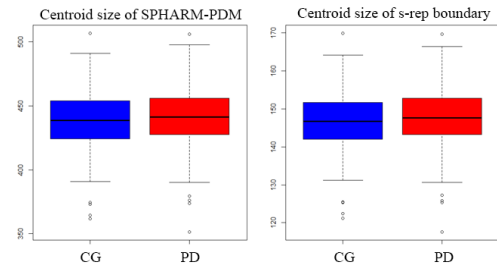


Table 5.1: T-test on centroid size.

and CG. The reason is the existence of at least one minimal p -value. In the non-parametric approach, tests with Fisher, Liptak, Logistic Liptak, and Tippet as combining functions indicate a significant difference but not with the Mahalanobis quadratic form.

Global test for parametric approach for shapes				
	Bonferroni ($\alpha = 0.05$)		Fisher combination	Fisher combination on adjusted p -values
	Min p -value	α/n		
SPHARM-PDM ($n = 1002$)	3.47e-05	4.99e-05	0	<0.001
Middle Surface PDM ($n = 1002$)	2.82e-05	4.99e-05	0	<0.001
s-rep ($n = 297$)	5.89e-05	1.68e-4	<0.001	<0.001

s-rep global test for non-parametric approach					
Combining Function	Fisher	Liptak	Logistic Liptak	Tippet	Mahalanobis quadratic
p-value	<0.001	<0.001	<0.001	0	0.4532

Table 5.2: Global tests.

5.3.2 s-rep analysis

In Section 4.4, we discussed different methods to calculate mean s-rep. Figure 5.3 compares CPNG and CPNS performance with CG data and shows there is no advantage in using PNS or PNG instead of PGA, to calculate mean s-rep. In fact, since the data is concentrated, PNS, PNG, PGA, and even PCA capture the same amount of data. Besides, the good performance of PCA in comparison to PNS and PNG underline the fact that the spherical data (i.e., skeletal positions on hypersphere and spokes' directions on S^2) are very concentrated and possibly normally distributed. The two identical plots (denoted by a double headed arrow) in Figure 5.3 show that the sequential test within PNS always chooses great spheres for the fitting, so the PNS on the skeletal PDMs is identical to PNG.

According to the fact that PNS and PNG are computationally expensive and for the current data their performance is not better than PGA, it is logical to calculate mean s-rep by using

GPA and Fréchet means (i.e., GPA mean for skeletal positions + Fréchet mean for directions + Geometric mean for spokes length) in the permutation test.

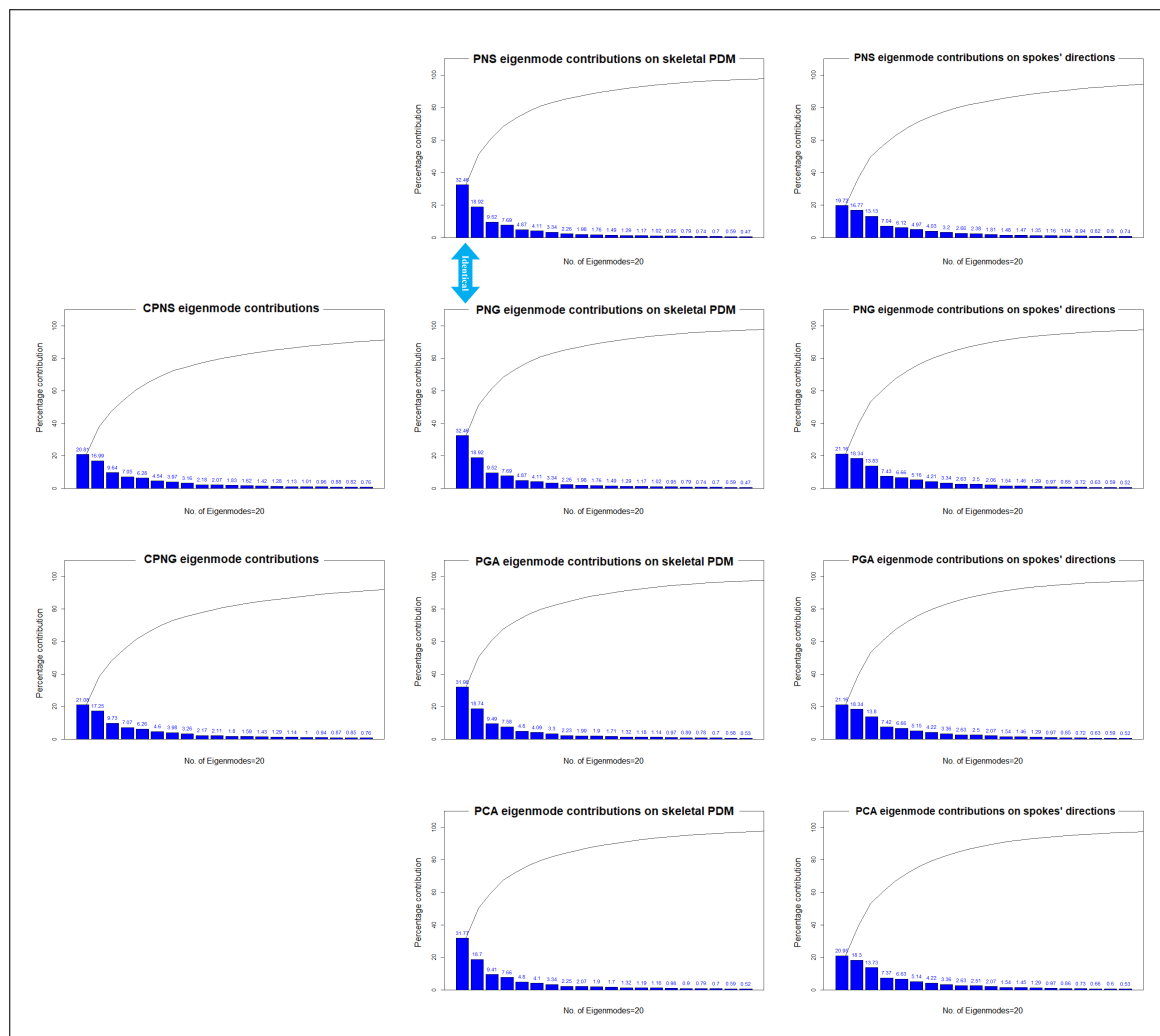


Figure 5.3: Eigenmode contributions of CG s-reps.

Left column corresponds to the eigenmode contributions of CPNS and CPNG. Middle and right columns are associated with Eigenmode contributions of PNS, PNG, PGA, PCA on skeletal positions and spokes' directions. Double-headed arrow indicates identical plots.

Hypothesis tests with alignment

As we mentioned, there is no significant difference between the s-reps' scale factors, and the results of s-rep shape analysis and size-and-shape analysis are similar. To prevent repetition, we only consider shape analysis. In Section 2.5.3 we discussed s-rep alignment by GPA. Here we remove scale and align the s-reps based on spokes' tips and tails. Figure 5.4 illustrate the result of parametric and non-parametric approaches where the significant GOPs are depicted in red. To control false positives we use Benjamini-Hochberg method (Section 4.2.1, FDR=0.05) and Schulz correction method (Section 4.5.1) for the non-parametric approach, and for the parametric approach we use Benjamini-Hochberg method (FDR=0.05).

The results of parametric and non-parametric approaches are comparable. Still, since we use different metrics for the tests (e.g., geodesic distance in non-parametric approach and Mahalanobis distance in parametric approach for the spokes' directions) nonconformity of the results is predictable.

Although we observe plenty of significant GOPs, the similarity of mean s-reps (see Figure 5.5) brings up the discussion about the reason. Seemingly, we have concentrated data both in positions and directions. Concentrated data have small variances, usually, in multivariate analysis, minor variations increase the test statistic and consequently decrease the p -value, so we may encounter significant p -values even if the means are very close (see Equations (4.9) and (4.10)). In this regard, considering multivariate confidence interval for future study could be helpful (definition of the confidence interval for multivariate data is not a straightforward task, some attempts can be found in (Šidák, 1967; Korpela et al., 2017)).

Effect of weighted alignment

The Procrustes distance between mean skeletal of PD and CG is 0.013, which indicates that the skeletal mean shapes are very close, and weighted alignment (see Section 2.5.6) might be unnecessary. But to study the effect of weighted alignment (see Section 2.5.5), we assume a small threshold $\delta = 0.006$ and extract unsuspecting points with the elimination algorithm 2.2.

In Figure 5.6(b) blue indicates unsuspecting points. Weighted alignment causes more stability and less variation in unsuspecting points, which is evident in the left side of Figure 5.6(d). Figure 5.7 compares the result of the Hotelling T^2 test on the Procrustes and weighted Procrustes alignment. Note that in this part, we align skeletal positions without considering spokes' tips, so the result of the Procrustes alignment is slightly different from the result in Figure 5.4.

Test by distance matrices

We apply Euclidean distance matrices for the skeletal positions and geodesic distance matrices for the spokes' directions. To compare the result with the test results from the SPHARM-PDM, we consider skeletal positions with and without removing scale (see Figure 5.8). By removing scale, the differences become more obvious. Test for the spokes' lengths in Section 5.3.2 was independent of the alignment, so we exclude spokes' lengths in this part.

Note that scaling does not affect spokes' directions.

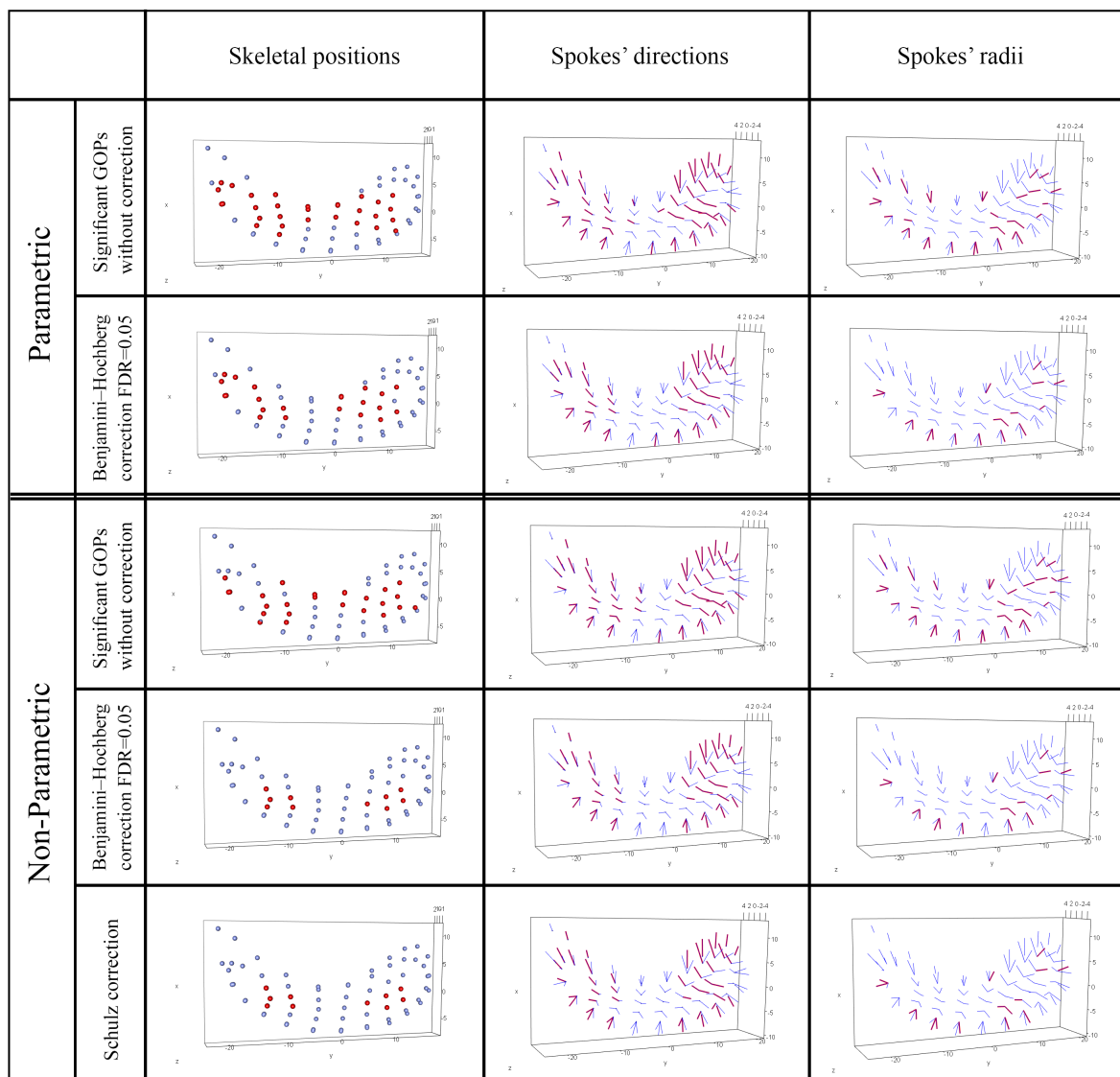


Figure 5.4: s-rep results.

Red color indicates significant GOPs. FDR=0.05 for Benjamini-Hochberg correction. Method of alignment is GPA on spokes' tips and tails

5.3.3 SPHARM-PDM analysis

Hypothesis test with alignment

Each SPHARM-PDM of the left hippocampus includes 1002 points. Here we implement the parametric hypothesis test (with the multivariate normality assumption) and analyze SPHARM-PDMs' shapes and size-and-shapes based on GPA alignment. Bear in mind that, without correction, we have a considerable amount of significant points on the bottom side of the hippocampi. To highlight the more significant area, we apply the Benjamini-Hochberg correction as described in Section 4.2.1 and decrease FDR to 0.01 (see Figure 5.9). Seemingly, by removing scale, the differences become more observable. Figure 5.10 reflects the fact that sorted p -values

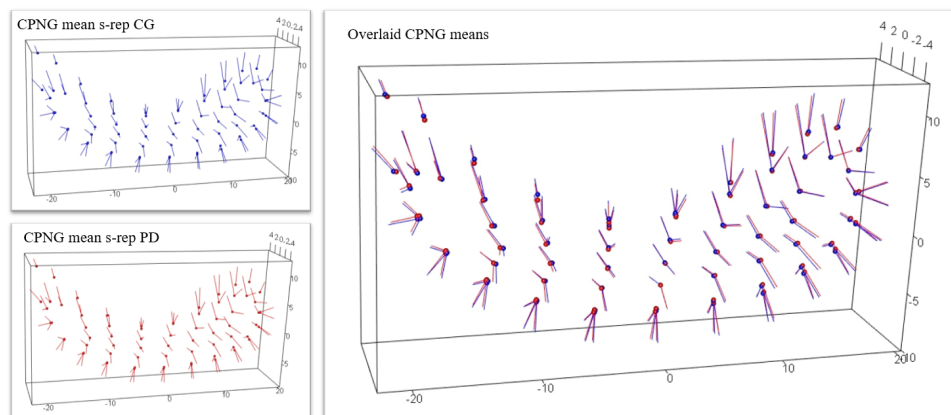


Figure 5.5: s-rep CPNG means.

Left: Mean s-rep of CG and PD. Right: Overlaid mean s-reps. Red and blue associate with PD and CG respectively.

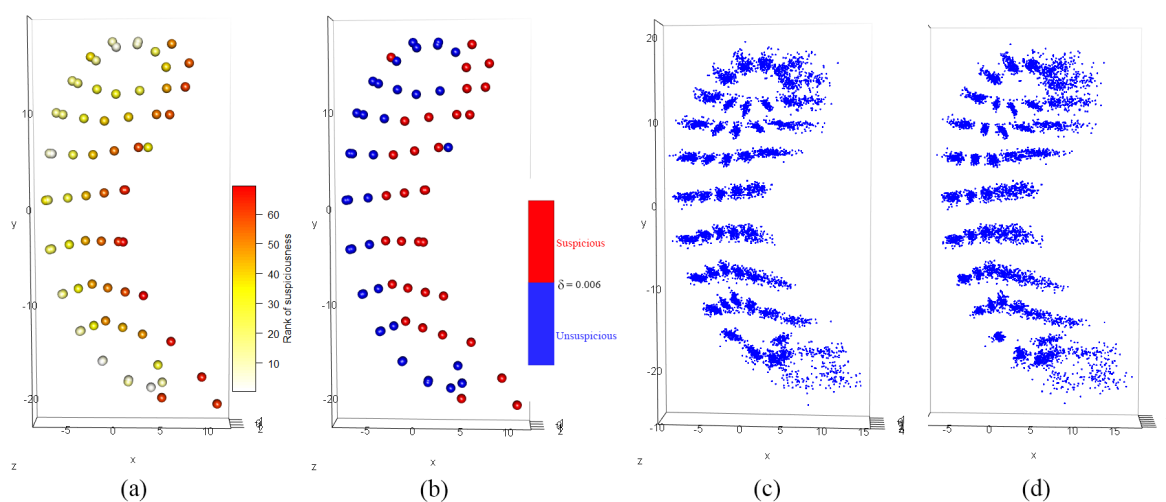


Figure 5.6: Effect of weighted alignment on s-rep skeletal.

(a) Ranked skeletal positions based on their level of suspiciousness. (b) Classified suspicious and unsuspecting points with threshold $\delta = 0.006$. (c) Aligned CG by GPA. (d) Aligned CG by weighted GPA where data variation in unsuspecting points (left side) is less than suspicious points (right side).

and consequently adjusted p -values from the size-and-shape analysis have relatively higher values compared with shape analysis.

Generalized Shapiro-Wilk test (Villasenor Alva and Estrada, 2009) for multivariate normality on aligned SPHARM-PDMs indicates about 20% and 45% of the points in CG, and PD respectively do not have multivariate normal distributions, so implementing non-parametric hypothesis test as described in (Styner et al., 2006) might be necessary. However, in practice, the outcome of permutation is similar to the parametric approach. The result of the permutation test on SPHARM-PDMs can be found in Appendix A.1.1.

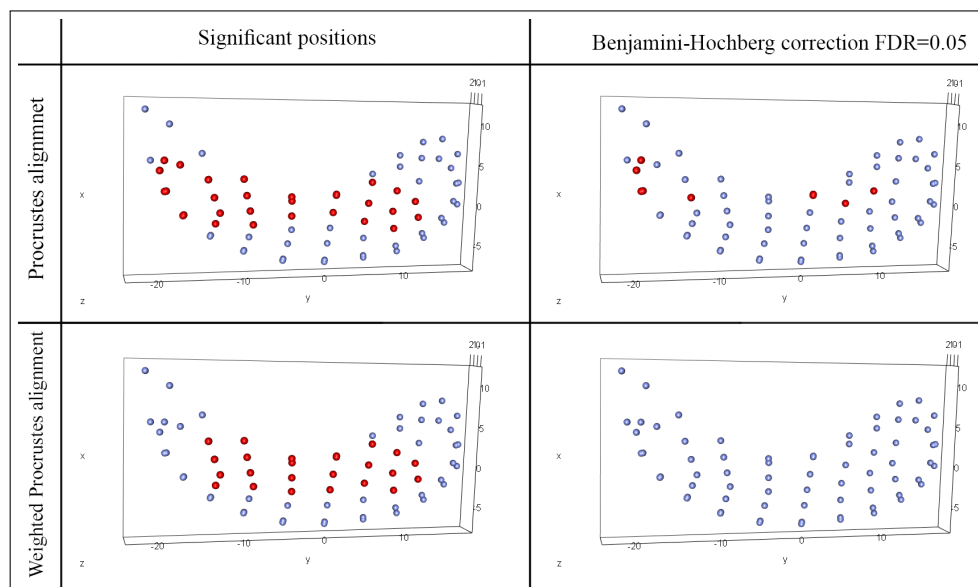


Figure 5.7: Significant positions by GPA alignment vs. weighted Procrustes alignment. Red indicate significant points.

Test by distance matrices

The outcome of the hypothesis test by distance matrices from Section 4.7.1 is similar to the test with alignment. Again, after removing scale, groups dissimilarity turn out to be more explicit. In Figure 5.11, darker colors indicate high-rank points. Obviously, there is a close relationship between the result of SPHARM-PDM and skeletal positions since both of them point to the same hippocampal area as a significant region (see Figures 5.8 and 5.11).

5.3.4 Further analysis

For further analysis and results, including the non-parametric approach for SPHARM-PDM, analysis of the boundary normal vectors, analysis of Middle Surface PDM, and analysis of SPHARM-PDM plus skeletal PDM, see Appendix A.1.

5.3.5 s-rep classification

For the s-rep classification based on CPNG, the dimension of feature space is higher than the sample size (for s-rep with 114 spokes the dimension of the feature space is 549 i.e., $69 \times 3 - 1$ and 114×2 for PNG residuals of positions and directions respectively, 114 for lengths and 1 for scale factor). Therefore, we encounter an HDLSS problem. For the classification, we use DWD and SVM, as discussed in Section 3.5.2. As we expected, we have data piling problem in SVM, thus DWD might be a better choice (see Figure 5.12(b) and 5.12(c)). By using half of the samples as training data, the result of classification is not promising (see Table 5.3).

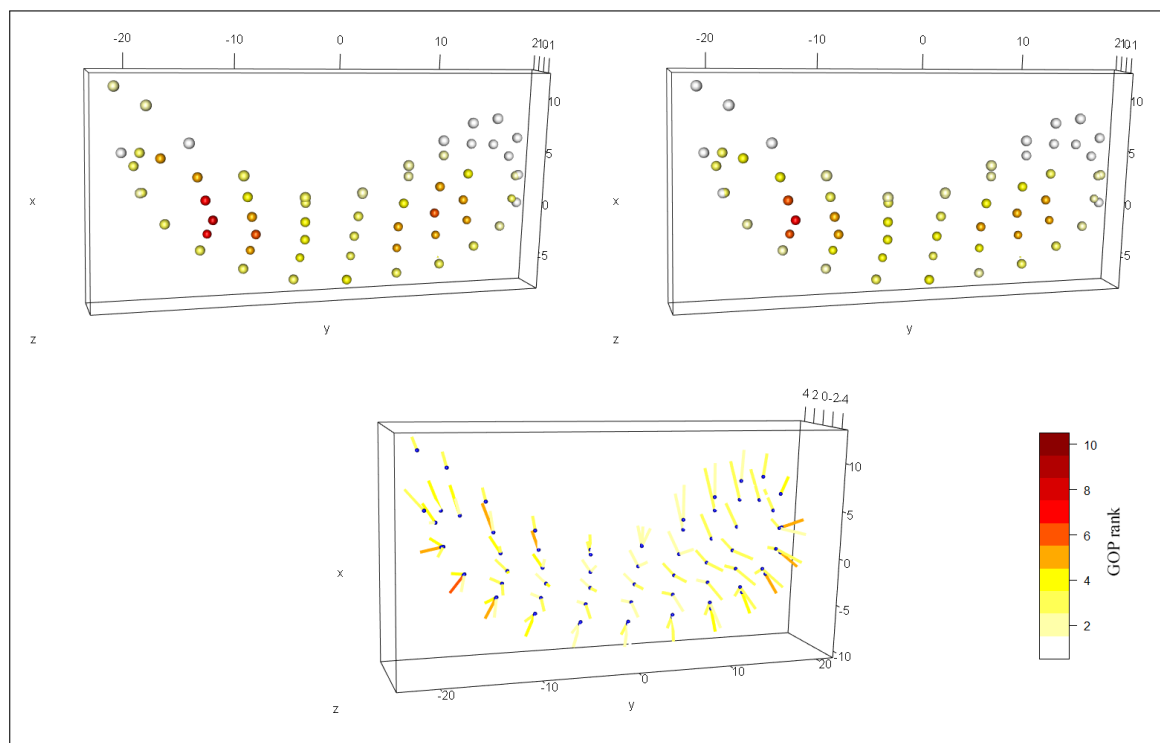


Figure 5.8: s-rep hypothesis test independent from alignment.

Color map ranks GOPs from 1 to 10 based on the KDEs' magnitude. Top left: Ranked skeletal positions after removing scale. Top right: Ranked skeletal positions without removing scale. Bottom: Ranked spokes' directions.

Note that the CPNG data subordinate the alignment. With proper alignment, the data may become separated in one or more features, which can improve the classification. To elaborate more, in the cat and dog example (see Figure 2.6), by the weighted alignment we can easily distinguish a muzzle landmark belongs to a cat or a dog, leading to better classification. But by GPA alignment, the landmarks become very close, and it would be challenging to classify dogs and cats (based on the muzzle or any other landmarks). In fact, GPA hides the groups' differences. We have the same situation in CPNG, where GPA alignment makes the spokes' information of different groups similar.

Average performance on test data by 100 random sampling (training set $\frac{1}{2}$ samples)				
Classifier	Accuracy	Sensitivity	Specificity	Kappa
DWD	0.59	0.84	0.20	0.05
SVM	0.54	0.61	0.41	0.03

Table 5.3: DWD vs. SVM.

5.3.6 PD with dementia

So far, we detected GOPs' mean differences between PD and CG, but the dissimilarity may come from other factors such as dementia rather than PD. At the baseline, we have no PD

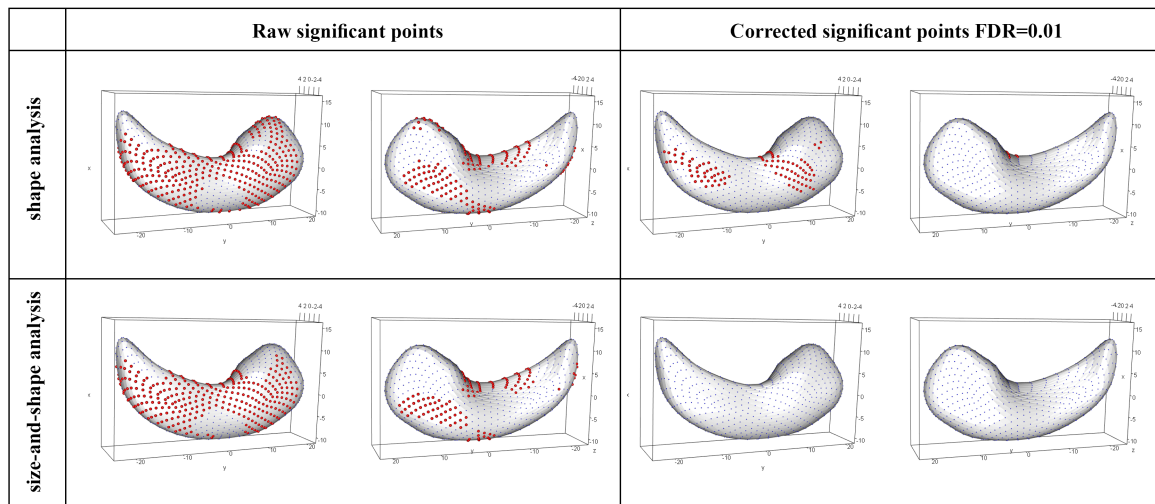


Figure 5.9: Significance plot of SPHARM-PDMs by Procrustes alignment.

Visualization of Hotelling's T^2 test on SPHARM-PDMs of the left hippocampi from two angles. The left column indicates significant points ($\alpha = 0.05$) in red without correction. The right column indicates significant points after correction (FDR=0.01). Top and bottom rows are associated with shape analysis and size-and-shape analysis, respectively.

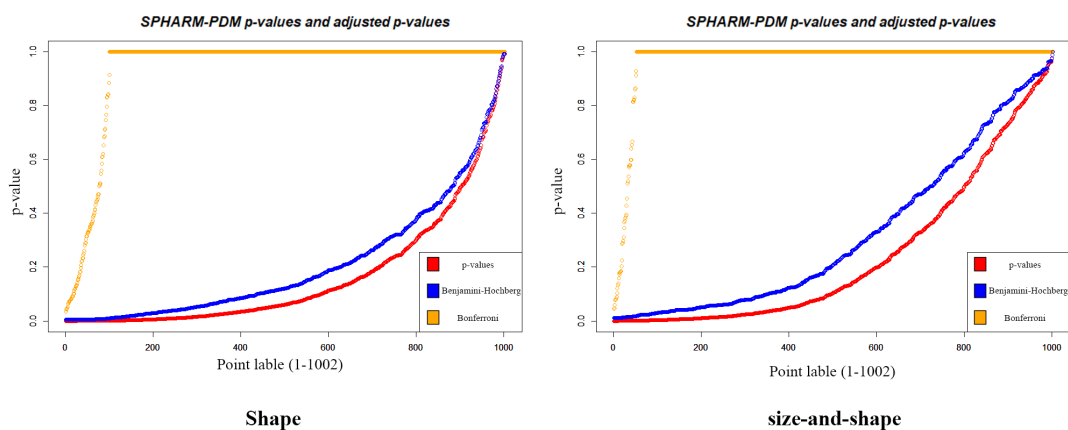


Figure 5.10: Ordered p -values and adjusted p -values of SPHARM-PDM.

p -values in red, adjusted p -values by the Benjamini-Hochberg and Bonferroni methods in blue and orange respectively.

with dementia (PDD), but the longitudinal study shows 11 and 44 cases expanded dementia before and after year 3, respectively. The t -test on SPHARM-PDM centroid size of CG and PDD before year 3 shows a significant difference (p -value=0.008 and 95% CI=(4.55, 26.74)). Thus, it is possible that the deformation between PD and CG caused by dementia. On the other hand, PD without dementia has significantly larger centroid size than CG (p -value=0.01 and 95% CI=(1.91, 15.05)). Generally, as we mentioned in Sections 1.1, the size of the hippocampus can be affected by different parameters (e.g., aging) and this needs to be checked in further studies. The overlaid mean SPHARM-PDM of CG and PDD before year 3 in Figure 5.13(a) shows the difference in their size-and-shape means. Box plots in Figure 5.13(b) illustrate the centroid size distributions in different groups.

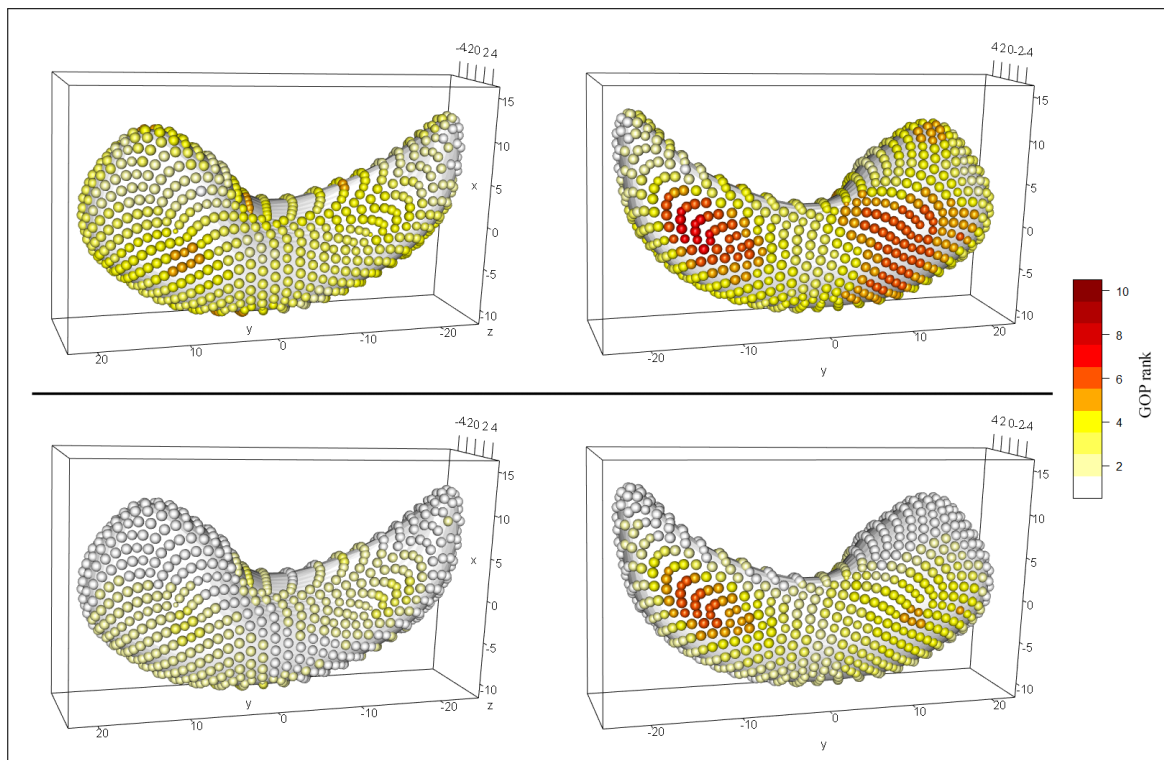


Figure 5.11: SPHARM-PDM hypothesis test independent from alignment.

Top row associates with test after removing scale and bottom row corresponds to test with preserving scale. (KDE bandwidth $h = 0.05$ and the maximum possible value $\delta = 7.97$)

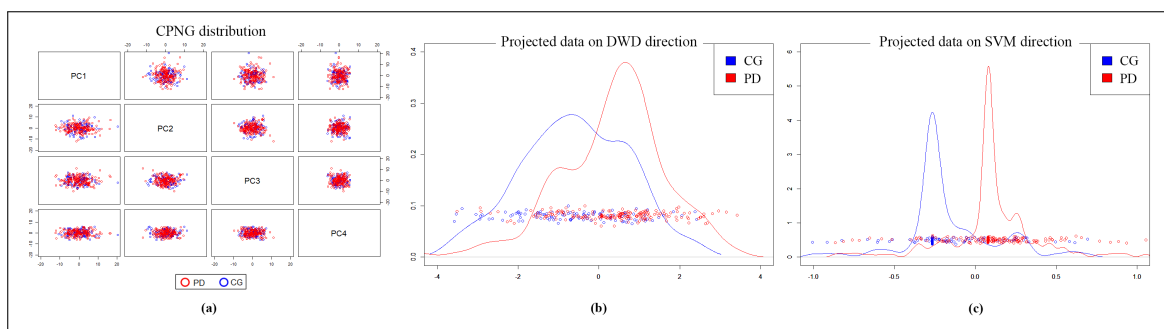


Figure 5.12: s-rep classification.

(a) Paired plot of PCA on CPNG data of CG and PD. (b) KDE plot for the projection of classified data on DWD direction. (c) KDE plot for the projection of classified data on SVM direction.

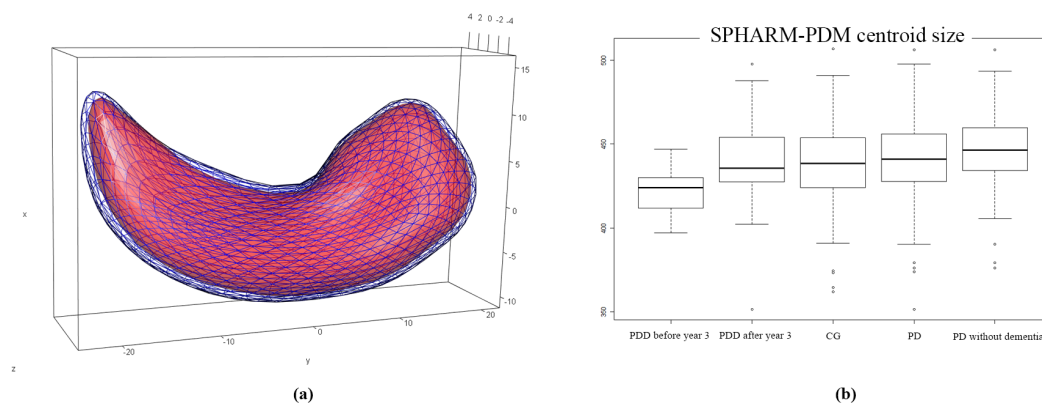


Figure 5.13: PD with dementia.

(a) Overlaid SPHARM-PDM mean shapes from size-and-shape analysis, mean PDD in red mesh and mean CG in blue wired mesh. (b) SPHARM-PDM centroid size box plots of CG, PD and PDD.

Chapter 6

Discussion and Future Work

6.1 Discussion

We applied two sensitive shape representations, s-rep, and SPHARM-PDM to study the morphological difference of the left hippocampus between PD and CG. For the hypothesis test, finding mean s-rep was necessary. Therefore, we discussed dimensionality reduction methods for spherical data such as PNG and its expansion CPNG to obtain mean s-rep. Besides, we proposed NLPGA on the rotating tangent space of S^2 to analyze spherical data. Although we ignored using NLPGA in this work, we believe it can be advantageous for the study of directional data because of its simplicity.

Each s-rep of this study had 114 spokes and 69 skeletal positions, so the pre-shape space of skeletal positions was a 68-dimensional hypersphere. Thus, the implementation of PNS to find the mean shape was computationally costly. Also, in the s-reps, spherical GOPs had concentrated distributions, so the CPNG had no superiority over PGA. To boost the speed of the permutation test, we defined mean s-rep as a combination of GPA mean for skeletal positions, Fréchet mean for spokes' directions, and the geometric mean for spokes' lengths. Besides, we introduced a parametric approach based on multivariate Hotelling's T^2 test (with normality assumption) and controlled the false positive with the Benjamini-Hochberg method. The analysis of GOPs exposed statistically global and partial significant differences.

Both the parametric and non-parametric approaches are highly dependent on the method of alignment. Hence, we explained weighted alignment and introduced the elimination algorithm, which classifies points in two suspicious and unsuspecting groups. Then we defined an appropriate covariance matrix for the weighted Procrustes alignment based on unsuspecting points. The skeletal means of two groups were extremely similar, and applying the elimination algorithm was unnecessary. Nevertheless, to study the effect of weighted alignment, we considered a small

threshold $\delta = 0.006$ to extract a set of unsuspecting points. By the weighted alignment, we stabilized unsuspecting points and displayed the effect of this act on the result.

We discussed a hypothesis testing method independent from the alignment. This method is constructed on the distance matrices and is applicable for PDMs and shapes with directional data like s-rep. We mentioned how to use KDE on the mirrored p -values to rank the GOPs and assign a color map for the visualization. Also, we explained a similar test with angles for planner shapes.

The application of different tests without controlling false positive indicates a similar significant area of the hippocampi. Nonetheless, by correcting false positives, depending on the method we choose for the correction (e.g., FDR), differences in size-and-shapes may disappear. Likewise, the analysis of shape and size-and-shape illuminates the fact that removing scale exposes more differences.

Further analysis of normal vectors and size of the corresponding SPHARM-PDM triangulated surfaces after correction showed no significant differences even with a large FDR (e.g., FDR=0.15). Besides, the analysis outcome of the Middle Surface PDM supports the obtained results of the analysis of s-rep skeletal positions.

Implementation of DWD and SVM took place on the CPNG data to classify s-reps with half of the samples as the training set. Although DWD had a better performance, result of the classification (on the test set) was not promising. On average, the classification accuracy we reached was less than 60%.

Finally, despite the dissimilarity between PD and CG, we briefly discussed the differences might come from dementia rather than PD, but further investigation is needed.

6.2 Future work

6.2.1 Analysis of other brain parts and their relationship

In this thesis, we focused on the left hippocampus and applied some advanced shape representations such as s-rep and SPHARM-PDM for the analysis. The same procedure is considered for other brain structures, including Caudate, Thalamus, Putamen, Amygdala, and Pallidum. Figure 6.1 shows the fitted s-rep for four particular brain sections.

On the other hand, we can study the multi-object shape model, which considers multiple objects and their relationship instead of focusing on one specific brain segment.

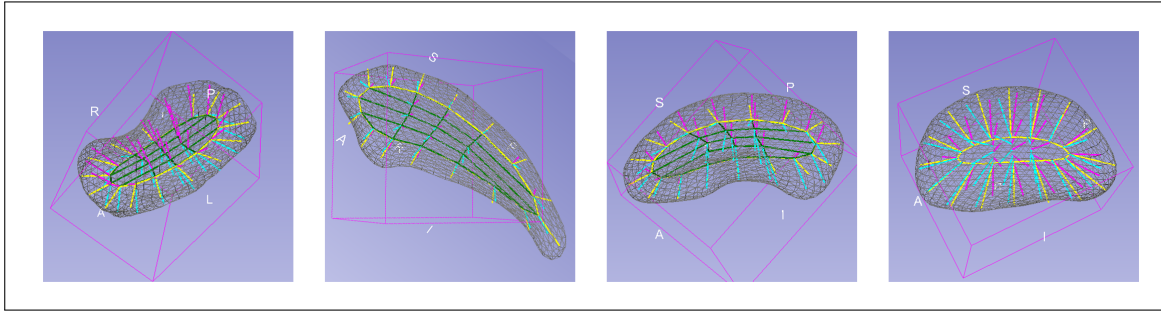


Figure 6.1: Samples of s-rep for different brain parts.

Left to right: Fitted s-rep samples of Left-Amygdala, Left-Caudate, Left-Putamen, and Left-Thalamus.

6.2.2 Distance and Force

Different shape representations provide us with a variety of intrinsic shape properties, which could be a matter of future research. For instance, in Section 4.7.1 for landmark modeling, we discussed distance and force between landmarks as intrinsic properties. The distance is a general concept and can be defined in different ways. One possible way to measure the distance, as explained in (Gattone et al., 2017), is to add uncertainty level to the landmarks and consider each landmark as a multivariate density function (6.1). Thus, to calculate the distance between two landmarks in a shape we can use Fisher information distance (Fisher–Rao metric) (Costa et al., 2015).

Another intrinsic property could be the force. In reality, the force decreases when we move away from the source. This concept is expressible by statistical distributions, for example in multivariate normal distribution the probability density function

$$p(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{m/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right), \quad (6.1)$$

decreases exponentially when we move away from the mean (more straight forward approach is to replace Σ with the identity matrix I). In this sense, we consider each landmark as a mass or a center of the multivariate normal distribution. Thus, in the feature space, force could be the value of the (6.1), and we can build force matrix $(F)_{ij}, i, j = 1, \dots, k$ the same way we introduced the distance matrix at (4.11). Each element of F (i.e., $F(\mathbf{x}_i, \mathbf{x}_j)$) represent the force that \mathbf{x}_j imposes on \mathbf{x}_i .

It is also possible to think of the overall forces that all the landmarks impose on a specific one. Therefore, we can assign a KDE value (4.12) to each landmark. Assume n landmarks $\mathbf{x}_1, \dots, \mathbf{x}_n$ as data points then, the overall force in coordinate \mathbf{x} can be estimated by (4.12). Figure 6.2 illustrates the KDE value of the feature space with the projection of skeletal points to xy plane as data points.

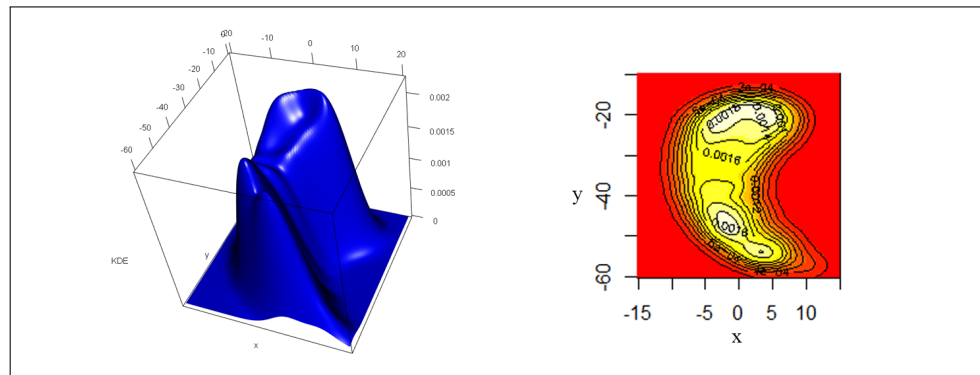


Figure 6.2: KDE.

Left: The KDE value of the feature space with the projected skeletal points to the xy plane as data points. Right: Contour map of the left figure

6.2.3 Classification improvement

In the s-rep analysis, we usually encounter HDLSS. To deal with the data and classify s-reps (Hong, 2018) applied DWD as a classifier and tried to improve outcomes based on the Bayes' theorem. Another strategy could be dimensionality reduction. Dimensionality reduction can be made by the classical feature selection method or PCA 3.1. Since we introduced the fast approach of hypothesis testing, in binomial classification, it is reasonable to select features that are significantly different between the two groups. Seemingly, fisher information distance could help us to choose features with adequate separation. This strategy can improve the classification as we avoid the curse of dimensionality.

6.2.4 Alignment

With the alignment, locational shape differences is a relative concept, and it is necessary to develop methods independent from the alignment or improve the methods of alignment to expose real differences.

The elimination algorithm we introduced is computationally expensive because, in each iteration, the algorithm needs to calculate the mean shape, so the implementation for shapes with a large number of points such as SPHARM-PDM is not rational. Further, the alignment (to find mean shape) occurs within the algorithm, which can lead to inappropriate point selection. Therefore, developing suitable alignment methods applicable to the hypothesis testing merit further study.

List of Figures

2.1	Landmark model and m-rep.	7
2.2	Workflow of s-rep model fitting.	9
2.3	SPHARM-PDM.	11
2.4	Middle Surface PDM.	12
2.5	Suspicious landmarks.	26
2.6	Comparison of alignments.	27
2.7	Procrustes vs. weighted alignment by elimination.	28
3.1	Illustration of NLPCA.	35
3.2	Rotation of the tangent space.	36
3.3	PNS vs. NLPGA.	37
4.1	Distance matrices.	57
4.2	Workflow of the hypothesis test independent from alignment.	58
4.3	Symmetric objects.	60
4.4	Test with alignment vs. Test with distance and angle.	62
4.5	KDE on mirrored p -values.	62
5.1	ParkWest data.	64
5.2	SPHARM-PDM smooth triangle mesh.	65
5.3	Eigenmode contributions of CG s-reps.	67
5.4	s-rep results.	69
5.5	s-rep CPNG means.	70
5.6	Effect of weighted alignment on s-rep skeletal.	70
5.7	Significant positions by GPA alignment vs. weighted Procrustes alignment.	71
5.8	s-rep hypothesis test independent from alignment.	72
5.9	Significance plot of SPHARM-PDMs by Procrustes alignment.	73
5.10	Ordered p -values and adjusted p -values of SPHARM-PDM.	73
5.11	SPHARM-PDM hypothesis test independent from alignment.	74
5.12	s-rep classification.	74
5.13	PD with dementia.	75
6.1	Samples of s-rep for different brain parts.	78
6.2	KDE.	79
A.1	Significance plot of SPHARM-PDMs by permutation.	82
A.2	Triangle mesh analysis.	83
A.3	Middle Surface PDM	83
A.4	Significance plot of SPHARM-PDMs by permutation	84

List of Tables

2.1	Elimination examples.	26
4.1	Multiple testing.	43
4.2	Mean s-rep computation time.	48
4.3	Non-parametric combination.	53
4.4	Quadratic combination.	53
5.1	T-test on centroid size.	66
5.2	Global tests.	66
5.3	DWD vs. SVM.	72

Appendix A

Appendix

A.1 Analysis

A.1.1 Non-parametric result for SPHARM-PDM

Result of the 10000 permutation with Hotelling T^2 metric on size-and-shape of SPHARM-PDMs. Figure A.1 visualizes the results.

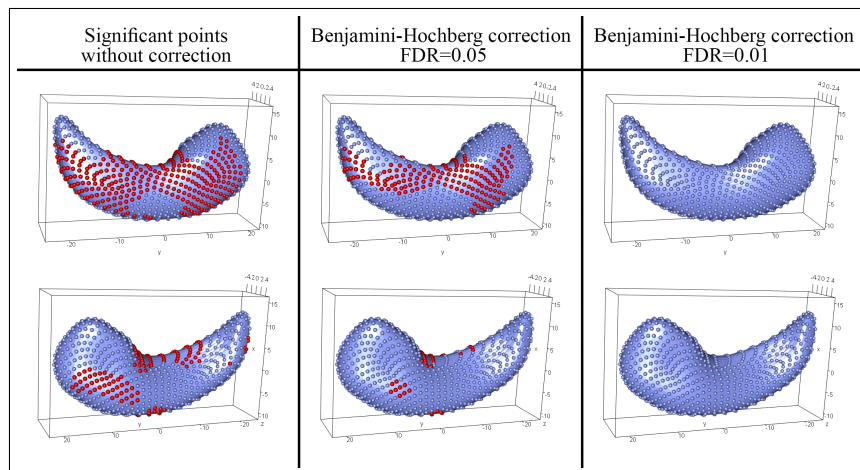


Figure A.1: Significance plot of SPHARM-PDMs by permutation.

Visualization of permutation test on SPHARM-PDMs of the left hippocampi from two angles. Red color indicates significant points. Left column shows significant points without controlling false positive ($\alpha = 0.05$). Middle column shows significant points after correction with Benjamini-Hochberg method $FDR=0.05$. Right column indicates significant points after correction with Benjamini-Hochberg method $FDR=0.01$.

A.1.2 Analysis of boundary normal vectors

We introduced SPHARM-PDM triangulated surfaces, their size and normal direction in Section 2.1.4. Thus, we hypothesize the mean direction of normal vectors of CG and PD hippocampus

meshes by Hotelling- T^2 on tangent space same as the parametric approach for s-rep spokes' directions in Section 4.6.2 and then we test the triangles' sizes by T-test. Although we observe a considerable amount of significant directions and sizes (see Figure A.2), by Benjamini–Hochberg controlling $FDR=0.15$ literally there is no difference between two groups neither in directions nor in sizes.

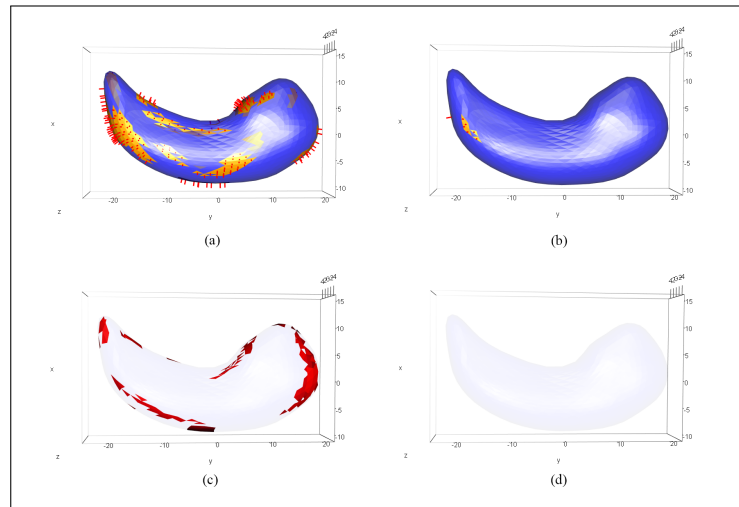


Figure A.2: Triangle mesh analysis.

(a) Significant unit normal vectors in red accompanied by their brightened corresponding triangles (b) Significant normal vectors in red after correction $FDR=0.15$ (c) Significant triangles in size depicted in red (d) No significant triangles in size after correction $FDR=0.15$

A.1.3 Middle Surface PDM analysis

Analysis of Middle Surface PDMs supports the result of s-rep skeletal analysis as we see the same significant area (see Figure A.3).

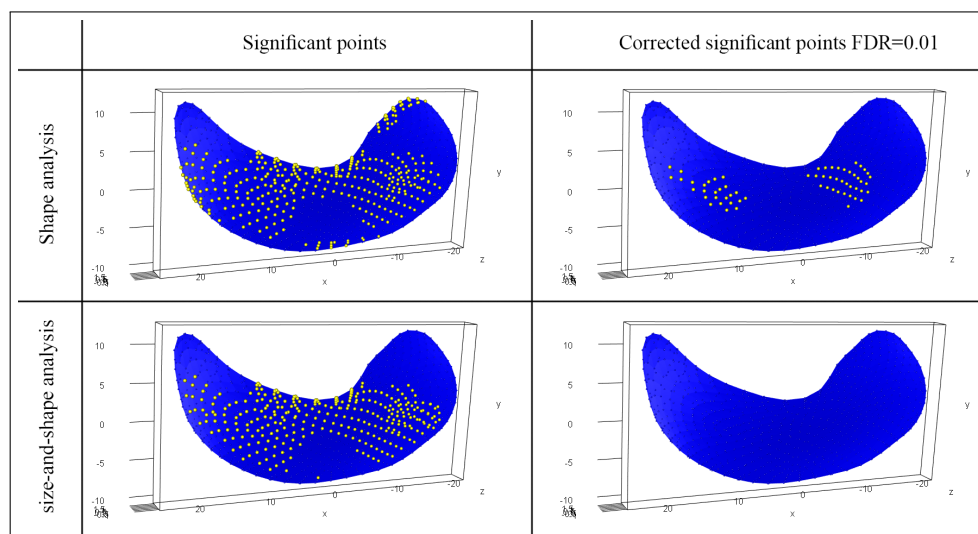


Figure A.3: Middle Surface PDM
Yellow indicates significant points.

A.1.4 Analysis of SPHARM-PDM plus skeletal PDM

In Section 2.1.5 we introduced a PDM as a combination of SPHARM-PDM and skeletal PDM. Outcome of the analysis on this PDM is very similar to the analysis outcome of SPHARM-PDM and s-rep skeletal individually. However, it was tempting to align meshes' points (i.e. SPHARM-PDMs) based on the skeletal position with the weighted alignment and study the outcome. Seemingly this attempt add more variation to the SPHARM-PDMs and as a result we see more significant points on the mesh. Figure A.4 illustrate the results.

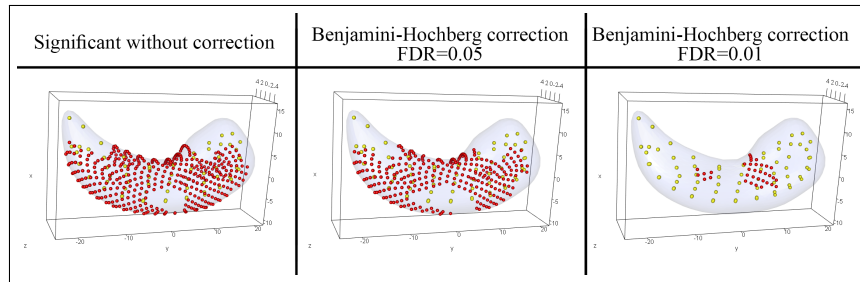


Figure A.4: Significance plot of SPHARM-PDMs by permutation
Red indicates SPHARM-PDM significant points and yellow indicates skeletal PDM

A.2 Example

A.2.1 Helmert submatrix

E.g., for $k = 4$ landmarks the Helmert submatrix would be,

$$H = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{-1}{\sqrt{12}} & \frac{-1}{\sqrt{12}} & \frac{-1}{\sqrt{12}} & \frac{3}{\sqrt{12}} \end{bmatrix}.$$

(Dryden and Mardia, 2016, ch.2)

Bibliography

- Acton, Q.
2012. *Advances in Dementia Research and Treatment: 2012 Edition*. ScholarlyEditions.
- Adhikari, M.
2016. *Basic Algebraic Topology and its Applications*. Springer India.
- Apostolova, L., G. Alves, K. S. Hwang, S. Babakchianian, K. S. Bronnick, J. P. Larsen, P. M. Thompson, Y.-Y. Chou, O. B. Tysnes, H. K. Vefring, et al.
2012. Hippocampal and ventricular changes in parkinson's disease mild cognitive impairment. *Neurobiology of aging*, 33(9):2113–2124.
- Barber, R., A. Gholkar, P. Scheltens, C. Ballard, I. McKeith, and J. O'Brien
2000. Mri volumetric correlates of white matter lesions in dementia with lewy bodies and alzheimer's disease. *International journal of geriatric psychiatry*, 15(10):911–916.
- Benjamini, Y. and Y. Hochberg
1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.
- Beyer, M. K., K. S. Bronnick, K. S. Hwang, N. Bergsland, O. B. Tysnes, J. P. Larsen, P. M. Thompson, J. H. Somme, and L. G. Apostolova
2013. Verbal memory is associated with structural hippocampal changes in newly diagnosed parkinson's disease. *J Neurol Neurosurg Psychiatry*, 84(1):23–28.
- Bonferroni, C.
1936. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8:3–62.
- Bookstein, F. L. et al.
1986. Size and shape spaces for landmark data in two dimensions. *Statistical science*, 1(2):181–222.
- Brechbühler, C., G. Gerig, and O. Kübler
1995. Parametrization of closed surfaces for 3-d shape description. *Computer vision and image understanding*, 61(2):154–170.

Breuß, M., A. Bruckstein, and P. Maragos

2013. *Innovations for Shape Analysis: Models and Algorithms*, Mathematics and Visualization. Springer Berlin Heidelberg.

Brignell, C. J., I. L. Dryden, and W. J. Browne

2016. Covariance weighted procrustes analysis. In *Riemannian Computing in Computer Vision*, Pp. 189–209. Springer.

Brück, A., T. Kurki, V. Kaasinen, T. Vahlberg, and J. Rinne

2004. Hippocampal and prefrontal atrophy in patients with early non-demented parkinson's disease is related to cognitive impairment. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(10):1467–1469.

Budson, A. and P. Solomon

2011. *Memory Loss E-Book: A Practical Guide for Clinicians*. Elsevier Health Sciences.

Budson, A. and P. Solomon

2015. *Memory Loss, Alzheimer's Disease, and Dementia E-Book: A Practical Guide for Clinicians*. Elsevier Health Sciences.

Candes, E.

2018. Stats 300c: Lecture on theory of statistics. Preprint on webpage at <https://statweb.stanford.edu/~candes/teaching/stats300c/lectures.html>.

Christensen, O.

2010. *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis. Birkhäuser Boston.

Costa, S. I., S. A. Santos, and J. E. Strapasson

2015. Fisher information distance: a geometrical reading. *Discrete Applied Mathematics*, 197:59–69.

Csernansky, J., L. Wang, J. Swank, J. Miller, M. Gado, D. McKeel, M. Miller, and J. Morris

2005. Preclinical detection of alzheimer's disease: hippocampal shape and volume predict dementia onset in the elderly. *Neuroimage*, 25(3):783–792.

Dalaker, T. O., J. P. Larsen, N. Bergsland, M. K. Beyer, G. Alves, M. G. Dwyer, O.-B. Tysnes, R. H. Benedict, A. Kelemen, K. Bronnick, et al.

2009. Brain atrophy and white matter hyperintensities in early parkinson's disease. *Movement disorders: official journal of the Movement Disorder Society*, 24(15):2233–2241.

Damon, J. and J. Marron

2014. Backwards principal component analysis and principal nested relations. *Journal of Mathematical Imaging and Vision*, 50(1-2):107–114.

- Davison, A., D. Hinkley, R. Gill, B. Ripley, S. Ross, M. Stein, D. Williams, and B. Silverman
1997. *Bootstrap Methods and Their Application*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- de la Monte, S. M., S. E. Wells, E. T. Hedley-Whyte, and J. H. Growdon
1989. Neuropathological distinction between parkinson's dementia and parkinson's plus alzheimer's disease. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 26(3):309–320.
- de Winter, P. and P. Cahusac
2014. *Starting out in Statistics: An Introduction for Students of Human Health, Disease, and Psychology*. Wiley.
- Dhikav, V. and K. Anand
2011. Potential predictors of hippocampal atrophy in alzheimer's disease. *Drugs & aging*, 28(1):1–11.
- Dryden, I. and K. Mardia
1998. *Statistical Shape Analysis*, Wiley Series in Probability & Statistics. Wiley.
- Dryden, I. and K. Mardia
2016. *Statistical Shape Analysis: With Applications in R*, Wiley Series in Probability and Statistics. Wiley.
- Dryden, I. L.
2018. `shapes` package. R Foundation for Statistical Computing, Vienna, Austria. Contributed package, Version 1.2.4.
- Fletcher, P. T., C. Lu, S. M. Pizer, and S. Joshi
2004. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE transactions on medical imaging*, 23(8):995–1005.
- Fong, C.
2015. Analytical methods for squaring the disc. *arXiv preprint arXiv:1509.06344*.
- Fox, N. C., P. A. Freeborough, and M. N. Rossor
1996. Visualisation and quantification of rates of atrophy in alzheimer's disease. *The Lancet*, 348(9020):94–97.
- Fréchet, M. R.
1948. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'institut Henri Poincaré*, 10(4):215–310.

- Gattone, S. A., A. De Sanctis, T. Russo, and D. Pulcini
2017. A shape distance based on the fisher–rao metric and its application for shapes clustering. *Physica A: Statistical Mechanics and its Applications*, 487:93–102.
- Gerig, G., M. Styner, D. Jones, D. Weinberger, and J. Lieberman
2001. Shape analysis of brain ventricles using spharm. In *Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2001)*, Pp. 171–178. IEEE.
- Goldstein, H., C. Poole, and J. Safko
2013. *Classical Mechanics*. Pearson.
- Goodall, C.
1991. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):285–321.
- Hastie, T. and W. Stuetzle
1989. Principal curves. *Journal of the American Statistical Association*, 84(406):502–516.
- Helse Stavanger
2020. Stavanger university hospital. <https://helse-stavanger.no/>.
- Hong, J. H.
2018. *Classification of Neuroanatomical Structures Based on Non-Euclidean Geometric Object Properties*. PhD thesis, University of North Carolina.
- Hotelling, H.
1933. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417.
- Ikeda, M., H. Tanabe, Y. Nakagawa, H. Kazui, H. Oi, H. Yamazaki, K. Harada, and T. Nishimura
1994. Mri-based quantitative assessment of the hippocampal region in very mild to moderate alzheimer’s disease. *Neuroradiology*, 36(1):7–10.
- Jolliffe, I.
2013. *Principal Component Analysis*, Springer Series in Statistics. Springer New York.
- Jorgensen, B.
1993. *Theory of Linear Models*, Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.
- Jung, S.
2011. *Asymptotics for high dimension low sample size data and analysis of data on manifolds*. PhD thesis, University of North Carolina.

Jung, S.

2014. Composite principal nested spheres for s-reps. CPNS implementation.

Jung, S., I. L. Dryden, and J. Marron

2012. Analysis of principal nested spheres. *Biometrika*, 99(3):551–568.

Junqué, C., B. Ramírez-Ruiz, E. Tolosa, C. Summerfield, M.-J. Martí, P. Pastor, B. Gómez-Ansón, and J. M. Mercader

2005. Amygdalar and hippocampal mri volumetric reductions in parkinson’s disease with dementia. *Movement disorders: official journal of the Movement Disorder Society*, 20(5):540–544.

Kendall, D. G.

1977. The diffusion of shape. *Advances in applied probability*, 9(3):428–430.

Kendall, D. G.

1984. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121.

Kim, J. and H. Bang

2016. Three common misuses of p values. *Dental hypotheses*, 7(3):73.

Korpela, J., E. Oikarinen, K. Puolamäki, and A. Ukkonen

2017. Multivariate confidence intervals. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, Pp. 696–704. SIAM.

Kruger, U., J. Zhang, and L. Xie

2008. Developments and applications of nonlinear principal component analysis—a review. In *Principal manifolds for data visualization and dimension reduction*, Pp. 1–43. Springer.

Mardia, K., J. Bibby, and J. Kent

1982. *Multivariate analysis*, Probability and mathematical statistics. Acad. Press.

Marron, J. S., M. J. Todd, and J. Ahn

2007. Distance-weighted discrimination. *Journal of the American Statistical Association*, 102(480):1267–1271.

McClure, R. K., M. Styner, E. Maltbie, J. A. Lieberman, S. Gouttard, G. Gerig, X. Shi, and H. Zhu

2013. Localized differences in caudate and hippocampal shape are associated with schizophrenia but not antipsychotic type. *Psychiatry Research: Neuroimaging*, 211(1):1–10.

Montgomery, D., E. Peck, and G. Vining

2015. *Introduction to Linear Regression Analysis*, Wiley Series in Probability and Statistics. Wiley.

Mosteller, F. and R. A. Fisher

1948. Questions and answers. *The American Statistician*, 2(5):30–31.

Nobis, L., S. G. Manohar, S. M. Smith, F. Alfaro-Almagro, M. Jenkinson, C. E. Mackay, and M. Husain

2019. Hippocampal volume across age: Nomograms derived from over 19,700 people in uk biobank. *NeuroImage: Clinical*, 23:101904.

Panaretos, V. M., T. Pham, and Z. Yao

2014. Principal flows. *Journal of the American Statistical Association*, 109(505):424–436.

ParkWest

2020. Parkwest study. <http://www.parkvest.no/>.

Pennec, X., S. Sommer, and T. Fletcher

2019. *Riemannian Geometric Statistics in Medical Image Analysis*. Elsevier Science.

Perl, D. P.

2010. Neuropathology of alzheimer’s disease. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine: A Journal of Translational and Personalized Medicine*, 77(1):32–42.

Pesarin, F.

2001. *Multivariate Permutation Tests: With Applications in Biostatistics*. Wiley.

Pfeiffer, R., Z. Wszolek, and M. Ebadi

2012. *Parkinson’s Disease*. CRC Press.

Pizer, S. M., J. Hong, J. Vicory, Z. Liu, J. Marron, H.-y. Choi, J. Damon, S. Jung, B. Paniagua, J. Schulz, et al.

2020. Object shape representation via skeletal models (s-reps) and statistical analysis. In *Riemannian Geometric Statistics in Medical Image Analysis*, Pp. 233–271. Elsevier.

Pizer, S. M., S. Jung, D. Goswami, J. Vicory, X. Zhao, R. Chaudhuri, J. N. Damon, S. Huckemann, and J. Marron

2013. Nested sphere statistics of skeletal models. In *Innovations for Shape Analysis*, Pp. 93–115. Springer.

Rizzo, M.

2007. *Statistical Computing with R*, Chapman & Hall/CRC The R Series. Taylor & Francis.

SALMASO, L. and C. Brombin

2013. *Permutation Tests in Shape Analysis*, SpringerBriefs in Statistics. Springer New York.

Scholz, M.

2002. Nichtlineare hauptkomponentenanalyse auf basis neuronaler netze. Master's thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II.

Schulz, J.

2013. *Statistical Analysis of Medical Shapes and Directional Data*. PhD thesis, UiT Norges arktiske universitet.

Schulz, J., S. M. Pizer, J. Marron, and F. Godtlielsen

2016. Non-linear hypothesis testing of geometric object properties of shapes applied to hippocampi. *Journal of Mathematical Imaging and Vision*, 54(1):15–34.

Šidák, Z.

1967. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633.

Siddiqi, K. and S. Pizer

2008. *Medial Representations: Mathematics, Algorithms and Applications*, Computational Imaging and Vision. Springer Netherlands.

Snowdon, D. A., L. H. Greiner, J. A. Mortimer, K. P. Riley, P. A. Greiner, and W. R. Markesbery

1997. Brain infarction and the clinical expression of alzheimer disease: the nun study. *Jama*, 277(10):813–817.

Srivastava, A. and E. Klassen

2002. Monte carlo extrinsic estimators of manifold-valued parameters. *IEEE Transactions on Signal Processing*, 50(2):299–308.

Styner, M., I. Oguz, S. Xu, C. Brechb, D. Pantazis, J. Levitt, M. Shenton, and G. Gerig

2006. Framework for the statistical shape analysis of brain structures using spharm-pdm release 1.00. *Insight J*, 1071.

Styner, M., I. Oguz, S. Xu, D. Pantazis, and G. Gerig

2007. Statistical group differences in anatomical shape analysis using Hotelling T2 metric. In *Medical Imaging 2007: Image Processing*, J. P. W. Pluim and J. M. Reinhardt, eds., volume 6512, Pp. 1326 – 1336. International Society for Optics and Photonics, SPIE.

Terrell, G. R. and D. W. Scott

1992. Variable kernel density estimation. *The Annals of Statistics*, Pp. 1236–1265.

Theodoridis, S.

2015. *Machine Learning: A Bayesian and Optimization Perspective*, .NET Developers Series. Elsevier Science.

- Thompson, P. M., K. M. Hayashi, G. I. De Zubicaray, A. L. Janke, S. E. Rose, J. Semple, M. S. Hong, D. H. Herman, D. Gravano, D. M. Doddrell, et al.
2004. Mapping hippocampal and ventricular change in alzheimer disease. *Neuroimage*, 22(4):1754–1766.
- Tu, L., M. Styner, J. Vicory, S. Elhabian, R. Wang, J. Hong, B. Paniagua, J. C. Prieto, D. Yang, R. Whitaker, et al.
2017. Skeletal shape correspondence through entropy. *IEEE transactions on medical imaging*, 37(1):1–11.
- Villasenor Alva, J. A. and E. G. Estrada
2009. A generalization of shapiro–wilk’s test for multivariate normality. *Communications in Statistics—Theory and Methods*, 38(11):1870–1883.
- Wasserstein, R. L., N. A. Lazar, et al.
2016. The asa’s statement on p-values: context, process, and purpose. *The American Statistician*, 70(2):129–133.